The Role of Standing Genetic Variation in Rapid Adaptation - Insights from Ancient and Large-scale Contemporary Threespine Stickleback Genomes

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Melanie Kirch

aus Fürth

Tübingen

2022

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	28.02.2023
Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatterin:	Dr. Felicity Jones
2. Berichterstatter:	Prof. Dr. Oliver Kohlbacher

Acknowledgements

I am sincerely grateful to my PhD advisor Felicity Jones. Thank you for challenging me to outgrow myself and for believing in me and my skills. You are one of the most scientifically brilliant persons I have ever met and I am deeply grateful for the chance to discuss scientific questions on a daily basis with you. Thank you for your empathy and your guidance.

I would also like to thank the members of my thesis advisory committee: Dr. Detlef Weigel as well as my university supervisor Prof. Oliver Kohlbacher. Thank you for your valuable feedback and for accompanying the process of my PhD from the beginning to the end.

A special thanks goes to Andrew Foote. I am deeply grateful for your scientific and personal guidance. Thank you for your honesty, your patience and your empowerment.

I am also grateful to Frank Chan for all the valuable discussions and feedback. Your perspective and out-of-the-box questions have always challenged me and broadened my knowledge as well as my skills.

My time in the lab without the help and support of numerous past and current lab mates is unimaginable. Two persons' contribution has been absolutely crucial for the success of the 10 000 marine sticklebacks project. Elena Avdievich and Kavita Venkataramani, thank you both so much for the months of work and all the nerves you dedicated to this project to make it possible! Many more persons helped to tackle the challenge of processing several thousand fish in the lab. A special thanks goes to Robin Thome and George Philippos, who supported me as Hiwis with the lab work. Thank you for all your power of endurance. I would like to thank Stefano Lazarrano, Layla Hiramatsu, Andreaa Dreau, Muhua Wang, Rahmiye Kürkcü, Vrinda Venu, Ludmila Gaspar, Jonas Schwickert, Moritz Peters, Volker Soltys, Cholpon Zhakshylykova, Marek Kucka, Dingwen Su and Julia Hagauer for their helping hands, but also for scientific discussions as well as advice and feedback. I would like to point out Dingwen Su and Julia Hagauer who especially supported me during the writing process – with insightful discussions, valuable feedback and scientific as well as emotional support.

The Friedrich Miescher Laboratory and the Max Planck Institute of Biology provided amazing facilities and a unique work environment. I benefited tremendously from the Institute cluster which greatly facilitated parallel computing to process tera bytes of data. The in-house Genome Center was also amazingly convenient. Thank you to Christa Lanz and Heike Budde for processing the samples for sequencing and to Ilja Bezrukov for providing the sequencing data.

I also want to thank Herta Soffel, who was secretary of the Friedrich Miescher Laboratory during the majority of my PhD - thank you for helping to organize all the travelling and the shipping of numerous samples. Thanks also Claudia Jahn, her successor.

I am deeply grateful to our former PhD coordinator Sarah Danes for her welcoming reception at the institute and for her support throughout my PhD, especially during the hard times. Thank you also to Jeanette Müller and Sibylle Patheiger for their help.

Some special persons have sweetened my everyday life at the institute. Thank you to Christian Kubica, Tess Renahan, Or Shalev, Alexander Blaessle, Maria Almuedo-Castillo, Luciano Marcon and Jelena Raspopovic. Last, but not least, I would like to thank Bridgit Waithaka. Bridgit, thank you for countless talks and walks, and for being such an inspiring human being.

Finally, I am sincerely grateful for all my friends outside of the institute, and especially my roommates, who have supported me in small and big ways during the time of my PhD. I want to thank all the strong and candid persons, who have filled my life with inspiration, joy and strength during the last years. I am also beholden to my brothers for their love and support.

Table of contents

Abstract

The climate crisis will cause tremendous environmental change in the upcoming decades and will challenge many organisms to adapt rapidly to a new or changing environment. A major objective in evolutionary biology is to understand the factors promoting and/or restraining rapid adaptation in order to disentangle how we can maintain biodiversity on our planet. Two molecular key mechanisms of rapid adaptation include adaptation via newly arising mutations and via the reuse of pre-existing standing genetic variation (SGV), with the latter mechanism predicted to outperform *de novo* mutations in most cases. Despite its major role in rapid adaptation, little is known about the availability and maintenance of SGV in natural populations. This thesis aims to fill this gap of knowledge by taking advantage of a prime model system in evolutionary biology, the threespine sticklebacks (*Gasterosteus aculeatus*). Since the end of the Pleistocene, sticklebacks have undergone a recent adaptive radiation with marine sticklebacks repeatedly and independently colonising freshwater habitats and rapidly adapting to the new environment. This system has been well-studied in the past and hence, the genomic basis of parallel adaptation has been identified, large sample sizes are accessible and many genetic as well as genomic tools such as a high-quality reference genome are available turning this organism into a unique platform to investigate SGV and its role in rapid adaptation. In this thesis, I investigate the availability of SGV, how SGV can facilitate and/or constrain rapid adaptation and how SGV can be maintained in a population based on ancient as well as large-scale contemporary samples.

This thesis comprises two published and one unsubmitted manuscript encompassing different genomic approaches: a paleogenomics approach studying an ancient genome extracted from 11 000 to 13 000 years old stickleback bones; contemporary evolutionary time-series data of stickleback adaptation to multiple freshwater environments; individual based genotyping data as well as targeted enrichment sequencing data for thousands of marine sticklebacks in order to characterise SGV in contemporary marine populations.

The paleogenomics approach uncovers the presence of freshwater-adaptive SGV in an ancient marine stickleback genome indicating adaptive SGV is similarly found in both ancient and present-day marine populations. Contemporary marine populations are therefore likely to be good representatives for ancient founders of present-day freshwater population. In addition, some freshwater adaptive alleles present in the ancient genome are lacking in the present-day lake population from which the ancient marine fish was sampled indicating that demographic processes resulted in stochastic loss of beneficial variants over time. Contemporary time-series data revealed that

freshwater adaptive variants in three newly founded Alaskan lake populations rose to high allele frequencies within a time span of less than 30 generations confirming that marine sticklebacks can rapidly adapt to freshwater environments. Genotyping data for >750 marine individuals from the marine source population at >350 SNPs showed further that freshwater adaptive alleles are present at low frequencies in the marine source populations and highlighted that data encompassing samples an order of magnitude larger are required to estimate properties of SGV such as the linkage disequilibrium among adaptive alleles reliably. To overcome this obstacle, pull-down sequencing at individual resolution was performed for more than 9 000 marine sticklebacks enriching the data for loci previously identified to be divergent between freshwater and marine sticklebacks. The frequency of freshwater adaptive variants determined in two marine stickleback populations form Alaska and Washington is low and frequency distributions differ significantly among the two populations. Marine sticklebacks do not correspond to simulated F1 hybrids of freshwater and marine sticklebacks, but rather to multigenerational backcrossed F1s with marine individuals. The large-scale empirical data shows that each marine individual carriers multiple freshwater adaptive alleles and that inter-chromosomal linkage disequilibrium is slightly elevated compared to random expectations indicating that the availability of SGV in the marine population has the potential to facilitate future rapid re-assembly of adaptive alleles in the freshwater environment. Comparisons to genomic datasets of global freshwater sticklebacks demonstrated that freshwater adaptive haplotypes present in the marine populations studied cluster with North-American, European as well as Californian individuals supporting the hypothesis that SGV in the marine population originates from multiple diverse source populations. Furthermore, whole genome sequencing data of a freshwater population uncovered evidence of both soft and hard sweeps at different adaptive loci emphasising the crucial role of pre-existing SGV for rapid adaptive evolution in freshwater populations.

By comprising both ancient and present-day genomic approaches this thesis delivers insight into how SGV can facilitate rapid adaptive evolution, how the availability and maintenance of SGV may be promoted and/or restrained by evolutionary processes and thus, into the evolutionary potential of biodiversity.

Zusammenfassung

Die Klimakrise wird in den kommenden Jahrzehnten zu einem beispiellosen Ausmaß an Umweltveränderungen führen, wodurch viele Organismen dazu gezwungen sein werden, sich rapide an neue oder sich ändernde Lebensräume anzupassen. Um effektive Strategien zum Erhalt von Biodiversität zu entwickeln ist ein wichtiges Ziel der Evolutionsbiologie, zu verstehen, welche Faktoren zugunsten oder entgegen schneller evolutionärer Anpassung wirken. Die zwei wichtigsten molekularen Mechanismen, die schnelle Anpassung ermöglichen, sind das Auftreten neuer genetischer (*de novo*) Mutationen und die Wiederverwendung von bestehender genetischer Variation (*standing genetic variation*, SGV), wobei letzteres in den meisten Fällen effektiver sein sollte. Trotz ihrer wichtigen Rolle in schnellen Anpassungsprozessen ist bisher wenig zu Verfügbarkeit und Aufrechterhaltung von SGV in natürlichen Populationen bekannt. Diese Doktorarbeit versucht diese Wissenslücke zu füllen und nutzt dafür einen erstklassigen Modelorganismus der Evolutionsbiologie, den dreistachligen Stichling (*Gasterosteus aculeatus*). Seit dem Ende des Pleistozäns haben marine Stichlinge eine adaptive Radiation erlebt — sie haben wiederholt und unabhängig voneinander Süßwassergewässer besiedelt und sich dabei schnell an die neue Umgebung angepasst. Dieses Modelsystem wurde in der Vergangenheit gut untersucht und daher wurde die genomische Grundlage der parallelen Anpassung identifiziert; es ist möglich, große Probengrößen zu sammeln; und viele genetische sowie genomische Werkzeuge wie beispielsweise ein hochwertiges Referenzgenom sind verfügbar, und machen diesen Organismus zu einer einzigartigen Plattform für die Untersuchung von SGV und ihrer Rolle bei schneller Anpassung. In dieser Arbeit untersuche ich basierend auf sehr alten Proben sowie einer großen Anzahl von zeitgenössischen Proben die Verfügbarkeit von SGV, wie SGV eine schnelle Anpassung fördern und/oder einschränken kann und wie SGV in einer Population aufrechterhalten werden kann.

Diese Dissertation umfasst zwei veröffentlichte und ein nicht eingereichtes Manuskript und beinhaltet verschiedene genomische Ansätze: einen paläogenomischen Ansatz, der ein sehr altes Genom untersucht, das aus 11.000 bis 13.000 Jahre alten Stichlingsknochen extrahiert wurde; Daten einer zeitgenössischen evolutionären Zeitreihe der Anpassung des Stichlings an mehrere Süßwasserumgebungen; und Genotypisierung sowie Sequenzierung gezielt angereichter Genomregionen auf Individualebene für Tausende marine Stichlingen, um SGV in heutigen marinen Stichlingspopulationen zu charakterisieren.

Der paläogenomische Ansatz deckt das Vorhandensein von Süßwasser-adaptiver SGV in einem sehr alten marinen Stichlingsgenom auf, was darauf hindeutet,

dass adaptive SGV sowohl in alten als auch in heutigen Meerespopulationen vorhanden war und ist. Daher können zeitgenössische Populationen wahrscheinlich als guter Ersatz für die Gründerpopulationen der heutige Süßwasserpopulationen verwendet werden. Darüber hinaus fehlen einige im alten Genom vorhandene Süßwasser-Allele in der heutigen Seepopulation, die aus demselben See stammt wie das alte Genom. Dies deutet darauf hin, dass demografische Prozesse im Laufe der Zeit zu einem stochastischen Verlust nützlicher Allele führten. Die Zeitreihendaten zeigten, dass adaptive Süßwasservarianten in drei neu gegründeten Seepopulationen Alaskas innerhalb einer Zeitspanne von weniger als 30 Generationen zu hohen Allelfrequenzen anstiegen, was bestätigt, dass marine Stichlinge sich schnell an Süßwasserumgebungen anpassen können. Genotypdaten von >750 marinen Individuen aus der ursprünglichen marinen Population mit >350 SNPs zeigten ferner, dass Süßwasserallele in geringen Häufigkeiten in der marinen Gründerpopulation vorhanden sind, und hoben damit hervor, dass um einen Faktor 10 größere Probengrößen notwendig sind, um Eigenschaften von SGV wie etwa das Kopplungsungleichgewicht zwischen adaptiven Allelen zuverlässig ermitteln zu können. Um diese Limitierungen zu überwinden, wurde eine Pulldown-Sequenzierung für mehr als 9.000 marine Stichlinge mit individueller Auflösung durchgeführt, bei der die Daten für Genomregionen angereichert wurden, die zuvor als divergierend zwischen Süßwasser- und marinen Stichlingen identifiziert wurden. Die Häufigkeit adaptiver Süßwasserallele, die in zwei marinen Stichlingspopulationen aus Alaska und Washington bestimmt wurde, ist gering und die Häufigkeitsverteilungen unterscheiden sich signifikant zwischen den beiden Populationen. Marine Stichlinge entsprechen nicht simulierten F1 Hybriden von Süßwasser- und marinen Individuen, sondern mehrfach mit marinen Individuen rückgekreuzten F1s. Die auf tausenden Individuen basierenden empirischen Daten zeigen, dass jedes marine Individuum mehrere adaptive Süßwasserallele trägt und dass das interchromosomale Kopplungsungleichgewicht im Vergleich zu zufälligen Erwartungen leicht erhöht ist. Die Verfügbarkeit von SGV in der marinen Population hat dadurch das Potenzial, eine zukünftige schnelle Wiederzusammensetzung des kompletten Ensembles an adaptiven Allelen in der Süßwasserumgebung zu erleichtern. Vergleiche mit genomischen Datensätzen globaler Süßwasser-Stichlinge zeigten, dass adaptive Süßwasser-Haplotypen, die in den untersuchten marinen Populationen vorkommen, in eine Gruppe mit nordamerikanischen, europäischen sowie kalifornischen Individuen fallen. Dies stützt die Hypothese, dass SGV in der marinen Population aus mehreren unterschiedlichen Quellpopulationen stammt. Darüber hinaus deckten Sequenzierdaten des gesamten Genoms einer Süßwasserpopulation Hinweise auf sowohl weiche als auch harte *Sweeps* an verschiedenen adaptiven Genomregionen auf,

was die entscheidende Rolle von bereits existierender SGV für die schnelle adaptive Evolution in Süßwasserpopulationen betont.

Diese Dissertation umfasst die Analyse sehr alter sowie zeitgenössischer Genome und liefert somit einen Einblick, wie SGV schnelle adaptive Evolution erleichtern kann, wie die Verfügbarkeit und Aufrechterhaltung von SGV durch evolutionäre Prozesse gefördert und/oder gehemmt werden kann und damit auch in das evolutionäre Potenzial von Biodiversität.

List of abbreviations

Genes/proteins

Eda	Ectodysplasin A
Kitlg	Kit ligand
Mc1r	melanocortin-1 receptor
Pel	pelvic enhancer upstream of *Pitx1*
Pitx1	Pituitary homeobox transcription factor 1
Wnt7b	Wnt Family Member 7B

Stickleback populations

CH	Cheney Lake, Alaska, USA
LB	Loberg Lake, Alaska, USA
LITC	Little Campbell River, Canada
PGTS	Puget Sound, Washington, USA
RABS	Rabbit Slough, Alaska, USA
SC	Scout Lake, Alaska, USA

Others

bp	base pair
chr	chromosome
CGV	cryptic genetic variation
CSS	cluster separation score
DNA	deoxyribonucleic acid
e.g.	for example
et al.	and others
FDR	false discovery rate
GWAS	genome wide association study
HMM	Hidden Markov Model
i.e.	that is
IPCC	Intergovernmental Panel on Climate Change
kb	kilo base pair
LD	Linkage disequilibrium
Mb	Mega base pair
N_e	effective population size
QTL	quantitative trait loci

SOM/HMM	self-organizing map-based iterative Hidden Markov Model
SGV	standing genetic variation
SNP	single nucleotide polymorphism

1. General introduction

The climate crisis is a major challenge for our generation and will shape our future tremendously. This anthropogenic phenomena has already altered various ecosystems including terrestrial, freshwater as well as marine ecosystems with an increasing amount of irreversible damage (Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, 2022). In order to persist in the face of this environmental change, organisms must adapt via evolutionary change, or alternatively face extinction. Global warming is predicted to lead to a significant decline in biodiversity in the next 100 years (called the “sixth mass extinction”), with the rate of extinction more than 1000 fold higher than background (Pimm *et al.*, 2014; Ceballos *et al.*, 2015; Thomas *et al.*, 2004; Loarie *et al.*, 2009; Pimm, 2009; Warren *et al.*, 2013; Urban, 2015; Warren *et al.*, 2018). A major challenge facing evolutionary biologists is therefore to understand the factors that facilitate and/or constrain the ability of organisms to adapt to rapidly changing environments. This will provide insight for management of biodiversity, agricultural crops, and have relevance for our understanding of environmentally mediated disease.

Species may initially respond to climate change through plastic responses such as changes in distribution or range. For example, about half of the species assessed globally in the IPCC22 have already shifted their habitats towards the poles or, for terrestrial species, to higher altitudes (Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, 2022). However, such plastic responses are likely to be both constrained (e.g. by physical geography), and insufficient to avoid extinction. It is therefore important to understand adaptive evolutionary processes that enable species to persist in new environmental conditions via heritable variation that is passed on to subsequent generations.

Deciphering the heritable basis of adaptation to novel or changing environments is a key goal of evolutionary geneticists. What types of mutations underlie adaptation? How many are there? Is there a single or a few loci or is the genetic basis of adaptation polygenic?; What are the mutations’ effect sizes on an organism’s fitness? Do adaptive mutations show large or small effect sizes?; What is their genetic architecture? Are adaptive mutations carried in linkage and what is the recombination rate among adaptive mutations?; What is their frequency in the population? Are mutations already available in the founder population as (standing) genetic variation (SGV) or do the mutations arise *de novo*? The answers to these questions, including the latter question which is the focus of this thesis, help us understand constraints on the rate of adaptive evolution; predict how populations respond to selection pressures in

the wild; and enable us to better manage biodiversity, crop breeding and environmentally mediated diseases (e.g. diabetes) in the face of rapid environmental change.

Conservationists have emphasized the importance of available SGV in particular to avert extinction and even claim that neutral diversity predicts adaptive potential (DeWoody *et al.*, 2021; Kardos *et al.*, 2021). This view stems in part from population genetic understanding of the extremely low rate of *de novo* germline mutation rates. Mutation rates in humans, for example, are estimated to 10^{-8} per bp per generation (Durbin *et al.*, 2010) indicating that 30 nucleotide mutations are expected in each human gamete. The effect of any mutation on an organism's fitness and survival is most likely to be slightly deleterious or 'nearly neutral' (Hartl & Clark, 2007), and populations facing rapid environmental change over the course of just a few generations are therefore unlikely to evolve a beneficial mutation through *de novo* mutation in sufficient time to adapt (Figure 1.1).

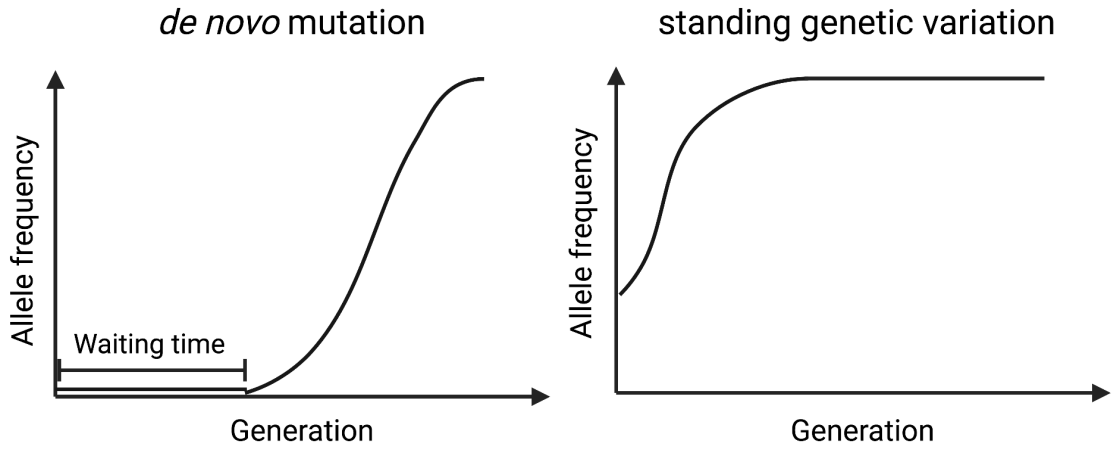


Figure 1.1: Adaptation from *de novo* mutation versus standing genetic variation. It might take some time until a beneficial *de novo* mutation occurs in a population and rises to high frequencies. In contrast, standing genetic variation is already available in the source population and can immediately rise to high allele frequencies in a population. Figure reproduced based on Chevalier *et al.*, 2019.

In contrast, the "ready-availability" of SGV, provides considerably richer wealth of mutations. For instance, the pairwise nucleotide diversity in humans is ~0.1% or 2 to 3 million base pair differences between any two human genomes (Jorde & Wooding, 2004; Schneider *et al.*, 2003). Even if only a small fraction of these variants are slightly beneficial to individual fitness and survival, they are instantly available genetic variants upon which natural selection can act in the course of rapid adaptation (Figure 1.1, and see 1.2 for further discussion). Even though SGV seems to play an important role for rapid adaptation, it is largely understudied primarily due to a lack of knowledge of which mutations and loci underlie adaptation, and the challenges of sampling large numbers of individuals to detect rare, but adaptively significant alleles. This thesis is motivated by the need to fill this gap. By leveraging the power of the threespine stickleback fish system, which provides a high-resolution map of adaptive loci and

the availability of large sample sizes, this thesis offers an insight into the availability of SGV, into the mechanisms and constraints of rapid adaptation and into the evolutionary potential of biodiversity.

1.1. Hard sweeps, soft sweeps and polygenic adaptation

The two fields of population genetics and quantitative genetics developed greatly disparate approaches and viewpoints on the genetic basis of phenotypic variation: a “one locus for one trait” versus an “infinitely many loci for one trait” model (Csilléry *et al.*, 2018). Population geneticists used approaches based on the classical hitch-hiking model introduced by Maynard-Smith and Haugh to detect genomic signatures of selection, known as hard and soft selective sweeps (Barton, 1998; Hermisson & Pennings, 2017; Kaplan *et al.*, 1989; Maynard-Smith & Haigh, 1974). A hard sweep refers to the selective advantage of one particular haplotype over all others in the population leading to a large increase in allele frequency for this beneficial haplotype. The hitch-hiking model describes that the rapid fixation of a haplotype, e.g. a beneficial mutation along with linked variants, results in detectable molecular signatures of hard sweeps in the genome (Hermisson & Pennings, 2005; Maynard-Smith & Haigh, 1974). The selective advantage of several different haplotypes concurrently relative to other haplotypes present in the population is referred to as soft sweep. Soft sweeps can originate from SGV and lead to “subtle” shifts in allele frequency. Thus, soft sweeps normally show weaker signals of selection than hard sweeps and are harder to detect (Pennings & Hermisson, 2006; Prezeworski *et al.*, 2005; Teshima *et al.*, 2006). Numerous loci across diverse organism have been found to show remarkable sweep signals fitting the predictions of a hard sweep (Lamason *et al.*, 2005, 2005; Williamson *et al.*, 2007). However, empirical (Colosimo *et al.*, 2005; Hamblin & Di Rienzo, 2000; Jeong *et al.*, 2008; Scheinfeldt *et al.*, 2009) as well as theoretical (Hermisson & Pennings, 2005; Innan & Kim, 2004; Pennings & Hermisson, 2006; Prezeworski *et al.*, 2005) studies emphasised the potential importance of “soft sweeps”. With the beginning of the genomic era, a large number of genome-wide association studies (GWAS) have been performed revealing that sweeps in general are not as common as expected by population genetics and that mechanisms of adaptation are more complicated than previously assumed (Visscher *et al.*, 2017).

In contrast, quantitative genetics assumed that most traits are highly polygenic and focused on phenotypic differences (Barton *et al.*, 2017; Barton & Keightley, 2002; Falconer & Mackay, 1996; Fisher, 1918). Quantitative trait locus (QTL) mapping studies as well as genome-wide association studies (GWAS), classical breeding and selection experiments demonstrate that many traits are polygenic (Castro *et al.*, 2019; Johnston *et al.*, 2011; Laporte *et al.*, 2015; Peichel & Marques, 2017; Yuan *et al.*, 2018), and that adaptation proceeds via

selection acting upon SGV at many loci in parallel (Coop *et al.*, 2009; Jones *et al.*, 2012). Moreover, most adaptation events seem to happen via small allele frequency shifts at multiple loci without the complete fixation of a single beneficial haplotype (Pritchard *et al.*, 2010; Pritchard & Di Rienzo, 2010). The interaction of adaptive alleles through metabolic and signalling pathways can have an additive effect on phenotype and the response to selection (Carter *et al.*, 2005). Furthermore, some studies argue that epistatic interaction between polygenic loci plays a meaningful role in adaptation (Hansen, 2013; D. B. Stern *et al.*, 2022). Molecular signatures of polygenic adaptation in the genome are challenging to detect with population genetics' approaches and, in the absence of independent evolutionary replicates (provided e.g. by parallel adaptation), might be indistinguishable from signatures resulting from random genetic drift (Berg *et al.*, 2019; Sohail *et al.*, 2019). Importantly, polygenic adaptation might allow a population to adapt rapidly to a new environment based on SGV already present in the founder population (Pritchard *et al.*, 2010).

While single loci of large phenotypic effect have been found to play important roles underlying adaptive traits in some organisms (Colosimo *et al.*, 2005; Chan *et al.*, 2010; Hoekstra *et al.*, 2006), it is likely they represent the "easy to study" tail of the distribution, and many more small effect loci contribute to adaptation and collectively play a significant role in adaptive evolution. In threespine sticklebacks, for instance, independent evolutionary replicates of parallel adaptation from marine to freshwater environments offered the opportunity to identify 242 divergent loci representing the polygenic basis of parallel adaptation (Jones *et al.*, 2012, FDR=0.05). The availability of this polygenic basis as SGV in founder populations as well as its role in the adaptive process will be explored in this thesis.

1.2. Rapid adaptation from standing genetic variation

Rare, neutral or even deleterious variants can be harboured in populations for millions of generations as pre-existing, or so-called "standing genetic variation" (SGV) and can then be used for rapid adaptation to changing environmental conditions. Research in the past two decades has started to elucidate the important role of SGV in rapid adaptation to novel environments and why SGV might outperform *de novo* mutations (Barrett *et al.*, 2008).

1.2.1. Factors affecting adaptation via SGV

SGV may outperform *de novo* mutations due to three primary criteria (Barrett & Schluter, 2008; Reid *et al.*, 2021). Pre-existing variation is, firstly, more likely to rise to high frequencies, secondly, will do so quicker and, thirdly, the adaptive haplotype has potentially been pre-screened and optimised (Reid *et al.*, 2021).

Probability of fixation

A newly arising *de novo* mutation is normally present as a single copy in the whole population and can easily be lost by genetic drift. SGV, however, is already present in multiple copies resulting in a higher probability of fixation (frequency=1) within the population (Hermisson & Pennings, 2005). Independent of the number of copies present at the beginning, fixation of SGV becomes more likely with increasing beneficial fitness effect size and with a higher effective population size, as large populations experience smaller stochastic effects. For a lot of selective effect sizes, the corresponding allele is likely to fix when originating from SGV, whereas the fixation probability for the same allele starting as a *de novo* mutation is negligible (Hermisson & Pennings, 2005). In particular, small effect mutations are much more likely to fix from SGV and thus fixed small effect mutations should originate to a greater extent from SGV than from *de novo* mutations (Barrett & Schluter, 2008).

Speed of adaptation

Adaptation via SGV can accelerate adaptation, since alleles beneficial in the new environment are already present and can immediately act as the heritable substrate for natural selection. The waiting time for advantageous newly arising mutations is thereby rendered unnecessary (Reid *et al.*, 2021). Moreover, the average fixation time is shortened for SGV due to the higher starting frequency of beneficial alleles (Hermisson & Pennings, 2005; Figure 1.1). Therefore, rapid adaptation should be dominated by alleles originating from SGV in most scenarios (Barrett & Schluter, 2008).

Optimised haplotypes

SGV can already be present for millions of years and has presumably been tested by natural selection in an analogue environment. Besides, adaptive haplotypes can be refined over time by accumulating multiple beneficial mutations with small fitness effects and resulting in haplotypes with large fitness effects in total (McGregor *et al.*, 2007; Reid *et al.*, 2021). Pre-tested SGV alleles with large fitness effects are more likely to help gain the fitness optimum without overreaching compared to *de novo* mutations (Reid *et al.*, 2021).

1.2.2. Examples of rapid adaptation from standing genetic variation

While theory predicts that adaptation from SGV is common, there are relatively few empirical examples where SGV has been shown to play a major role. One challenge is to detect whether adaptive alleles are novel or whether they origin from pre-existing genetic variation. One way to overcome this challenge is to study adaptive alleles involved in parallel adaptation.

As an example, it has been shown that the loss of armour plates in freshwater sticklebacks has globally evolved via re-use of the same freshwater haplotype at the

Ectodysplasin locus present as SGV in marine source populations (Colosimo *et al.*, 2005) (discussed further in chapter 1.4.2). Similarly, examples of SGV have been demonstrated based on parallel adaptation in mammals. As an example, old-field mice (*Peromyscus polionotus*) have adapted in parallel to white-sand beaches in the south-eastern United States exhibiting a colour shift from dark brown dorsal coats in mainland mice to pale coats in beach mice (Wooldridge *et al.*, 2022; Domingues *et al.*, 2012). The light fur serves as camouflage from predators for the mice living at sandy beaches and has evolved independently in Florida's Gulf and Atlantic coasts less than 3 000 years ago (Wooldridge *et al.*, 2022; Domingues *et al.*, 2012). One allele strongly associated with the colour variation is an enhancer of the Agouti signalling protein and is fixed in both beach mice lineages (Wooldridge *et al.*, 2022). The adaptive allele has most likely been introduced by the source population during independent colonisation events to both habitats facilitating rapid adaptation via the re-use of pre-existing variation (Wooldridge *et al.*, 2022).

Biological replicates available due to parallel evolution can also help to identify the polygenic basis of adaptation which is often present as SGV in the ancestral population and re-used in the derived population. In sticklebacks, for instance, 242 regions were identified to be consistently divergent between global freshwater and marine sticklebacks presumably representing the global polygenic basis of stickleback adaptation (Jones *et al.*, 2012; discussed further in chapter 1.4.2). These loci are most likely present as SGV in the marine population and thus spread across the Northern Hemisphere. Studying a marine-freshwater hybrid zone in River Tyne, Scotland, revealed further that 35% of the top divergent regions in this hybrid zone were globally shared, whereas the rest probably corresponds to recent or locally arising variants (Jones *et al.*, 2012). These variants could locally be available as SGV or could represent adaptive variants arising from *de novo* mutations.

Similar to sticklebacks, the polygenic basis of adaptation was identified in other parallelly evolving species. For instance, 151 genes were identified as genomic basis of parallel phenotypic differentiation following independent alpine colonisations of *Arabidopsis arenosa* and *Arabidopsis halleri* based on seven ecotype pairs (Bohutínská *et al.*, 2021). However, even though the genomic basis is known in multiple organisms, it is often to be uncovered whether adaptation occurs via re-use of SGV or via newly arising mutations.

Cichlid fish have undergone a recent rapid adaptive radiation in the African Great Lakes resulting in plenty of new species which persisted in sympatry in the lakes over the last 15 000 years and offering a great platform to study adaptation from SGV (Brawand *et al.*, 2014; McGee *et al.*, 2020; Witte & van Oijen, 1990). McGee *et al.* argued that this divergence relies crucially on the remarkable genetic variation in the genomes of cichlids (Brawand *et al.*, 2014; McGee *et al.*, 2020). Large-scale comparative methods revealed that 33% of the indel polymorphisms in Lake Victoria are also segregating in at least one outgroup taxa suggesting that these indels

have already existed for up to 10 million years, were most likely present as SGV in ancient populations and were recombined in numerous ways in order to create the recent radiation (McGee *et al.*, 2020). Without the pre-existing genomic variation and depending solely on the occurrence of new mutations, this radiation would potentially never have happened in this short period of time (McGee *et al.*, 2020).

Based on empirical examples, SGV appears to be important for adaptation in species with high levels of population interconnectivity and gene flow across disparate microhabitats and for species that experienced rapid local environmental change or substructure as shown in cichlid fish.

1.2.3. Maintenance of standing genetic variation

SGV can be ancient and up to millions of years old (Colosimo *et al.*, 2005; Louis *et al.*, 2021), which indicates that it has been maintained for a long time. However, the mechanisms for this maintenance have been little explored. In fact, it has been a long-standing goal in population genetics to understand the interaction of mutation, selection, migration, recombination and genetic drift and how they manage to maintain SGV in wild populations (Connallon & Clark, 2014). Theories explaining the maintenance of SGV in a population include migration-selection balance, balancing selection and cryptic genetic variation.

Migration-selection balance

SGV can be maintained in a population through a balance between migration and selection. While geneflow on its own can reduce genetic differentiation between populations and lead to a homogenous population, selection can counteract and stabilize the differentiation between populations (Barton & Hewitt, 1985; Ronce & Kirkpatrick, 2001). A migration-selection balance can thus maintain SGV at low frequencies.

Individuals migrating into the ancestral population carry their set of alleles in complete linkage disequilibrium (LD). The allele ensemble gets broken down by recombination during successive generations of backcrosses between the migrants and individuals from the ancestral population (introgression). Hence, descendants carry only parts of the initial migrant haplotypes. The migrant allele ensemble is presumably maladaptive in the new environment and the deleterious alleles present a genetic load in the population. The selective pressure on an allele is higher with other alleles under selection linked to it. Under the assumption that the migrant alleles are deleterious, selection will thus strongly operate against first generation hybrids which are carrying alleles with high LD (Harrison, 1993). Subsequent backcrosses, however, will experience weaker selective pressure. Scattered on several individuals, migrant alleles can then be harboured and maintained as SGV in the ancestral population.

Balancing selection

In contrast to the classical view which describes the removal of genetic variation by purifying selection, the balancing selection view means that variation itself is adaptive within a population and this variation is maintained by selection favouring a mix of alleles (Fijarczyk & Babik, 2015). Balancing selection summarises different mechanisms such as frequency-dependent selection, heterozygote advantage by overdominance, selection varying over time or space, or sexually antagonistic selection (Bernatchez, 2016). The mode under which an allele has higher fitness at lower frequencies is called negative frequency-dependent selection. Overdominance describes when heterozygotes show higher fitness than either homozygote and are thus present at higher frequencies in populations. The term sexually antagonistic selection specifies when an allele is advantageous to one sex, but detrimental to the other. Evidence for this mechanism has been provided in several wild populations e.g. sexually antagonistic selection for horn size in Soay sheep (Foerster *et al.*, 2007; Johnston *et al.*, 2013; Rice, 1992). The importance of balancing selection for the maintenance of SGV is still unclear.

Charlesworth states that long-term balancing selection is unlikely (Charlesworth, 2006) and Asthana *et al.* substantiates that ancestral genetic variation in chimpanzees is low or absent and that balancing selection plays an inferior role for long-term maintenance of SGV (Asthana *et al.*, 2005). With the new genomic era and the development of more potent methods of analysis for identifying molecular signatures of balancing selection, an increasing number of studies is recently exhibiting the effect of long-term balancing selection (Andrés *et al.*, 2009; Bitarello *et al.*, 2018; Leffler *et al.*, 2013; Liu *et al.*, 2021; Siewert & Voight, 2017). For instance, long-term balancing selection seems to affect a considerable part of the human genome including immunity genes and genes encoding keratins and membrane channels (Andrés *et al.*, 2009; Bitarello *et al.*, 2018). These recent studies implicate that long-term balancing selection within a population can contribute to the maintenance of SGV and potentially has been incorrectly neglected.

Cryptic genetic variation

Another aspect affecting the maintenance of SGV, is described by so-called cryptic genetic variation (CGV). CGV describes alleles that are not affecting the phenotype under certain environmental and genetic conditions and are thus masked from selection. When a new allele is introduced e.g. via introgression or the environment is changing, previously "hidden" variation might aid to produce a novel phenotype (Barrett & Schluter, 2008; Gibson & Dworkin, 2004). For instance, parallel adaptation of dark-coated oldfield mice (*Peromyscus polionotus*) to sand dunes of Florida's Gulf Coast has revealed how the effects of an allele can be hidden in the ancestral environment. The derived beach mice populations exhibit much lighter fur supposedly as an effect of selection for camouflage in the sandy environment

(Steiner *et al.*, 2007). Most of the divergent fur colour in oldfield mice and beach populations is controlled by two candidate genes. The genes include melanocortin-1 receptor (*Mc1r*) and the Agouti signalling protein (Agouti) acting as the antagonist of *Mc1r* (Hoekstra *et al.*, 2006; Steiner *et al.*, 2007). Mice which are homozygous for the ancestral Agouti allele always display dark fur independent of their *Mc1r* genotype. Hence, the strong epistatic interaction between the two loci can mask the effects of *Mc1r* allele, facilitating the maintenance of SGV at this locus in the ancestral mainland population. When the mainland mice colonise the beach habitat, selection will act upon the Agouti allele and the derived, light pigment Agouti allele will rise to high frequencies. Subsequently, the effects of the *Mc1r* alleles will be unveiled and the light pigment allele will become subject of selection (Barrett & Schluter, 2008). In summary, a part of SGV might be cryptic and easily maintained in the ancestral population.

1.3. Adaptation from *de novo* mutations

In spite of the increasing evidence for the importance of SGV in rapid adaptation, some studies show that *de novo* mutations can also play a crucial role (Bomblies & Peichel, 2022). This is surprising since *de novo* germline mutation rates are estimated to be very low, e.g. 10^{-8} per base pair per generation in humans (Durbin *et al.*, 2010), and the effect of a randomly arising mutation is expected to be slightly deleterious for an organism in the majority of the cases (Hartl & Clark, 2007). Therefore, a population has to wait for a beneficial *de novo* mutation to arise and to rise to high frequencies in the population (Figure 1.1). This time delay is highly disadvantageous if the population is challenged to adapt rapidly to a new or quickly changing environment. Nevertheless, repeatedly and independently arising *de novo* mutations were detected across taxa (Chan *et al.*, 2010; Cockram *et al.*, 2007; Lenski, 2017). A classic example includes the *de novo* deletions near the *Pitx1* locus which happened independently in freshwater stickleback populations and repeatedly led to the loss of pelvic spines in freshwater populations (Chan *et al.*, 2010). This example is highly relevant for this thesis and will be discussed in more detail in chapter 1.4.2. Repeated and independent small *de novo* deletions have also been reported at the vernalization 1 locus of plants (Cockram *et al.*, 2007). The deletions were often immediately flanked by short repeated sequences whose base compositions displayed motifs which were significantly associated with deletion breakpoints (Cockram *et al.*, 2007). These results suggest that the features of a DNA sequence, e.g. increased DNA fragility, might increase the likelihood of structural changes and might thus facilitate the repeated and independent occurrence of novel mutations at a locus (Chan *et al.*, 2010; Xie *et al.*, 2019). Although the parallel adaptation from *de novo* mutations has been shown to happen in nature, its importance and prevalence compared to adaptation via SGV is still largely understudied.

1.4. Threespine sticklebacks – an excellent model system for studies of rapid adaptation

Here we chose the threespine stickleback fish (*Gasterosteus aculeatus*), a premier model organism to study evolutionary biology, to investigate factors affecting rapid adaptation such as the availability of SGV.

1.4.1. Stickleback background and ecotypes

Threespine sticklebacks show several attributes which make it an appealing model organism. The teleost fish displays a small body size, high fecundity and a generation time of one to three years in the wild or of six months in the laboratory (Bell & Foster, 1994). Sticklebacks are further highly abundant and widely spread across the Northern Hemisphere occupying diverse marine as well as freshwater habitats (Bell, 1984; Bell & Foster, 1994). Threespine sticklebacks are ancestrally a marine anadromous species and have repeatedly colonised freshwater environments for at least 10 million years (Bell & Foster, 1994). The fish displays two major types of life history: marine anadromous (hereafter called marine) sticklebacks are present in the oceanic environment and move upstream rivers for spawning showing migratory behaviour, whereas freshwater sticklebacks are resident and show an absence of migratory behaviour (Moyle, 1976; Wootton, 1984). At the end of the Last Glacial Maximum (~10 000 - 20 000 years ago) with the retreat of the Pleistocene ice-sheet, diverse freshwater habitats formed and were colonised by marine anadromous populations leading to a burst of rapid diversification in sticklebacks (Bell & Foster, 1994). This evolutionary process continues today as new freshwater habitats emerge from geological events such as glacial retreat and earthquakes (Lescak *et al.*, 2015). As a result, in the last 10 000 - 20 000 years freshwater sticklebacks have evolved repeatedly and independently from marine sticklebacks all over the Northern Hemisphere, providing powerful biological replicates of the evolutionary rapid adaptation process and rendering these fish a potent organism to study parallel adaptation (Bassham et al., 2018; Bell et al., 2004; Lescak et al., 2015; Taylor & McPhail, 1999; Terekhanova et al., 2014; Withler & McPhail, 1985).

Stickleback populations exposed to different environmental conditions evolve differences in phenotype, including morphology, physiology and behaviour, and are also referred to as ‘ecotypes’ (Figure 1.2; Hendry *et al.*, 2009; McKinnon & Rundle, 2002; McPhail, 1993). Regarding morphology, for instance, marine sticklebacks show a large body size, silver coloration, a pelvic spine and a complete set of armour plates, whereas freshwater sticklebacks are smaller and exhibit a green-brown dorsal coloration. Freshwater sticklebacks have a reduced pelvic spine or have lost it completely, and show a reduced number of armour plates. The loss and reduction of bony defensive structures in freshwater sticklebacks is

thought to be related to the different predation pressure and/or a lower ionic content of the water resulting in higher cost of mineralizing bones (Bell *et al.*, 1993; Giles, 1983; Marchinko, 2009; Vamosi & Schluter, 2004). Multiple behavioural traits differ between marine and freshwater sticklebacks including migratory behaviour, aggressiveness, male courtship and mating choice (Di-Poi *et al.*, 2014; McPhail & Hay, 1983; Moyle, 1976; Vines & Schluter, 2006; Wootton, 1984).

Figure 1.2: Stickleback marine-freshwater ecotype pair from Little Campbell River, Canada. Example images of a stickleback ecotype pair from Little Campbell River. Freshwater and marine stickleback ecotypes differ in various phenotypic traits such as body size, armour plates and presence of the pelvic spine. Anadromous marine sticklebacks are sampled about 0.5 km upstream, whereas freshwater sticklebacks are sampled about 23 km upstream. Figure courtesy: Stanley Neufeld.

Even though marine and freshwater ecotypes differ tremendously in various traits, they can still hybridise and produce viable offspring. This can happen in the wild, when marine fish swim up rivers to spawn in freshwater and encounter breeding freshwater sticklebacks. In fact, F1 hybrids have been reported to be common in hybrid zones (Hagen, 1967; Jones *et al.*, 2006; Vines *et al.*, 2016). However, habitat preference and mate choice act as prezygotic barriers for hybridization (Lackey & Boughman, 2017). Besides, selection against hybrid zygotes or fry could act as postzygotic barrier and survival of hybrids could further be impeded by selection over winter, competition for food and predation pressures (Hagen, 1967; Jones *et al.*, 2006; Rundle, 2002; Rundle *et al.*, 2003; Vines *et al.*, 2016). Homogenisation by gene flow is thus prevented and the integrity of ecotypes remains maintained (Ravinet *et al.*, 2015).

1.4.2. Stickleback genomics

In the past decade, threespine sticklebacks have developed into a powerful model system for evolutionary genomics with many genetic as well as genomic tools available for their study (Jones et al., 2012; Kingsley et al., 2004).

High quality reference genome

A high-quality reference genome encompassing 463 Mb is available for sticklebacks, adding to the numerous attributes which render this teleost fish a premier model organism. The repeat and duplication content of the stickleback genome is relatively low which facilitated the generation of a highly contiguous anchored genome assembly. The first draft of a stickleback genome called gasAcu1 was published by Jones *et al.* in 2012 (Jones *et al.*, 2012). It was assembled based on paired-end Sanger sequencing of multiple genomic libraries of a single stickleback female from Bear Paw Lake, Alaska. 86.5% (400.4 Mb) of the genome were assembled in 21 chromosome-level scaffolds with the remaining 13.5% (60.7 Mb) of the genome being represented by unplaced scaffolds. The first genome assembly for sticklebacks already comprised larger contig and scaffold sizes than other published teleost genome assemblies (Aparicio *et al.*, 2002; Jaillon *et al.*, 2004; Kasahara *et al.*, 2007; Star *et al.*, 2011). Subsequent revisions corrected some scaffold orientations and incorporated in total 10.7% more genome content by assigning previously unanchored scaffolds to chromosomes leaving only 1.8% (8.4 Mb) of the genome unanchored (Glazer *et al.*, 2015; Peichel *et al.*, 2017; Roesti *et al.*, 2013). These improvements as well as two minor corrections including the mitochondrial genome and a previously missing region on chrVII were included in a new genome assembly called gasAcu1-4 as introduced in Appendix 8.2 (Kingman *et al.*, 2021). This newly available, revised genome assembly was used for the analysis presented in chapters 5.2 and 5.3.

Adaptation from standing genetic variation

The parallel evolution of reduced lateral armour plate phenotypes in freshwater sticklebacks independent of their geographic location was the first evidence of a common genetic basis for repeated freshwater adaptation. The underlying genetic basis of this trait is an ancient haplotype at the *Ectodysplasin* (*Eda*) locus that diverged from the high plated haplotype more than 2 million years ago (Colosimo *et al.*, 2005; Colosimo *et al.*, 2004; Cresko *et al.*, 2004). Miller *et al.* showed a similar phylogenetic pattern near the Kit ligand (*Kitlg*), a gene responsible for lighter skin pigmentation in freshwater populations (Miller *et al.*, 2007). This parallel sequence divergence suggests that the *Kitlg* gene was also re-used for adaptation of pigmentation in freshwater habitats. Both freshwater adaptive alleles were presumably present at low frequencies in the marine population colonising the freshwater habitats resulting in parallel evolution of freshwater ecotypes via SGV, rather than independently accumulating the same *de novo* mutations in parallel in each new freshwater population (Colosimo *et al.*, 2005). Indeed, the freshwater *Eda* allele was found at a frequency of 3.8% in 109 marine sticklebacks at the outlet of the Navarro River in California (Colosimo *et al.*, 2005), whereas the derived *Kitlg* allele was found at a frequency of 12% in 107 marine sticklebacks from the same site supporting this hypothesis (Miller *et al.*, 2007). These first studies on *Eda* and *Kitlg*

highlight that adaptation via the re-use of pre-existing genetic variation is a crucial mechanism underlying rapid adaptation in sticklebacks. However, both loci were identified as highly significant quantitative trait loci (QTLs), exhibit large effect sizes and most likely only represent the easy detectable tip of the iceberg.

The most parsimonious way for sticklebacks to adapt repeatedly and in parallel to freshwater environments would be to use one single major locus controlling all differing traits. However, as outlined in chapter 1.1, quantitative genetics theory assumed that most traits are not controlled by a single locus, but rather highly polygenic. Numerous GWAS studies have additionally revealed that many traits are polygenic and multiple studies have proposed that adaptation happens via selection acting upon many loci in parallel with mechanisms of adaptation being more complex than expected under parsimony (Castro *et al.*, 2019; Coop *et al.*, 2009; Johnston *et al.*, 2011; Jones *et al.*, 2012; Laporte *et al.*, 2015; Peichel & Marques, 2017; Visscher *et al.*, 2017; Yuan *et al.*, 2018). The genomic basis of adaptation in sticklebacks was hence also expected to be polygenic including loci with a range of effect sizes and presumably present as SGV in the founder population. The studies on *Eda* and *Kitlg* are largely biased by their large effect sizes and the loci are most likely not representative for other adaptive variants. In order to explore the role of SGV during stickleback adaptation, it was thus fundamental to identify the polygenic basis of adaptation.

242 freshwater-marine divergent regions associated with parallel adaptation (Jones et al., 2012)

Jones *et al.* published the first study looking at the genomic basis of freshwater adaptation using low-coverage whole genome sequencing data of ten geographic pairs of freshwater and marine sticklebacks from both Atlantic and Pacific origin. The study identified a polygenic and dispersed architecture revealing that a fish needs many variants across their genome in order to successfully adapt to a freshwater habitat. Genetic regions consistently associated with parallel evolution of distinct marine and freshwater ecotypes (hereafter called freshwater-marine divergent regions) were identified with two different methods. Firstly, a Hidden Markov Model (SOM/HMM) was used to identify genetic relationships among individuals and secondly, a genetic distance approach was applied by calculating cluster separation scores (CSS). CSS as well as the corresponding p-values obtained by permutation tests are continuous measures. Specific genomic coordinates for divergent regions were determined by introducing a p-value cut-off based on a false discovery rate (FDR) of 0.05. SOM/HMM and CSS determined together a total of 242 loci (0.5% of the genome) displaying global freshwater-marine divergence. 147 such loci (0.2% of the genome) were identified by both approaches (Figure 1.3). Identified freshwater-marine divergent regions represent adaptive variants which have been repeatedly used across global freshwater populations. The

loci themselves harbour strongly divergent, presumably ancient haplotypes with up to 4% sequence divergence. This pattern of highly divergent haplotype blocks ranges in size from 175 bp to 130 kb, includes chromosome inversions and is accumulated on a few chromosomes. For instance, 62 out of all 242 detected divergent loci are located on chromosome IV including the well-studied *Eda* locus as well as the developmental signalling gene *Wnt7b* (Jones *et al.*, 2012).

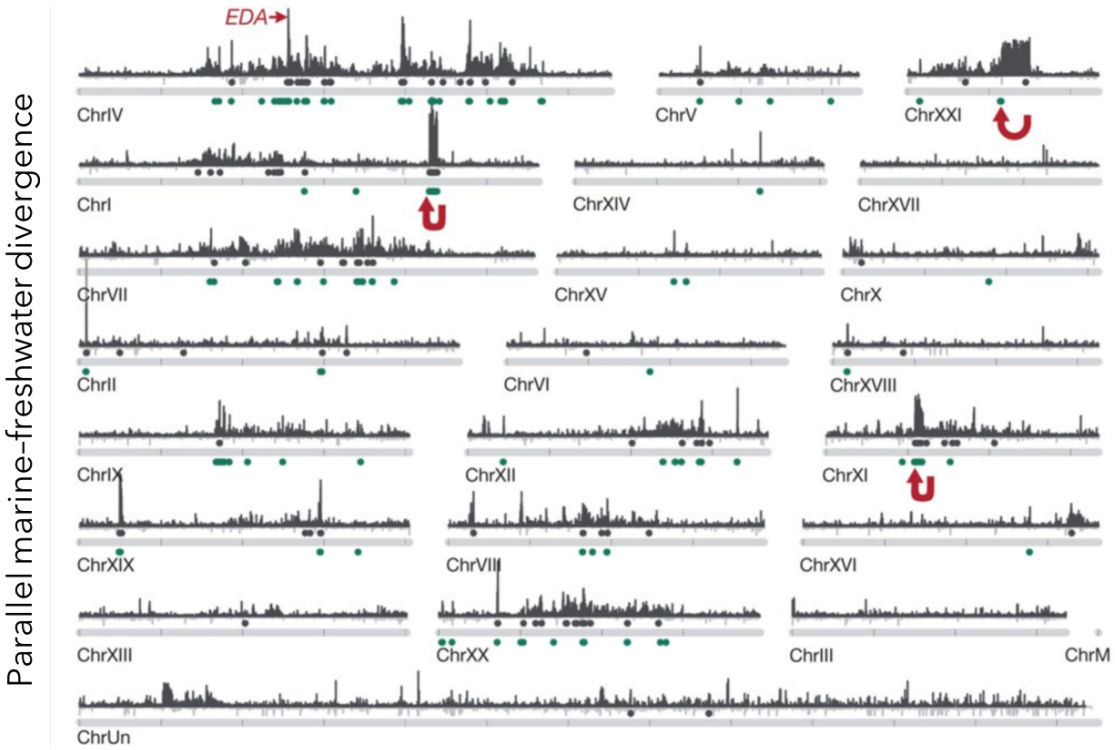


Figure 1.3: Genome-wide distribution of freshwater-marine divergent loci. Marine-freshwater divergent loci were identified based on CSS and SOM/HMM analysis. Divergent loci detected by CSS (FDR <0.05) are displayed as grey peaks with grey points above chromosomes and those detected by SOM/HMM as green points below chromosomes (Jones *et al.*, 2012). The red arrows on chromosome I, XI and XXI indicate inversions. Unanchored scaffolds are represented as "ChrUn". Adapted from Jones *et al.* 2012 and displayed in gasAcu1 coordinates. Adapted from Jones *et al.* (2012).

The "flanking" regions of identified divergent regions show neutral patterns of limited parallel divergence at a global level. A recently published study based on genome-wide SNP data of an expanded sample size of 166 sticklebacks from the Northern Hemisphere revealed that the proportion of the genome associated with global parallel marine-freshwater differentiation (0.2%) is seven to ten times lower than solely in Eastern Pacific populations (2.1%) (Fang *et al.*, 2020). Forward-in-time simulations suggest a stochastic loss of freshwater-adaptive alleles due to founder events when marine sticklebacks from the Eastern Pacific populations colonised the rest of the Northern Hemisphere before the closing of the Bering Strait (38-40 ka) (Fang *et al.*, 2020). Eastern Pacific sticklebacks might thus have a greater capability of adapting repeatedly and rapidly to new freshwater environments than elsewhere

(Klepaker, 1993). However, the divergent loci identified only in Pacific sticklebacks might also be neutral variants hitchhiking when near-by adaptive variants were rising to high frequencies during adaptation and might not be relevant for stickleback adaptation in the rest of the world (Maynard-Smith & Haigh, 1974). Being relevant for stickleback adaptation across the globe, the 242 freshwater-marine divergent regions representing the polygenic basis of global parallel adaptation in sticklebacks were used and studied further in this thesis.

A gene ontology analysis on all identified divergent regions showed significant enrichment of genes involved in cellular responses to signals, behavioural interactions between organisms, amine and fatty acid metabolism, cell-cell junctions as well as WNT signalling. Studying the 64 most significantly divergent regions revealed that only 17% of them are protein-coding sequences with ecotype-specific amino acid substitutions. 41% of the regions mapped entirely to non-coding regions indicating regulatory changes. The remaining 43% contained both non-coding and coding sequences, but lacked non-synonymous mutations consistently differing between ecotypes. The latter probably also represents regulatory alterations. Overall, Jones *et al.* shows that coding as well as regulatory differences play a role in parallel stickleback evolution with regulatory changes appearing in the majority of the divergent regions (Jones *et al.*, 2012).

Protein coding regions and the impact of coding changes on protein structure have been studied elaborately in the past across taxa and a big part of the genomics community focussed on exome sequencing (Karczewski *et al.*, 2017). However, Jones *et al.*'s study in sticklebacks and a few years later Grossman *et al.*'s study in humans revealed that the non-coding parts of the genome matter for fitness and survival in natural populations and furthermore that signatures of selection are predominantly found within these parts of the genome (Grossman *et al.*, 2013; Jones *et al.*, 2012). Sean Carroll previously outlined that regulatory DNA and especially *cis*-regulatory elements offer some features which make them more tuneable than other regions and thus important targets for adaptation (Carroll, 2000). Firstly, the organisation of *cis*-regulatory systems is modular enabling single *cis*-elements to act and evolve independently (Arnone & Davidson, 1997). *Cis*-regulatory regions of developmental regulatory genes encompass many elements suggesting that these systems have expanded and diversified playing an important role in evolution (Carroll, 2000). Secondly, regulatory DNA is more tolerant to mutational change than coding sequences (Carroll, 2000). Mutations with greater pleiotropic effects are expected to be more deleterious for organisms and are thus probably present at lower frequencies than mutations with less pleiotropic effects (Carroll, 2000; Stern, 2000). For instance, a mutation in the coding region of a transcription factor operating in several tissues could lead to altered expression of all genes regulated by this factor. A mutation in a single *cis*-regulatory element of a gene, however, results in altered gene expression in the domain affected by that element and is thus much more tuneable

(Carroll, 2000). Although, regulatory genes have repeatedly been proposed to be important for morphological evolution, Jones *et al.* provided evidence for polygenic adaptation in general depending mainly on noncoding and regulatory regions (Carroll, 2008; King & Wilson, 1975). The importance of noncoding regions for future research was hence reinforced.

Transporter hypothesis

As discussed in the two latter paragraphs, previous studies have outlined that the use of pre-existing SGV presumably plays a crucial role in the adaptation from sticklebacks to freshwater environments (Colosimo *et al.*, 2005; Miller *et al.*, 2007; Jones *et al.*, 2012; Terekhanova *et al.*, 2014). Colosimo *et al.* showed that the same freshwater haplotype was used for parallel adaptation to freshwater environments at geographically disparate locations across the Northern Hemisphere and proposes that freshwater-adaptive alleles are "transported" to different freshwater habitats by the marine population (Colosimo *et al.*, 2005; Figure 1.4). Alleles from freshwater populations have been shown to leak into marine populations via hybridization (Jones *et al.*, 2006; Vines *et al.*, 2016). The freshwater-adaptive alleles are subsequently hypothesised to be harboured in the marine population at low frequencies and later to be recycled in future adaptations facilitating rapid adaptation from sticklebacks to freshwater environments. This model was later termed "transporter hypothesis" by Schluter and Conte (Schluter & Conte, 2009). Galloway *et al.* implemented the transporter model *in silico* by individual-based, spatially explicit forward simulations. Realistic levels of gene flow in the model give results suiting hitherto existing empirical genomic studies in sticklebacks supporting the transporter model as an explanation for the adaptation process in sticklebacks. Surprisingly, most freshwater-adaptive alleles in freshwater populations originated from the original colonists of the site in contrast to subsequent migrants from the ocean (Galloway *et al.*, 2020).

Even though the "transporter hypothesis" provides a reasonable explanation for the spread of the same freshwater haplotypes throughout the Northern Hemisphere, it raises a lot of questions that remain to be answered. One emerging conundrum is whether freshwater-adaptive alleles present a genetic load in the marine environment. Since divergence between freshwater and marine sticklebacks is maintained despite ongoing geneflow (Jones *et al.*, 2006), marine alleles seem to present an advantage in the marine populations. Furthermore, this could also indicate that freshwater-adaptive alleles are deleterious in the marine environment. So far the selection pressures acting on freshwater-adaptive alleles in the marine environment have been unknown. The "transporter hypothesis" suggests that freshwater variants are harboured in the marine population at low frequencies. Does this mean that adaptive alleles are introduced by introgression, are deleterious in the marine environment and thus, simultaneously purged by selection? Hence, is SGV maintained in the marine population

by a migration-selection balance? An alternative explanation could be that freshwater-adaptive alleles are neutral in the marine environment – for example by a dominance modifier that changes the additive state of freshwater alleles to a recessive state when in the marine environment. This leads us to the question whether the advantage of carrying a marine allele in the marine environment is sufficient to prevent neutral, freshwater adaptive alleles to rise to high frequencies in the marine population. Haenel *et al.* points out that the maintenance of SGV cannot be solely explained by a long-term migration-selection balance and proposes selective neutrality as an additional mechanism (Haenel *et al.*, 2022). Hence, freshwater-adaptive alleles can persist at low frequencies in the marine population under neutrality (Haenel *et al.*, 2022). This study represents a first clue for solving this enigma, but further investigations are required to understand how SGV can be maintained in the marine population.

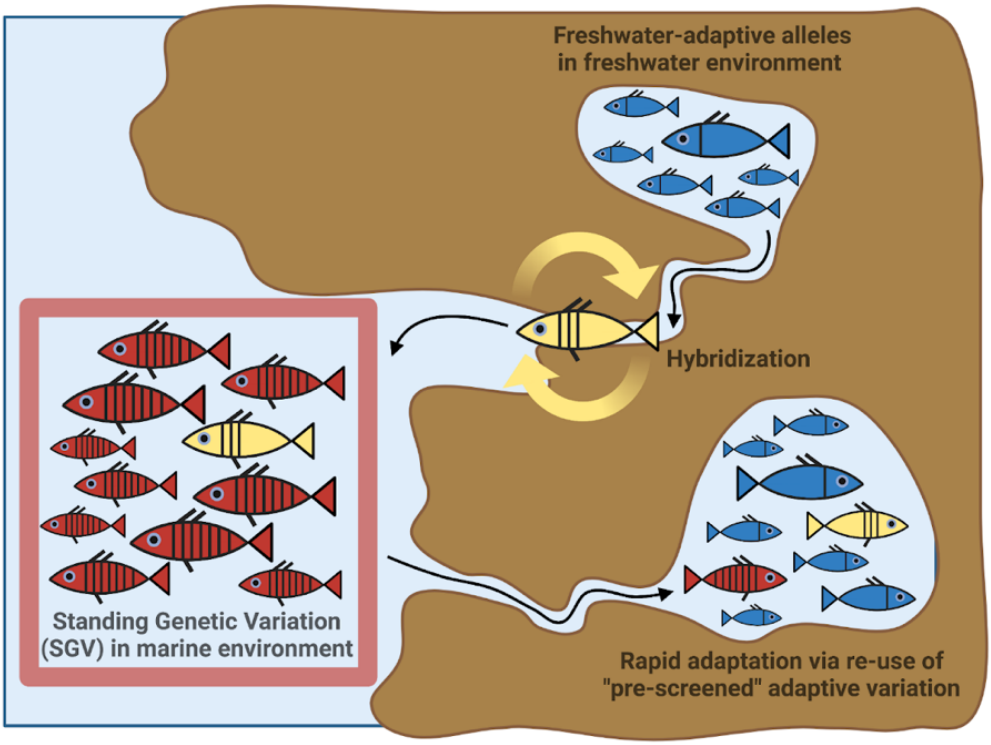


Figure 1.4: Transporter Hypothesis. Rapid adaptation in sticklebacks is hypothesised to be promoted by adaptive variation in the marine population. Freshwater-adaptive alleles can leak into the marine population via hybridisation of freshwater and marine sticklebacks. The adaptive alleles are present in the marine population at low frequencies as so-called standing genetic variation. When colonising a new freshwater pond, the available genetic variation can be utilised for rapid adaptation (Colosimo *et al.*, 2005; Schluter & Conte, 2009).

The “transporter hypothesis” states that freshwater adaptive alleles are present at low frequencies in the marine population. But in fact, the availability of SGV is still largely unknown. Apart from studies on a few loci with large effect sizes such as the *Eda* locus, the frequencies of freshwater adaptive alleles in the marine environment have not been quantified.

Moreover, it is unclear whether freshwater adaptive alleles are present due to recent introgression or persist for a longer time in the marine population. Recent introgression would lead to a few marine individuals carrying a huge part of the freshwater adaptive allele ensemble and hence displaying strong linkage among adaptive alleles. In contrast, freshwater adaptive alleles could also be dispersed in the marine population with each individual carrying some freshwater adaptive alleles. The linkage among adaptive alleles present as SGV affects the process of re-assembly of adaptive alleles in a newly colonised freshwater habitat. Large freshwater adaptive allele blocks carried in linkage by a few colonisers could accelerate the re-assembly of the complete freshwater-adaptive allele ensemble, whereas the re-assembly of small adaptive allele blocks dispersed among a large number of individuals would take more generations. Only very few studies started to address this conundrum. Miller *et al.* studied 109 Californian individuals and showed that none of the carriers of a freshwater adaptive allele at *Eda* and *Kitlg* carried both adaptive alleles together suggesting that no F1 hybrids are present in the marine environment (Miller *et al.*, 2007). Besides, Hohenlohe *et al.* has detected modest linkage disequilibrium among adaptive alleles albeit in a very small number of Alaskan marine individuals. This hints that a few "jackpot carriers" could carry an excess of freshwater adaptive alleles (Hohenlohe *et al.*, 2012). This study is however confounded by small sample sizes of 34 individuals and by disregarded population structure with two different marine populations pooled for this analysis. As a result, so far, evidence supporting linkage among adaptive alleles is limited. In order to elucidate how rapid adaptation can be facilitated by SGV, further studies on the availability of SGV, including the frequency of freshwater adaptive alleles as well as the linkage disequilibrium among them, are required.

Lastly, the "transporter hypothesis" proposes that freshwater adaptive alleles are introduced by freshwater populations. Freshwater adaptive alleles have been shown to leak into the marine population in hybrid zones (Jones *et al.*, 2006; Vines *et al.*, 2016). With the same central freshwater *Eda* haplotype being re-used in freshwater populations all over the Northern Hemisphere, the question arises if this freshwater haplotypes (and others) were introduced by a single hybrid zone and then spread through interconnectivity between marine populations across the globe. Hence, is the SGV present in a marine population only originating from local hybrid zones or is it harbouring a compilation of freshwater haplotypes originating from geographical disparate sources. And which haplotypes are then important for adaptation – are local as well as global freshwater haplotypes rising to high frequencies in a newly colonised freshwater population?

All in all, many questions about SGV remain unanswered and will be addressed throughout this thesis.

Rapid adaptation to freshwater environments

When a marine stickleback population colonises a freshwater pond, the fish are exposed to new selective forces, including predators (Marchinko, 2009; Reimchen, 2000), nutrient availability (McIntyre & Flecker, 2010), and ionic concentration (Bell *et al.*, 1993). Selection pressures can be high, with selection coefficients up to 0.50 (Barrett *et al.*, 2008; Schluter *et al.*, 2021), one order of magnitude higher than those normally reported in humans (Bersaglieri *et al.*, 2004; Tishkoff *et al.*, 2001). Particularly in the northern Pacific, several studies report that marine sticklebacks manage to evolve the same freshwater ecotype within decades after invading the new habitat (Aguirre & Bell, 2012; Bassham et al., 2018; Bell & Aguirre, 2013; Lescak et al., 2015). These young freshwater populations differ from their known or presumably ancestral marine population across the same loci previously identified to be divergent in lakes colonised up to thousands of years ago (Bassham *et al.*, 2018; Lescak *et al.*, 2015; Terekhanova *et al.*, 2014). The allele frequency of *Eda* in a freshwater population in Loberg Lake rises e.g. from 0.48 in 1992 to 0.96 in 2010 after a re-colonisation by marine sticklebacks of the lake between 1983 and 1989 (Schluter *et al.*, 2021). These findings lead to the fascinating hypothesis that the majority of stickleback adaptation to freshwater occurs within the first decades (generations) after colonisation of the novel environment (Lescak *et al.*, 2015). Hence, the SGV available plays an important role in shaping this population's future evolutionary trajectory. Given the polygenic architecture of adaptive loci there are relatively few opportunities (i.e. generations) for recombination to break up linkage between beneficial and deleterious alleles and to combine beneficial alleles together on the same chromosome. Linkage disequilibrium among adaptive alleles can largely accelerate this process and will be estimated based on large sample sizes in this thesis (chapter 5.3). Finally, there may be some "relatively" unique biology that may further facilitate rapid polygenic adaptation in the context of a few generations: namely, unlike mammals, fish have clutch sizes up to two orders of magnitude larger providing a platform for much more meiotic recombination opportunities to shuffle allelic combinations and arrive at the correct set. Given the same amount of generation time to achieve adaptation, this biological feature further facilitates the ability of sticklebacks to adapt compared to humans or other mammals.

Adaptation from de novo mutations

Although SGV seems to be the predominant mechanism of rapid adaptation in threespine sticklebacks, alternative adaptation pathways might also play a relevant role in adaptation. For instance independent *de novo* deletions of the pelvic enhancer (*Pel*) upstream of the homeodomain transcription factor gene (*Pitx1*) resulted in the recurrent loss of the pelvic spine in freshwater stickleback populations (Chan *et al.*, 2010; Shapiro *et al.*, 2006; Xie *et al.*, 2019). The deletions are located in the telomeric region of chrVII which contains several

stretches of TG-repeats with the ability to form alternative DNA structures *in vitro*. The resulting DNA fragility can stimulate chromosome breakage and deletions of the full pelvic enhancer region (Xie *et al.*, 2019). The probability of a deletion to occur at the *Pel* region is 104 times higher than a point mutation (Xie *et al.*, 2019). The *Pel* region demonstrates that *de novo* independently arising mutations can facilitate repeated and rapid freshwater adaptation and provide an example of an alternative adaptation mechanism that is not based on SGV.

Not being part of the global marine-freshwater divergent set of loci detected by Jones *et al.* (Jones *et al.*, 2012) the question arises, why the deletion near Pitx1 has not been globally used for adaptation and why it has not evolved by SGV? Is gene flow from freshwater into marine populations at this locus inhibited by physical, pre- or postzygotic barriers? Or is the deletion presenting a genetic load in the marine environment and therefore, quickly purged from the population? These questions remain to be answered and require further investigations. Another conundrum is whether adaptation via *de novo* mutations happens rarely in parallel or whether it is a quite common mechanism with many more loci affected. Pitx1 could represent a relatively easy detectable large effect loci. More parallelly and independently arising *de novo* mutations could exhibit smaller effect sizes and could be challenging to identify. Thousands of other positions in stickleback genome show sequence features which are associated with DNA fragility in the *Pel* region (Xie *et al.*, 2019) and TG-repeats are enriched in other deletions occurring in freshwater adaptation in sticklebacks (Lowe *et al.*, 2018). These pattern suggest that the deletion near *Pitx1* might not be the only repeatedly arising *de novo* mutation in freshwater populations. However, a genomic survey for examining the general importance of this adaptation mechanism is difficult since sequencing through repetitive regions is challenging (Reid *et al.*, 2021). More research is needed to disentangle the abundance of this process further and to assess how important and prevalent it is compared to adaptation via SGV.

2. Research objectives

The overall aim of this study is to elucidate the mechanisms and constraints affecting the rate of adaptation to novel or changing environments. Rapid adaptation can happen via the arousal of novel mutations or via the re-use of pre-existing standing genetic variation (SGV) with the latter very likely outperforming novel mutations in most cases. SGV can operate as a substrate for rapid adaptive change and is expected to play an important role for rapid adaptation. Nevertheless, we have a relatively poor understanding of what SGV looks like in a natural population. For instance, the frequencies of adaptive alleles in the ancestral population, the linkage disequilibrium among them, their geographic origin as well as their age are still to be uncovered. These features of SGV can provide important insight into the future evolutionary potential of a population, with implications for conservation, agriculture management and environmental diseases. This thesis addresses these topics by applying genomics in three different approaches across ancient and present-day populations of threespine sticklebacks (*Gasterosteus aculeatus*).

Our inability to look into the past means population genetics has traditionally relied on models and inference based on contemporary populations. With new technologies in the field of paleogenomics, we can see what genomes in the past actually looked like. This window into the past offers the amazing opportunity to look at empirical SGV in ancient stickleback genomes as presented in chapter 5.1 (Kirch *et al.*, 2021). In this study, DNA was extracted from 11 000 - 13 000-year-old stickleback bones found in the sediments of lakes providing unique insight into the adaptive alleles carried across the genome by ancestral sticklebacks more than 10 000 years ago.

Evolution is traditionally viewed as taking thousands of generations (Darwin & Kebler, 1859). With environments changing rapidly as in the current anthropogenic era, the question arises whether organisms can adapt in time. Threespine sticklebacks demonstrate that organisms can adapt extremely rapidly to a new or changing environment with marine sticklebacks adapting to newly colonised freshwater environments within only 30 years corresponding to 30 generations. In the study presented in chapter 5.2, contemporary time-series data of three stickleback populations offered the possibility to track adaptation and allele frequency changes over time (Kingman *et al.*, 2021). The alleles showing significant changes in allele frequency only partially overlapped with global freshwater-marine divergent loci emphasising the importance to study locally available SGV. Subsequently, individual based genotyping of >1600 marine individuals from three local marine populations at >350 SNPs was conducted providing information about the availability of local SGV and about which alleles are carried together as haplotype. By enabling the rapid genotyping of thousands of fish at hundreds of variants across the genome this study identified SGV to be present at very low

frequency and with detectable LD among loci. However, given the very low frequencies detected and the relatively small number of SNPs targeted estimates of frequencies and LD among rare alleles were underpowered and formed a major motivation for the study of SGV in samples of marine fish an order of magnitude larger.

Even though SGV plays a major role in rapid adaptation, it is largely understudied. The investigation of SGV is challenging as the polygenic basis of adaptation present as SGV in the founder population is often unknown and large sample sizes are required to assess SGV at low frequencies. Both obstacles are overcome here by using threespine sticklebacks as study organism. The genomic basis of parallel adaptation in sticklebacks was previously identified enabling the characterisation of SGV at a known set of loci (Jones *et al.*, 2012). However, up to the present the collected empirical data on SGV in sticklebacks was insufficient, missing either individual data or involving small data sets impeding a reliable assessment of SGV. In the third study included in this thesis (chapter 5.3), a large empirical data set was collected in the field and used to fill this gap of knowledge. Emerging developments in genome sequence technology and bioinformatics algorithms as well as highly optimised protocols in the laboratory allowed pull-down sequencing at individual resolution of more than 9 000 marine threespine sticklebacks. A Bayesian probability as well as a Hidden Markov Model approach were adopted in order to detect carriers of freshwater adaptive alleles in this largescale data set and to empirically quantify properties of SGV. These encompassed the allele frequency of freshwater-adaptive alleles in the marine population, the number of alleles carried per individual and linkage disequilibrium among loci. Additionally, haplotype tagged sequencing data of freshwater-adaptive haplotypes carried by marine fish was compared to whole genome sequencing data of global freshwater populations in order to assign the geographic origin of adaptive haplotypes. The analysis of a freshwater population exposed further the presence of soft and hard sweeps at divergent loci and supported the hypothesis of rapid adaptation relying greatly on already existing SGV in the source population.

By providing both ancient and present-day insight this thesis gives a deeper picture of how SGV can facilitate rapid adaptation as well as of the evolutionary processes and factors that may promote or constrain the availability of SGV and its future evolutionary potential.

3. List of publications included in the thesis

Accepted papers

I. Ancient and modern stickleback genomes reveal the demographic constraints on adaptation.

Kirch M, Romundset A, Gilbert MTP, Jones FC, Foote AD.

Current Biology 2021 May 10;31(9):2027-2036.e8.

doi: 10.1016/j.cub.2021.02.027. Epub 2021 Mar 10. PMID: 33705715.

II. Predicting future from past: The genomic basis of recurrent and rapid stickleback evolution

Roberts Kingman GA, Vyas DN, Jones FC, Brady SD, Chen HI, Reid K, Milhaven M, Bertino TS, Aguirre WE, Heins DC, von Hippel FA, Park PJ, **Kirch M**, Absher DM, Myers RM, Di Palma F, Bell MA, Kingsley DM, Veeramah KR.

Science Advances. 2021 Jun 18;7(25):eabg5285.

doi: 10.1126/sciadv.abg5285. PMID: 34144992; PMCID: PMC8213234.

Unsubmitted manuscript

III. Adaptive standing genetic variation from disparate sources provides a substrate for rapid adaptation via hard and soft sweeps

Kirch M, Avdievich E, Venkataramani K, Kucka M, Philippos G, Thome R, Gaspar L, Bell MA, Peichel K, Archambeault S, Kingsley DM, Veeramah KR, Jones FC

4. Personal contributions to collaborative publications

I. Ancient and modern stickleback genomes reveal the demographic constraints on adaptation.

Kirch M, Romundset A, Gilbert MTP, Jones FC, Foote AD. Current Biology 2021 May 10;31(9):2027-2036.e8. doi: 10.1016/j.cub.2021.02.027. Epub 2021 Mar 10. PMID: 33705715.

Author contributions: Kirch M processed the genomic data of the ancient as well as the contemporary samples, conducted the majority of the genomic data analysis with the help of Foote A, performed simulations in SLiM, designed figures, drafted the initial manuscript with Foote A and contributed to the manuscript writing.

Relevance to the collective work: The genetic basis of adaptation is normally studied by comparing contemporary populations or by tracking adaptation over time. The empirical look at the genome of an ancestor of contemporary populations is a unique opportunity and provides a to date missing timepoint. The analysis of the genomic data thus forms the core of the manuscript. Besides, simulations performed by Kirch M. complemented and supported the empirical results.

Co-author contributions: Romundset A collected the sediment cores with the ancient stickleback bones, conducted all geological analysis and contributed to the figures as well as manuscript writing. Gilbert T helped with the ancient DNA lab work and contributed to the manuscript writing. Jones FC conducted the probability of ancestry analysis, designed figures and contributed to the manuscript writing. Foote A designed the experimental set-up and coordinated the study. He sampled the contemporary samples in the field, performed the ancient DNA lab work, helped Kirch M with the genomic data analyses, sketched the initial manuscript with Kirch M, designed figures and contributed to the manuscript writing.

II. Predicting future from past: The genomic basis of recurrent and rapid stickleback evolution

Roberts Kingman GA, Vyas DN, Jones FC, Brady SD, Chen HI, Reid K, Milhaven M, Bertino TS, Aguirre WE, Heins DC, von Hippel FA, Park PJ, **Kirch M**, Absher DM, Myers RM, Di Palma F, Bell MA, Kingsley DM, Veeramah KR. Science Advances. 2021 Jun 18;7(25):eabg5285. doi: 10.1126/sciadv.abg5285. PMID: 34144992; PMCID: PMC8213234.

Author contributions: Kirch M analysed the SNP array genotyping data. GenomeStudio was used to call genotypes, normalise and cluster the data. Due to the noisiness of the data, Kirch M reviewed each of the 384 SNPs individually and re-adjusted the clusters if necessary. The data was exported and prepared for further population genetic analysis of freshwater-adaptive allele frequency, linkage disequilibrium and haplotype blocks.

Relevance to the collective work: The majority of the data in the study is pooled sequencing data lacking information about the haplotypes carried by each individual. The SNP array genotyping data provides individual genotype information for 1,643 samples and gives the opportunity to examine the frequency, linkage disequilibrium as well as the size of haplotypes in three Alaskan marine populations. Sticklebacks from Loberg Lake sampled at different time points were also genotyped for tracking rapid adaptation over time. The SNP array genotyping data showed that EcoPeaks can be linked in multi-megabase haplotypes, which has to be considered in rapid stickleback adaptation.

Co-author contributions: Kingsley DM and Bell MA designed the experimental set-up. Kingsley DM guided the project conceptually. Bell performed the field work in Alaska including (tissue) sampling, population founding as well as collection of morphological data. Park PJ, Von Hippel FA, Aguirre WE and Heins DC sampled Alaskan sticklebacks. Park PJ helped with the population founding, Von Hippel FA provided logistical support and Aguirre WE conducted morphological data collection. Roberts Kingman GA analysed the genomic data and identified genomic properties predictive of evolution. Vyas DN, Milhaven M and Bertino TS performed lab work including DNA extraction, quantitation, selection and pooling of Alaskan samples. Jones FC designed the geographic sampling and the SNP array as well as contributed to the SNP array analysis. Brady SD organized the samples, prepared DNA and performed the SNP array calling. Absher DM and Myers RM performed the SNP array genotyping. Chen HI transferred and visualized the genome annotations. Di Palma F sequenced the geographic populations. Veeramah KR and Reid K analysed and modelled contemporary populations. The manuscript was written by Roberts Kingman GA and Veeramah KR with input from all authors.

III. Adaptive standing genetic variation from disparate sources provides a substrate for rapid adaptation via hard and soft sweeps

Kirch M, Avdievich E, Venkataramani K, Kucka M, Thome R, Philippos G, Gaspar L, Peichel C, Archambeault S, Bell M, Veeramah K, Kingsley DM, Jones FC

Author contributions: Kirch M coordinated the study together with Jones FC, selected and collected BAC clones from BAC libraries, conducted the field work in Alaska together with Avdievich E, coordinated the processing of 10,000 samples in the lab, performed tissue sampling, DNA extractions and nextera library preparations, processed the genomic data, adopted the Bayesian Probability approach to identify freshwater-adaptive carriers, performed the majority of the genomic data analysis with the help of Jones FC, designed figures, drafted the initial manuscript with Jones FC and wrote the majority of the manuscript with the help of Jones FC.

Relevance to the collective work: In order to elucidate and quantify the availability of standing genetic variation in natural populations, processing of thousands of samples is required. This high-throughput study offers the unique opportunity to study the origin, availability and maintenance of standing genetic variation based on individual sequencing data of thousands of samples. Kirch M accompanied and coordinated the study from data collection to the final manuscript and contributed to each step on the way.

Co-author contributions: Jones FC designed the experimental set-up, coordinated the study together with Kirch M, adopted the HMM approach to identify freshwater-adaptive carriers, contributed and helped with the genomic analysis, sketched the initial manuscript together with Kirch M, designed figures and contributed to the manuscript writing. Avdievich E performed Alaskan sampling, prepared BAC probes, optimised and performed pull-down experiments. Venkataramani K performed DNA extractions and nextera library preparations. Thome R DNA and prepared nextera libraries. Philippos G performed tissue sampling and DNA extractions. Gaspar L collected BAC clones in Stanford. Kucka M optimised the nextera library protocol and performed haplo-tagging on 96 samples. Peichel K and Archambeault S provided the samples from Puget Sound, Washington. Bell MA helped with the field work in Alaska. Kingsley D helped with access to BAC clones used in creating targeted pulldown libraries. Both, Kingsley D and Veeramah K contributed to insightful discussions on the topic of standing genetic variation.

5. Summary of results

5.1. Synopsis for Kirch *et al.* (2021): Ancient and modern stickleback genomes reveal the demographic constraints on adaptation

There are several possibilities to study adaptation. Two quite common approaches are to compare contemporary populations or to track adaptation with real-time experiments. Paleogenomics offer a third approach to enlighten the processes of adaptation. They provide a window into the past and add data for a timepoint that can normally only be interfered by modern DNA. This paper exhibits the first paleogenomics approach to study rapid adaptation in threespine sticklebacks.

In this study, 11 000 - 13 000-year-old stickleback bones were found in layers of sediment cores obtained from two Norwegian isolation lakes. The age of the ancient bones was determined by radiocarbon dating of organic matter from the same layer. Furthermore, the bones were collected from a layer corresponding to the brackish phase when the lakes got separated from the marine sea and were transitioning to freshwater. Genomes extracted from the ancient bones represent thus most likely marine ancestors of the contemporary freshwater population in the respective lake. In a principal component analysis both ancient genomes showed closest affinity with the contemporary local marine population and not with local freshwater populations suggesting that contemporary marine populations are a good proxy for the ancestor population of nearby freshwater lakes.

One of the ancient genomes provided enough data to assign probability of freshwater ancestry to 34 genomic regions which were previously identified to be divergent between marine and freshwater sticklebacks (Jones *et al.*, 2012; Figure 5.1). A quarter of these divergent regions showed a high probability of freshwater ancestry suggesting that freshwater-adaptive standing genetic variation (SGV) was already present in ancient marine sticklebacks. However, contemporary marine samples from a nearby location did not carry the same freshwater-adaptive regions compared to the ancient genome, which could mean that standing genetic variation in the marine population is not constant, but changing over time. Although freshwater-adaptive regions were present in one of their ancestors, contemporary samples from the same lake showed a higher probability of marine ancestry for half of these regions implying that

freshwater-adaptive loci once present in this lake were lost over time. This emphasizes the relevance of stochasticity in the course of parallel adaptation.

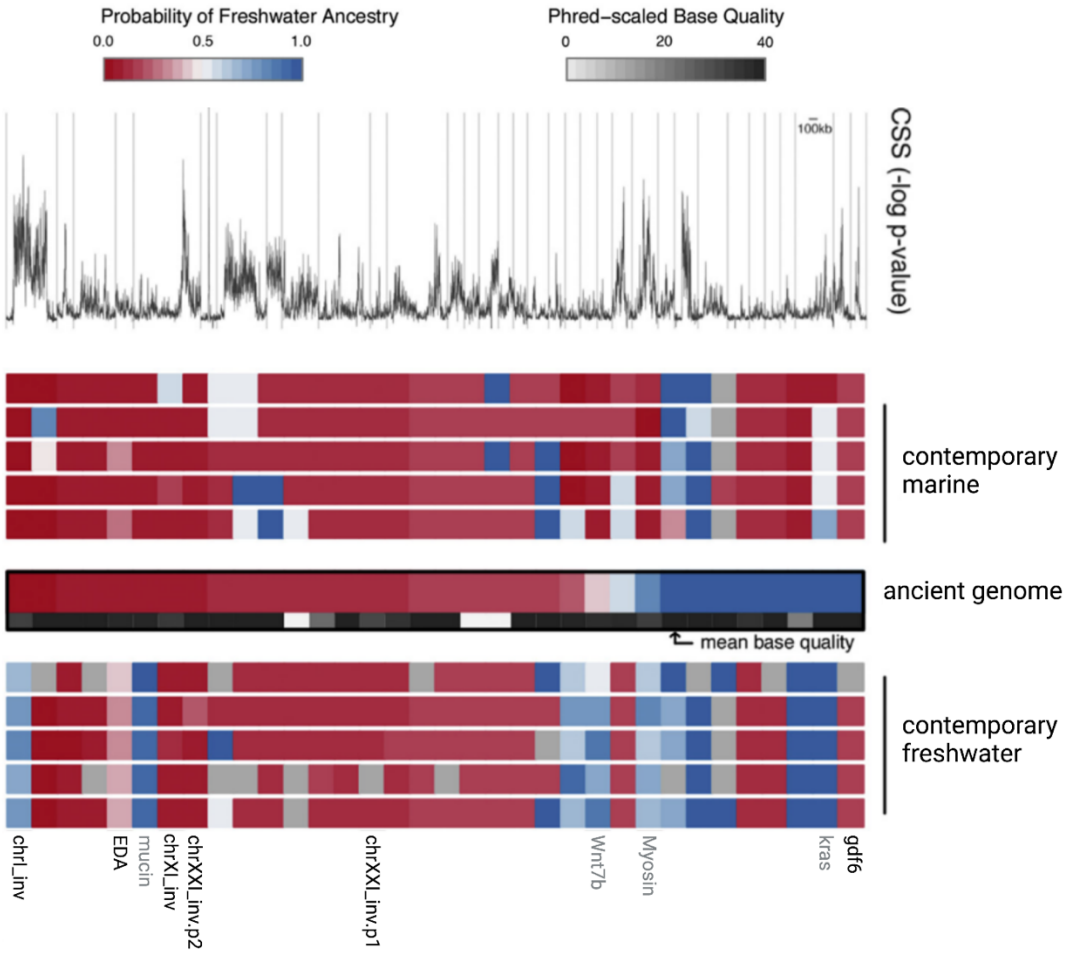


Figure 5.1: The probability of freshwater ancestry at 34 genomic regions is plotted for the ancient genome as well as contemporary marine and freshwater samples from the same lake. Genomic regions have been previously identified to be divergent between global marine and freshwater stickleback pairs using cluster separation score (CSS) (Jones *et al.*, 2012). The divergent regions are ranked by the probability of freshwater ancestry in the ancient genome. Mean mapDamage-rescaled base qualities of the ancient genome at each genomic region are included below its corresponding probability of freshwater ancestry score. Adapted from Kirch *et al.* (2021).

To our knowledge, this study includes genomic data of the oldest fish bones published until now (Oosting *et al.*, 2019). By doing that it adds a completely new timepoint to fish genomics, provides a first glimpse into the genomic long-term past of sticklebacks and highlights the power of ancient genomics to shed light on the evolutionary process.

Working with ancient DNA is challenging. The sparse data obtained from ancient DNA, typically due to limited availability of starting material and DNA degradation, precludes population-level analyses and thus our ability to confidently extrapolate to

wider evolutionary and demographic processes occurring in paleo times. However, insight into genetic variation present in one ancient genome highlights both the long-term significance of adaptive standing genetic variation and stochasticity of evolutionary processes that occurred since that time point.

This pilot study sets the basis for a new field in stickleback research. For future explorations, it will be important to collaborate with geologists in order to get access to sediment cores and thus, potentially contained stickleback bones. One possible follow-up study includes studying ancient material from multiple individuals obtained from the same sediment core layer in the same lake. This allows a better estimate of ancient standing genetic variation at one timepoint in the past and a more elaborate comparison to contemporary standing genetic variation. Another possibility is to study multiple ancient samples extracted from different sediment core layers of the same lake, enabling us to track adaptation over a long period of time. This data would give access to studying evolutionary processes at the population level during a time not accessible in present-day experimental real-time studies. Moreover, comparing ancient genomes of material obtained from the brackish layers of isolation lakes at different geographic locations could elucidate whether similar patterns of ancient standing genetic variation can be discovered in all lakes.

5.2. Synopsis for Kingman *et al.* (2021): Predicting future from past: The genomic basis of recurrent and rapid stickleback evolution

The previous chapter presented a paleogenomic approach to study adaptation by investigating two 11 000 - 13 000 year-old stickleback genomes. However, adaptive evolution can happen rapidly. For instance, marine sticklebacks can adapt to a new freshwater pond within a few decades, roughly 30 generations. In order to capture adaptation on these short time scales, this paper studied contemporary time series of marine threespine sticklebacks adapting to newly colonized freshwater habitats. The study of contemporary evolution brings a few benefits. In contrast to the limited data availability in ancient studies, large sample sizes from contemporary populations allow the estimation of allele frequencies and the application of a broad suite of population genetic tools and models. Standing genetic variation (SGV), for instance, is often present at low frequencies in a population and large sample sizes are thus needed to determine its availability. Here, more than a thousand marine sticklebacks were used to investigate SGV in marine populations nearby the studied lakes in order to elucidate the substrate available for adaptive evolution.

Populations in three Alaskan lakes that were recently founded by anadromous sticklebacks are examined for this study. The populations were sampled 7-8 times within the first 30 years since their colonization by anadromous sticklebacks (Figure 5.2 A). The samples for each year were pooled and then subjected to genomic library preparation and sequencing to track both the change of allele frequencies over time in each lake and the process of adaptation in real-time. For convenience, the terms “TempoPeaks” and “EcoPeaks” are used to describe the identified differential signals. TempoPeaks refer to single nucleotide polymorphisms (SNPs) with significant allele frequency changes. These SNPS are presumably important for adapting to the new freshwater environment as they imply directional selection (Figure 5.2 C). One example is the Ectodysplasin (*Eda*) SNP (O’Brown *et al.*, 2015). It is associated with the armor plate differences in stickleback ecotypes (Colosimo *et al.*, 2005) and rapidly rises from an initial allele frequency of <1% to fixation within 15 years (Figure 5.2 B).

EcoPeaks are genomic regions that are divergent between marine and freshwater sticklebacks, which are also putatively essential for adaptation. EcoPeaks were estimated based on (subsets of) 227 global freshwater and marine stickleback genomes. The different sets of identified EcoPeaks included up to 90% of the global divergent regions previously identified in Jones *et al.* 2012 (Jones *et al.*, 2012). EcoPeaks identified based on the Pacific subset of individuals (Pacific EcoPeaks) span

approximately seven-fold more of the genome compared to EcoPeaks identified by using all 227 global individuals (global EcoPeaks). While TempoPeaks largely coincided with Pacific EcoPeaks, they surprisingly only overlapped by a quarter with the global EcoPeaks. This indicates that SGV can differ in geographically disparate populations and suggests that the availability of SGV in the founder population affects which alleles rise to high frequencies under selection. The study of locally available SGV is thus important for understanding the mechanisms of rapid adaptation.

In order to uncover to what extent SGV can affect the speed of adaptation, it is crucial to investigate whether adaptive alleles are carried in linkage in the founder population. When adaptation occurs extremely rapidly, alleles on the same chromosome are not transmitted independently, but rather as linked haplotype blocks. The pooled whole-genome sequencing data obtained from the contemporary time series misses individual data and hence, haplotype blocks cannot be resolved. In order to explore what alleles are carried in linkage, we applied individual based genotyping of 1643 samples originating from three Alaskan marine populations which represent SGV in the founder population of the studied freshwater populations. The genotyping data allowed us to estimate the frequency of freshwater-adaptive alleles, the linkage disequilibrium among them as well as the size of ancient freshwater-adaptive haplotype blocks in the marine sticklebacks.

Frequencies of freshwater-adaptive alleles in the three marine populations were overall low, ranging from 0-0.14 with median values below 0.01 and for one of the populations (Rabbit Slough) even below 0.001. Freshwater-adaptive alleles were usually present in heterozygous form at one out of the 44 adaptive loci included in the analysis. Due to the low frequencies of freshwater-adaptive alleles, some allelic associations were largely affected by very rare carriers. Hence, the analysis of linkage disequilibrium was underpowered, and only little evidence for inter-chromosomal linkage disequilibrium among adaptive loci was found. While the majority of the marine fish carried small freshwater-adaptive haplotype blocks, some individuals carried multi-mega-base freshwater-adaptive haplotype blocks spanning multiple EcoPeaks (Figure 5.3). The partial assembly and linkage of freshwater-adaptive alleles in these rare large haplotype blocks can facilitate rapid adaptation to new freshwater habitats. Therefore, it is crucial for better understanding rapid adaptation to account for the complex linkage between adaptive loci present as SGV and not to treat them independently. Even though the large haplotype block sizes suggest recent introgression, none of the individuals was identified as a F1 hybrid or a freshwater migrant. The set of freshwater-adaptive alleles carried by an individual differs not only among individuals but also between populations. In fact, carriers of large freshwater-adaptive haplotype blocks were not present in Rabbit Slough.

The ancient freshwater-adaptive haplotype blocks detected in Rabbit Slough were all short which indicated that the ancient blocks in this population have been broken down by recombination and are already present for many generations in the marine environment. Moreover, the differences in frequencies and haplotype block sizes between populations suggest that the factors affecting the availability of SGV in the populations are different. These factors include the degree of introgression, which introduces freshwater-adaptive material into the marine population, and/or the selective pressure purging freshwater-adaptive alleles from the population. The population in Rabbit Slough, for instance, probably experiences lower introgression and/or stronger selection resulting in lower frequencies of freshwater-adaptive alleles and shorter freshwater-adaptive haplotype blocks.

The results of the study highlight that the availability of SGV is relevant for rapid adaptation and that the factors promoting or restraining maintenance of SGV can differ in different populations.

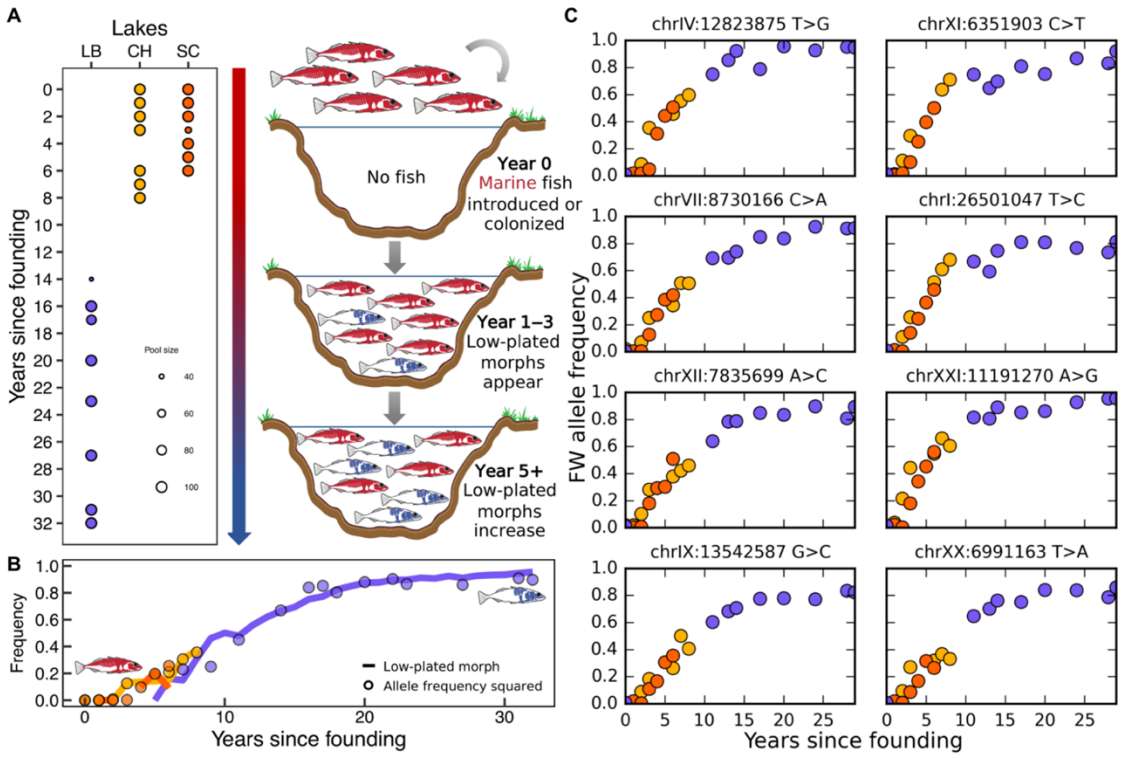


Figure 5.2: Contemporary time series data of three Alaskan populations. (A) On the left: the timepoints as well as sizes of samples from three Alaskan lake populations Loberg Lake (LB), Cheney Lake (CH), and Scout Lake (SC)] which were recently founded by anadromous sticklebacks. On the right: schematic displays rapid adaptation of anadromous sticklebacks to new freshwater habitat. Red fish represent the highly-plated sticklebacks, whereas blue fish represent the typical freshwater phenotype with reduced armor plates. (B) Frequency of armor-reduced phenotype in each lake over time as well as the allele frequency trajectory of the *Eda* SNP (O'Brown *et al.*, 2015), an allele coding for the armor plates in sticklebacks (Colosimo *et al.*, 2005). (C) Allele frequency over time for eight SNPs within TempoPeaks on different chromosomes. Adapted from Kingman *et al.* (2021).

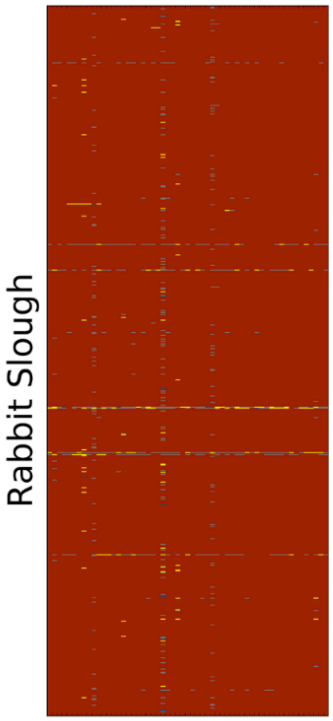


Figure 5.3: Genotypes across chrIV obtained by SNP array from 750 Rabbit Slough individuals. Depicted SNPs were identified as being divergent between marine and freshwater sticklebacks. Marine homozygous, red; heterozygous, yellow; and freshwater homozygous genotypes, blue; respectively. Adapted from Kingman *et al.* (2021).

This study introduces a contemporary time-series with timepoints over three decades. Although threespine sticklebacks are a well-studied organism in evolutionary biology, such a dataset was not available until now and allows us to track allele frequency changes over real-time. To investigate the genomic substrate available for this adaptive evolution, SGV in nearby marine populations was assessed based on genotyping data of several hundred marine individuals. For the first time, the data with individual resolution provided information about freshwater-adaptive haplotype block sizes and linkage in a relatively large population of marine sticklebacks. Both analyses open new ways to study rapid adaptation of threespine sticklebacks.

The study revealed that SGV can be harboured in a population at extremely low frequencies, e.g. with a median below 0.001 for the marine population sampled at Rabbit Slough. The SNP genotyping array for estimating SGV was performed with sample sizes ranging from 237 to 751 sticklebacks per marine population. With an allele frequency of 0.001, less than two carriers are expected from a study with a sample size around 800. Therefore, to be able to reliably determine allele frequencies of freshwater-adaptive alleles in marine populations, large sample sizes are required, e.g. 10-fold larger than the sample sizes in this analysis. Likewise, the assessment of linkage disequilibrium was

also underpowered due to the available sample size in this study and a larger sample size is required for an analysis with sufficient power.

Freshwater-adaptive haplotype blocks with a size of multiple mega base pairs were detected in this study. However, this analysis is based on genotyping data of 235 divergent SNPs. These SNPs were manually selected in order to capture divergence, which could introduce an ascertainment bias. For instance, the detected multi-mega-base freshwater-adaptive haplotype blocks were determined based on data of SNPs at scattered positions and might actually not represent one huge block but several smaller freshwater-adaptive haplotype blocks. Thus, the sparse data may simplify results which might be more complex in reality. For future studies, it is important to obtain unascertained data, e.g. whole genome sequencing data, in order to reliably determine the size of ancient freshwater-adaptive haplotype blocks carried by marine individuals.
In general, whole genome sequencing data at individual resolution for marine founder populations as well as for time series data is crucial to better understand the process of rapid adaptation. Individual data will enable us to obtain information on haplotypes and investigate the re-assembly of freshwater-adaptive alleles from SGV over time under selection in the freshwater environment. Besides, it is important to characterise the available freshwater-adaptive material in the founder population or a nearby marine population representing the founder population. Since SGV is sometimes present at extremely low frequencies in the marine population, a large amount of individuals is required to assess its availability. To tackle this challenge and get a more reliable estimate of SGV, we determined SGV in more than 9 000 marine sticklebacks as presented in the next chapter.

5.3. Synopsis for Kirch *et al.* (unpublished): Adaptive standing genetic variation from disparate sources provides a substrate for rapid adaptation via hard and soft sweeps

Theory suggests that standing genetic variation (SGV) often outrivals *de novo* mutations in the process of adaptation and plays a pivotal role in rapid adaptation (Barrett & Schluter, 2008; Hermisson & Pennings, 2005). Despite its importance SGV is vastly understudied. Many challenges are associated with studying SGV complicating progress in this field. In order to determine SGV at adaptive loci, the genomic basis of adaptation needs to be known. Although many empirical studies focus on finding regions involved in adaptation, the majority of studies identify candidate genes with large effect sizes potentially missing a large amount of small effect regions which contribute to the polygenic basis of adaptation. Moreover, SGV can be present at extremely low frequencies. Hence, characterisation of SGV in natural populations requires sampling of several thousand individuals in the wild representing another impediment. Threespine sticklebacks represent a premium system for evolutionary biology and enabled us to overcome these obstacles. Firstly, the polygenic basis of parallel adaptation from marine sticklebacks to freshwater environment has been identified by Jones *et al.* setting the foundation for characterising SGV in marine sticklebacks (Jones *et al.*, 2012). In addition, contemporary samples of marine sticklebacks are accessible at large scale providing the required number of individuals. Thousands of marine sticklebacks swim upstream for spawning during spring time which gave us the opportunity to sample 5 100 sticklebacks from a stream at Rabbit Slough in Alaska during field work in 2018.

The study presented in the previous chapter 5.2 assessed the availability of SGV in less than 1600 marine sticklebacks originating from three different populations. As stated at the end of chapter 5.2, the amount of individuals as well as the sparse data of the SNP array confined the informative value of the determined availability of SGV, e.g. the estimate of linkage disequilibrium among adaptive alleles was underpowered. In order to overcome the limitations of Kingman *et al.*, we took advantage of the progress in the field of genome sequencing and bioinformatic algorithms and performed targeted enrichment sequencing for more than 9 000 marine sticklebacks capturing genomic regions previously identified to be divergent between global marine and freshwater sticklebacks (Jones *et al.*, 2012, FDR 0.05). High-throughput processing of thousands of samples in the laboratory was largely facilitated by highly optimised protocols as well as many persons contributing to different steps of the workflow.

1.7 TB of sequencing data was obtained and processed via parallel computing available at the Max Planck facilities in Tübingen. The empirical data encompassed sequencing data at individual resolution for ~3 300 marine sticklebacks from Puget Sound (PGTS), Washington, and ~5 500 marine sticklebacks from Rabbit Slough (RABS), Alaska, at 240 previously identified adaptive loci (Jones *et al.*, 2012, FDR 0.05). Two statistical approaches were adopted in order to identify marine carriers of freshwater adaptive alleles. A Hidden Markov Model approach and a Bayesian Probability approach applied on sliding windows segmented the genomic regions of each individual into freshwater, marine and heterozygous ancestry states. Based on the proportional state assignment each fish was then classified into a non-carrier, a heterozygous or homozygous carrier at each divergent region. The results of both approaches concurred for 85% of the data and the results shown in the following were obtained based on the Bayesian Probability approach.

As expected, frequencies of freshwater adaptive alleles were detected to be low in the marine populations with mean allele frequencies of 0.10 and 0.05 in PGTS and RABS, respectively (Figure 5.4A). PGTS displayed significantly higher allele frequencies indicating that RABS experiences less introgression and/or higher selective pressure against freshwater adaptive alleles. Hence, the environmental conditions for PGTS seem to facilitate the availability and maintenance of SGV in the population. These results suggest that different populations provide varying substrate for evolutionary potential and might differ in their eligibility as “source” population for conservation management.

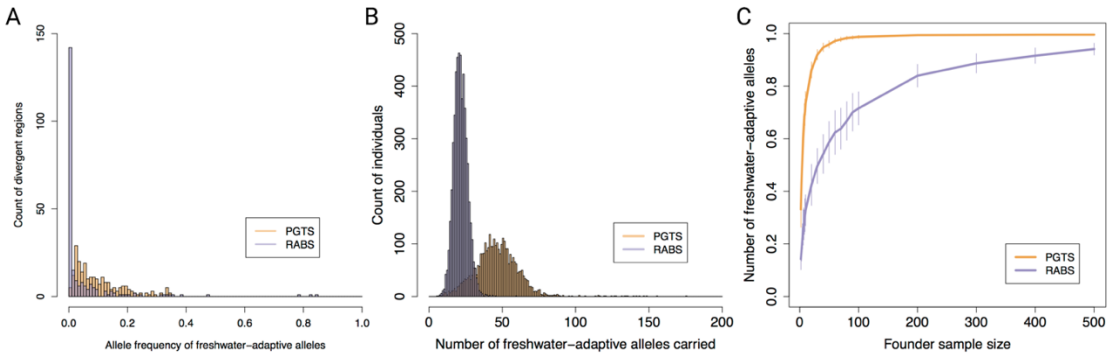


Figure 5.4: (A) Histogram of allele frequencies of freshwater adaptive alleles in PGTS and RABS for 232 divergent loci. (B) Histogram of freshwater adaptive alleles carried by individuals from PGTS and RABS at 232 divergent loci. (C) Percentage of freshwater adaptive alleles available in different founder population sizes. Individuals were randomly sampled from PGTS and RABS, respectively, and results are based on 100 iterations. Results for PGTS are shown in orange and for RABS in purple.

With freshwater adaptive alleles being present at low frequencies in marine populations, the question arises whether a few rare, recently introgressed individuals carry the complete freshwater adaptive genetic material present and show a high abundance of alleles or whether the freshwater adaptive material is more or less equally distributed among all marine individuals. Our study supports the latter and shows that adaptive alleles were distributed among all individuals in both populations with each individual carrying at least six adaptive alleles. PGTS and RABS individuals carried a mean of 47 and 22 freshwater adaptive alleles, respectively (Figure 5.4B). No F1s were detected in both marine populations and further analysis suggested that the presence of freshwater adaptive genetic material in both marine populations can be traced back to multi-generational introgression of freshwater alleles. Although simulations pointed to small founder sizes already providing a nearly complete freshwater-adaptive allele ensemble (Figure 5.4C), these results indicated that SGV is mostly dissembled into pieces and needs to be re-assembled when a new freshwater environment is colonised.

Inter- as well as intra-chromosomal linkage disequilibrium (LD) among adaptive alleles was estimated to explore whether freshwater adaptive alleles are carried in linkage by marine individuals. Even though strong LD was detected for loci within the same inversion and some proximate adaptive loci, intra-chromosomal LD was relatively low in general further supporting the disassembly of adaptive allele blocks into pieces in the marine environment. Adaptation of sticklebacks to freshwater environments is reported to happen within as quickly as 30 generations (Bell *et al.*, 2004; Lescak *et al.*, 2015; Terekhanova *et al.*, 2014; Bassham *et al.*, 2018). In contrast, the re-assembly of broken down adaptive allele blocks might be time-consuming and require several generations of recombination. However, large clutch sizes in sticklebacks provide a platform for diverse meiotic recombination within one generation potentially largely accelerating the process of re-assembly. All in all, this poses the conundrum whether the re-use of SGV during adaptation to freshwater environments predominantly relies on a few founder individuals carrying an excess of freshwater-adaptive alleles (as previously suggested in Kingman *et al.*, 2021; Hohenlohe *et al.*, 2012), on the collective presence of SGV in the founder population or on both.

Short-read sequencing including haplotype information termed as haplotype tagging (Meier *et al.*, 2021) was performed for 57 RABS carriers of adaptive alleles on chrIV and within the chrI inversion as well as 39 individuals from Arness Lake, an Alaskan freshwater population. Freshwater haplotypes present in RABS were projected on a principal component analysis (PCA) performed with whole genome sequencing data of global freshwater populations and were associated with several North American, European as well as a Californian freshwater haplotype cluster(s) (Figure 5.5). The

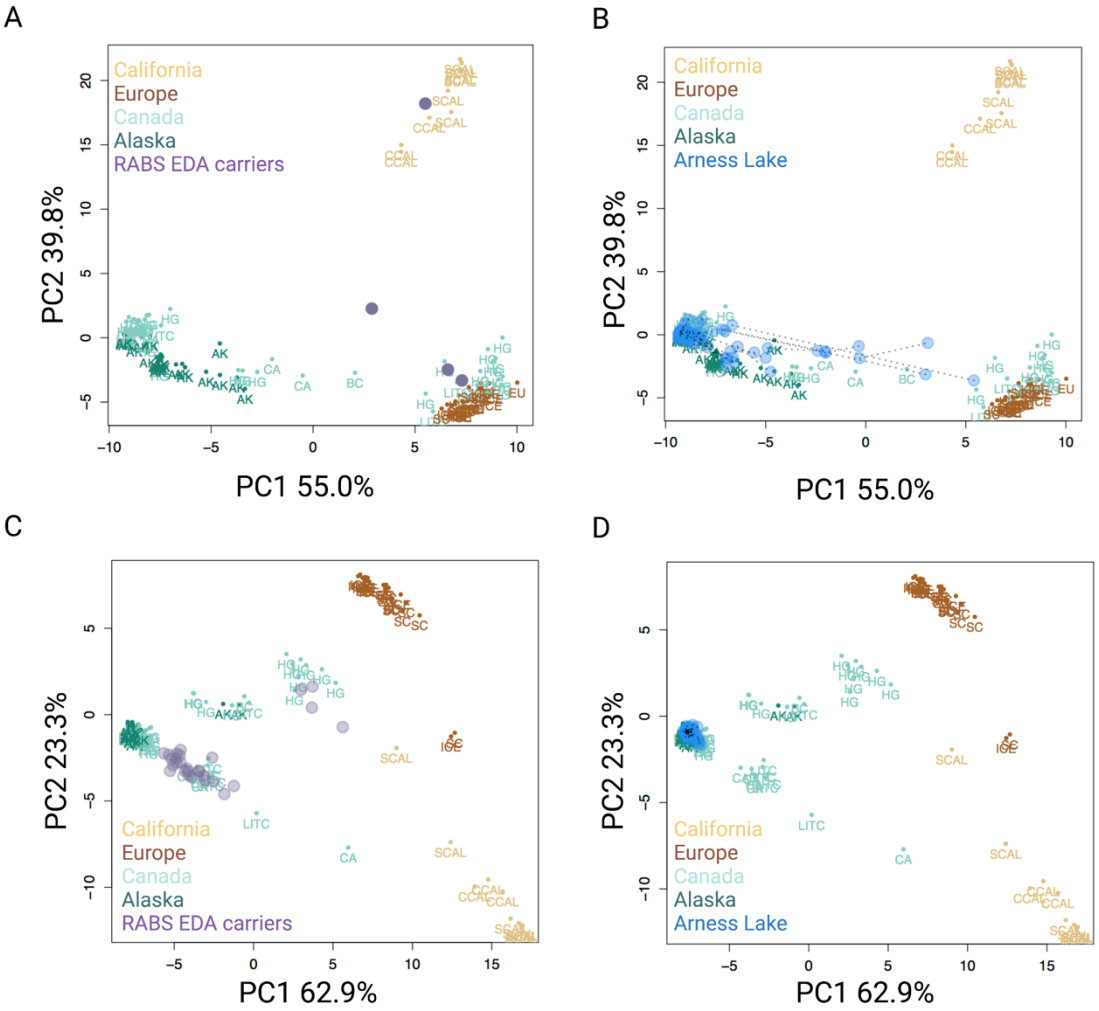


Figure 5.5: Principal component analysis (PCA) performed for 103 global freshwater individuals from Alaska, Canada, Europe and California at a central 6 kb block at the *Eda* region (A,B) and at a 35 kb block within chrI inversion spanning the ATPase locus (C,D). (A,C) Freshwater haplotypes of RABS carriers at *Eda* locus were projected on the PCA for *Eda* or ATPase, respectively, and are associated with samples from North America, Europe and California. (B,D) Freshwater haplotypes of freshwater individuals from Arness Lake were projected on the PCA and are associated with North-American haplotype groups depicting a soft (C) as well as a hard sweep (D).

marine population seems to harbour a conglomeration of ancient alleles from diverse geographic sources presumably introduced by local hybrid zones as well as via interconnectivity between marine populations. Freshwater haplotypes of Alaskan freshwater individuals projected on the same PCA cluster only with North-American and mostly Alaskan freshwater haplotype groups (Figure 5.5). Even though diverse freshwater haplotypes are available in RABS, representing a potential founder population for Arness Lake, the freshwater population depicts predominantly the (re-)use of local Alaskan haplotypes suggesting that local haplotypes are most beneficial for adaptation and that effective population sizes of the founder population were sufficiently large to bypass stochastic loss of haplotypes due to demographic processes. In addition,

the results exhibit the presence of hard as well as soft sweeps in the freshwater population indicating that rapid adaptation relies remarkably on pre-existing freshwater adaptive haplotypes in the founder population. Collectively, this analysis points out that especially adaptive alleles introduced by local gene flow into the marine population seem to enhance its future adaptive potential as source population.

Threespine sticklebacks are a highly suitable model for studying SGV. Nevertheless, previous studies on SGV in sticklebacks have been fragmentary. In many cases, small sample sizes were processed precluding the quantification of low allele frequencies as well as the determination of linkage disequilibrium among adaptive alleles (Colosimo *et al.*, 2005; Miller *et al.*, 2007; Hohenlohe *et al.*, 2010). Although large sample sizes were used in the study presented in chapter 5.2, pooled sequencing as well as SNP arrays limited the findings. The pooled sequencing data impeded resolution at individual level and as a result also information about linkage and haplotypes. In the SNP arrays performed on around 1600 marine sticklebacks, sparse data for less than 300 divergent SNPs was obtained (Kingman *et al.*, 2021). The data presented in this manuscript shows significant improvements compared to previous studies. The data is gathered for more than 9 000 marine sticklebacks with resolution at individual level for 5.8% of their genome and provides the unique opportunity to characterise SGV in many marine sticklebacks.

Nonetheless, the analysis in this study is confined by targeted enrichment sequencing. Firstly, frequencies of freshwater-adaptive alleles might be underestimated due to pulldown bias. In addition, the targeted regions in this study only include globally freshwater-marine divergent regions and disregard divergent regions identified based on Pacific sticklebacks which cover sevenfold more of the stickleback genome (Kingman *et al.*, 2021) not accounting for local SGV potentially important for adaptation to close-by freshwater environments. Lastly, the low mean read depth coverage of 4.6x does not suffice for reliable phasing of haplotypes which was only possible for 57 whole genome sequenced RABS carriers.

The empirical characterization of SGV in this manuscript sets the foundation for future research. One potential continuation of this study includes simulations to assess whether SGV is maintained by a selection-migration balance or whether adaptive alleles present at low frequencies are neutral in the marine population as proposed by Haenel *et al.* (Haenel *et al.*, 2022). Besides, future studies could explore how adaptive alleles available as SGV are re-assembled in the freshwater environment. To do so, one could perform haplotype tagging (Meier *et al.*, 2021) for a contemporary time series in a newly colonized lake as well as for the respective founder population. The new arrangement of

adaptive alleles in the genomes could thus be tracked over time disentangling how the available SGV facilitates or constrains rapid adaptation to a new environment.

6. Discussion

To disentangle the genetic basis and the molecular mechanisms underlying adaptation to a new or changing environment is a fundamental objective of evolutionary biology and there are different ways to approach this conundrum. Firstly, adaptation can be tracked in experiments over generations in “contemporary evolution time-series"; secondly, the evolutionary past of organisms can be inferred retrospectively using approaches such as coalescent theory to trace mutations in the genomes of contemporary individuals back to common ancestral states; and thirdly, paleogenomics can open a window to the past and provide empirical data of ancestors that are thousands of years old. Each of these approaches present various challenges or requirements such as a short generation time of the studied organism for experiments or access to ancient material, e.g. bones containing ancient DNA for paleogenomics. The majority of evolutionary model systems offer the possibility to use one or two of these approaches. However, as presented in this thesis, threespine sticklebacks, a super model organism for evolutionary biology, give us the unique opportunity to use all three approaches for studying adaptation.

The rapid adaptive radiation of marine sticklebacks into freshwater ecotypes following Pleistocene ice sheet retreat 10 000 to 20 000 years ago occurred repeatedly and in parallel across the Northern Hemisphere providing many biological replicates to identify consistent molecular mechanisms associated with rapid adaptive evolution (Bassham *et al.*, 2018; Bell *et al.*, 2004; Bell & Foster, 1994; Lescak *et al.*, 2015; Terekhanova *et al.*, 2014). Notably, stickleback fish have a generation time of one to three years in the wild, enabling the assessment of time-series (Bell & Foster, 1994). Moreover, their bony body composition has led to ancient bones being present in sediments of lakes laying the foundation for the first paleogenomics study in threespine sticklebacks, which is included in chapter 5.1 of this thesis.

In using various approaches to elucidate the processes of rapid adaptation, this thesis has a special focus on the availability of standing genetic variation (SGV). Besides *de novo* mutations, adaptation via pre-existing SGV is one of the main mechanisms for rapid adaptation to a new or changing environment (Hermisson & Pennings, 2005). Population genetics theory predicts that SGV largely facilitates rapid adaptation and conservationists have considered SGV as a crucial factor for maintaining biodiversity for decades (Hermisson & Pennings, 2005; Crandall *et al.*, 2000; Moritz, 2002). Various empirical studies including threespine sticklebacks have outlined the crucial role of SGV in rapid adaptation (Colosimo *et al.*, 2005; Meier *et al.*, 2017; Barrett *et al.*, 2008). Despite

its importance, few empirical studies have been performed to characterise SGV and the forces that shape its availability for future adaptation events.

In this chapter, the major findings of this thesis as well as potential evolutionary implications are summarised. Furthermore, future steps, required to further understand the processes of rapid adaptation and the factors promoting or impeding the availability of SGV for rapid adaptation, are outlined.

6.1. Standing genetic variation and its role in rapid adaptation

6.1.1. Adaptation to new environments can occur very rapidly

In Kingman *et al.* 2021 (chapter 5.2) three stickleback-free Alaskan lakes were colonised with marine sticklebacks and the populations were tracked over time for a maximum span of 30 years (30 generations). The time-series data displayed remarkable increases in allele frequencies of freshwater-adaptive alleles over very few generations and highlighted that adaptation of threespine sticklebacks can happen very rapidly. This underlines that selective pressures are strong in the freshwater environment, as previously reported with selection coefficients up to 0.5 (Barrett *et al.*, 2008; Schluter *et al.*, 2021). The striking speed of adaptation suggests further that freshwater adaptive alleles can have a large fitness effect in the freshwater environment and that individuals carrying them have elevated chances to survive and reproduce. As the Anthropocene progresses we anticipate very large environmental shifts within a few decades (Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, 2022). Threespine sticklebacks show that it is possible to adapt to a new environment within this short period of time (Bassham *et al.*, 2018; Bell *et al.*, 2004; Kingman *et al.*, 2021; Lescak *et al.*, 2015; Terekhanova *et al.*, 2014). However, the question arises, whether other organisms can adapt similarly rapidly. Some biological features differ across taxa including generation time, clutch size, recombination rate and gene flow among different populations and/or ecotypes. All these features facilitate rapid adaptation in sticklebacks. Therefore, our findings might apply to taxa which show similarly to sticklebacks short generation time, large clutch sizes, high recombination rate and interconnectivity between populations (as elaborated in section 6.3.2).

6.1.2. SGV important for rapid adaptation

Kirch *et al.* (unpublished) detected hard as well as soft sweeps in a freshwater population. While one freshwater adaptive haplotype was fixed at the ATPase locus on

chromosome I, several different haplotype versions were present at *Eda*. Although hard sweeps can arise from *de novo* mutations, both sweep types can also originate from SGV. A soft sweep indicates that multiple distinct haplotypes showed an increase in allele frequency in the new environment. This means that multiple haplotypes were introduced via SGV in the marine population emphasising the important role of SGV in rapid adaptation. With multiple haplotypes present in the freshwater population, the question arises whether these haplotypes have equivalent beneficial value or whether one is better than the other(s). There are two different explanations for why predominantly two distinct Alaskan *Eda* haplotypes were present in the freshwater population. In the first scenario, the effective population size was sufficiently large to prevent the fixation of one haplotype at the expense of others due to demographic stochasticity. A soft sweep occurred based on the SGV present in the founder population leading to several different freshwater haplotypes in the freshwater population. Another explanation includes temporal displacement in the arrival of the haplotypes. Under this scenario, the second haplotype may be introduced into a population which was already fixed for the first haplotype. If the second haplotype confers a stronger fitness advantage than the first haplotype, it may invade the population, outrivalling the first haplotype. The time the population was sampled could thus represent the time when the first haplotype was still present and the second haplotype was on the way to rise to high allele frequency. Future studies tracking this population over time would provide a way to disentangle these two possible explanations for the existence of two or several distinct versions of freshwater *Eda* haplotypes at high frequency. If one haplotype is sufficiently more beneficial than the other(s) (enough to overcome fluctuations due to demographic stochasticity), then we would expect to see an increase in the frequency of that haplotype in subsequent generations. If both haplotypes are of equivalent beneficial value, we would expect to see random fluctuations over subsequent generations similar to other neutral loci in the genome.

6.1.3. SGV frequencies and geographic structure

The characterisation of SGV in a large amount of individuals revealed that SGV can be quite prevalent in one marine population, and present at extremely low frequencies in another population. For instance, freshwater-adaptive alleles in RABS were quantified to be mostly present at low frequencies in Kingman *et al.* (2021) as well as in Kirch *et al.* (unpublished). In both studies, the other investigated populations showed higher frequencies, suggesting that RABS experiences lower introgression of freshwater adaptive alleles and/or higher selective pressures against them compared to

the other populations investigated. Other marine populations from Cook Inlet show higher allele frequencies of freshwater adaptive alleles than RABS, but the frequencies of the three populations from Cook Inlet, Alaska, correlate indicating some local geographic structure. PGTS, a marine population from Washington more than 1000 km apart from RABS displays limited correlation in allele frequencies compared to RABS. These results highlight that selective pressure and/or introgression via hybrid zones might differ for the same loci at different locations. Even though freshwater adaptive alleles are present at low frequencies in RABS, Kingman *et al.* 2021 showed that populations in Alaskan lakes which were founded by 3 000 RABS individuals have undergone a successful rapid adaptation. This poses the question how frequent SGV has to be in a founder population in order to ensure a successful adaptation to a new or changing environment. Using simulations in Kirch *et al.* (unpublished) we showed that 231.7 out of 232 freshwater adaptive alleles were present in 3 000 RABS individuals indicating that freshwater adaptive SGV present at low frequencies in a population can still be highly prevalent in a founder population with sufficiently big population size. Simulations with PGTS even showed that 231 out of 231 freshwater adaptive alleles were already present in 500 individuals consistent with the higher allele frequencies reported in PGTS. These simulations assume a single founding event. In nature, however, adaptive alleles could arrive over subsequent generations and hybridisation could allow continuous introduction of freshwater adaptive material from SGV in the marine population providing additional sources for adaptive alleles. This scenario poses interesting questions as to whether the order of arrival of freshwater-adaptive alleles influences the rate at which a population can adapt to a new environment. In contrast, simulations of the transporter model from Galloway *et al.* proposed that the majority of freshwater adaptive alleles in freshwater populations is actually originating from the first colonists and only minorly from subsequent migrants emphasising the importance of SGV available in the founder population (Galloway *et al.*, 2020). Consistent with simulations of Galloway *et al.*, our empirical contemporary evolution study of allele frequency changes in replicate Alaskan lakes founded by a larger number (N=3000) of RABS fish show very similar trajectories of allele frequency changes among populations suggesting any minor sampling differences in the SGV available upon initial founding has minimal effect on rates of adaptation (Kingman *et al.*, 2021).

6.1.4. SGV present and old

Threespine sticklebacks experienced a burst of adaptive radiation 10 000 to 20 000 years ago, when many freshwater habitats got available due to glacier melting

(Bell & Foster, 1994). Has this ancient adaptive evolution been facilitated by SGV meaning that SGV has been present in marine stickleback populations for thousands of years? Colosimo *et al.* previously showed that almost all *Eda* haplotypes present in global freshwater populations clustered in one distinct clade and shared a common ancestor being 2 to 10 million years of age suggesting that SGV in the marine population is even up to 10 million years old (Colosimo *et al.*, 2005). Kirch *et al.* (2021) found SGV in a single more than 12 000 years old, ancient stickleback genome. For the first time in stickleback research, the here analysed ancient genome provided a glimpse into the past proving that SGV in marine populations was already present for thousands of years. The ancient genome showed freshwater ancestry for 10 out of 34 adaptive loci. Assuming that these freshwater adaptive alleles are carried in heterozygous state, the heterozygosity and freshwater ancestry of the ancient genome are estimated to be 0.29 and 0.15. These values correspond to F1-Marine backcrosses with marine ancestors based on simulations performed in Kirch *et al.* (unpublished) (Figure 6) and fall within the empirical distribution obtained from PGTS individuals. The ancient genome shows higher freshwater ancestry and heterozygosity compared to RABS individuals and also elevated values compared to contemporary samples from nearby Altafjord (heterozygosity: 0.15-0.24 and freshwater ancestry 0.07-0.12). In general, individuals from contemporary samples match predominantly F1-Marine backcrosses with marine ancestors and in case of PGTS additionally F1-Marine backcrosses based on simulations. Showing a similar degree of introgression, these results indicate that contemporary marine individuals are a good representative for ancient marine individuals and hence, for founder populations of contemporary freshwater populations.

6.1.5. Linkage among adaptive alleles

In theory, rapid adaptation would be accelerated the most if one or several founding individuals carried a complete set of freshwater adaptive alleles (an “adaptive ensemble”). If the adaptive ensemble was already introduced in linkage, the necessity of re-assembling adaptive alleles during several generations of recombination would be abolished and the rate of adaptation would be largely increased. In addition, the size of the founder population providing a complete set of freshwater adaptive genetic material would be decreased. Hence, beside the frequency of freshwater adaptive alleles in the marine population, the linkage exerts a major influence on the speed of adaptation. For instance, adaptive alleles might be carried in linkage due to recent introgression of freshwater adaptive material into the marine population. Kirch *et al.* (unpublished) reported a lack of F1 hybrids of marine and freshwater sticklebacks in more than 8 800

marine individuals belonging to two different marine populations. Further simulations revealed that marine individuals in both populations correspond to (one or) multigenerational backcrosses of F1s with marine individuals. Even though SGV can be very rare at some adaptive loci, the study further showed that each individual in a marine population is carrying some freshwater-adaptive alleles contributing to the SGV present in the population. These results indicate that freshwater adaptive SGV introduced by a marine founder population into a new environment is mostly not present as adaptive allele ensemble, but rather uncoupled and distributed among many individuals. Despite these indications, neighbouring loci could still be carried as allele blocks in linkage and functional epistasis in the marine environment could promote the presence of inter-chromosomal freshwater adaptive allele combinations facilitating the re-assembly of the "adaptive ensemble" in the freshwater environment (as suggested by Hohenlohe *et al.*, (2012)). In order to explore the inter- and intra-chromosomal linkage disequilibrium (LD) among adaptive loci, large sample sizes are needed. For instance, Kingman *et al.* (2021) was underpowered to estimate LD for rare adaptive alleles with data encompassing less than 800 individuals per population. Kirch *et al.* (unpublished), however, managed to fill this gap with sample sizes of 3 273 PGTS and 5 490 RABS individuals. Although strong LD was estimated among some adaptive loci in proximity and especially for those within the same inversion, intra-chromosomal LD estimated in general in this study was rather low. LD among adaptive alleles located on different chromosomes was also determined to be low but significantly higher than expected by chance. These results suggest that adaptive ensembles are broken down into smaller allele blocks in the marine population. The detected linkage among adaptive alleles on different chromosomes might accelerate the process of reassembling adaptive alleles in the freshwater environment. Adaptive allele blocks within the same chromosome might present a strong genetic load on the marine population, providing a constraint on the maintenance of adaptive SGV and thus potentially retarding the rate of future adaptive evolution upon colonisation of a new freshwater habitat.

6.1.6. Re-assembly of freshwater-adaptive alleles

When a new lake population is founded by marine individuals, all marine individuals potentially contribute to the freshwater-adaptive allele pool in the population. The large clutch sizes of 40 to 450 eggs allow for various forms of meiotic recombination leading to large spectrum of new allele combinations in each generation upon which selection can act (Bell & Foster, 1994). The process of re-shuffling adaptive material might largely facilitate rapid adaptation. However, in case many neighbouring adaptive

alleles are uncoupled, several generations of recombination might be required to re-assemble the complete adaptive ensemble. In Kingman *et al.* (2021) and Kirch *et al.* (unpublished), we thus propose that rare individuals carrying adaptive alleles in linkage blocks might significantly contribute to the speed of adaptation accelerating the re-assembly of adaptive alleles. In addition, a large starting population might enable adaptation to occur via various routes in parallel using distinct freshwater-adaptive haplotypes present in the founder population and leading to soft sweeps. Kirch *et al.* (unpublished) exhibited different freshwater-adaptive haplotypes present in a freshwater population presumably originating from various carriers.

Kirch *et al.* 2021, however, depicted that the re-assembly of freshwater-adaptive alleles is not always successful. The ancient genome showed the presence of some freshwater-adaptive alleles that were not present in the contemporary lake population indicating that adaptive material can get lost over time. The founder population of this lake potentially didn't provide enough available SGV in the first place. For instance, frequencies of adaptive alleles could have been low and the founder population size too small. Furthermore, nearby adaptive loci could have been uncoupled retarding the re-assembly of adaptive alleles. In general, these results show that the re-assembly of adaptive alleles may be constrained by demographic stochasticity and that adaptation can be promoted but also impeded by the availability of SGV.

6.1.7. Geographic origin of adaptive alleles

Kirch *et al.* (unpublished) exhibited that various freshwater-adaptive haplotypes present in the marine population shared origin with global freshwater populations including Californian as well as Atlantic populations. Previous studies concluded that the Atlantic was colonised by Pacific sticklebacks in the Late Pleistocene 44.6 thousand years ago and concluded that there is no sign of trans-oceanic admixture following this colonisation (Fang *et al.*, 2018; Artamonova *et al.*, 2022; Fang *et al.*, 2020). Further Kingman *et al.* (2021) showed that divergent regions identified between Pacific marine and freshwater individuals spanned sevenfold more of the genome compared to those identified in global individuals. The colonisation of the Atlantic has thus presumably led to a loss of SGV. Accordingly, the Atlantic haplotype in the Alaskan population is probably not originating from the Atlantic, but rather vice versa: the freshwater-adaptive haplotype travelled to the Atlantic present as SGV in the marine population and was subsequently used for adaptation to Atlantic freshwater environments. This again highlights that freshwater-adaptive variants are ancient and have been present in marine stickleback populations for thousands of years. How haplotypes with genetic similarity to

those found in present day Californian populations came to be present in Alaskan marine sticklebacks is unknown, though is consistent with the transporter hypothesis and the spread of freshwater adaptive alleles among populations via introgression in hybrid zones at lower reaches of rivers (Jones et al., 2006; Schluter & Conte, 2009). Since the land in the Northern Hemisphere was largely covered by ice sheets during the late Pleistocene up to 20 000 years ago, introgression from local hybrid zones into Pacific marine populations was probably not occurring (Leverett, 1929). It is possible that freshwater adaptive SGV was introduced into marine populations via hybrid zones in California and was transported into Pacific populations via interconnectivity between marine populations. These results highlight that gene flow between different ecotypes via hybrid zones as well as interconnectivity between marine populations is likely to play an important role in the maintenance of SGV in a population.

6.1.8. Maintenance of SGV

Since some freshwater-adaptive haplotypes carried by marine individuals share origin with Californian or Atlantic freshwater populations, they are unlikely introduced into the marine population via local hybrid zones. Hence, the alleles are probably present in the marine populations for a long time, presumably thousands of years, and have not been removed from the population by selection during this time. Marine freshwater divergence at adaptive loci could be maintained by selection-migration balance: migration leads to neutral alleles homogenising rapidly among the ecotypes, whereas selection maintains differences at adaptive loci affecting fitness by purging alleles that are deleterious in the wrong environment (Barton & Hewitt, 1985; Ronce & Kirkpatrick, 2001). This means that freshwater adaptive alleles are maintained at high frequencies in the freshwater environment since marine alleles are deleterious in the freshwater environment, and marine alleles are maintained at high frequencies in the marine environment since freshwater alleles are deleterious in the marine environment. Marine individuals carrying freshwater adaptive SGV would thus have lower fitness in the marine environment and present a genetic load to the marine population. The effective population size of marine populations is considered to be very large meaning that selection is not obfuscated by drift, but should rather efficiently purge maladaptive alleles from the population (Liu *et al.*, 2016). Nevertheless, in Kirch *et al.* (unpublished) we have shown that all marine individuals carry a significant number of freshwater adaptive alleles, leading us to a fundamental conundrum: how are these individuals able to persist in the marine environment?

As suggested by Haenel et al, long-term selection-migration balance does not completely explain the persistence of adaptive alleles in the marine environment (Haenel *et al.* 2022). This study further proposes that selection abates with decreasing allele frequency enabling the neutral persistence of adaptive alleles.

242 regions were identified to be divergent between marine and freshwater sticklebacks thus suggesting a polygenic basis of adaptation in threespine sticklebacks (Jones *et al.*, 2012). Although the effect size of a single locus has to be still large enough to be parallel selected across multiple freshwater populations, it could be relatively small. If the freshwater-adaptive allele ensemble is sufficiently broken down the effect of carrying even several adaptive alleles might thus be a negotiable disadvantage in the marine environment. Since freshwater adaptive alleles are predominantly carried in heterozygous states as shown in Kirch *et al.* (unpublished), recessive alleles could hide in the marine population as cryptic genetic variation. The respective phenotype is not manifested if recessive alleles are carried in heterozygous state, and the trait is thus "invisible" for selection. Additive alleles, in contrast, should be rapidly purged from the marine population, according to population genetics theory (see e.g. Falconer & Mackay, 1996). Since both alleles affect a trait, selection can distinguish between homozygotes and heterozygotes and act efficiently upon the respective allele. Consistent with population genetics theory, there is some QTL mapping evidence suggesting that most skeletal trait differences between marine and freshwater sticklebacks are controlled by mutations that behave additively or partially additive with a minor bias towards recessivity (Miller *et al.*, 2014). Additionally, Verta and Jones have shown that gene expression differences between ecotypes are cis-regulated with additive cis regulatory effects (Verta & Jones, 2019). Combined these studies provide accumulating evidence that the adaptive divergence of sticklebacks to different habitats primarily involves mutations that behave additively, i.e. additive freshwater adaptive alleles are "visible" to selection in the marine environment and do present a genetic load to the population. Hence, the enigma still remains how freshwater adaptive alleles can persist and "hide" in the marine population. One possible explanation of this conundrum are environmentally dependent "dominance modifiers", i.e. alleles behave additively on trait variation in the freshwater environment, but recessively in the marine environment (Billiard et al., 2021; Fisher, 1928; Proulx & Teotónio, 2022; Schluter et al., 2021). This concept could apply to physiological, metabolic and behavioural traits. However, since marine sticklebacks are anadromous and migrate to freshwater during their life history, it is unlikely that morphological traits are affected by "dominance modifiers".

Another possible explanation could be that adaptive alleles do not hide in the marine population as cryptic genetic variation, but rather that balancing selection is

acting upon them. Selection on adaptive alleles might for instance vary over time and space. As anadromous marine sticklebacks migrate from the ocean to freshwater for spawning, freshwater adaptive alleles could be beneficial for breeding in the lower freshwater reaches of rivers, but deleterious in the ocean and could be maintained in the population due to contrarian selective pressures.

Two large effect loci of stickleback adaptation to freshwater environments are the *Eda* and the *Pitx1* locus. Deletion of a tissue-specific enhancer of the *Pitx1* gene has repeatedly caused the loss of pelvic spines in freshwater stickleback populations throughout the Northern Hemisphere and a *cis*-acting regulatory change at the *Eda* locus has repeatedly led to the loss of armour plates in global freshwater populations (Chan *et al.*, 2010; Colosimo *et al.*, 2005; O'Brown *et al.*, 2015). *Eda* is present as SGV in marine populations, but surprisingly the deletion near the *Pitx1* locus is not present. The mutation at the enhancer seems to appear *de novo* and not from pre-existing SGV with inherent structural features facilitating the *de novo* deletion (Chan *et al.*, 2010; Xie *et al.*, 2019). Pelvis reduction in sticklebacks has already been reported over a temporal sequence in fossils, suggesting that the mutation has been present for a long time and was theoretically available for introgression into the marine environment (Giles, 1983; Bell *et al.*, 1993; Hunt *et al.*, 2008). While both loci have been shown to underlie marine freshwater differences in major skeletal traits, the *Eda* locus is highly pleiotropic, affecting numerous other traits (Peichel & Marques, 2017). Dominance modifiers cannot easily act on morphological traits and hence, *Pitx1* remains additive and costly in the marine environment. In fact, *Pitx1* seems to be too deleterious in the marine environment to be harboured even at low frequencies in a marine population. In contrast, it is possible that dominance modifiers can act on *Eda*'s "other" functions which are not affecting morphology, enabling the haplotype to hide in marine environment for longer.

Our comprehensive empirical study displayed further that no F1s were present in large sample sizes of two marine populations despite high prevalence of SGV in one of them (PGTS). These results indicate that F1s are under strong selection in the marine environment. The completely linked freshwater-adaptive allele ensemble apparently presents a significant genetic load in the marine environment and thus leads to purging of F1s from the marine population.

6.2. What next in sticklebacks

Already Socrates stated: "For I was conscious that I knew practically nothing." and this also applies to this thesis (Plato, translated by Harold North Fowler, 1966). Although the major findings described in the previous paragraphs represent some

progress in understanding rapid adaptation, still a lot of questions remain unanswered. In order to disentangle them, more studies are required. Ideas for future research in sticklebacks have been elaborated in each chapter of this thesis. In the following paragraphs, I am presenting some general proposals for further studies resulting from the knowledge gained in this thesis.

6.2.1. Future research on SGV in general

In this thesis, SGV was characterised at a set of globally divergent regions. Kingman *et al.* (2021) identified Pacific divergent regions spanning sevenfold more of the genome compared to the globally divergent regions. Since the marine populations studied with large sample sizes in Kirch *et al.* (unpublished) are both Pacific populations, it would be highly interesting to perform whole genome sequencing with these samples and assess the SGV present at the Pacific divergent regions in these populations. Since the Pacific freshwater-adaptive alleles were probably introduced via local hybrid zones into the marine population, the introgression might have occurred more recently. Moreover, less haplotype diversity sharing origin with global populations would be expected in the Pacific SGV.

In addition, haplotype-phased data for many marine individuals would allow for a more precise estimate of haplotype block sizes as well as LD among adaptive alleles in the marine population. Until now, estimates in Kirch *et al.* (unpublished) rely on haplotype phasing interfered from genotyping data which could bias current results. To avoid this bias, potential re-sequencing of marine individuals could benefit from a novel method for short-read sequencing encompassing haplotype information called haplotype tagging (Meier *et al.*, 2021). Haplotype tagging was already used for studying freshwater individuals as well as a small set of marine carriers in this thesis (see chapter 5.3).

Another informative study would be the characterisation of SGV in many Atlantic marine stickleback to test the hypothesis that the haplotype diversity for freshwater-adaptive alleles present as SGV in the marine population was lost during the colonisation of the Atlantic (Fang et al., 2020). On the contrary, the diversity could be present in an Atlantic marine population, but not be utilized and present in Atlantic freshwater populations.

6.2.2. Future research for re-assembly

One topic that is in need of more elaborate research in the future is the re-assembly of freshwater-adaptive alleles during the adaptation of marine stickleback fish to freshwater environments. In general, this field of study is accessible for theoretical as

well as empirical studies. The availability of SGV empirically assessed in this thesis represents a great starting point for forward-in-time simulations. The information about the frequency as well as linkage of adaptive loci in a founder population in combination with the selection coefficients estimated for divergent regions in freshwater environments (Kingman *et al.*, 2021) and the available recombination maps for threespine sticklebacks can be used to simulate the possible re-assemble of freshwater-adaptive alleles under selection. A potential empirical approach could use haplotype tagging (Meier *et al.*, 2021) for further study the re-assembly of freshwater-adaptive alleles in stickleback adaptation to freshwater environments. Haplotype-phased whole genome sequencing data for the founder population of a freshwater lake as well as time-series data tracking the evolution of this adaptation would enable to track the re-assembly of freshwater-adaptive alleles over time. This data could elucidate to what extent SGV is used in rapid adaptation of sticklebacks and whether a few "jackpot carrier" or the majority of the founder population contribute to the adaptation process (Hohenlohe et al., 2012).

6.2.3. Selection in marine environment

The majority of empirical studies in stickleback adaptation have been focused on the selection pressures acting upon stickleback fish in freshwater environments (e.g. Barrett et al., 2008; Bassham et al., 2018; Kingman et al., 2021; Schluter et al., 2021). Accordingly, little is known about selection acting upon individuals in the marine environment. However, the importance of the availability of SGV for rapid adaptation in freshwater environments highlights that understanding the factors affecting this availability are relevant. Unfortunately, the study of selection in the marine environment is not trivial. Rearing freshwater sticklebacks or F1s in the laboratory in a tank with salinities according to potential salinities in the wild does not capture the full suite of biotic and abiotic differences and selection pressures between the habitats throughout the course of an individual's lifetime. Many environmental factors aside from salinity might be of importance for natural stickleback populations including temperature profile, the presence of predators and parasites and the available nutrients (Bell & Foster, 1994). A proper estimate of selection can only be obtained in a natural or naturalistic enough environment.

In order to overcome the challenge to simulate natural conditions in the laboratory, time series experiments could be performed in the wild. For instance, marine lagoons could be seeded with freshwater sticklebacks, a combination of freshwater and marine sticklebacks or F1s, and the populations could be tracked over time. Selection coefficients for divergent loci could subsequently be inferred from allele frequencies

changing over time. Furthermore, the break-down of the freshwater-adaptive allele ensemble in the marine environment could be traced over time.

Another way to study natural selection on marine individuals circumventing the difficulties of time-series data in marine environments, is to look at the genomes of marine sticklebacks sampled from the wild. Selection leaves molecular signatures in genomes which can be detected by various developed statistical tests (Kim & Stephan, 2002; Nielsen, 2005; Tajima, 1989). Genomic regions with loss of heterozygosity and loss of haplotype diversity might origin from long-term selection, whereas genomic regions showing extended haplotype block sizes and/or an excess of derived alleles might result from more recent selection (Sabeti *et al.*, 2006). The latter might be challenging to detect in sticklebacks, since marine populations cannot be used for determining the ancestral state. Genomes of outgroups are thus needed to define the ancestral and derived state. Besides, molecular signatures of selective sweeps can be detected based on their site frequency spectrum, for instance with Tajima's D (Kim & Stephan, 2002; Nielsen, 2005; Tajima, 1989).

Theoretical studies could complement empirical studies for selection in the marine environment. Different degrees of introgression and selection coefficients at the divergent loci could be simulated for a marine population and the consequential results could be compared to the assessed empirical availability of SGV.

6.2.4. Maintenance of SGV via dominance modifiers

One major conundrum in stickleback evolution is how adaptive SGV, i.e. potentially deleterious alleles, can be maintained in the marine environment. One possible mechanism, which I proposed in section 6.1.8, includes environmentally sensitive dominance modifiers affecting the additivity or dominance of alleles (as indicated in Schluter et al., 2021). In order to test the presence and importance of environmentally sensitive dominance modifiers, freshwater homozygous, marine homozygous parents as well as F1 hybrids could be raised at different salinities in the laboratory. Dominance modifiers are not expected to act upon morphological traits, thus excluding morphology as measurable trait to assess additivity and dominance. Since physiological, metabolic and behavioural traits are expected to be affected by dominance modifiers, gene expression would be a good candidate as measurable phenotype underlying all affected traits. Loci of interest are those with divergent gene expression levels in marine and freshwater parents living in their respectively favoured environment. Additivity as well as dominance coefficients can subsequently be calculated based on gene expression levels in F1 hybrids compared to the divergent expression levels of their

parents (Falconer & Mackay, 1996). Exactly intermediate expression levels in F1 hybrids would for instance correspond to a dominance coefficient h of 0.5 and pure additivity. Assessing dominance and additivity for F1 hybrids raised in low as well as high salinity, would enable us to disentangle whether these parameters are sensitive to divergent environmental conditions. Such an experiment could help to uncover the presence and importance of environmentally sensitive dominance modifiers in the maintenance of SGV.

6.3. Zoom out from sticklebacks to the broader community

6.3.1. Bioinformatic tools with sticklebacks as prototype

The findings in various study organisms contribute all to elucidating the mechanisms of evolution, but also promote the research progress for each other by developing universal tools. The bioinformatic tools developed in Kirch *et al.* (unpublished) for characterising SGV in populations can also be applied to other organisms. Specifically, the methods developed to assign ancestry states within divergent regions based on Hidden Markov Models or Bayesian probability, respectively, are transferable for studying genetic variation in other evolutionary model systems that experienced introgression from another ecotype or habitat.

6.3.2. Transferability of results to other organisms

The findings on the availability of SGV and its role for rapid adaptation in sticklebacks are applicable across taxa. While theory predicts that adaptation from SGV is common, little is known about the factors shaping and constraining the availability of SGV. This thesis, however, shows, consistent with previous studies in sticklebacks, that genetic variation provided by introgression can fuel rapid adaptation (Colosimo *et al.*, 2005; Schluter & Conte, 2009). This phenomena might apply for organisms that live in more than one type of habitat, e.g. in at least two types of divergent habitats, and whose habitat differences appear repeatedly. In fact, adaptation from SGV would be irrelevant if the two divergent habitats occurred exactly once each. Besides, the availability of SGV is rendered possible by gene flow happening across the distribution of the species despite potentially strong selection pressures. It is thus important that the species' habitats are interconnected and that the respective populations are not separated by prezygotic isolation barriers or intrinsic postzygotic isolation barriers.

The interconnectivity between habitats is for instance met by organisms showing hybrid zones. Multiple examples of hybrid zones have been reported (Barton & Hewitt,

1985) and this phenomena seems to be specifically abundant at the geographic midpoints between glacial refugia due to range expansion at the end of the last Ice Age, at the foothills of mountains and due to anthropogenic changes in the landscape (Remington, 1968). Hybrid zones have been uncovered for a wide range of species including insects (e.g. butterflies, grasshoppers), amphibia (e.g. frogs, lizard), birds, mammals (e.g. mice, seal) and plants (e.g. monkeyflowers) (Barton & Hewitt, 1985; Ottenburghs *et al.*, 2017; Shurtliff, 2013; Stankowski & Streisfeld, 2015) indicating that SGV is present in numerous organisms due to gene flow. Given the availability of SGV the question remains, whether these species can adapt as rapidly to new environments, e.g. within 10s of generations, as reported in sticklebacks. Rapid adaptation in sticklebacks is facilitated by several factors. Highly parallel habitat divergence has persisted for more than 15 million years for sticklebacks allowing many rounds of improving freshwater adaptive haplotypes (Nelson & Cresko, 2018). Besides, sticklebacks show short generation time of 1 to 3 years, large clutch sizes (40 to 450 eggs) , large effective population sizes as well as high recombination rates (Bell & Foster, 1994; S. Liu *et al.*, 2016; Roesti *et al.*, 2013; Wootton, 1984). These factors driving adaptation are also present in some insect species. For instance, *Heliconius* butterflies show short generation times of minimally one month and medium clutch sizes and have repeatedly and independently colonised montane habitats and rapidly adapted to the new environment presumably using pre-existing SGV (Kronforst, 2008; Montejo-Kovacevich et al., 2022; Reed, 2003). As displayed in *Heliconius* butterflies, hybrid zones at the foothills of mountains show often parallel habitat differences allowing SGV to act as source of variation in independent, repeated colonisations of higher altitudes promoting rapid adaptation.

SGV has been maintained in marine stickleback populations over millions of years, i.e. over geological timescales. How many other organisms have experienced such a long-term habitat divergence enabling the optimisation of adaptive haplotypes? As an example, continuous gene flow in *Heliconius* butterflies has persisted for similar timescales, i.e. for millions of years (Merrill *et al.*, 2015). Such introgressive hybridisation can expand the evolutionary potential as well as fuel rapid adaptation by introducing genetic variation in the source population. (Marques et al., 2019; Oziolor et al., 2019)

Aside from the age of SGV, some other factors facilitating rapid adaptation in sticklebacks, might not be available for a large range of species indicating that the speed of adaptation could be retarded for them. Firstly, large clutch sizes are not at disposal for all organisms. In particular, mammals and birds often show clutch sizes which are two magnitude smaller than in sticklebacks(Charnov & Ernest, n.d.; Godfray et al., 1991; Zammuto, 1986). Large clutch sizes provide a platform for multiple allelic re-assemblies

due to meiotic recombination in only generation. Taxa with large reproductive output are thus predicted to reassemble polygenic alleles rapidly over short number of generations. Organisms with small clutch sizes, in contrast, could require up to two magnitudes of generations more to achieve the same result given the same recombination rate.

Furthermore, species might comprise small effective population sizes, e.g. due to habitat fragmentation or inbreeding (Fahrig, 2003). In this thesis, we showed that adaptation can be constrained by small effective population sizes leading to stochastic loss of adaptive alleles. Small effective population sizes of organisms can result in incomplete or unsuccessful adaptation and, in the worst case, lead to extinction (Frankham et al., 2014; Franklin, 1980; Franklin et al., 2014; Pérez-Pereira et al., 2022; Soulé, 1980). This aspect will be elaborated in section 6.3.3.

Knowing that several features affect the evolutionary potential of an organism, the question arises whether we can actually predict the speed and success of adaptation. In order to test theoretical predictions, the genetic architecture of rapidly evolving traits needs to be uncovered (Whiting *et al.*, 2022). In Kingman *et al.*, we showed that the magnitude of ecotypic differentiation is the best predictor for the rate of evolution. Alleles that are broadly and repeatedly positively selected, are being selected more quickly in a new environment (Kingman *et al.*, 2021). In sticklebacks, 242 loci have been identified to comprise the polygenic genomic basis of parallel adaptation of marine sticklebacks to freshwater environments (Jones *et al.*, 2012). In general, polygenic genetic architecture might promote rapid adaptation allowing populations to shift fitness optima by many small frequency changes of alleles already present as SGV in the source population (Jain & Stephan, 2017). Furthermore, regions overlapping major QTLs displayed strong association with rapid evolution in Kingman *et al.* (2021). The maintenance of large effect loci as SGV in the maladaptive environment depends on the severity of their genetic load. When these maladaptive alleles are for instance highly deleterious in the environment, it is unlikely that they persist and stay available for future adaptations. In addition, adaptive alleles carried in linkage blocks could accelerate the re-assembly of adaptive alleles to a complete allele ensemble in the new environment and thus facilitate rapid adaptation. Exploring the underlying genetic architecture of adaptive evolution might help to predict the evolutionary potential of an organism.

6.3.3. The role of SGV in conservation genomics

Being able to predict how taxa (including agricultural species) will respond to rapid environmental changes is fundamental to understanding broader ecosystem changes. Can we predict which species are likely to be able to adapt rapidly enough?

For decades, SGV has been considered as a fundamental factor maintaining the evolutionary potential of organisms (Crandall *et al.*, 2000; Moritz, 2002). With the rise of the genomic era and the availability of performing whole genome sequencing for many samples and organisms, conservation genomics has become an important part of conservation (Supple & Shapiro, 2018). However, since the field is rather new, it is challenging to transform the information about genomes of an organism into an estimate for its risk of extinction (Khan *et al.*, 2016). Effective population size (N_e) is used to determine the degree to which a population or species is threatened with extinction and to infer breeding policies for wild populations as well as livestock (Jiménez-Mena *et al.*, 2016). Management strategies predominantly encompass the maximisation of N_e and try to prevent the loss of genetic diversity (Ballou & Lacy, 1995). Lower limits for N_e of originally 50 and 500 (and later 100 and 1000) have been proposed to impede inbreeding depression for the short term and preserve evolutionary potential of a population in the long run, respectively (Frankham *et al.*, 2014; Franklin, 1980; Soulé, 1980). However, these values remain the subject of a controversial discussion since it is unclear which portion of the genetic diversity is beneficial for future adaptations (Frankham *et al.*, 2014; Franklin *et al.*, 2014; Rosenfeld, 2014). The minimum size for a viable population further depends on the reproductive rate of an organism (Pérez-Pereira et al., 2022).

It appears that SGV is most useful for rapid adaptation if the SGV is originating from populations which have experienced similar selective pressure in the past, i.e. SGV is relevant for future adaptations. In contrast, if we started with a clonal stickleback population, mutagenized the individuals so that they carry random SGV, and then put them into a new freshwater habitat, they would be unlikely to adapt anywhere near as rapidly as what sticklebacks have been observed to do. In addition, despite millions of years of evolution, sticklebacks are only occurring in the Northern Hemisphere and have apparently not managed to colonise warmer equatorial habitats (Bell & Foster, 1994). Although SGV has been maintained across geological timescales in sticklebacks, the SGV has not been relevant for the adaptation to warmer habitats and the abundance of the species has thus been limited to the Northern Hemisphere. An objective in conservation management should thus be to not only maximise the effective population size and therefore, the genetic diversity in a species in general, but rather to maintain interconnectivity especially among populations of a species that might provide relevant SGV for future adaptations . One remaining question is how much relevant SGV needs to be available for facilitating rapid adaptation and what size a founder population needs to encompass, in order to assure the success of adaptation to a new or changing environment. The rapid adaptation of RABS individuals to the freshwater environment in multiple lakes shown in this thesis, highlights for instance that low availability of SGV

does not contradict rapid adaptation. Further research is required to identify species missing sufficient relevant SGV and large effective population sizes for future rapid adaptation prematurely and to find ways to counteract their extinction and ensure the maintenance of biodiversity on this planet.

6.4. Take home message

This thesis provides insights into the mechanisms of rapid adaptation in natural populations and emphasises the evolutionary importance of SGV. The findings presented here set a foundation for future studies on rapid adaptation and encourage amongst others the further investigation of the re-assembly of adaptive alleles based on the availability of SGV. This research forms the basis of conservation genomics and is crucial for ensuring the maintenance of biodiversity.

7. References

Aguirre, W. E., & Bell, M. A. (2012). Twenty years of body shape evolution in a threespine stickleback population adapting to a lake environment: STICKLEBACK BODY SHAPE EVOLUTION. *Biological Journal of the Linnean Society*, *105*(4), 817–831. https://doi.org/10.1111/j.1095-8312.2011.01825.x

Andrés, A. M., Hubisz, M. J., Indap, A., Torgerson, D. G., Degenhardt, J. D., Boyko, A. R., Gutenkunst, R. N., White, T. J., Green, E. D., Bustamante, C. D., Clark, A. G., & Nielsen, R. (2009). Targets of balancing selection in the human genome. *Molecular Biology and Evolution*, *26*(12), 2755–2764. https://doi.org/10.1093/molbev/msp190

Aparicio, S., Chapman, J., Stupka, E., Putnam, N., Chia, J.-M., Dehal, P., Christoffels, A., Rash, S., Hoon, S., Smit, A., Gelpke, M. D. S., Roach, J., Oh, T., Ho, I. Y., Wong, M., Detter, C., Verhoef, F., Predki, P., Tay, A., … Brenner, S. (2002). Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes. *Science (New York, N.Y.)*, *297*(5585), 1301–1310. https://doi.org/10.1126/science.1072104

Arnone, M. I., & Davidson, E. H. (1997). The hardwiring of development: Organization and function of genomic regulatory systems. *Development (Cambridge, England)*, *124*(10), 1851–1864. https://doi.org/10.1242/dev.124.10.1851

Artamonova, V. S., Bardukov, N. V., Aksenova, O. V., Ivanova, T. S., Ivanov, M. V., Kirillova, E. A., Koulish, A. V., Lajus, D. L., Malyutina, A. M., Pashkov, A. N., Reshetnikov, S. I., & Makhrov, A. A. (2022). Round-the-World Voyage of the Threespine Stickleback (Gasterosteus aculeatus): Phylogeographic Data Covering the Entire Species Range. *Water*, *14*(16), Article 16. https://doi.org/10.3390/w14162484

Asthana, S., Schmidt, S., & Sunyaev, S. (2005). A limited role for balancing selection. *Trends in Genetics: TIG*, *21*(1), 30–32. https://doi.org/10.1016/j.tig.2004.11.001

Ballou & Lacy. (1995). Identifying genetically important individuals for management of genetic variation in pedigreed populations. *Population Management for Survival and Recovery*. https://cir.nii.ac.jp/crid/1570291225569333632

Barrett, R. D. H., Rogers, S. M., & Schluter, D. (2008). Natural Selection on a Major Armor Gene in Threespine Stickleback. *Science*, *322*(5899), 255–257. https://doi.org/10.1126/science.1159978

Barrett, R. D. H., & Schluter, D. (2008). Adaptation from standing genetic variation. *Trends in Ecology & Evolution*, *23*(1), 38–44. https://doi.org/10.1016/j.tree.2007.09.008

Barton, N. H. (1998). The effect of hitch-hiking on neutral genealogies. *Genetical Research*, *72*(2), 123–133. https://doi.org/10.1017/S0016672398003462

Barton, N. H., Etheridge, A. M., & Véber, A. (2017). The infinitesimal model: Definition, derivation, and implications. *Theoretical Population Biology*, *118*, 50–73. https://doi.org/10.1016/j.tpb.2017.06.001

Barton, N. H., & Hewitt, G. M. (1985). Analysis of Hybrid Zones. *Annual Review of Ecology and Systematics*, *16*, 113–148.

Barton, N. H., & Keightley, P. D. (2002). Understanding quantitative genetic variation. *Nature Reviews Genetics*, *3*(1), Article 1. https://doi.org/10.1038/nrg700

Bassham, S., Catchen, J., Lescak, E., von Hippel, F. A., & Cresko, W. A. (2018). Repeated Selection of Alternatively Adapted Haplotypes Creates Sweeping Genomic Remodeling in Stickleback. *Genetics*, genetics.300610.2017. https://doi.org/10.1534/genetics.117.300610

Bell, M. A. (1984). Evolutionary Phenetics and Genetics. In B. J. Turner (Ed.), *Evolutionary Genetics of Fishes* (pp. 431–528). Springer US. https://doi.org/10.1007/978-1-4684-4652-4_9

Bell, M. A., & Aguirre, W. E. (2013). Contemporary evolution, allelic recycling, and adaptive radiation of the threespine stickleback. *Evolutionary Ecology Research*, *15*, 377–411.

Bell, M. A., Aguirre, W. E., & Buck, N. J. (2004). Twelve Years of Contemporary Armor Evolution in a Threespine Stickleback Population. *Evolution*, *58*(4), 814–824. https://doi.org/10.1111/j.0014-3820.2004.tb00414.x

Bell, M. A., & Foster, S. A. (1994). *The Evolutionary Biology of the Threespine Stickleback*. Oxford University Press.

Bell, M. A., Ortí, G., Walker, J. A., & Koenings, J. P. (1993). Evolution of Pelvic Reduction in Threespine Stickleback Fish: A Test of Competing Hypotheses. *Evolution*, *47*(3), 906–914. https://doi.org/10.1111/j.1558-5646.1993.tb01243.x

Berg, J. J., Harpak, A., Sinnott-Armstrong, N., Joergensen, A. M., Mostafavi, H., Field, Y., Boyle, E. A., Zhang, X., Racimo, F., Pritchard, J. K., & Coop, G. (2019). Reduced signal for polygenic adaptation of height in UK Biobank. *ELife*, *8*, e39725. https://doi.org/10.7554/eLife.39725

Bernatchez, L. (2016). On the maintenance of genetic variation and adaptation to environmental change: Considerations from population genomics in fishes. *Journal of Fish Biology*, *89*(6), 2519–2556. https://doi.org/10.1111/jfb.13145

Bersaglieri, T., Sabeti, P. C., Patterson, N., Vanderploeg, T., Schaffner, S. F., Drake, J. A., Rhodes, M., Reich, D. E., & Hirschhorn, J. N. (2004). Genetic Signatures of

Strong Recent Positive Selection at the Lactase Gene. *The American Journal of Human Genetics*, *74*(6), 1111–1120. https://doi.org/10.1086/421051

Billiard, S., Castric, V., & Llaurens, V. (2021). The integrative biology of genetic dominance. *Biological Reviews*, *96*(6), 2925–2942. https://doi.org/10.1111/brv.12786

Bitarello, B. D., de Filippo, C., Teixeira, J. C., Schmidt, J. M., Kleinert, P., Meyer, D., & Andrés, A. M. (2018). Signatures of Long-Term Balancing Selection in Human Genomes. *Genome Biology and Evolution*, *10*(3), 939–955. https://doi.org/10.1093/gbe/evy054

Bohutínská, M., Vlček, J., Yair, S., Laenen, B., Konečná, V., Fracassetti, M., Slotte, T., & Kolář, F. (2021). Genomic basis of parallel adaptation varies with divergence in Arabidopsis and its relatives. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(21), e2022713118. https://doi.org/10.1073/pnas.2022713118

Bomblies, K., & Peichel, C. L. (2022). Genetics of adaptation. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(30), e2122152119. https://doi.org/10.1073/pnas.2122152119

Brawand, D., Wagner, C. E., Li, Y. I., Malinsky, M., Keller, I., Fan, S., Simakov, O., Ng, A. Y., Lim, Z. W., Bezault, E., Turner-Maier, J., Johnson, J., Alcazar, R., Noh, H. J., Russell, P., Aken, B., Alföldi, J., Amemiya, C., Azzouzi, N., … Di Palma, F. (2014). The genomic substrate for adaptive radiation in African cichlid fish. *Nature*, *513*(7518), 375–381. https://doi.org/10.1038/nature13726

Carroll, S. B. (2000). Endless Forms: The Evolution of Gene Regulation and Morphological Diversity. *Cell*, *101*(6), 577–580. https://doi.org/10.1016/S0092-8674(00)80868-5

Carroll, S. B. (2008). Evo-devo and an expanding evolutionary synthesis: A genetic theory of morphological evolution. *Cell*, *134*(1), 25–36. https://doi.org/10.1016/j.cell.2008.06.030

Carter, A. J. R., Hermisson, J., & Hansen, T. F. (2005). The role of epistatic gene interactions in the response to selection and the evolution of evolvability. *Theoretical Population Biology*, *68*(3), 179–196. https://doi.org/10.1016/j.tpb.2005.05.002

Castro, J. P., Yancoskie, M. N., Marchini, M., Belohlavy, S., Hiramatsu, L., Kučka, M., Beluch, W. H., Naumann, R., Skuplik, I., Cobb, J., Barton, N. H., Rolian, C., & Chan, Y. F. (2019). An integrative genomic analysis of the Longshanks selection experiment for longer limbs in mice. *ELife*, *8*, e42014. https://doi.org/10.7554/eLife.42014

Ceballos, G., Ehrlich, P. R., Barnosky, A. D., García, A., Pringle, R. M., & Palmer, T. M. (2015). Accelerated modern human–induced species losses: Entering the sixth mass extinction. *Science Advances*, *1*(5), e1400253. https://doi.org/10.1126/sciadv.1400253

Chan, Y. F., Marks, M. E., Jones, F. C., Villarreal, G., Shapiro, M. D., Brady, S. D., Southwick, A. M., Absher, D. M., Grimwood, J., Schmutz, J., Myers, R. M., Petrov, D., Jónsson, B., Schluter, D., Bell, M. A., & Kingsley, D. M. (2010). Adaptive Evolution of Pelvic Reduction in Sticklebacks by Recurrent Deletion of a Pitx1 Enhancer. *Science*, *327*(5963), 302–305. https://doi.org/10.1126/science.1182213

Charlesworth, D. (2006). Balancing Selection and Its Effects on Sequences in Nearby Genome Regions. *PLoS Genetics*, *2*(4), e64. https://doi.org/10.1371/journal.pgen.0020064

Charnov, E. L., & Ernest, S. K. M. (n.d.). *The Offspring-Size/Clutch-Size Trade-Off in Mammals.* 5.

Chevalier, F. D., Clec'h, W. L., McDew-White, M., Menon, V., Guzman, M. A., Holloway, S. P., Cao, X., Taylor, A. B., Kinung'hi, S., Gouvras, A. N., Webster, B. L., Webster, J. P., Emery, A. M., Rollinson, D., Djirmay, A. G., Mashikhi, K. M. A., Yafae, S. A., Idris, M. A., Moné, H., … Anderson, T. J. C. (2019). Oxamniquine resistance alleles are widespread in Old World Schistosoma mansoni and predate drug deployment. *PLOS Pathogens*, *15*(10), e1007881. https://doi.org/10.1371/journal.ppat.1007881

Cockram, J., Mackay, I. J., & O'Sullivan, D. M. (2007). The Role of Double-Stranded Break Repair in the Creation of Phenotypic Diversity at Cereal *VRN1* Loci. *Genetics*, *177*(4), 2535–2539. https://doi.org/10.1534/genetics.107.074765

Colosimo, P. F., Hosemann, K. E., Balabhadra, S., Villarreal, G., Dickson, M., Grimwood, J., Schmutz, J., Myers, R. M., Schluter, D., & Kingsley, D. M. (2005). Widespread Parallel Evolution in Sticklebacks by Repeated Fixation of Ectodysplasin Alleles. *Science*, *307*(5717), 1928–1933. https://doi.org/10.1126/science.1107239

Colosimo, P. F., Peichel, C. L., Nereng, K., Blackman, B. K., Shapiro, M. D., Schluter, D., & Kingsley, D. M. (2004). The Genetic Architecture of Parallel Armor Plate Reduction in Threespine Sticklebacks. *PLoS Biology*, *2*(5), e109. https://doi.org/10.1371/journal.pbio.0020109

Connallon, T., & Clark, A. G. (2014). Balancing Selection in Species with Separate Sexes: Insights from Fisher's Geometric Model. *Genetics*, *197*(3), 991–1006. https://doi.org/10.1534/genetics.114.165605

Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [H.-O. Pörtner, D.C. Roberts, M. Tignor, E.S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, B. Rama (eds.)]. (2022). *IPCC, 2022: Climate Change 2022: Impacts, Adaptation, and Vulnerability.* Cambridge University Press.

Coop, G., Pickrell, J. K., Novembre, J., Kudaravalli, S., Li, J., Absher, D., Myers, R. M., Cavalli-Sforza, L. L., Feldman, M. W., & Pritchard, J. K. (2009). The Role of Geography in Human Adaptation. *PLOS Genetics*, *5*(6), e1000500. https://doi.org/10.1371/journal.pgen.1000500

Crandall, K. A., Bininda-Emonds, O. R., Mace, G. M., & Wayne, R. K. (2000). Considering evolutionary processes in conservation biology. *Trends in Ecology & Evolution*, *15*(7), 290–295. https://doi.org/10.1016/s0169-5347(00)01876-0

Cresko, W. A., Amores, A., Wilson, C., Murphy, J., Currey, M., Phillips, P., Bell, M. A., Kimmel, C. B., & Postlethwait, J. H. (2004). Parallel genetic basis for repeated evolution of armor loss in Alaskan threespine stickleback populations. *Proceedings of the National Academy of Sciences*, *101*(16), 6050–6055. https://doi.org/10.1073/pnas.0308479101

Csilléry, K., Rodríguez-Verdugo, A., Rellstab, C., & Guillaume, F. (2018). Detecting the genomic signal of polygenic adaptation and the role of epistasis in evolution. *Molecular Ecology*, *27*(3), 606–612. https://doi.org/10.1111/mec.14499

Darwin, C., & Kebler, L. (1859). *On the origin of species by means of natural selection, or, The preservation of favoured races in the struggle for life .* London: J. Murray.

DeWoody, J. A., Harder, A. M., Mathur, S., & Willoughby, J. R. (2021). The long-standing significance of genetic diversity in conservation. *Molecular Ecology*, *30*(17), 4147–4154. https://doi.org/10.1111/mec.16051

Di-Poi, C., Lacasse, J., Rogers, S. M., & Aubin-Horth, N. (2014). Extensive Behavioural Divergence following Colonisation of the Freshwater Environment in Threespine Sticklebacks. *PLoS ONE*, *9*(6), e98980. https://doi.org/10.1371/journal.pone.0098980

Domingues, V. S., Poh, Y.-P., Peterson, B. K., Pennings, P. S., Jensen, J. D., & Hoekstra, H. E. (2012). Evidence of adaptation from ancestral variation in young populations of beach mice. *Evolution; International Journal of Organic Evolution*, *66*(10), 3209–3223. https://doi.org/10.1111/j.1558-5646.2012.01669.x

Durbin, R. M., Altshuler, D., Durbin, R. M., Abecasis, G. R., Bentley, D. R., Chakravarti, A., Clark, A. G., Collins, F. S., De La Vega, F. M., Donnelly, P., Egholm, M., Flicek, P., Gabriel, S. B., Gibbs, R. A., Knoppers, B. M., Lander, E. S., Lehrach,

H., Mardis, E. R., McVean, G. A., … The Translational Genomics Research Institute. (2010). A map of human genome variation from population-scale sequencing. *Nature*, *467*(7319), Article 7319. https://doi.org/10.1038/nature09534

Fahrig, L. (2003). Effects of Habitat Fragmentation on Biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, *34*(1), 487–515. https://doi.org/10.1146/annurev.ecolsys.34.011802.132419

Falconer, D. S., & Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics* (4th edition). Longman.

Fang, B., Kemppainen, P., Momigliano, P., Feng, X., & Merilä, J. (2020). On the causes of geographically heterogeneous parallel evolution in sticklebacks. *Nature Ecology & Evolution*. https://doi.org/10.1038/s41559-020-1222-6

Fang, B., Merilä, J., Ribeiro, F., Alexandre, C. M., & Momigliano, P. (2018). Worldwide phylogeny of three-spined sticklebacks. *Molecular Phylogenetics and Evolution*, *127*, 613–625. https://doi.org/10.1016/j.ympev.2018.06.008

Fijarczyk, A., & Babik, W. (2015). Detecting balancing selection in genomes: Limits and prospects. *Molecular Ecology*, *24*(14), 3529–3545. https://doi.org/10.1111/mec.13226

Fisher, R. A. (1918). *009: The Correlation Between Relatives on the Supposition of Mendelian Inheritance.* https://digital.library.adelaide.edu.au/dspace/handle/2440/15097

Fisher, R. A. (1928). The Possible Modification of the Response of the Wild Type to Recurrent Mutations. *The American Naturalist*, *62*(679), 115–126. https://doi.org/10.1086/280193

Foerster, K., Coulson, T., Sheldon, B. C., Pemberton, J. M., Clutton-Brock, T. H., & Kruuk, L. E. B. (2007). Sexually antagonistic genetic variation for fitness in red deer. *Nature*, *447*(7148), 1107–1110. https://doi.org/10.1038/nature05912

Frankham, R., Bradshaw, C. J. A., & Brook, B. W. (2014). 50/500 rules need upward revision to 100/1000 – Response to Franklin et al. *Biological Conservation*, *176*, 286. https://doi.org/10.1016/j.biocon.2014.05.006

Franklin, I. R. (1980). *Evolutionary change in small populations.* Sunderland, Massachusetts, Sinauer Associates, U.S.A. https://publications.csiro.au/rpr/pub?list=BRO&pid=procite:46c4045a-9333-43d5-973b-6f55ad9a9cc8

Franklin, I. R., Allendorf, F. W., & Jamieson, I. G. (2014). The 50/500 rule is still valid – Reply to Frankham et al. *Biological Conservation*, *176*, 284–285. https://doi.org/10.1016/j.biocon.2014.05.004

Galloway, J., Cresko, W. A., & Ralph, P. (2020). A Few Stickleback Suffice for the Transport of Alleles to New Lakes. *G3 (Bethesda, Md.)*, *10*(2), 505–514. https://doi.org/10.1534/g3.119.400564

Gibson, G., & Dworkin, I. (2004). Uncovering cryptic genetic variation. *Nature Reviews Genetics*, *5*(9), 681–690. https://doi.org/10.1038/nrg1426

Giles, N. (1983). The possible role of environmental calcum levels during the evolution of phenotypic diversity in Outer Hebridean populations of the Three-spined stickleback, Gasterosteus aculeatus. *Journal of Zoology*, *199*(4), 535–544. https://doi.org/10.1111/j.1469-7998.1983.tb05104.x

Glazer, A. M., Killingbeck, E. E., Mitros, T., Rokhsar, D. S., & Miller, C. T. (2015). Genome Assembly Improvement and Mapping Convergently Evolved Skeletal Traits in Sticklebacks with Genotyping-by-Sequencing. *G3 Genes|Genomes|Genetics*, *5*(7), 1463–1472. https://doi.org/10.1534/g3.115.017905

Godfray, H. C. J., Partridge, L., & work(s):, P. H. H. R. (1991). Clutch Size. *Annual Review of Ecology and Systematics*, *2*, 409–429.

Grossman, S. R., Andersen, K. G., Shlyakhter, I., Tabrizi, S., Winnicki, S., Yen, A., Park, D. J., Griesemer, D., Karlsson, E. K., Wong, S. H., Cabili, M., Adegbola, R. A., Bamezai, R. N. K., Hill, A. V. S., Vannberg, F. O., Rinn, J. L., Lander, E. S., Schaffner, S. F., & Sabeti, P. C. (2013). Identifying Recent Adaptations in Large-Scale Genomic Data. *Cell*, *152*(4), 703–713. https://doi.org/10.1016/j.cell.2013.01.035

Haenel, Q., Guerard, L., MacColl, A. D. C., & Berner, D. (2022). The maintenance of standing genetic variation: Gene flow vs. selective neutrality in Atlantic stickleback fish. *Molecular Ecology*, *31*(3), 811–821. https://doi.org/10.1111/mec.16269

Hagen, D. W. (1967). Isolating Mechanisms in Threespine Sticklebacks (Gasterosteus). *Journal of the Fisheries Research Board of Canada*, *24*(8), 1637–1692. https://doi.org/10.1139/f67-138

Hamblin, M. T., & Di Rienzo, A. (2000). Detection of the signature of natural selection in humans: Evidence from the Duffy blood group locus. *American Journal of Human Genetics*, *66*(5), 1669–1679.

Hansen, T. F. (2013). Why Epistasis Is Important for Selection and Adaptation. *Evolution*, *67*(12), 3501–3511. https://doi.org/10.1111/evo.12214

Harrison, E. by R. G. (Ed.). (1993). *Hybrid Zones and the Evolutionary Process*. Oxford University Press.

Hartl, D. L., & Clark, A. G. (2007). *Principles of Population Genetics* (4th ed. 2007). Sinauer.

Hendry, A. P., Bolnick, D. I., Berner, D., & Peichel, C. L. (2009). Along the speciation continuum in sticklebacks. *Journal of Fish Biology*, *75*(8), 2000–2036. https://doi.org/10.1111/j.1095-8649.2009.02419.x

Hermisson, J., & Pennings, P. S. (2005). Soft Sweeps: Molecular population genetics of adaptation from standing genetic variation. *Genetics*, *169*(4), 2335–2352. https://doi.org/10.1534/genetics.104.036947

Hermisson, J., & Pennings, P. S. (2017). Soft sweeps and beyond: Understanding the patterns and probabilities of selection footprints under rapid adaptation. *Methods in Ecology and Evolution*, *8*(6), 700–716. https://doi.org/10.1111/2041-210X.12808

Hoekstra, H. E., Hirschmann, R. J., Bundey, R. A., Insel, P. A., & Crossland, J. P. (2006). A Single Amino Acid Mutation Contributes to Adaptive Beach Mouse Color Pattern. *Science*, *313*(5783), 101–104. https://doi.org/10.1126/science.1126121

Hohenlohe, P. A., Bassham, S., Currey, M., & Cresko, W. A. (2012). Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1587), 395–408. https://doi.org/10.1098/rstb.2011.0245

Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., & Cresko, W. A. (2010). Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags. *PLoS Genetics*, *6*(2), e1000862. https://doi.org/10.1371/journal.pgen.1000862

Hunt, G., Bell, M. A., & Travis, M. P. (2008). EVOLUTION TOWARD A NEW ADAPTIVE OPTIMUM: PHENOTYPIC EVOLUTION IN A FOSSIL STICKLEBACK LINEAGE. *Evolution*, *62*(3), 700–710. https://doi.org/10.1111/j.1558-5646.2007.00310.x

Innan, H., & Kim, Y. (2004). Pattern of polymorphism after strong artificial selection in a domestication event. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(29), 10667–10672. https://doi.org/10.1073/pnas.0401720101

Jaillon, O., Aury, J.-M., Brunet, F., Petit, J.-L., Stange-Thomann, N., Mauceli, E., Bouneau, L., Fischer, C., Ozouf-Costaz, C., Bernot, A., Nicaud, S., Jaffe, D., Fisher, S., Lutfalla, G., Dossat, C., Segurens, B., Dasilva, C., Salanoubat, M., Levy, M., … Roest Crollius, H. (2004). Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype. *Nature*, *431*(7011), Article 7011. https://doi.org/10.1038/nature03025

Jain, K., & Stephan, W. (2017). Modes of Rapid Polygenic Adaptation. *Molecular Biology and Evolution*, *34*(12), 3169–3175. https://doi.org/10.1093/molbev/msx240

Jeong, S., Rebeiz, M., Andolfatto, P., Werner, T., True, J., & Carroll, S. B. (2008). The evolution of gene regulation underlies a morphological difference between two Drosophila sister species. *Cell*, *132*(5), 783–793. https://doi.org/10.1016/j.cell.2008.01.014

Jiménez-Mena, B., Hospital, F., & Bataillon, T. (2016). Heterogeneity in effective population size and its implications in conservation genetics and animal breeding. *Conservation Genetics Resources*, *8*(1), 35–41. https://doi.org/10.1007/s12686-015-0508-5

Johnston, S. E., Gratten, J., Berenos, C., Pilkington, J. G., Clutton-Brock, T. H., Pemberton, J. M., & Slate, J. (2013). Life history trade-offs at a single locus maintain sexually selected genetic variation. *Nature*, *502*(7469), 93–95. https://doi.org/10.1038/nature12489

Johnston, S. E., McEWAN, J. C., Pickering, N. K., Kijas, J. W., Beraldi, D., Pilkington, J. G., Pemberton, J. M., & Slate, J. (2011). Genome-wide association mapping identifies the genetic basis of discrete and quantitative variation in sexual weaponry in a wild sheep population. *Molecular Ecology*, *20*(12), 2555–2566. https://doi.org/10.1111/j.1365-294X.2011.05076.x

Jones, F. C., Brown, C., Pemberton, J. M., & Braithwaite, V. A. (2006). Reproductive isolation in a threespine stickleback hybrid zone. *Journal of Evolutionary Biology*, *19*(5), 1531–1544. https://doi.org/10.1111/j.1420-9101.2006.01122.x

Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., Swofford, R., Pirun, M., Zody, M. C., White, S., Birney, E., Searle, S., Schmutz, J., Grimwood, J., Dickson, M. C., Myers, R. M., Miller, C. T., Summers, B. R., Knecht, A. K., … Kingsley, D. M. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, *484*(7392), 55–61. https://doi.org/10.1038/nature10944

Jorde, L. B., & Wooding, S. P. (2004). Genetic variation, classification and 'race'. *Nature Genetics*, *36*(11), Article 11. https://doi.org/10.1038/ng1435

Kaplan, N. L., Hudson, R. R., & Langley, C. H. (1989). The ``hitchhiking Effect'' Revisited. *Genetics*, *123*(4), 887–899.

Karczewski, K. J., Weisburd, B., Thomas, B., Solomonson, M., Ruderfer, D. M., Kavanagh, D., Hamamsy, T., Lek, M., Samocha, K. E., Cummings, B. B., Birnbaum, D., The Exome Aggregation Consortium, Daly, M. J., & MacArthur, D. G. (2017). The ExAC browser: Displaying reference data information from over

60 000 exomes. *Nucleic Acids Research*, *45*(D1), D840–D845. https://doi.org/10.1093/nar/gkw971

Kardos, M., Armstrong, E. E., Fitzpatrick, S. W., Hauser, S., Hedrick, P. W., Miller, J. M., Tallmon, D. A., & Funk, W. C. (2021). The crucial role of genome-wide genetic variation in conservation. *Proceedings of the National Academy of Sciences*, *118*(48), e2104642118. https://doi.org/10.1073/pnas.2104642118

Kasahara, M., Naruse, K., Sasaki, S., Nakatani, Y., Qu, W., Ahsan, B., Yamada, T., Nagayasu, Y., Doi, K., Kasai, Y., Jindo, T., Kobayashi, D., Shimada, A., Toyoda, A., Kuroki, Y., Fujiyama, A., Sasaki, T., Shimizu, A., Asakawa, S., … Kohara, Y. (2007). The medaka draft genome and insights into vertebrate genome evolution. *Nature*, *447*(7145), Article 7145. https://doi.org/10.1038/nature05846

Khan, S., Nabi, G., Ullah, M. W., Yousaf, M., Manan, S., Siddique, R., & Hou, H. (2016). Overview on the Role of Advance Genomics in Conservation Biology of Endangered Species. *International Journal of Genomics*, *2016*, 3460416. https://doi.org/10.1155/2016/3460416

Kim, Y., & Stephan, W. (2002). Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics*, *160*(2), 765–777. https://doi.org/10.1093/genetics/160.2.765

King, M. C., & Wilson, A. C. (1975). Evolution at two levels in humans and chimpanzees. *Science (New York, N.Y.)*, *188*(4184), 107–116. https://doi.org/10.1126/science.1090005

Kingman, G. A. R., Vyas, D. N., Jones, F. C., Brady, S. D., Chen, H. I., Reid, K., Milhaven, M., Bertino, T. S., Aguirre, W. E., Heins, D. C., Hippel, F. A. von, Park, P. J., Kirch, M., Absher, D. M., Myers, R. M., Palma, F. D., Bell, M. A., Kingsley, D. M., & Veeramah, K. R. (2021). Predicting future from past: The genomic basis of recurrent and rapid stickleback evolution. *Science Advances*, *7*(25), eabg5285. https://doi.org/10.1126/sciadv.abg5285

Kingsley, D. M., Zhu, B., Osoegawa, K., De Jong, P. J., Schein, J., Marra, M., Peichel, C., Amemiya, C., Schluter, D., Balabhadra, S., Friedlander, B., Cha, Y. M., Dickson, M., Grimwood, J., Schmutz, J., Talbot, W. S., & Myers, R. (2004). New Genomic Tools for Molecular Studies of Evolutionary Change in Threespine Sticklebacks. *Behaviour*, *141*(11/12), 1331–1344.

Kirch, M., Romundset, A., Gilbert, M. T. P., Jones, F. C., & Foote, A. D. (2021). Ancient and modern stickleback genomes reveal the demographic constraints on adaptation. *Current Biology*, *31*(9), 2027-2036.e8. https://doi.org/10.1016/j.cub.2021.02.027

Klepaker, T. (1993). Morphological changes in a marine population of threespined stickleback, Gasterosteus aculeatus, recently isolated in fresh water. *Canadian Journal of Zoology*, *71*(6), 1251–1258. https://doi.org/10.1139/z93-171

Kronforst, M. R. (2008). Gene flow persists millions of years after speciation in Heliconiusbutterflies. *BMC Evolutionary Biology*, *8*(1), 98. https://doi.org/10.1186/1471-2148-8-98

Lackey, A., & Boughman, J. (2017). Evolution of reproductive isolation in stickleback fish. *Evolution*, *71*, 357–372. https://doi.org/10.1111/evo.13114

Lamason, R. L., Mohideen, M.-A. P. K., Mest, J. R., Wong, A. C., Norton, H. L., Aros, M. C., Jurynec, M. J., Mao, X., Humphreville, V. R., Humbert, J. E., Sinha, S., Moore, J. L., Jagadeeswaran, P., Zhao, W., Ning, G., Makalowska, I., McKeigue, P. M., O'donnell, D., Kittles, R., … Cheng, K. C. (2005). SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. *Science (New York, N.Y.)*, *310*(5755), 1782–1786. https://doi.org/10.1126/science.1116238

Laporte, M., Rogers, S. M., Dion-Côté, A.-M., Normandeau, E., Gagnaire, P.-A., Dalziel, A. C., Chebib, J., & Bernatchez, L. (2015). RAD-QTL Mapping Reveals Both Genome-Level Parallelism and Different Genetic Architecture Underlying the Evolution of Body Shape in Lake Whitefish (*Coregonus clupeaformis*) Species Pairs. *G3 Genes|Genomes|Genetics*, *5*(7), 1481–1491. https://doi.org/10.1534/g3.115.019067

Leffler, E. M., Gao, Z., Pfeifer, S., Ségurel, L., Auton, A., Venn, O., Bowden, R., Bontrop, R., Wall, J. D., Sella, G., Donnelly, P., McVean, G., & Przeworski, M. (2013). Multiple Instances of Ancient Balancing Selection Shared Between Humans and Chimpanzees. *Science*, *339*(6127), 1578–1582. https://doi.org/10.1126/science.1234070

Lenski, R. E. (2017). Experimental evolution and the dynamics of adaptation and genome evolution in microbial populations. *The ISME Journal*, *11*(10), Article 10. https://doi.org/10.1038/ismej.2017.69

Lescak, E. A., Bassham, S. L., Catchen, J., Gelmond, O., Sherbick, M. L., von Hippel, F. A., & Cresko, W. A. (2015). Evolution of stickleback in 50 years on earthquake-uplifted islands. *Proceedings of the National Academy of Sciences*, *112*(52), E7204–E7212. https://doi.org/10.1073/pnas.1512020112

Leverett, F. (1929). Pleistocene Glaciations of the Northern Hemisphere. *Science*, *69*(1783), 231–239.

Liu, Q., Mishra, M., Saxena, A. S., Wu, H., Qiu, Y., Zhang, X., You, X., Ding, S., & Miyamoto, M. M. (2021). Balancing selection maintains ancient polymorphisms

at conserved enhancers for the olfactory receptor genes of a Chinese marine fish. *Molecular Ecology*, *30*(16), 4023–4038. https://doi.org/10.1111/mec.16016

Liu, S., Hansen, M. M., & Jacobsen, M. W. (2016). Region-wide and ecotype-specific differences in demographic histories of threespine stickleback populations, estimated from whole genome sequences. *Molecular Ecology*, *25*(20), 5187–5202. https://doi.org/10.1111/mec.13827

Loarie, S. R., Duffy, P. B., Hamilton, H., Asner, G. P., Field, C. B., & Ackerly, D. D. (2009). The velocity of climate change. *Nature*, *462*(7276), 1052–1055. https://doi.org/10.1038/nature08649

Louis, M., Galimberti, M., Archer, F., Berrow, S., Brownlow, A., Fallon, R., Nykänen, M., O'Brien, J., Roberston, K. M., Rosel, P. E., Simon-Bouhet, B., Wegmann, D., Fontaine, M. C., Foote, A. D., & Gaggiotti, O. E. (2021). Selection on ancestral genetic variation fuels repeated ecotype formation in bottlenose dolphins. *Science Advances*, *7*(44), eabg1245. https://doi.org/10.1126/sciadv.abg1245

Lowe, C. B., Sanchez-Luege, N., Howes, T. R., Brady, S. D., Daugherty, R. R., Jones, F. C., Bell, M. A., & Kingsley, D. M. (2018). Detecting differential copy number variation between groups of samples. *Genome Research*, *28*(2), 256–265. https://doi.org/10.1101/gr.206938.116

Marchinko, K. B. (2009). Predation's Role in Repeated Phenotypic and Genetic Divergence of Armor in Threespine Stickleback. *Evolution*, *63*(1), 127–138. https://doi.org/10.1111/j.1558-5646.2008.00529.x

Marques, D. A., Meier, J. I., & Seehausen, O. (2019). A Combinatorial View on Speciation and Adaptive Radiation. *Trends in Ecology & Evolution*, *34*(6), 531–544. https://doi.org/10.1016/j.tree.2019.02.008

Maynard-Smith, J., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetics Research*, *23*(1), 23–35. https://doi.org/10.1017/S0016672300014634

McGee, M. D., Borstein, S. R., Meier, J. I., Marques, D. A., Mwaiko, S., Taabu, A., Kishe, M. A., O'Meara, B., Bruggmann, R., Excoffier, L., & Seehausen, O. (2020). The ecological and genomic basis of explosive adaptive radiation. *Nature*, *586*(7827), 75–79. https://doi.org/10.1038/s41586-020-2652-7

McGregor, A. P., Orgogozo, V., Delon, I., Zanet, J., Srinivasan, D. G., Payre, F., & Stern, D. L. (2007). Morphological evolution through multiple cis-regulatory mutations at a single gene. *Nature*, *448*(7153), 587–590. https://doi.org/10.1038/nature05988

McIntyre, P. B., & Flecker, A. S. (2010). *Ecological Stoichiometry as an Integrative Framework in Stream Fish Ecology*. 20.

McKinnon, J. S., & Rundle, H. D. (2002). Speciation in nature: The threespine stickleback model systems. *Trends in Ecology & Evolution*, *17*(10), 480–488. https://doi.org/10.1016/S0169-5347(02)02579-X

McPhail, J. D. (1993). Ecology and evolution of sympatric sticklebacks (Gasterosteus): Origin of the species pairs. *Canadian Journal of Zoology*, *71*(3), 515–523. https://doi.org/10.1139/z93-072

McPhail, J. D., & Hay, D. E. (1983). Differences in male courtship in freshwater and marine sticklebacks (Gasterosteus aculeatus). *Canadian Journal of Zoology*, *61*(2), 292–297. https://doi.org/10.1139/z83-039

Meier, J. I., Marques, D. A., Mwaiko, S., Wagner, C. E., Excoffier, L., & Seehausen, O. (2017). Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature Communications*, *8*(1), Article 1. https://doi.org/10.1038/ncomms14363

Meier, J. I., Salazar, P. A., Kučka, M., Davies, R. W., Dréau, A., Aldás, I., Box Power, O., Nadeau, N. J., Bridle, J. R., Rolian, C., Barton, N. H., McMillan, W. O., Jiggins, C. D., & Chan, Y. F. (2021). Haplotype tagging reveals parallel formation of hybrid races in two butterfly species. *Proceedings of the National Academy of Sciences*, *118*(25), e2015005118. https://doi.org/10.1073/pnas.2015005118

Merrill, R. M., Dasmahapatra, K. K., Davey, J. W., Dell'Aglio, D. D., Hanly, J. J., Huber, B., Jiggins, C. D., Joron, M., Kozak, K. M., Llaurens, V., Martin, S. H., Montgomery, S. H., Morris, J., Nadeau, N. J., Pinharanda, A. L., Rosser, N., Thompson, M. J., Vanjari, S., Wallbank, R. W. R., & Yu, Q. (2015). The diversification of Heliconius butterflies: What have we learned in 150 years? *Journal of Evolutionary Biology*, *28*(8), 1417–1438. https://doi.org/10.1111/jeb.12672

Miller, C. T., Beleza, S., Pollen, A. A., Schluter, D., Kittles, R. A., Shriver, M. D., & Kingsley, D. M. (2007). Cis-Regulatory Changes in Kit Ligand Expression and Parallel Evolution of Pigmentation in Sticklebacks and Humans. *Cell*, *131*(6), 1179–1189. https://doi.org/10.1016/j.cell.2007.10.055

Miller, C. T., Glazer, A. M., Summers, B. R., Blackman, B. K., Norman, A. R., Shapiro, M. D., Cole, B. L., Peichel, C. L., Schluter, D., & Kingsley, D. M. (2014). Modular Skeletal Evolution in Sticklebacks Is Controlled by Additive and Clustered Quantitative Trait Loci. *Genetics*, *197*(1), 405–420. https://doi.org/10.1534/genetics.114.162420

Montejo-Kovacevich, G., Meier, J. I., Bacquet, C. N., Warren, I. A., Chan, Y. F., Kucka, M., Salazar, C., Rueda-M, N., Montgomery, S. H., McMillan, W. O., Kozak, K. M., Nadeau, N. J., Martin, S. H., & Jiggins, C. D. (2022). Repeated genetic adaptation

to altitude in two tropical butterflies. *Nature Communications*, *13*(1), Article 1. https://doi.org/10.1038/s41467-022-32316-x

Moritz, C. (2002). Strategies to Protect Biological Diversity and the Evolutionary Processes That Sustain It. *Systematic Biology*, *51*(2), 238–254. https://doi.org/10.1080/10635150252899752

Moyle, P. B. (1976). Fish introductions in California: History and impact on native fishes. *Biological Conservation*, *9*(2), 101–118. https://doi.org/10.1016/0006-3207(76)90043-4

Nelson, T. C., & Cresko, W. A. (2018). Ancient genomic variation underlies repeated ecological adaptation in young stickleback populations. *Evolution Letters*, *2*(1), 9–21. https://doi.org/10.1002/evl3.37

Nielsen, R. (2005). Molecular Signatures of Natural Selection. *Annual Review of Genetics*, *39*(1), 197–218. https://doi.org/10.1146/annurev.genet.39.073003.112420

O'Brown, N. M., Summers, B. R., Jones, F. C., Brady, S. D., & Kingsley, D. M. (2015). A recurrent regulatory change underlying altered expression and Wnt response of the stickleback armor plates gene EDA. *ELife*, *4*, e05290. https://doi.org/10.7554/eLife.05290

Oosting, T., Star, B., Barrett, J. H., Wellenreuther, M., Ritchie, P. A., & Rawlence, N. J. (2019). Unlocking the potential of ancient fish DNA in the genomic era. *Evolutionary Applications*, *12*(8), 1513–1522. https://doi.org/10.1111/eva.12811

Ottenburghs, J., Kraus, R. H. S., van Hooft, P., van Wieren, S. E., Ydenberg, R. C., & Prins, H. H. T. (2017). Avian introgression in the genomic era. *Avian Research*, *8*(1), 30. https://doi.org/10.1186/s40657-017-0088-z

Oziolor, E. M., Reid, N. M., Yair, S., Lee, K. M., Guberman VerPloeg, S., Bruns, P. C., Shaw, J. R., Whitehead, A., & Matson, C. W. (2019). Adaptive introgression enables evolutionary rescue from extreme environmental pollution. *Science*, *364*(6439), 455–457. https://doi.org/10.1126/science.aav4155

Peichel, C. L., & Marques, D. A. (2017). The genetic and molecular architecture of phenotypic diversity in sticklebacks. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1713), 20150486. https://doi.org/10.1098/rstb.2015.0486

Peichel, C. L., Sullivan, S. T., Liachko, I., & White, M. A. (2017). Improvement of the Threespine Stickleback Genome Using a Hi-C-Based Proximity-Guided Assembly. *Journal of Heredity*, *108*(6), 693–700. https://doi.org/10.1093/jhered/esx058

Pennings, P. S., & Hermisson, J. (2006). Soft Sweeps III: The Signature of Positive Selection from Recurrent Mutation. *PLOS Genetics*, *2*(12), e186. https://doi.org/10.1371/journal.pgen.0020186

Pérez-Pereira, N., Wang, J., Quesada, H., & Caballero, A. (2022). Prediction of the minimum effective size of a population viable in the long term. *Biodiversity and Conservation*, *31*(11), 2763–2780. https://doi.org/10.1007/s10531-022-02456-z

Pimm, S. L. (2009). Climate Disruption and Biodiversity. *Current Biology*, *19*(14), R595–R601. https://doi.org/10.1016/j.cub.2009.05.055

Pimm, S. L., Jenkins, C. N., Abell, R., Brooks, T. M., Gittleman, J. L., Joppa, L. N., Raven, P. H., Roberts, C. M., & Sexton, J. O. (2014). The biodiversity of species and their rates of extinction, distribution, and protection. *Science*, *344*(6187), 1246752. https://doi.org/10.1126/science.1246752

Plato & translated by Harold North Fowler. (1966). *Plato in Twelve Volumes*. Harvard University Press.

Prezeworski, M., Coop, G., & Wall, J. D. (2005). The Signature of Positive Selection on Standing Genetic Variation. *Evolution*, *59*(11), 2312–2323. https://doi.org/10.1111/j.0014-3820.2005.tb00941.x

Pritchard, J. K., & Di Rienzo, A. (2010). Adaptation – not by sweeps alone. *Nature Reviews Genetics*, *11*(10), Article 10. https://doi.org/10.1038/nrg2880

Pritchard, J. K., Pickrell, J. K., & Coop, G. (2010). The genetics of human adaptation: Hard sweeps, soft sweeps, and polygenic adaptation. *Current Biology: CB*, *20*(4), R208-215. https://doi.org/10.1016/j.cub.2009.11.055

Proulx, S. R., & Teotónio, H. (2022). Selection on modifiers of genetic architecture under migration load. *PLOS Genetics*, *18*(9), e1010350. https://doi.org/10.1371/journal.pgen.1010350

Ravinet, M., Hynes, R., Poole, R., Cross, T. F., McGinnity, P., Harrod, C., & Prodöhl, P. A. (2015). Where the Lake Meets the Sea: Strong Reproductive Isolation Is Associated with Adaptive Divergence between Lake Resident and Anadromous Three-Spined Sticklebacks. *PLOS ONE*, *10*(4), e0122825. https://doi.org/10.1371/journal.pone.0122825

Reed, R. D. (2003). Gregarious Oviposition and Clutch Size Adjustment by a Heliconius Butterfly. *Biotropica*, *35*(4), 555–559.

Reid, K., Bell, M. A., & Veeramah, K. R. (2021). Threespine Stickleback: A Model System For Evolutionary Genomics. *Annual Review of Genomics and Human Genetics*, *22*, 357–383. https://doi.org/10.1146/annurev-genom-111720-081402

Reimchen, T. E. (2000). PREDATOR HANDLING FAILURES OF LATERAL PLATE MORPHS IN GASTEROSTEUS ACULEATUS: FUNCTIONAL IMPLICATIONS

FOR THE ANCESTRAL PLATE CONDITION. *Behaviour*, *137*(7–8), 1081–1096. https://doi.org/10.1163/156853900502448

Remington, C. L. (1968). Suture-Zones of Hybrid Interaction Between Recently Joined Biotas. In *Evolutionary Biology* (pp. 321–428). https://doi.org/10.1007/978-1-4684-8094-8_8

Rice, W. R. (1992). Sexually antagonistic genes: Experimental evidence. *Science (New York, N.Y.)*, *256*(5062), 1436–1439. https://doi.org/10.1126/science.1604317

Roesti, M., Moser, D., & Berner, D. (2013). Recombination in the threespine stickleback genome—Patterns and consequences. *Molecular Ecology*, *22*(11), 3014–3027. https://doi.org/10.1111/mec.12322

Ronce, O., & Kirkpatrick, M. (2001). When Sources Become Sinks: Migrational Meltdown in Heterogeneous Habitats. *Evolution*, *55*(8), 1520–1531. https://doi.org/10.1111/j.0014-3820.2001.tb00672.x

Rosenfeld, J. S. (2014). 50/500 or 100/1000? Reconciling short- and long-term recovery targets and MVPs. *Biological Conservation*, *176*, 287–288. https://doi.org/10.1016/j.biocon.2014.05.005

Rundle, H. D. (2002). A test of ecologically dependent postmating isolation between sympatric sticklebacks. *Evolution; International Journal of Organic Evolution*, *56*(2), 322–329. https://doi.org/10.1111/j.0014-3820.2002.tb01342.x

Rundle, H. D., Vamosi, S. M., & Schluter, D. (2003). Experimental test of predation's effect on divergent selection during character displacement in sticklebacks. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(25), 14943–14948. https://doi.org/10.1073/pnas.2036360100

Sabeti, P. C., Schaffner, S. F., Fry, B., Lohmueller, J., Varilly, P., Shamovsky, O., Palma, A., Mikkelsen, T. S., Altshuler, D., & Lander, E. S. (2006). Positive Natural Selection in the Human Lineage. *Science*, *312*(5780), 1614–1620. https://doi.org/10.1126/science.1124309

Scheinfeldt, L. B., Biswas, S., Madeoy, J., Connelly, C. F., Schadt, E. E., & Akey, J. M. (2009). Population Genomic Analysis of ALMS1 in Humans Reveals a Surprisingly Complex Evolutionary History. *Molecular Biology and Evolution*, *26*(6), 1357–1367. https://doi.org/10.1093/molbev/msp045

Schluter, D., & Conte, G. L. (2009). Genetics and ecological speciation. *Proceedings of the National Academy of Sciences*, *106*(Supplement 1), 9955–9962. https://doi.org/10.1073/pnas.0901264106

Schluter, D., Marchinko, K. B., Arnegard, M. E., Zhang, H., Brady, S. D., Jones, F. C., Bell, M. A., & Kingsley, D. M. (2021). Fitness maps to a large-effect locus in introduced stickleback populations. *Proceedings of the National Academy of*

Sciences of the United States of America, *118*(3), e1914889118. https://doi.org/10.1073/pnas.1914889118

Schneider, J. A., Pungliya, M. S., Choi, J. Y., Jiang, R., Sun, X. J., Salisbury, B. A., & Stephens, J. C. (2003). DNA variability of human genes. *Mechanisms of Ageing and Development*, *124*(1), 17–25. https://doi.org/10.1016/S0047-6374(02)00165-3

Shapiro, M. D., Bell, M. A., & Kingsley, D. M. (2006). Parallel genetic origins of pelvic reduction in vertebrates. *Proceedings of the National Academy of Sciences*, *103*(37), 13753–13758. https://doi.org/10.1073/pnas.0604706103

Shurtliff, Q. R. (2013). Mammalian hybrid zones: A review. *Mammal Review*, *43*(1), 1–21. https://doi.org/10.1111/j.1365-2907.2011.00205.x

Siewert, K. M., & Voight, B. F. (2017). Detecting Long-Term Balancing Selection Using Allele Frequency Correlation. *Molecular Biology and Evolution*, *34*(11), 2996–3005. https://doi.org/10.1093/molbev/msx209

Sohail, M., Maier, R. M., Ganna, A., Bloemendal, A., Martin, A. R., Turchin, M. C., Chiang, C. W., Hirschhorn, J., Daly, M. J., Patterson, N., Neale, B., Mathieson, I., Reich, D., & Sunyaev, S. R. (2019). Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies. *ELife*, *8*, e39702. https://doi.org/10.7554/eLife.39702

Soulé. (1980). Thresholds for survival:maintaining fitness and evolutionary potential. *Comservation Biology:An Evolutionary-Ecological Perspective*, 151–169.

Stankowski, S., & Streisfeld, M. A. (2015). Introgressive hybridization facilitates adaptive divergence in a recent radiation of monkeyflowers. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1814), 20151666. https://doi.org/10.1098/rspb.2015.1666

Star, B., Nederbragt, A. J., Jentoft, S., Grimholt, U., Malmstrøm, M., Gregers, T. F., Rounge, T. B., Paulsen, J., Solbakken, M. H., Sharma, A., Wetten, O. F., Lanzén, A., Winer, R., Knight, J., Vogel, J.-H., Aken, B., Andersen, Ø., Lagesen, K., Tooming-Klunderud, A., … Jakobsen, K. S. (2011). The genome sequence of Atlantic cod reveals a unique immune system. *Nature*, *477*(7363), Article 7363. https://doi.org/10.1038/nature10342

Steiner, C. C., Weber, J. N., & Hoekstra, H. E. (2007). Adaptive Variation in Beach Mice Produced by Two Interacting Pigmentation Genes. *PLOS Biology*, *5*(9), e219. https://doi.org/10.1371/journal.pbio.0050219

Stern, D. B., Anderson, N. W., Diaz, J. A., & Lee, C. E. (2022). Genome-wide signatures of synergistic epistasis during parallel adaptation in a Baltic Sea copepod. *Nature Communications*, *13*(1), Article 1. https://doi.org/10.1038/s41467-022-31622-8

Stern, D. L. (2000). Evolutionary developmental biology and the problem of variation. *Evolution; International Journal of Organic Evolution*, *54*(4), 1079–1091. https://doi.org/10.1111/j.0014-3820.2000.tb00544.x

Supple, M. A., & Shapiro, B. (2018). Conservation of biodiversity in the genomics era. *Genome Biology*, *19*(1), 131. https://doi.org/10.1186/s13059-018-1520-3

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, *123*(3), 585–595. https://doi.org/10.1093/genetics/123.3.585

TAYLOR, E. B., & McPHAIL, J. D. (1999). Evolutionary history of an adaptive radiation in species pairs of threespine sticklebacks (Gasterosteus): Insights from mitochondrial DNA. *Biological Journal of the Linnean Society*, *66*(3), 271–291. https://doi.org/10.1111/j.1095-8312.1999.tb01891.x

Terekhanova, N. V., Logacheva, M. D., Penin, A. A., Neretina, T. V., Barmintseva, A. E., Bazykin, G. A., Kondrashov, A. S., & Mugue, N. S. (2014). Fast Evolution from Precast Bricks: Genomics of Young Freshwater Populations of Threespine Stickleback Gasterosteus aculeatus. *PLOS Genetics*, *10*(10), e1004696. https://doi.org/10.1371/journal.pgen.1004696

Teshima, K. M., Coop, G., & Przeworski, M. (2006). How reliable are empirical genomic scans for selective sweeps? *Genome Research*, *16*(6), 702–712. https://doi.org/10.1101/gr.5105206

Thomas, C. D., Cameron, A., Green, R. E., Bakkenes, M., Beaumont, L. J., Collingham, Y. C., Erasmus, B. F. N., de Siqueira, M. F., Grainger, A., Hannah, L., Hughes, L., Huntley, B., van Jaarsveld, A. S., Midgley, G. F., Miles, L., Ortega-Huerta, M. A., Townsend Peterson, A., Phillips, O. L., & Williams, S. E. (2004). Extinction risk from climate change. *Nature*, *427*(6970), Article 6970. https://doi.org/10.1038/nature02121

Tishkoff, S. A., Varkonyi, R., Cahinhinan, N., Abbes, S., Argyropoulos, G., Destro-Bisol, G., Drousiotou, A., Dangerfield, B., Lefranc, G., Loiselet, J., Piro, A., Stoneking, M., Tagarelli, A., Tagarelli, G., Touma, E. H., Williams, S. M., & Clark, A. G. (2001). Haplotype diversity and linkage disequilibrium at human G6PD: Recent origin of alleles that confer malarial resistance. *Science (New York, N.Y.)*, *293*(5529), 455–462. https://doi.org/10.1126/science.1061573

Urban, M. C. (2015). Accelerating extinction risk from climate change. *Science*, *348*(6234), 571–573. https://doi.org/10.1126/science.aaa4984

Vamosi, S. M., & Schluter, D. (2004). Character Shifts in the Defensive Armor of Sympatric Sticklebacks. *Evolution*, *58*(2), 376–385. https://doi.org/10.1111/j.0014-3820.2004.tb01653.x

Verta, J.-P., & Jones, F. C. (2019). Predominance of cis-regulatory changes in parallel expression divergence of sticklebacks. *ELife*, *8*, e43785. https://doi.org/10.7554/eLife.43785

Vines, T. H., Dalziel, A. C., Albert, A. Y. K., Veen, T., Schulte, P. M., & Schluter, D. (2016). Cline coupling and uncoupling in a stickleback hybrid zone. *Evolution*, *70*(5), 1023–1038. https://doi.org/10.1111/evo.12917

Vines, T. H., & Schluter, D. (2006). Strong assortative mating between allopatric sticklebacks as a by-product of adaptation to different environments. *Proceedings of the Royal Society B: Biological Sciences*, *273*(1589), 911–916. https://doi.org/10.1098/rspb.2005.3387

Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., & Yang, J. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. *American Journal of Human Genetics*, *101*(1), 5–22. https://doi.org/10.1016/j.ajhg.2017.06.005

Warren, R., Price, J., Graham, E., Forstenhaeusler, N., & VanDerWal, J. (2018). The projected effect on insects, vertebrates, and plants of limiting global warming to 1.5°C rather than 2°C. *Science (New York, N.Y.)*, *360*(6390), 791–795. https://doi.org/10.1126/science.aar3646

Warren, R., VanDerWal, J., Price, J., Welbergen, J. A., Atkinson, I., Ramirez-Villegas, J., Osborn, T. J., Jarvis, A., Shoo, L. P., Williams, S. E., & Lowe, J. (2013). Quantifying the benefit of early climate change mitigation in avoiding biodiversity loss. *Nature Climate Change*, *3*(7), Article 7. https://doi.org/10.1038/nclimate1887

Whiting, J. R., Paris, J. R., Parsons, P. J., Matthews, S., Reynoso, Y., Hughes, K. A., Reznick, D., & Fraser, B. A. (2022). On the genetic architecture of rapidly adapting and convergent life history traits in guppies. *Heredity*, *128*(4), Article 4. https://doi.org/10.1038/s41437-022-00512-6

Williamson, S. H., Hubisz, M. J., Clark, A. G., Payseur, B. A., Bustamante, C. D., & Nielsen, R. (2007). Localizing Recent Adaptive Evolution in the Human Genome. *PLOS Genetics*, *3*(6), e90. https://doi.org/10.1371/journal.pgen.0030090

Withler, R. E., & McPhail, J. D. (1985). Genetic variability in freshwater and anadromous sticklebacks (Gasterosteus aculeatus) of southern British Columbia. *Canadian Journal of Zoology*, *63*(3), 528–533. https://doi.org/10.1139/z85-078

Witte, F., & van Oijen, M. J. P. (1990). Taxonomy, ecology and fishery of Lake Victoria haplochromine trophic groups. *Zoologische Verhandelingen*, *262*(1), 1–47.

Wooldridge, T. B., Kautt, A. F., Lassance, J.-M., McFadden, S., Domingues, V. S., Mallarino, R., & Hoekstra, H. E. (2022). An enhancer of *Agouti* contributes to

parallel evolution of cryptically colored beach mice. *Proceedings of the National Academy of Sciences*, *119*(27), e2202862119. https://doi.org/10.1073/pnas.2202862119

Wootton, R. J. (1984). *A Functional Biology of Sticklebacks*. University of California Press.

Xie, K. T., Wang, G., Thompson, A. C., Wucherpfennig, J. I., Reimchen, T. E., MacColl, A. D. C., Schluter, D., Bell, M. A., Vasquez, K. M., & Kingsley, D. M. (2019). DNA fragility in the parallel evolution of pelvic reduction in stickleback fish. *Science*, *363*(6422), 81–84. https://doi.org/10.1126/science.aan1425

Yuan, Y., Peng, D., Gu, X., Gong, Y., Sheng, Z., & Hu, X. (2018). Polygenic Basis and Variable Genetic Architectures Contribute to the Complex Nature of Body Weight—A Genome-Wide Study in Four Chinese Indigenous Chicken Breeds. *Frontiers in Genetics*, *9*. https://www.frontiersin.org/articles/10.3389/fgene.2018.00229

Zammuto, R. M. (1986). Life histories of birds: Clutch size, longevity, and body mass among North American game birds. *Canadian Journal of Zoology*, *64*(12), 2739–2749. https://doi.org/10.1139/z86-398

8. Appendix

The appendix includes

- Kirch et al. (2021): Manuscript and Supplementary Material
- Kingman et al. (2021): Manuscript and extract from Supplementary Material
- Kirch et al. (unpublished): Manuscript and Supplementary Material

Report

Current Biology

Ancient and modern stickleback genomes reveal the demographic constraints on adaptation

Highlights

- Sequencing of two threespine stickleback genomes dated to the Late Pleistocene
- These genomes show mostly marine-adapted ancestry
- Present-day lake stickleback genomes indicate ancestral founder bottleneck
- This demographic history constrains natural selection

Authors

Melanie Kirch, Anders Romundset, M. Thomas P. Gilbert, Felicity C. Jones, Andrew D. Foote

Correspondence

andrew.foote@ntnu.no

In brief

Kirch et al. sequence genomes of two Late Pleistocene stickleback collected from the sediment layer corresponding to the transition from marine to freshwater habitat. Comparison with present-day lake and fjord populations highlights that demography can constrain natural selection's action upon standing genetic variation in the ancestral gene pool.

Kirch et al., 2021, Current Biology *31*, 2027–2036
May 10, 2021
https://doi.org/10.1016/j.cub.2021.02.027

CellPress

Current Biology

CellPress

Report

Ancient and modern stickleback genomes reveal the demographic constraints on adaptation

Melanie Kirch,[1] Anders Romundset,[2] M. Thomas P. Gilbert,[3,4] Felicity C. Jones,[1,6] and Andrew D. Foote[4,5,6,7,*]
[1]Friedrich Miescher Laboratory of the Max Planck Society, Max-Planck-Ring 9, 72076 Tübingen, Germany
[2]Geological Survey of Norway, Trondheim, Norway
[3]Center for Evolutionary Hologenomics, The GLOBE Institute, University of Copenhagen, Øster Farimagsgade 5A, DK-1353 Copenhagen, Denmark
[4]Department of Natural History, Norwegian University of Science and Technology (NTNU), University Museum, 7491 Trondheim, Norway
[5]Molecular Ecology and Fisheries Genetics Laboratory, School of Biological Sciences, Bangor University, Bangor, UK
[6]Senior author
[7]Lead Contact
*Correspondence: andrew.foote@ntnu.no
https://doi.org/10.1016/j.cub.2021.02.027

SUMMARY

Adaptation is typically studied by comparing modern populations with contrasting environments. Individuals persisting in the ancestral habitat are typically used to represent the ancestral founding population; however, it has been questioned whether these individuals are good proxies for the actual ancestors.[1] To address this, we applied a paleogenomics approach[2] to directly access the ancestral genepool: partially sequencing the genomes of two 11- to 13,000-year-old stickleback recovered from the transitionary layer between marine and freshwater sediments of two Norwegian isolation lakes[3] and comparing them with 30 modern stickleback genomes from the same lakes and adjacent marine fjord, in addition to a global dataset of 20 genomes.[4] The ancient stickleback shared genome-wide ancestry with the modern fjord population, whereas modern lake populations have lost substantial ancestral variation following founder effects, and subsequent drift and selection. Freshwater-adaptive alleles found in one ancient stickleback genome have not risen to high frequency in the present-day population from the same lake. Comparison to the global dataset suggested incomplete adaptation to freshwater in our modern lake populations. Our findings reveal the impact of population bottlenecks in constraining adaptation due to reduced efficacy of selection on standing variation present in founder populations.

RESULTS AND DISCUSSION

The threespine stickleback *Gasterosteus aculeatus* genome is well characterized through mapping, sequencing, and transgenics studies, which have provided a strong basis for understanding the role of genes underlying phenotypic changes in skeletal armor, body shape, and other morphological and physiological changes associated with freshwater or marine adaptation.[5,6] Adaptation of marine sticklebacks to freshwater habitats can occur rapidly over tens of generations,[7,8] facilitated by the reuse of pre-existing genetic variation carried in ancient haplotype blocks within the marine population.[4,9,10] This standing genetic variation is thought to have been maintained over geological timescales[9,11] through recurrent migration between freshwater and marine populations,[12] and thus repeatedly acted upon by natural selection in freshwater populations.[4,13] The present adaptive radiation commenced during the transition between the Pleistocene and Holocene epochs, and from glacial to inter-glacial.[5] Investigating signatures of freshwater adaptation in the genomes of stickleback from this key time point in their evolutionary history has not been possible until now.

Here, we demonstrate the power of paleogenomics to provide novel temporal insights into evolutionary processes by sequencing partial genomes of two 11- to 13,000-year-old stickleback, which lived during the period immediately after the latest retreat of the Pleistocene Scandinavian Ice Sheet, when many freshwater coastal lakes were forming due to strong post-glacial land uplift. "Isolation basins" formed by glacio-isostatic rebound, elevating them above sea level, and rapidly changed from marine to freshwater ecosystems.[14] This ecological change causes a distinct sedimentary boundary between the marine and freshwater lacustrine sediment facies (Figures 1B and S1) and is used by geologists to reconstruct relative sea-level history.[14] We examined sediment cores collected from freshwater lakes formed from isolation basins in Finnmark, northernmost Norway (Figures 1A and S1), an area where post-glacial uplift has caused a net relative sea-level fall of 50–100 m since deglaciation.[3]

Stickleback bones, spines, and bony armor plates were found in the layers of cores from two lakes corresponding to the brackish phase when the lakes became isolated from the marine fjord and were transitioning to freshwater (Figures 1C and S1). Radiocarbon dating of organic matter from the same

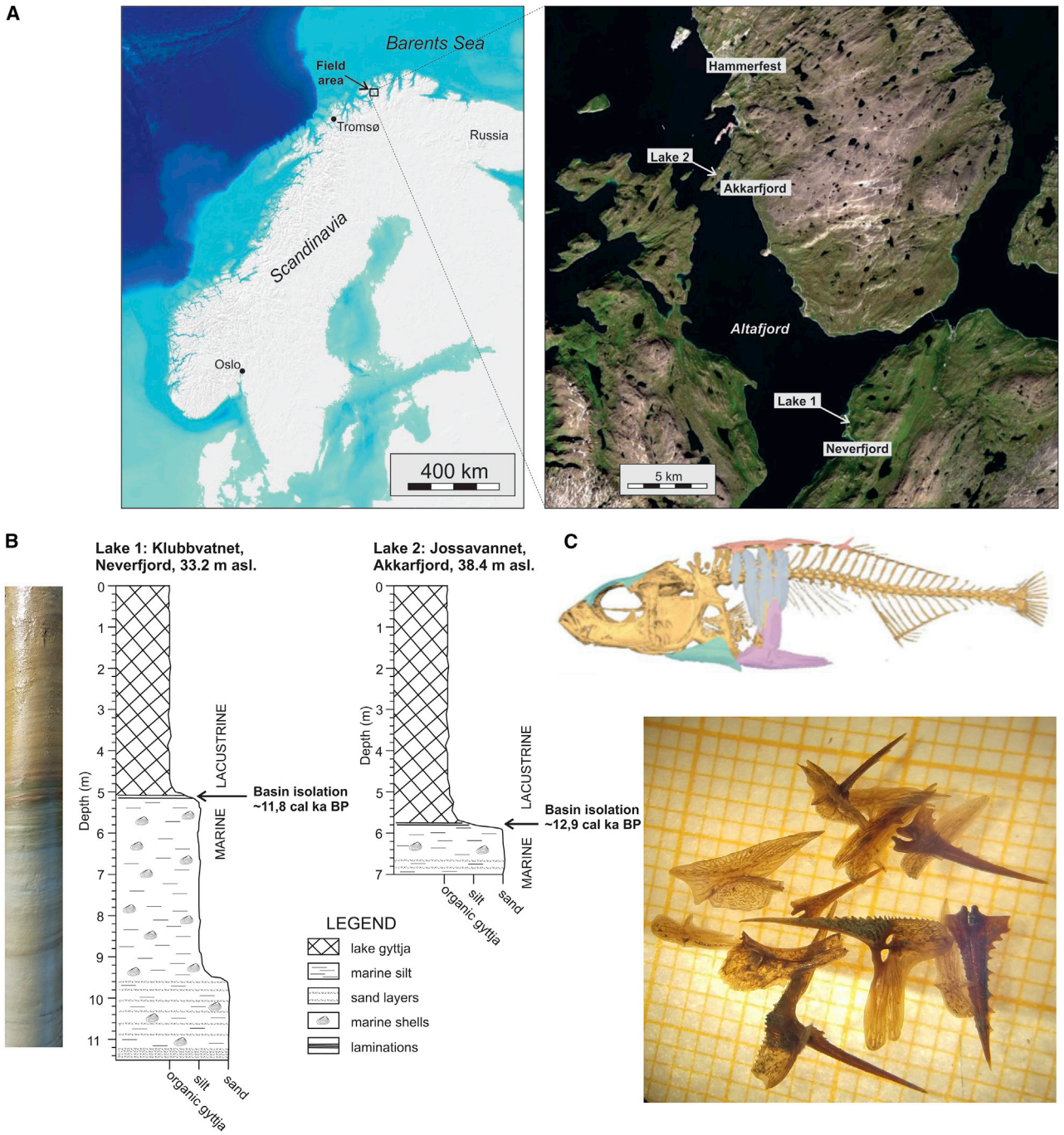


Figure 1. The ecological and geological context of Late Pleistocene stickleback remains

(A) Ancient and present-day samples were collected from Klubbvatnet freshwater lake (70° 36′ N, 23° 37′ E; hereafter Lake 1) and from Jossavannet freshwater lake (70° 27′ N, 23° 47′ E; hereafter Lake 2); additionally, samples of the marine ecotype were collected from the outer branch area (next to the lake sites) of Altafjord (70° 27′ N, 23° 46′ E).

(B) The ancient samples were found in the sediment layers of cores from the two lakes corresponding to the isolation phase, dated to ~11.8 and 12.9 KY BP, respectively. An example of the variation in core stratigraphy is shown on the left; schematic diagrams of the stratigraphy in the two study lakes are shown to the right.

(C) Bones, spines, and bony armor plates found in Lake 2. The background grid is mm-scale. Bone positions illustrated on an X-ray scan of modern freshwater fish (above).

See also Figure S1 and Methods S1.

stratigraphic depth placed the age of the stickleback bones as 12,040–11,410 cal year BP for Lake 1 and 13,070–12,800 cal year BP for Lake 2 (Figure 1B). Shotgun sequencing was performed on DNA extracted from a stickleback spine from Lake 1 and a bony armor plate from Lake 2 (Figures 1C and S1), resulting in coverage of $\geq$1× at 369,344 bp and 16,923,179 bp, respectively. Nucleotide misincorporations relative to the reference genome indicate an approximately 30% deamination rate of cytosine at the read-ends (Methods S1, section 1). Such post-mortem damage patterns are characteristic of degradation in ancient DNA samples that are thousands of years old,[15] and do not represent contamination from exogenous modern DNA. To the best of our knowledge, these are the oldest fish bones from which genomic data have been obtained.[16] While these data represent just a single ancestor from each lake, their genomes potentially contain ancestry from >1,000 individuals from the previous ten generations (equivalent to 10–20 years). Single samples have routinely been used as representative examples of ancestral populations in palaeogenomics[17–19] and even partial genomes have provided key and novel insights into our understanding of evolutionary histories.[20]

To understand the importance of the ancient samples in the chronology of freshwater adaptation, we first established their relationships to present-day sticklebacks. Uneven sampling of different demes can influence the inference of population clusters in principal component analysis (PCA), due to strong covariance in allele frequencies among samples from the same population.[21] We therefore included a single randomly sampled haploid genome, removing coverage bias, from each of 23 present-day populations from geographically distant locations across the Northern Hemisphere[4] (hereafter referred to as the "global" population dataset; Table S1), and considered only transversions to avoid post-mortem DNA damage patterns of excess C→T and A→G changes resulting from deamination.[15] Samples separate by geography into Pacific and Atlantic clusters on principal component 1 (PC1, $p < 0.001$; Figure 2A; Methods S1, section 2) when genomic regions underlying parallel marine-freshwater adaptive divergence[4] are excluded. Both ancient samples show closest affinity to present-day Atlantic populations, as expected. An analysis using only marine-freshwater parallel divergent genomic regions separated samples into marine and freshwater ecotypes on PC1 and revealed both ancient samples cluster with the globally sampled marine individuals ($p < 0.005$; Figure 2A; Methods S1, section 2). Under a polygenic model this suggests that the two ancient fish would have had a predominantly marine phenotype.

Focusing on the two focal freshwater lakes from which the ancient samples were recovered, and the adjacent marine fjord, we compared the two ancient samples to 15 present-day samples from Lake 1, 10 samples from Lake 2, and 5 present-day marine samples from Altafjord (hereafter, the "local" population dataset). Considering just sites covered in the ancient samples and estimating genotype likelihoods to account for any uncertainty in genotypes, we find strong covariance among present-day genomes within each lake, and likewise among genomes from the fjord (Figure 2C). The focal lakes in this study are found less than 300 m from the fjord on land that rises steeply to the post-uplift height of 33.2 and 38.4 m above sea level. Isostatic rebound in northern Norway occurred rapidly over a period of several hundred years.[3,14] The rapid uplift and steepness of the terrain would produce sudden and ongoing isolation of the freshwater lakes from the marine source population. It is therefore likely that present-day stickleback in these "isolation lakes" are descendants of early colonists that include the ancient samples. Accordingly, the strongest differentiation in allele frequencies is between genomes from Lake 1 and those from Lake 2 (PC1, $p < 0.005$), explaining 42.9% of the variance in the data (Figure 2C; Methods S1, section 2). The ancient genomes cluster most closely with the marine fjord samples, though the Lake 1 ancient sample is found between the fjord and lake samples along PC2 ($p < 0.001$; Figure 2C; Methods S1, section 2). A similar pattern is seen when considering the marine-freshwater divergent regions (Methods S1, section 2). Clustering patterns in the PCA are reflected in admixture plots, in which both ancient samples share ancestry components with the present-day fjord samples, while present-day lake samples retain just a lake-specific subset of this ancestral variation (Figure 2D).

The placement of the ancient samples relative to the present-day fjord and lake populations in the PCA, and the pattern of ancestry components in the admixture plots, indicates high covariance in allele frequencies, consistent with strong drift associated with the colonization of each lake population and subsequent demographic bottleneck. Such drift would genetically differentiate each lake population from the ancestral founder population, the contemporary marine population, and one another. The PCA in Figure 2C was generated from data that excluded regions of the genome associated with parallel marine-freshwater adaptation[4]; however, covariance in some locally adapted alleles could explain separation of lake and marine populations along PC2.

If the ancient samples represent the ancestral populations that first colonized the lake from the marine source population, then the ancient sample would be symmetrically related to the present-day lake and marine fjord populations (i.e., sharing approximately equal alleles with each). To formally test this hypothesis, we computed *D*-statistics corresponding to the population history *D* (*G. nipponicus*, (ancient, (modern fjord, modern lake))). To remove biases associated with coverage and post-mortem DNA damage, we considered only transversions and sites covered in modern and ancient samples, randomly drawing an allele from each sample. We focused our analyses on the Lake 2 sample, for which we had data from many unlinked genomic regions, providing jack-knife estimates of D from which to generate *Z* scores. The *D*-statistic tests are consistent with the ancient sample from Lake 2 being symmetrically related to the present-day marine fjord and freshwater Lake 2 populations ($-3 < |Z| < 3$; Figure 2E). Conversely, tests assigning the ancient sample to a clade with one of the present-day sticklebacks to the exclusion of the other, i.e., (*G. nipponicus*, (modern, (ancient, modern))), were all rejected ($Z > 3$). Thus, based on the sharing of derived alleles (Figure 2E), the ancient stickleback from Lake 2 is inferred to have lived close to the time of the divergence of the ancestral fjord and lake populations, and prior to the strong independent drift in the two lake populations, which causes correlated allele frequencies and drives the patterns in the PCA and admixture plots (Figures 2C and 2D).

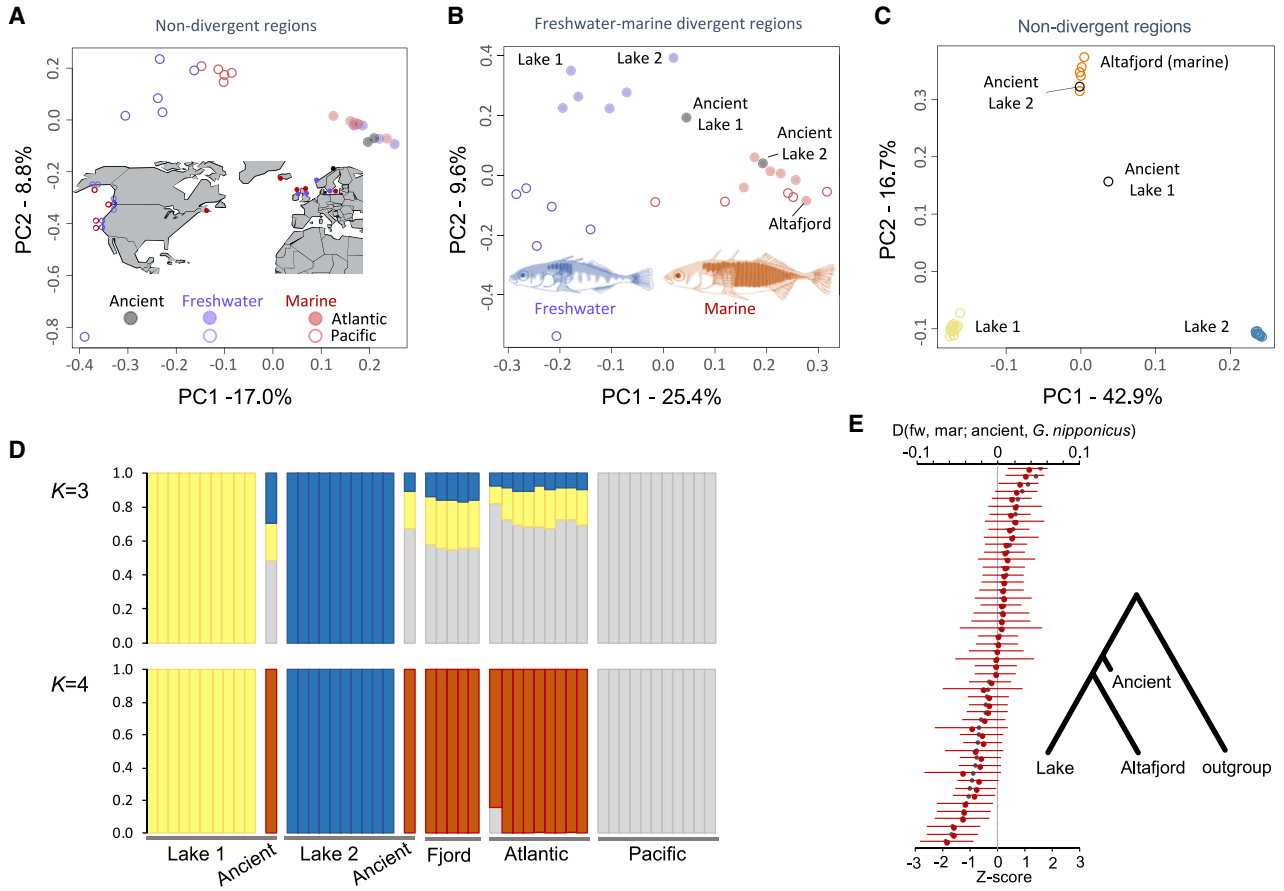


Figure 2. Relationships between ancient and present-day stickleback

(A and B) Principal component analyses (PCA) of the global dataset from Jones et al.,[4] a single modern sample from each of Altafjord, Lake 1, and Lake 2 and ancient samples from Lake 1 and Lake 2, based on (A) transversions in non-divergent regions and (B) transversions in freshwater-marine divergent regions identified by Jones et al.[4]

(C) PCA of local present-day and ancient samples using transversions in non-divergent regions.

(D) Admixture plots of combined global and local populations.

(E) *D*-statistics of the form (Lake 2, Altafjord; ancient, Japan Sea stickleback) testing whether the ancient sample shares more alleles with the present-day lake or fjord samples. The results were not significantly different from zero ($-3 < |Z| < 3$), suggesting the ancient sample is symmetrically related to both. Red markers show *D*-statistics (top axis) and horizontal bars show associated standard error. Black markers show *Z* scores (bottom axis). *D*-statistics therefore support the topology shown in the schematic.

See also Tables S1 and S2 and Methods S1.

Our findings of relative isolation and strong independent drift in the lake populations suggest reduced effective population size (N_e). This has implications for the ability of the population to adapt to freshwater, as the effectiveness with which natural selection fixes advantageous alleles in a population depends not only upon the selection coefficient(s) of an allele, but also on N_e.[22] To better understand how N_e and, by proxy, the efficacy of natural selection had varied through time, we reconstructed the demographic history of fjord and lake populations using the pairwise sequentially Markovian coalescent (PSMC) method.[23] Estimates of N_e for the fjord and lake populations overlap from 100 to 20 KY BP, but this shared demographic history diverges from 20 to 10 KY BP (Figure 3A; Methods S1, section 3)—approximately at the directly dated time isostatic rebound isolated the lakes from the marine population (Figure 1B). Following isolation from the marine population, we observed a steep decline in inferred N_e in both lake populations, consistent with studies of other freshwater lake populations.[24]

Runs of homozygosity (ROH) provide further support for decreasing N_e in recent demographic history.[25] The impact of smaller population size of the lake stickleback results in an increased proportion of the genome being identical by descent and in long ROH, particularly in Lake 2 (Figure 3B). After excluding regions of the genome associated with parallel marine-freshwater divergence,[4] the sum of ROH longer than 300 kb was significantly higher in both lake populations than in the fjord population (Wilcoxon rank-sum test, $p < 0.01$). ROH sum up to 83 Mb for stickleback from Lake 1 and 277 Mb for Lake 2 (Figure S2), corresponding to 18% and 60% of the genome, respectively, indicating that both populations had undergone intense genetic bottlenecks. Considering only regions associated with parallel marine-freshwater divergence[4] (Figure S2),

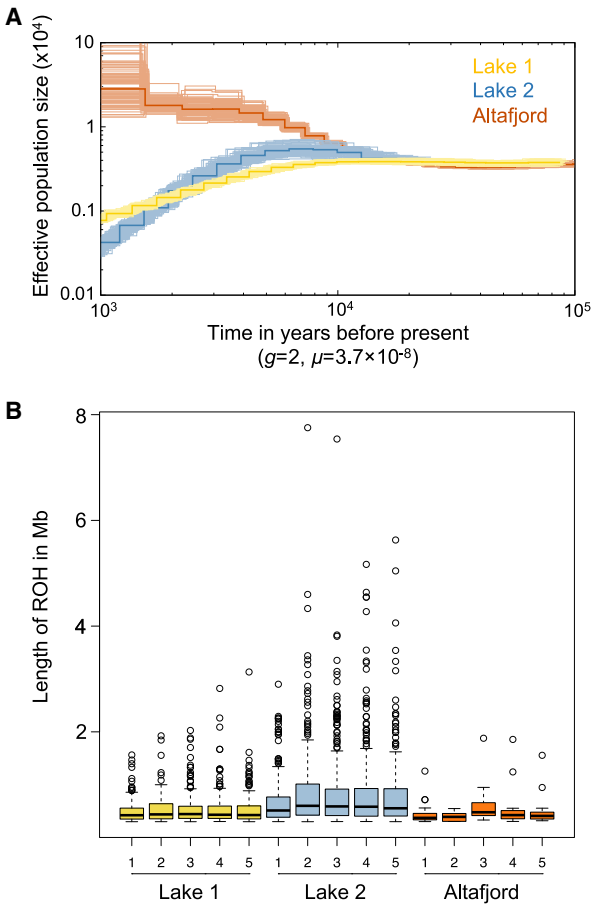


Figure 3. Demographic history of local marine and freshwater stickleback

(A) PSMC estimates of changes in effective population size (N_e) over time inferred from the autosomes of a Lake 1 (yellow), Lake 2 (blue), and Altafjord (orange) sample. Thick lines represent the median and thin light lines of the same color correspond to 100 rounds of bootstrapping.

(B) Distribution of the length of runs of homozygosity (ROH) greater than 0.3 Mb in the genomes of five samples each from Lake 1 (yellow), Lake 2 (blue), and Altafjord (orange). The thick black line shows the median. The bottom and top of the box represent the 1st (Q1) and 3rd (Q3) quartile. The upper whisker corresponds to the smaller value of the maximum length of ROH or the sum of Q3 and 1.5 times the size of the box (Q3-Q1). All values above the upper whisker are shown as black circles. The lower whisker shows the smallest length of ROH for the corresponding individual. The PSMC and ROH analyses exclude known freshwater-marine divergent and 100 kb flanking regions.

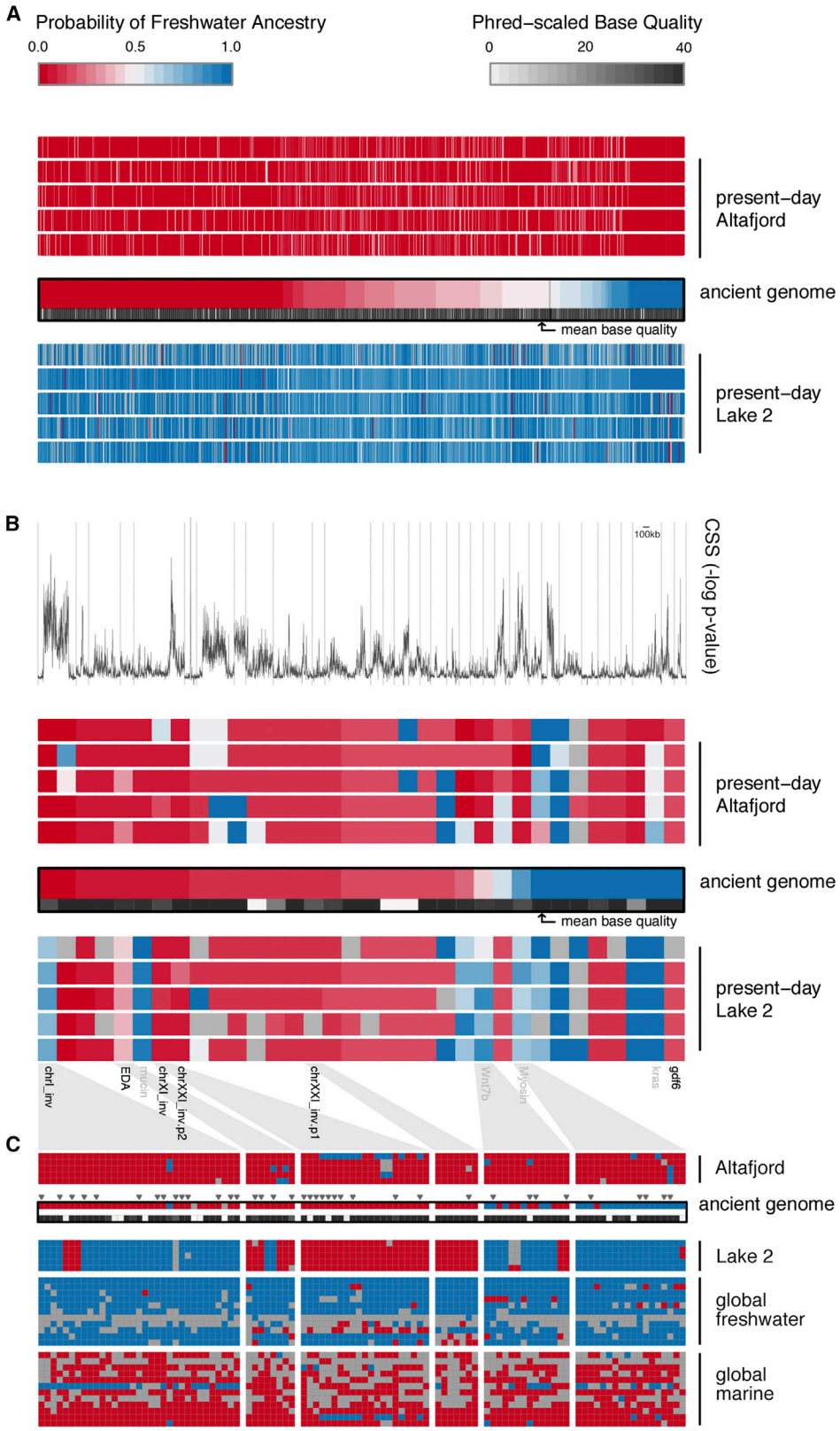
See also Figures S2 and S3 and Methods S1.

we find ROH clustered in known inversions and loci that show strong differentiation between fjord and lake populations (Figure S3; Methods S1, section 4), consistent with selective sweeps.[26]

Our population genetic results indicate an ongoing reduction in effective population size following the colonization of the isolation lakes. This would be expected to reduce efficacy of selection (N_es) on freshwater alleles.[22] In addition to the constraints imposed by reduced effective population size, the progression of adaptation will also be dependent upon the availability of freshwater adaptive alleles upon which selection acts. Our paleogenomic data provide the first opportunity to directly compare standing genetic variation present in a freshwater lake at the start of the freshwater adaptation process to present-day genetic variation. We focused our analyses on the Lake 2 sample, for which we had data from many unlinked genomic regions associated with marine-freshwater adaptation. Using genomic positions with data present in the ancient stickleback sample, and after down-sampling present-day Lake 2 and Altafjord genomes to equivalent levels, we identified 814 regions of the genome that contain sites with strong divergence among present-day freshwater Lake 2 and marine Altafjord fish (locally divergent regions; Supplemental Information). At these locally divergent regions, the ancient stickleback from Lake 2 predominantly shared alleles with the present-day fjord population (Figure 4A). Fixed alleles in these locally divergent regions could represent instances of fixation due to drift in the lake population, rather than having a functional role in freshwater-marine adaption.

At genomic regions underlying marine versus freshwater adaptation based on the parallel divergence among a global dataset of marine and freshwater populations,[4] the ancient genome also carries predominantly marine adapted genotypes, yet carries freshwater genotypes at a greater number of adaptive loci (~24%) than in the comparison between the local populations (Figure 4B). The present-day Lake 2 fish carry similar proportions of globally shared marine and freshwater genotypes to the ancient genome, suggesting incomplete freshwater adaptation. Some regions of the genome enriched for the freshwater alleles in the ancient sample were also enriched for freshwater alleles in the present-day lake population (Figure 4B). However, the overall composition of the adaptive alleles carried by the present-day freshwater stickleback in Lake 2 differs from those found in the ancient genome (Figure 4B). For example, the ancient genome carries marine versions of the chromosome I inversion harboring Na^+/K^+ ion transporter ATPase1a2 and the chromosome II mucin region, whereas the present-day Lake 2 stickleback carry freshwater alleles (Figure 4C). Of course, the ancient sample represents just a single individual, and the ancestral gene pool is expected to have contained a greater diversity of freshwater alleles including those found in the present-day lake stickleback. In contrast, the ancient genome carries freshwater adaptive alleles at some loci where both present-day lake and fjord populations carry marine alleles, e.g., the *GDF6* region, a major effect locus driving bony armor plate size,[27] on chromosome XX (Figure 4B), 17.3 Mb of chromosome IX, and 9.3 Mb of chromosome VIII. Therefore, it appears that some freshwater-adaptive haplotypes available as standing genetic variation during the founding of Lake 2 have subsequently been lost during the past 12,000 years.

There are clear phenotypic and genomic signatures of directional selection in the present-day populations. Stickleback sampled from Lake 1 and 2 were phenotypically low-plated, while stickleback sampled from the fjord were fully plated. The ectodysplasin (*EDA*) signaling pathway has a key role in the parallel evolution of the low-plated freshwater phenotype due to repeated selection of alleles derived from an ancestral low-plated haplotype.[9,28,29] There is strong evidence that these alleles persist as standing genetic variation in the marine population.[4,9,12,13,30] Consistent with the importance of *EDA* in



(legend on next page)

marine-freshwater phenotypic divergence, we find differentiation between the marine fjord and lake populations (Figures S3 and S4), and evidence that both lake populations share the core freshwater haplotype at the EDA locus (Figure S4). ROH are prevalent at this locus in the two lake populations (Figure S3), consistent with a "hard sweep" of an extended haplotype under selection in too short a time frame for recombination to restore genetic variation.[26] Inspecting the underlying genotypes, we find extended freshwater haplotypes shared among individuals within each lake, but differing between individuals in Lake 1 and Lake 2 (extending to the right and left flanks, respectively; Figure S4). In both populations the ancient freshwater haplotype is flanked by marine haplotypes (Figure S4). This further highlights the independence of freshwater adaptation by stickleback in lakes just 25 km apart colonized from a shared ancestral gene pool.

Studies have reported compelling evidence of adaptation of Pacific threespine stickleback to freshwater habitats over decadal timescales,[7,8] assuming that modern marine stickleback are a suitable proxy for the ancestral population. By comparing the genomes of Late Pleistocene stickleback to present-day Atlantic threespine stickleback, we find support for this assumption, with relatively little drift in allele frequencies from the ancient to the present-day marine specimens. We also find the populations in our two study isolation lakes show clear evidence of directional selection acting upon loci such as *EDA*, which underlie marine or freshwater adapted phenotypes. These findings highlight that selection can drive adaptive alleles to high frequency at loci of large effect, even under demographic constraints. However, we do find evidence for a more stochastic process in our isolation lake populations than previously described, resulting in a freshwater optimal genotype not being reached even after several millennia. Our empirical findings are supported by forward simulations (Methods S1, section 5), in which freshwater alleles present as low-frequency standing variation continue to rise to high frequency 10,000 generations after the colonization of "isolation lakes," with limited parallelism between lakes. Our simulations highlight that parallel adaptation is constrained by the frequency of freshwater alleles in the founding population and low migration,[31] and are consistent with our coalescent estimates of changes in effective population size indicating the stochastic loss of freshwater alleles through increased drift during founder-associated population bottlenecks. The stochastic loss of freshwater-adapted alleles during colonization of the Atlantic from the Pacific has been proposed to have reduced parallelism in freshwater adaptation in Atlantic stickleback.[32] Our results suggest these demographic processes also occur during and after the colonization of individual lakes, explaining observations of variation in parallelism of freshwater adaptation globally[32,33] and among geographically proximate lakes.[34,35] Thus, while the adaptation of threespine stickleback to freshwater is a highly deterministic process, we find a significant role for stochasticity in the progression of parallel adaptation.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials Availability
  - Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Ancient sample collection and geological analysis
  - Modern sample collection
- METHOD DETAILS
  - DNA extraction
  - Ancient DNA lab work
  - Mapping, filtering and masking
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Assessing postmortem DNA damage and contamination
  - Sexing
  - Principal Component Analysis
  - Individual Assignment and Admixture Analyses
  - *D*-statistics
  - Mitochondrial DNA analysis
  - Pairwise sequentially Markovian coalescent
  - Runs of homozygosity
  - F_{ST} statistics
  - Comparison of ancient and modern genotypes

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2021.02.027.

Figure 4. Marine versus freshwater adaptive alleles in the ancient and modern samples

(A) The probability of "Lake 2 ancestry" in the ancient sample is plotted for 814 locally divergent regions of the genome (windows containing variant(s) with fixed allele frequency difference among 5 fish from each of Altafjord and Lake 2). Windows are plotted from left to right according to the probability of freshwater ancestry in the ancient genome. For each genomic region, the mapDamage-rescaled base qualities are plotted in grayscale for the ancient genome. The probability of freshwater ancestry in five fish from each of present-day Altafjord and Lake 2 populations, respectively, is shown above and below the ancient genome probabilities.

(B) The probability of freshwater ancestry in the ancient genome is plotted for 34 genomic regions underlying marine versus freshwater adaptation. These regions were identified based on parallel divergence among global marine versus freshwater populations using cluster separation score (CSS).[4] Regions of the genome are plotted from left to right according to the probability of freshwater ancestry in the ancient genome. For each genomic region, the mean mapDamage-rescaled base qualities are plotted in grayscale below the corresponding probability scores for the ancient genome. The probability of freshwater ancestry in the present day Altafjord and Lake 2 populations is shown above and below the ancient genome probabilities, respectively.

(C) Underlying genotypes at a focal subset of adaptive loci. Rows represent individual fish; columns represent individual SNPs; red boxes indicate marine alleles; blue boxes indicate freshwater alleles; gray boxes are missing data. The mean mapDamage-rescaled base qualities are plotted in grayscale below the corresponding ancient genome probabilities for each site. Transversions, which are less prone to DNA damage, are marked by gray triangles.

See also Figure S4 and Methods S1.

ACKNOWLEDGMENTS

We thank two anonymous reviewers whose comments greatly improved this manuscript. We thank Mark Ravinet, Per-Arne Amundsen, and Ian Mayer for help with permitting advice; Christian Carøe for advice on extraction and library build; Mike Martin for logistical support; Lina Gislefoss and Thomas Lakeman for help with lake coring; Camilla Scharff-Olsen and Marta Ciucani for ancient DNA lab support; Mette Juul Jacobsen and Lasse Vinner of the National High-throughput DNA Sequencing Centre, University of Copenhagen, Denmark, for assistance with Illumina sequencing and associated data processing of ancient samples; Enni Harjunmaa for assigning the ancient stickleback bones to their corresponding position in an X-ray scan of a modern stickleback; Kavita Venkataramani for DNA extraction and library preparation of the modern samples; Heike Budde, Katrin Fritschi, and Ilja Bezrukov for assistance with Illumina sequencing and associated data processing; and Katrin Töpner for consultation for creating SLiM simulations. M.K. is supported by the International Max Planck Research School "From Molecules to Organisms" and the German Research Foundation (DFG). F.C.J. is supported by the Deutsche Forschungsgemeinschaft (DFG) SPP1819- JO1316 and the European Research Council FP7 CoG617279. A.D.F. and the ancient DNA lab work and sequencing were supported by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement no. 663830. The geological field work was funded jointly by NGU and the archaeological research project "Stone Age Demographics" at the University of Tromsø.

AUTHOR CONTRIBUTIONS

A.D.F. conceived and coordinated the study. Collection of sediment cores and all geological analyses were conducted by A.R. Modern stickleback were sampled in the field by A.D.F. Ancient DNA lab work was conducted by A.D.F. and M.T.P.G. Modern DNA lab work was conducted by M.K. and F.C.J. Genomic data analyses were conducted by M.K., F.C.J., and A.D.F. M.K. and A.D.F. drafted the initial manuscript with all authors contributing toward the writing of this paper.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 27, 2020
Revised: January 5, 2021
Accepted: February 12, 2021
Published: March 10, 2021

REFERENCES

1. Morris, M.R.J., Bowles, E., Allen, B.E., Jamniczky, H.A., and Rogers, S.M. (2018). Contemporary ancestor? Adaptive divergence from standing genetic variation in Pacific marine threespine stickleback. BMC Evol. Biol. *18*, 113.
2. Dehasque, M., Ávila-Arcos, M.C., Díez-Del-Molino, D., Fumagalli, M., Guschanski, K., Lorenzen, E.D., Malaspinas, A.S., Marques-Bonet, T., Martin, M.D., Murray, G.G.R., et al. (2020). Inference of natural selection from ancient DNA. Evol Lett *4*, 94–108.
3. Romundset, A., Bondevik, S., and Bennike, O. (2011). Postglacial uplift and relative sea level changes in Finnmark, northern Norway. Quat. Sci. Rev. *19–20*, 2398–2421.
4. Jones, F.C., Grabherr, M.G., Chan, Y.F., Russell, P., Mauceli, E., Johnson, J., Swofford, R., Pirun, M., Zody, M.C., White, S., et al.; Broad Institute Genome Sequencing Platform & Whole Genome Assembly Team (2012). The genomic basis of adaptive evolution in threespine sticklebacks. Nature *484*, 55–61.
5. Bell, M.A., and Foster, S.A. (1994). The Evolutionary Biology of the Threespine Stickleback (Oxford University Press).
6. Peichel, C.L., and Marques, D.A. (2017). The genetic and molecular architecture of phenotypic diversity in sticklebacks. Philos. Trans. R. Soc. Lond. B Biol. Sci. *372*, 20150486.
7. Lescak, E.A., Bassham, S.L., Catchen, J., Gelmond, O., Sherbick, M.L., von Hippel, F.A., and Cresko, W.A. (2015). Evolution of stickleback in 50 years on earthquake-uplifted islands. Proc. Natl. Acad. Sci. USA *112*, E7204–E7212.
8. Bell, M.A., Aguirre, W.E., and Buck, N.J. (2004). Twelve years of contemporary armor evolution in a threespine stickleback population. Evolution *58*, 814–824.
9. Colosimo, P.F., Hosemann, K.E., Balabhadra, S., Villarreal, G., Jr., Dickson, M., Grimwood, J., Schmutz, J., Myers, R.M., Schluter, D., and Kingsley, D.M. (2005). Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. Science *307*, 1928–1933.
10. Bassham, S., Catchen, J., Lescak, E., von Hippel, F.A., and Cresko, W.A. (2018). Repeated selection of alternatively adapted haplotypes creates sweeping genomic remodeling in stickleback. Genetics *209*, 921–939.
11. Nelson, T.C., and Cresko, W.A. (2018). Ancient genomic variation underlies repeated ecological adaptation in young stickleback populations. Evol Lett *2*, 9–21.
12. Schluter, D., and Conte, G.L. (2009). Genetics and ecological speciation. Proc. Natl. Acad. Sci. USA *106* (*Suppl 1*), 9955–9962.
13. Roesti, M., Gavrilets, S., Hendry, A.P., Salzburger, W., and Berner, D. (2014). The genomic signature of parallel adaptation from shared genetic variation. Mol. Ecol. *23*, 3944–3956.
14. Romundset, A., Lakeman, T.R., and Høgaas, F. (2018). Quantifying variable rates of postglacial relative sea level fall from a cluster of 24 isolation basins in southern Norway. Quat. Sci. Rev. *197*, 175–192.
15. Sawyer, S., Krause, J., Guschanski, K., Savolainen, V., and Pääbo, S. (2012). Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. PLoS ONE *7*, e34131.
16. Oosting, T., Star, B., Barrett, J.H., Wellenreuther, M., Ritchie, P.A., and Rawlence, N.J. (2019). Unlocking the potential of ancient fish DNA in the genomic era. Evol. Appl. *12*, 1513–1522.
17. Skoglund, P., Ersmark, E., Palkopoulou, E., and Dalén, L. (2015). Ancient wolf genome reveals an early divergence of domestic dog ancestors and admixture into high-latitude breeds. Curr. Biol. *25*, 1515–1519.
18. Palkopoulou, E., Mallick, S., Skoglund, P., Enk, J., Rohland, N., Li, H., Omrak, A., Vartanyan, S., Poinar, H., Götherström, A., et al. (2015). Complete genomes reveal signatures of demographic and genetic declines in the woolly mammoth. Curr. Biol. *25*, 1395–1400.
19. Ramos-Madrigal, J., Smith, B.D., Moreno-Mayar, J.V., Gopalakrishnan, S., Ross-Ibarra, J., Gilbert, M.T.P., and Wales, N. (2016). Genome sequence of a 5,310-year-old maize cob provides insights into the early stages of maize domestication. Curr. Biol. *26*, 3195–3201.
20. Noonan, J.P., Coop, G., Kudaravalli, S., Smith, D., Krause, J., Alessi, J., Chen, F., Platt, D., Pääbo, S., Pritchard, J.K., and Rubin, E.M. (2006). Sequencing and analysis of Neanderthal genomic DNA. Science *314*, 1113–1118.
21. McVean, G. (2009). A genealogical interpretation of principal components analysis. PLoS Genet. *5*, e1000686.
22. Charlesworth, B. (2009). Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. Nat. Rev. Genet. *10*, 195–205.
23. Li, H., and Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. Nature *475*, 493–496.
24. Liu, S., Hansen, M.M., and Jacobsen, M.W. (2016). Region-wide and ecotype-specific differences in demographic histories of threespine stickleback populations, estimated from whole genome sequences. Mol. Ecol. *25*, 5187–5202.
25. Ceballos, F.C., Joshi, P.K., Clark, D.W., Ramsay, M., and Wilson, J.F. (2018). Runs of homozygosity: windows into population history and trait architecture. Nat. Rev. Genet. *19*, 220–234.

26. Hermisson, J., and Pennings, P.S. (2017). Soft sweeps and beyond: understanding the patterns and probabilities of selection footprints under rapid adaptation. Methods Ecol. Evol. *8*, 700–716.
27. Indjeian, V.B., Kingman, G.A., Jones, F.C., Guenther, C.A., Grimwood, J., Schmutz, J., Myers, R.M., and Kingsley, D.M. (2016). Evolving new skeletal traits by cis-regulatory changes in bone morphogenetic proteins. Cell *164*, 45–56.
28. Peichel, C.L., Nereng, K.S., Ohgi, K.A., Cole, B.L.E., Colosimo, P.F., Buerkle, C.A., Schluter, D., and Kingsley, D.M. (2001). The genetic architecture of divergence between threespine stickleback species. Nature *414*, 901–905.
29. Colosimo, P.F., Peichel, C.L., Nereng, K., Blackman, B.K., Shapiro, M.D., Schluter, D., and Kingsley, D.M. (2004). The genetic architecture of parallel armor plate reduction in threespine sticklebacks. PLoS Biol. *2*, E109.
30. Barrett, R.D.H., and Schluter, D. (2008). Adaptation from standing genetic variation. Trends Ecol. Evol. *23*, 38–44.
31. Galloway, J., Cresko, W.A., and Ralph, P. (2019). A few stickleback suffice for the transport of alleles to new lakes. G3 *10*, 505–514.
32. Fang, B., Kemppainen, P., Momigliano, P., Feng, X., and Merilä, J. (2020). On the causes of geographically heterogeneous parallel evolution in sticklebacks. Nat. Ecol. Evol. *4*, 1105–1115.
33. DeFaveri, J., Shikano, T., Shimada, Y., Goto, A., and Merilä, J. (2011). Global analysis of genes involved in freshwater adaptation in threespine sticklebacks (Gasterosteus aculeatus). Evolution *65*, 1800–1807.
34. Liu, S., Ferchaud, A.-L., Grønkjaer, P., Nygaard, R., and Hansen, M.M. (2018). Genomic parallelism and lack thereof in contrasting systems of three-spined sticklebacks. Mol. Ecol. *27*, 4725–4743.
35. Terekhanova, N.V., Barmintseva, A.E., Kondrashov, A.S., Bazykin, G.A., and Mugue, N.S. (2019). Architecture of parallel adaptation in ten lacustrine threespine stickleback populations from the White Sea area. Genome Biol. Evol. *11*, 2605–2618.
36. Schubert, M., Lindgreen, S., and Orlando, L. (2016). AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res. Notes *9*, 88.
37. Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: analysis of next generation sequencing data. BMC Bioinformatics *15*, 356.
38. Wang, L., Zhang, W., and Li, Q. (2020). AssocTests: an R package for genetic association studies. J. Stat. Soft. *94*, https://doi.org/10.18637/jss.v094.i05.
39. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al.; 1000 Genomes Project Analysis Group (2011). The variant call format and VCFtools. Bioinformatics *27*, 2156–2158.
40. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.
41. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P.L.F., and Orlando, L. (2013). mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics *29*, 1682–1684.
42. Skotte, L., Korneliussen, T.S., and Albrechtsen, A. (2013). Estimating individual admixture proportions from next generation sequencing data. Genetics *195*, 693–702.
43. Meisner, J., and Albrechtsen, A. (2018). Inferring population structure and admixture proportions in low-depth NGS data. Genetics *210*, 719–731.
44. Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. *52*, 696–704.
45. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.
46. Skoglund, P., Northoff, B.H., Shunkov, M.V., Derevianko, A.P., Pääbo, S., Krause, J., and Jakobsson, M. (2014). Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. Proc. Natl. Acad. Sci. USA *111*, 2229–2234.
47. Smit, A., Hubley, R., and Green, P. (2015). RepeatMasker Open-4.0. http://www.repeatmasker.org.
48. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.
49. Haller, B.C., and Messer, P.W. (2019). SLiM 3: forward genetic simulations beyond the Wright-Fisher Model. Mol. Biol. Evol. *36*, 632–637.
50. Morgulis, A., Gertz, E.M., Schäffer, A.A., and Agarwala, R. (2006). WindowMasker: window-based masker for sequenced genomes. Bioinformatics *22*, 134–141.
51. Jowsey, P.C. (1966). An improved peat sampler. New Phytol. *65*, 245–248.
52. Bennike, O. (1995). Palaeoecology of two lake basins from Disko, West Greenland. J. Quaternary Sci. *10*, 149–155.
53. Romundset, A., Fredin, O., and Høgaas, F. (2015). A Holocene sea-level curve and revised isobase map based on isolation basins from near the southern tip of Norway. Boreas *44*, 383–400.
54. Björck, S., Bennike, O., Ingólfsson, Ó., Barnekow, L., and Penney, D.N. (2008). Lake Boksehandsken's earliest postglacial sediments and their palaeoenvironmental implications, Jameson Land, East Greenland. Boreas *23*, 459–472.
55. Carøe, C., Gopalakrishnan, S., Vinner, L., Mak, S.S.T., Sinding, M.H.S., Samaniego, J.A., Wales, N., Sicheritz-Pontén, T., and Gilbert, M.T.P. (2018). Single-tube library preparation for degraded DNA. Methods Ecol. Evol. *9*, 410–419.
56. Schubert, M., Ginolhac, A., Lindgreen, S., Thompson, J.F., Al-Rasheid, K.A., Willerslev, E., Krogh, A., and Orlando, L. (2012). Improving ancient DNA read mapping against modern reference genomes. BMC Genomics *13*, 178.
57. Dabney, J., Meyer, M., and Pääbo, S. (2013). Ancient DNA damage. Cold Spring Harb. Perspect. Biol. *5*, a012567.
58. Peyrégne, S., and Prüfer, K. (2020). Present-day DNA contamination in ancient DNA datasets. BioEssays *42*, e2000081.
59. Fang, B., Merilä, J., Ribeiro, F., Alexandre, C.M., and Momigliano, P. (2018). Worldwide phylogeny of three-spined sticklebacks. Mol. Phylogenet. Evol. *127*, 613–625.
60. Tracy, C.A., and Widom, H. (1994). Level-spacing distributions and the airy kernel. Commun. Math. Phys. *159*, 151–174.
61. Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. *14*, 2611–2620.
62. Lawson, D.J., van Dorp, L., and Falush, D. (2018). A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. Nat. Commun. *9*, 3258.
63. Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. PLoS Genet. *2*, e190.
64. Janes, J.K., Miller, J.M., Dupuis, J.R., Malenfant, R.M., Gorrell, J.C., Cullingham, C.I., and Andrew, R.L. (2017). The K = 2 conundrum. Mol. Ecol. *26*, 3594–3602.
65. Cullingham, C.I., Miller, J.M., Peery, R.M., Dupuis, J.R., Malenfant, R.M., Gorrell, J.C., and Janes, J.K. (2020). Confidently identifying the correct K value using the ΔK method: when does K = 2? Mol. Ecol. *29*, 862–869.
66. Durand, E.Y., Patterson, N., Reich, D., and Slatkin, M. (2011). Testing for ancient admixture between closely related populations. Mol. Biol. Evol. *28*, 2239–2252.
67. Mäkinen, H.S., and Merilä, J. (2008). Mitochondrial DNA phylogeography of the three-spined stickleback (Gasterosteus aculeatus) in Europe-evidence for multiple glacial refugia. Mol. Phylogenet. Evol. *46*, 167–182.
68. Feulner, P.G.D., Chain, F.J.J., Panchal, M., Huang, Y., Eizaguirre, C., Kalbe, M., Lenz, T.L., Samonte, I.E., Stoll, M., Bornberg-Bauer, E., et al. (2015). Genomics of divergence along a continuum of parapatric population differentiation. PLoS Genet. *11*, e1004966.

69. Ceballos, F.C., Hazelhurst, S., and Ramsay, M. (2018). Assessing runs of homozygosity: a comparison of SNP array and whole genome sequence low coverage data. BMC Genomics *19*, 106.

70. Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics *27*, 2987–2993.

71. Weir, B.S., and Cockerham, C.C. (1984). Estimating F-statistics for the analysis of population structure. Evolution *38*, 1358–1370.

72. R Development Core Team (2010). R: a language and environment for statistical computing: reference index (R Foundation for Statistical Computing).

73. Hartl, D.L., and Clark, A.G. (2007). Principles of Population Genetics, 4th ed. (Sinauer).

74. Friedman, J.H. (1984). A variable span scatterplot smoother. Laboratory for Computational Statistics, Stanford University Technical Report No. 5., 30.

75. Friedman, J.H. (1984). SMART user's guide (Laboratory for Computational Statistics, Dept. of Statistics, Stanford University).

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples		
Stickleback fin clip	This study	SAMN17375686
Stickleback fin clip	This study	SAMN17377651
Stickleback fin clip	This study	SAMN17377676
Stickleback fin clip	This study	SAMN17377678
Stickleback fin clip	This study	SAMN17377679
Stickleback fin clip	This study	SAMN17377782
Stickleback fin clip	This study	SAMN17377785
Stickleback fin clip	This study	SAMN17377787
Stickleback fin clip	This study	SAMN17377788
Stickleback fin clip	This study	SAMN17377789
Stickleback fin clip	This study	SAMN17377790
Stickleback fin clip	This study	SAMN17377791
Stickleback fin clip	This study	SAMN17377792
Stickleback fin clip	This study	SAMN17377793
Stickleback fin clip	This study	SAMN17377794
Stickleback fin clip	This study	SAMN17377795
Stickleback fin clip	This study	SAMN17377796
Stickleback fin clip	This study	SAMN17377797
Stickleback fin clip	This study	SAMN17377798
Stickleback fin clip	This study	SAMN17377799
Stickleback fin clip	This study	SAMN17377800
Stickleback fin clip	This study	SAMN17377804
Stickleback fin clip	This study	SAMN17377805
Stickleback fin clip	This study	SAMN17377806
Stickleback fin clip	This study	SAMN17377807
Stickleback fin clip	This study	SAMN17377886
Stickleback fin clip	This study	SAMN17377887
Stickleback fin clip	This study	SAMN17377888
Stickleback fin clip	This study	SAMN17377889
Stickleback fin clip	This study	SAMN17377898
Osteological remain	This study	SAMN17514713
Osteological remain	This study	SAMN17514712
Chemicals, Peptides, and Recombinant Proteins		
AmpliTaq Gold	Thermo Fisher Scientific	Cat# N8080241
N-Lauroylsarcosine solution 30% 500ml	Dutscher	Cat# 348533
Proteinase K 100MG	Thermo Fisher Scientific	Cat# 10103533
H2O, Molecular Biology Grade	Thermo Fisher Scientific	Cat# 10490025
Tween 20 100ML	Thermo Fisher Scientific	Cat# 10113103
Ethanol, Absolute, Mol Biology Grade	Thermo Fisher Scientific	Cat# 10644795
5M Sodium Chloride 100ML	Thermo Fisher Scientific	Cat# 10609823

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
guanidine hydrochloride	Sigma-Aldrich	Cat# G3272
2-Propanol	Merck	Cat# 109634
NEBNext End Repair Module	New England Biolabs	Cat# E6050L
Bst DNA Polymerase	New England Biolabs	Cat# M0275L
NEBNext Quick Ligation Module	New England Biolabs	Cat# E6056L
ACCUPRIME PFX DNA POLYMERASE 100mL	Thermo Fisher Scientific	Cat# 10472482
Agencourt AMPure XP - 60ml	Beckman Coulter	Cat# A63881
Buffer PE	QIAGEN	Cat# 19065
Buffer PB	QIAGEN	Cat# 19066
Buffer EB	QIAGEN	Cat# 19086
DMSO molecular biol 250ML	Thermo Fisher Scientific	Cat# 10397841
dNTP Set 100mM 100mL	Thermo Fisher Scientific	Cat# 10336653
EDTA 0.5M pH 8.0 Fisher Bioreagents 500ML	Thermo Fisher Scientific	Cat# 10182903
Tris HCl, 1M, pH 8.0, 100ML	Thermo Fisher Scientific	Cat# 10336763
Q5 High-Fidelity DNA Polymerase	New England Biolabs	Cat# M0491L
Critical Commercial Assays		
MinElute PCR Purification kit	QIAGEN	Cat# 28006
Tapestation screenTape D1000 HS	Agilent	Cat# 5067-5584
Qubit dsDNA HS Assay Kit	Thermo Fisher Scientific	Cat# Q32854
Deposited Data		
Raw sequence data (fastq format)	This study	NCBI: PRJNA693136
Software and Algorithms		
AdapterRemoval v2	[36]	https://github.com/MikkelSchubert/adapterremoval
ANGSD	[37]	http://www.popgen.dk/angsd/index.php/ANGSD
AssocTests	[38]	https://cran.r-project.org/web/packages/AssocTests/index.html
bcftools	[39]	http://samtools.github.io/bcftools/bcftools.html
BWA	[40]	https://github.com/lh3/bwa
mapDamage	[41]	https://ginolhac.github.io/mapDamage/
NGSadmix	[42]	http://www.popgen.dk/software/index.php/NgsAdmixv2
PCAngsd	[43]	https://github.com/Rosemeis/pcangsd
PhyML	[44]	https://github.com/stephaneguindon/phyml
PLINK	[45]	http://zzz.bwh.harvard.edu/plink/
PMDtools	[46]	https://github.com/pontussk/PMDtools

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
PSMC	[23]	https://github.com/lh3/psmc
RepeatMasker	[47]	http://www.repeatmasker.org
Samtools	[48]	http://www.htslib.org
SLiM 3.0	[49]	https://messerlab.org/slim/
VCFtools	[39]	https://vcftools.github.io/man_latest.html
WindowMasker	[50]	https://github.com/goeckslab/WindowMasker

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources, material and reagents should be addressed and will be fulfilled by the lead contact, Andrew Foote (andrew.foote@ntnu.no).

Materials Availability

Raw sequence data and BioSample details are available at the National Centre for Biotechnology Information (NCBI) under BioProject accession number: PRJNA693136.

Data and Code Availability

The accession number for the genomic data generated for this study is NCBI: PRJNA693136. Previously published genomic data of 20 global samples from Jones et al.[4] were downloaded from NCBI in SRA format (SAMN00627549-SAMN00627550; SAMN00627914-SAMN00630301). The code used in this study is listed in the Key Resources Table and at https://github.com/Stickle-Back-in-Time.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Ancient sample collection and geological analysis

Sediment core samples from a number of lakes in the Finnmark region were collected in late spring 2018 and 2019, as part of a study on postglacial relative sea-level changes. Both of the lakes presented in this paper, are relatively shallow and were cored with a "Russian-type" peat corer.[51] The one meter-long, half-cylinder-shaped samples of lake deposits were collected and transported to the laboratory. As part of the sediment analysis, bulk samples of core material were carefully subsampled and wet-sieved at 125-μm for analysis of macroscopic remains of biota. During basin isolation from the sea, fundamental environmental changes lead to a complete replacement of floral and faunal assemblages, which is apparent in sediment core biostratigraphy.[14] Marine to lacustrine transitions are often visually distinct as laminated facies and can usually be further determined to within a few centimeters, using preserved remains from certain marine, brackish and freshwater organisms. One such organism which has often been found indicative of a lake in the isolation phase is the three-spine stickleback.[14,52–54] When found, stickleback bones were carefully picked out from the wet-sieved residual material, cleaned and dried at low temperature, before being sent for biological analysis.

Marine-lacustrine transitions were identified using biostratigraphy,[14] before sediment cores were subsampled for material suitable for radiocarbon dating. 1-cm-thick slices of the core were wet-sieved and residual terrestrial plant remains were identified, picked and dried overnight before being submitted to the Poznan Radiocarbon Laboratory, Poland. For Klubbvatnet, this comprised one sample (Lab no Poz-111167) of seeds (not identified to species), a small twig and some mosses, weighing in total 12 mg, yielded a radiocarbon age of 10140 ± 50 years, which corresponds to a calendar age interval of 12040-11410 cal year BP (2 sigma). For Jossavannet, a series of four samples were dated across the transition. Most importantly, a sample of Salix leaves (Poz-115352, 7 mg) picked at the boundary yielded 11080 ± 50 radiocarbon years, calibrated to 13070-12800 cal year BP. This is supported by a marine sample of algae found 7 cm deeper (large sample Pox-115492, weighing 61 mg), which yielded 11780 ± 50, calibrated to 13330-13070 cal year BP. It is further supported by a third sample of Salix leaves (Poz-115350, 5 mg) found shortly above the transition, yielding 11420 ± 50, calibrated to 12920-12700 cal year BP. All calibration using OxCal software with IntCal.

Modern sample collection

Adult threespine stickleback specimens were collected using minnow traps from the two lakes from which our ancient stickleback samples originated: fifteen individuals from the 1600 m^2 Klubbvatnet freshwater lake above the village of Neverfjord (70° 36′ N, 23°

37′ E; Lake 1), and ten individuals from 9800 m^2 Jossavannet freshwater lake (70° 27′ N, 23° 47′ E; Lake 2). Additionally, five samples of the marine ecotype were collected from the outer branch area (next to the lake sites) of Altafjord (70° 27′ N, 23° 46′ E). Samples were collected under permit (201300202-62) from Finnmark Fylkeskommune. Upon sampling, stickleback were euthanized and stored in 95% ethanol.

METHOD DETAILS

DNA extraction

Lab work was performed at the Friedrich Miescher Laboratory of the Max Planck Society in Tübingen. Fin clips of the collected sticklebacks were used for standard Proteinase K digestion (New England Biolabs GmbH, Frankfurt am Main, Germany). Each sample was first incubated for 5 h at 58°C in 400 μL lysis buffer (50mM Tris-HCl pH = 8, 0.1M NaCl, 10mM EDTA pH = 8, 0.8% SDS, 15 μg proteinase K) and for 30 min at 37°C with additional 2 μg RNase A. After adding 150 μL 5M potassium acetate, the sample was stored at 4°C overnight and centrifuged at 4000 rpm for 30 min. Extracted DNA was purified from the supernatant via AmpureXP bead purification (Beckman Coulter GmbH, Krefeld, Germany).

Nextera library preparation was performed with assembled Tn5 bound to magnetic beads. 5 mL of Hydrophilic Streptavidin Magnetic Beads (NEB) were transferred into a falcon tube, placed on a magnet and the bead-storage buffer was removed. The beads were washed with 20 mL streptavidin binding buffer (0.6 M NaCl, 10 mM Tris pH = 8.0, 0.5 mM EDTA, 0.1% Triton X- 100). Afterward 30 mL streptavidin binding buffer as well as 400 μL assembled Tn5 were added to the magnetic beads. The mixture was inverted and rotated at 10 rpm at RT for 30 min. The falcon tube was again placed on the magnet, the buffer was removed and replaced by 30 mL of Dialysis buffer (50 mM HEPES-KOH pH 7.2, 0.2 M NaCl, 0.2 mM EDTA, 2 mM DTT, 0.2% Triton X-100, 20% glycerol). The suspension was rotated for another 5 min at 10 rpm, the buffer was removed and 20 mL Dialysis buffer were added for storage of Tn5-on-beads.

2 μL extracted DNA was used for tagmentation with 10 μL Tn5-on-beads in 1x TAPS-DMF buffer (10 mM TAPS, 5 mM $MgCl_2$, 10% DMF) for 15 min at 55°C (total volume 20 μl). 50 μl SDS wash buffer (10 mM Tris pH = 8, 30 mM NaCl, 0.1% Triton X-100, 0.3% SDS) were added to the tagmented product for SDS stripping and the suspension was incubated at 55°C for further 4 min. To remove SDS, the samples were placed on a magnet, the supernatant was removed and the tagmented DNA bound to the beads was washed twice with wash buffer (10 mM Tris pH = 8, 30 mM NaCl, 0.1% Triton X-100).The library was dual indexed and amplified in a 9- cycle PCR using Q5 High Fidelity Polymerase (Biolabs New England). 50 μL PCR-Mastermix (200μM dNTPs, 300μM of each nextera primer, 1 U Q5 High Fidelity Polymerase, 1x Q5 buffer) was therefore added to each sample. PCR temperature profile included an initial step at 72°C for 5 min to fill up the 9 basepair long gaps made by Tn5 during the tagmentation, an activation step at 98°C for 30 s, followed by 9 cycles of denaturation at 98°C for 15 s, annealing at 65°C for 20 s and elongation at 72°C for 90 s. Amplified DNA was then purified with AmpureXP bead purification (Beckman Coulter GmbH, Krefeld, Germany).

Ancient DNA lab work

Ancient DNA lab work was conducted at the dedicated ancient DNA facilities at the Centre for GeoGenetics, University of Copenhagen. DNA was extracted using a silica-based method, where each individual bone or spine was incubated overnight under motion at 55°C in 500 μL extraction buffer (0.45 M EDTA, 0.1 M UREA, 100 μg proteinase K). Each sample was then centrifuged at 2300 rpm for 5 min and the supernatant was collected and concentrated and purified using a Zymo-Spin V reservoir (Zymo Research Irvine, CA, USA) and QIAGEN MinElute spin column (QIAGEN, Valencia, CA, USA). To maximize library complexity by reducing the number of DNA purification steps during library preparation, an Illumina library was constructed using the blunt-end single tube (B.E.S.T.) method.[55] The library was dual indexed and amplified in either a 15-cycle (Lake1 sample) or 20-cycle (Lake 2 sample) PCR using AmpliTaq Gold (ThermoFisher Scientific). The 50 μL PCR reaction contained 15 μL of library, 25 μM dNTP, 1x PCR buffer, 2.5 mM $MgCl_2$, and was made up to 50 μL with molecular grade water. PCR temperature profile included an activation step at 95°C for 5 min, followed by 15/20 cycles (sample dependent: lake1/lake2) of denaturation at 95°C for 30 s, annealing at 55°C for 30 s and elongation at 72°C for 1 min, with a final extension step at 72°C for 7 min. PCR products were then purified using Agencourt AMPure XP beads (BeckmanCoulter). The dual index amplified library of the Lake 1 sample had mean insert size of 186bp, including 114bp of index and adapters, indicating mean DNA fragment size of 72bp. Peak molarity was 286 pmol/l, comprising 34% of the library (including adaptor dimer, lower and upper size markers) as quantified using an Agilent 2200 TapeStation instrument with D1000 High Sensitivity ScreenTape and reagents. The dual index amplified library of the Lake 2 sample had mean insert size of 186bp, including 114bp of index and adapters, indicating mean DNA fragment size of 72bp. Peak molarity was 26,100 pmol/l, comprising 85% of the library (including lower and upper size marker peaks). Extraction, library build and index PCR blanks were also included to evaluate potential contamination during the library building process. To achieve the minimum threshold for DNA molarity, the Lake 1 DNA library was then pooled with an ancient killer whale *Orcinus orca* DNA library and sequenced across two lanes of 80bp- SE sequencing of an Illumina HiSeq4000, and the Lake 2 library was run across an entire single lane of 80bp-SE sequencing of an Illumina HiSeq4000 at the Danish National High-throughput Sequencing Centre of Copenhagen University (https://globe.ku.dk/infrastructure-and-facilities/seqcenter/).

Mapping, filtering and masking

Sequencing data from modern samples generated for this study were demultiplexed and adapters removed using bcl2fastq. Sequencing data of 20 global samples (Bear Paw Lake sample was excluded as this was also the reference sample) from Jones

et al.[4] were downloaded from NCBI in SRA format (SAMN00627549-SAMN00627550; SAMN00627914 - SAMN00630301). SRA files were transformed to fastq files by using fastq-dump. Sequencing of the ancient samples resulted in 1,381,302 reads for Lake 1 and 627,366,889 million reads for Lake 2. AdapterRemoval[36] removed adapters from the single-end reads of the ancient samples of Lake 1 and Lake 2 and trimmed both Ns and low-quality bases from the reads. The generated fastq files of the 20 global samples, as well as trimmed sequencing data of both ancient samples, were aligned against the reference genome gasAcu1[4] by using the Burrows-Wheeler Alignment Tool (BWA) with the aln algorithm,[40] disabling seeding (option'-l 1024) to turn-off seeding thereby increasing mapped data by including reads with post mortem damage at the read ends.[56] Based on the proportion of mapped and unmapped reads, endogenous content was estimated at approximately 2%–4% for each sample. Sequencing data aligned to the stickleback reference genome encompasses 9,127 reads for the ancient sample from Lake 1 and 7,199,676 reads for the ancient sample from Lake 2, corresponding to coverage of $\geq 1\times$ at 369,344 bp and 16,923,179 bp from the Lake 1 and Lake 2 ancient samples respectively. The resulting bam files were sorted and merged by Samtools.[48] Sequencing data of all modern samples were subsequently aligned against the reference genome gasAcu1[4] using the BWA mem algorithm,[40] mean coverage for each sample is given in Table S1. Generated bam files were then sorted by Samtools.[48] All types of duplicates in all sorted bam files were identified by MarkDuplicates from Picard Tools (http://broadinstitute.github.io/picard/). Masked regions as well as the sex chromosome (chrXIX) were removed from the bam files. Masked regions encompassed interspersed repeats and low complexity DNA sequences detected by RepeatMasker[47] covering 3.72% of the stickleback genome as well as highly repetitive DNA sequences detected by WindowMasker[50] from the NCBI C++ toolkit covering 25.59% of the stickleback genome using -sdust true as setting. After removing duplicates and masked regions, 5,054 reads were left for the ancient sample of Lake 1 and 329,226 reads for Lake 2.

QUANTIFICATION AND STATISTICAL ANALYSIS

Assessing postmortem DNA damage and contamination

Analyses of potential nucleotide misincorporations using PMDtools[46] to compare with the modern reference genome revealed that sequencing reads exhibited characteristic post-mortem damage patterns,[15,57] specifically an excess of C→T transitions at the 5′ termini as expected from deamination, and the complementary G→A transitions at the 3′ termini. Therefore, except where otherwise stated, only transversions were considered in downstream analyses that included the ancient samples. Contamination from present-day DNA can be estimated from the number of heterozygous calls in haploid markers.[58] As both our ancient stickleback were females (see below), contamination could only be inferred from the mitochondrial genomes. Coverage was 1x across most sites, and we detected no heterozygous genotypes in either fish. Furthermore, all segregating sites conformed to the expected haplotype structure based on comparison with modern samples in the same clade. Taken together with the high proportion of reads with DNA damage patterns at the read ends, these results suggest our genome data represented endogenous DNA from the ancient stickleback bones.

Sexing

Sticklebacks have an XY sex-determination system, where males are the heterogametic sex. Males are therefore haploid for the X chromosome and diploid for the autosomes, while females are diploid for both the X chromosome and autosomes. The sex of the ancient samples was determined by comparing the mean coverage of the autosomes (excluding unassembled scaffolds), the pseudoautosomal region (chrXIX:1-3300000 and chrXIX:12270000-20240660), and the sex-determining region (chrXIX:3300000-12270000) among 4 female and 4 male modern genomes down-sampled to comparable coverage with each ancient genome. Both ancient samples had approximately equivalent coverage across all three regions implying that both are diploid for the sex-determining region of the X chromosome, i.e., are female (Methods S1, section 6). For each modern individual the down-sampling was repeated 10 times with different random seeds to initiate the down-sampling.

Principal Component Analysis

We used pseudo-haploid genotype calls of globally distributed modern marine and freshwater genomes[4] and the ancient samples. First, we sampled autosomal regions outside of known freshwater-marine divergence associated regions as identified by Jones et al.[4] to compare across geographically informative markers. Then, we compared covariance within freshwater- marine divergence associated regions among samples using ecology informative markers. The ancient samples were included in the PC computations and not projected onto PCs of modern samples, which has the advantage of providing a quality control measure. For example, if the ancient samples were impacted by sequencing- or sequence data processing errors, the samples would appear as outliers in the PCA. Instead they cluster with Atlantic and marine samples respectively in the first two PCA plots (Figure 2A). In the first PCA, both ancient samples were at the extreme edge of PC1 showing closest affinity to the Atlantic samples. Inclusion of a single randomly selected modern sample from Altafjord, Lake 1 and Lake 2 showed that the modern samples also cluster at the edge of PC1 (Figure 2A). This suggests structure among the Atlantic samples, potentially reflecting past biogeographical processes[59] and gene flow during the Holocene between Pacific and Atlantic lineages, rather than artifacts of unmasked DNA damage or missingness of the data in the ancient samples. In Figure 2A, PC1 clearly separates out Atlantic from Pacific samples, with no sub-clustering of California samples, suggesting the signal was driven by the inclusion of multiple samples adjacent populations. Additional filtering steps included in these analyses were the removal of regions of poor mapping quality ($Q < 30$), removal of sites with low base quality scores ($q < 20$), calling only SNPs inferred with a likelihood ratio test (LRT) of $p < 0.000001$, a minimum allele frequency of 0.05 so that alleles had to be called in a minimum of two individuals, discarding reads that did not map uniquely, adjusting q-scores around indels,

adjusting mapping quality to 50 for excessive mismatches, discarding bad reads (flag > = 256), and the removal of transitions to avoid bias from C→T and A→G DNA damage patterns. The eigenvectors from the covariance matrix were generated with the R function "eigen", and significance was determined with a Tracy-Widom test[60] performed in the R-package AssocTest[38] to evaluate the statistical significance of each principal component.

The relationship of the 30 modern samples collected from the two lakes and fjord in Finnmark, and the two ancient samples was explored using PCAngsd, a Principal Component Analysis for low depth next-generation sequencing data using genotype likelihoods, thereby accounting for the uncertainty in the called genotypes which is inherently present in low-depth sequencing data.[43] We restricted the analyses to transversions covered in at least one of the two ancient samples, which restricted the dataset to 2,267 SNPs in autosomal chromosomes and outside of freshwater-marine divergent regions. Over 90% of the included SNPs were > 100,000 bp apart, thus reducing autocorrelation in covariance due to linkage disequilibrium. Filtering steps were as specified above. The statistical significance of each principal component was estimated from the eigenvalues as above.

Individual Assignment and Admixture Analyses

An individual-based assignment test was performed using NGSadmix,[42] a maximum likelihood method that bases its inference on genotype likelihoods. The input genotype likelihood values were the same as those use above for PCAngsd, as were the filtering steps. NGSadmix was run with the number of ancestral populations K set from 2–10. For each of these K values, NGSadmix was re-run five times for each value of K, and with different seeds to ensure convergence. The uppermost hierarchical level of structure, inferred from the greatest stepwise increase in log likelihood, ΔK,[61] identified two clusters. PCA and individual assignment- admixture models draw inference from similar information and therefore generate similar axes of variation.[62,63] Both methods typically identify the samples with the greatest population- specific drift that therefore share derived alleles that were rare in the source populations or have lost ancestral alleles from standing variation, as the major axes of structure.[62] Accordingly, both PCAngsd and NGSadmix identified the uppermost hierarchical level of structure within our dataset as being between the Lake 1 and Lake 2 with other samples sharing ancestry with both. However, the ΔK method is prone to over- or under-estimating population genetic structure[64] and performs poorly under scenarios with migration among populations at inferring hierarchical population structure.[65] Changes in likelihoods show the biggest jumps between $K = 2$ and $K = 3$, and between $K = 3$ and $K = 4$, before plateauing to a gradual rising rate in likelihood. Likelihood estimates for $K = 3$ and $K = 4$ were also highly consistent, having the lowest standard deviations. We therefore present the admixture plots for $K = 3$ and $K = 4$ in Figure 2D.

D-statistics

To investigate whether pairs of modern stickleback populations evenly shared derived alleles with the ancient samples, we estimated D-statistics. The test can be used to evaluate if the data are inconsistent with the null hypothesis that the tree (((Lake, Fjord), ancient), outgroup) is correct.

The D-statistic is based on counts of derived alleles shared by the ancient sample and the modern marine sample, but not the modern freshwater sample; and comparing this with counts of derived alleles shared by the ancient sample and the modern freshwater sample, but not the modern marine sample. We find the ancient sample shares a roughly equal number of derived alleles with both the modern marine and freshwater samples. This statistic excluded the known freshwater-marine divergent regions of the genome. The D-statistic is insensitive to drift in allele frequencies and is a test of 'treeness'. Our hypothesis is that the ancient sample represents an individual from the time that the ancestors of the present-day lake and present-day marine samples split. We can express this hypothetical relationship in newick format as (ancient, (lake, fjord)).

No mutations that occur along the branches to the lake or the fjord samples will count toward the D-statistic as these will not be shared with the ancient sample. Under neutral expectations, changes in allele frequencies should be random and the standing variation present in the ancient sample should be shared equally, when sampling a random allele from each SNP. We find this to be the case and the statistic is confirmation of the tree-like relationship above. We represent this tree in newick format as (outgroup, (ancient, (fjord, lake))).

If our hypothesized demographic history was incorrect, possible alternative topologies would be: (outgroup,(fjord,(ancient, lake))) or (outgroup,(lake,(ancient, fjord))). If one of these topologies were the true demographic history the ancient sample would share an excess of derived alleles with the lake population (left-topology) or the fjord population (right-topology). However, these topologies were rejected ($-3 > Z > 3$).

The definition used here is from Durand, Patterson, Reich, and Slatkin:[66]

$$D = \frac{\mathrm{nABBA} - \mathrm{nBABA}}{(\mathrm{nABBA} + \mathrm{nBABA})},$$

where in the tree given above, nABBA is the number of sites where lake and ancient samples share a derived allele, and the fjord samples has the ancestral allele; and nBABA is the number of sites where the fjord and ancient samples share a derived allele, and the lake sample has the ancestral allele. Under the null hypothesis that the given topology is the true topology, we expect an approximately equal proportion of ABBA and BABA sites and thus D = 0. The significance of the deviation from 0 was assessed using Z-scores, which are based on the assumption that the D-statistic (under the null hypothesis) is normally distributed with mean 0 and a standard error achieved using the jackknife procedure. The tests were implemented in ANGSD[37] and performed by sampling a single base at each position of the genome to remove bias caused by differences in sequencing depth at any genomic position, removing

transitions to avoid bias from C→T and A→G DNA damage patterns, and only considering sites covered in the ancient sample (only the Lake 2 sample was included). Further filtering steps were as specified above for PCA and excluding regions known to be associated with marine-freshwater divergence. The Japan Sea stickleback *Gasterosteus nipponicus* (NCBI: PRJDB5176) was used as the outgroup.

Mitochondrial DNA analysis

Both ancient samples were found to be female based on comparable read coverage across the autosomal, pseudo-autosomal and sex determining regions of the X chromosome. Therefore, we reconstructed the mitochondrial DNA phylogeny to test whether the ancient samples were directly matrilineally ancestral to either local marine or freshwater populations (Methods S1, section 6). The alignment of the mitogenomes of the 81 individuals included in the study by Liu et al.[24] was accessed via the dryad repository https://doi.org/10.5061/dryad.46fb1. Modern genomic data from our two lake and fjord study sites were then mapped to the *de novo* reference mitogenome sequence generated by Liu et al.,[24] which excluded the control region due to the high number of indels. Mapping criteria were the same as for the nuclear genome. We then assembled a consensus sequence of the mitochondrial genome of each ancient sample, manually inspecting mapped reads and removing those which had a T within 30 bp of the 5′ end where all other sequences had a C at the same site, and similarly removing those which had an A within 30 bp of the 3′ end where all other sequences had a G at the same site. As per Liu et al.,[24] the mitogenome of *G. wheatlandi* (GenBank: AB445129) was added as an outgroup and a Maximum Likelihood tree was constructed using PHYML 3.0,[44] applying the same settings as Liu et al.,[24] i.e., GTR model for substitution, allowing variable proportions of invariable sites and mutation rates across sites (GTR + I + gamma) and using 100 bootstraps.

The present-day freshwater sticklebacks formed two monophyletic clades corresponding to samples from Lake 1, and those from Lake 2, indicating a recent common maternal ancestor within each lake, and no detectable immigration into either lake (Methods S1, section 6). Mitochondrial genomes generated from the present-day marine samples collected from Altafjord, and the ancient samples also fell within the clade representing the so-called "European Lineage,"[67] but were intermixed among sequences from Denmark, Germany, Greenland and southern Norway. Thus, the ancient samples did not represent the direct mitochondrial ancestor of the local present-day freshwater population within the same lake. However, mitochondrial DNA represents a single genealogy and the ancient is a single sample, and the high haplotype diversity in our Altafjord sample set, and the lack of fine-scale phylogeographic patterns, suggests the ancestral marine founders of each lake population may have had similarly high mtDNA haplotype diversity.

Pairwise sequentially Markovian coalescent

The PSMC model estimates the Time to Most Recent Common Ancestor (TMRCA) of segmental blocks of the genome and uses information from the rates of the coalescent events to infer N_e at a given time, thereby providing a direct estimate of the past demographic changes of a population.[23] We selected the highest coverage modern genomes, building a consensus sequence of each bam file in fastq format sequentially using: first, SAMtools mpileup command with the –C50 option to reduce the effect of reads with excessive mismatches; second, bcftools view –c to call variants; lastly, vcfutils.pl vcf2fq to convert the vcf file of called variants to fastq format with further filtering to remove sites with less than a third or more than double the average depth of coverage and Phred quality scores less than 30. Furthermore, we excluded marine-freshwater divergence associated regions and 100 kb flanking either end. The PSMC inference was then carried out using the recommended input parameters as previously applied to threespine stickleback genomes,[24] i.e., 25 iterations, with maximum TMRCA (Tmax) = 15, number of atomic time intervals (n) = 64 (following the pattern (1*4 + 25*2 + 1*4 + 1*6), and initial theta ratio (r) = 5. Plots were scaled to real time as per,[24] assuming a generation time of 2 years and a neutral autosomal mutation rate of 3.7×10^{-8} substitutions/nucleotide/generation. To check for variance in N_e across the genome, we performed 100 bootstrap replicates, conducted by randomly sampling with replacement 5-Mb sequence segments obtained from the consensus genome sequence. To be assured that the inferred demographic history is not impacted by the sequence coverage, sequencing data of freshwater stickleback BS65 from Feulner et al.[68] was processed identically to our samples and downsampled to different coverages (Methods S1, section 3). The PSMC plot of these samples shows a consistent reduction in N_e for all samples, which suggests that lower coverage can affect the quantitative estimate of N_e, but not the overall pattern. In humans, PSMC has reduced power to estimate N_e for time periods less than 20 KYA from human genomes with an assumed generation time of 25 years, due to too few recombination events in the sequence representing this time interval.[23] We found that stickleback genomes, for which generation time is an order of magnitude lower, resulted in highly consistent PSMC bootstrap plots up to 2 KYA; representing an equivalent number of generations to that found for humans for which PSMC can accurately estimate ancestral N_e. Thus, PSMC has sufficient power to estimate changes in N_e up to, and after the colonization of each of our study lakes and throughout most of the Holocene.

Runs of homozygosity

Runs of homozygous genotypes (ROH) were identified using the window-based approach implemented in PLINK v1.07[45] from an input file of genotype likelihoods generated by ANGSD[37] with the following filtering settings: removing reads of poor mapping quality (MAPQ < 30), removing sites with low base quality scores (q < 20), calling only SNPs inferred with a likelihood ratio test of $p < 0.000001$, discarding reads that did not map uniquely, adjusting q-scores around indels, adjusting minimum quality score to 50 for excessive mismatches, and discarding bad reads (flag > = 256). We estimated ROH from pruned and unpruned data and found minimal qualitative difference with our data. Sliding window size was set to 300 kb, with a minimum of 50 SNPs at a minimum density of 1 SNP per 50 kb required to call a ROH. To account for genotyping errors, we allowed up to 4 heterozygote sites per 300 kb window

within called ROHs, as per Ceballos et al.[69] A length of 1,000 kb between two SNPs was required in order them to be considered in two different ROHs.

F_{ST} statistics

F_{ST} statistics between two populations at a time were performed with vcftools67 from an input file in variant call format (VCF) created by bcftools mpileup and bcftools call.[48,70] The reference stickleback genome gasAcu1[4] and modern samples sequenced at high coverage with five samples from each location were used as input for bcftools mpileup. The output file was subsequently piped into bcftools call which used the multiallelic-caller to output a VCF file for all fifteen modern samples. Thereafter, the resulting VCF file was used for calculating F_{ST} estimate per site between each two populations based on Weir and Cockerham's method[71] by using vcftools. The F_{ST} estimate was processed in R[72] and plotted in 10kb sliding windows with 5kb steps in Methods S1, section 4. For Figure S4, the goal was to visualize variation at a finer scale, we therefore opted to estimate F_{ST} per site. The VCF file was filtered with vcftools for sites with a minor allele frequency greater than or equal to 0.07, at least 80% of the data non- missing, a quality value above 30, and mean depth values as well as genotypes with a depth between 2 and 9 excluding indels. Subsequently, F_{ST} was estimated per site,[73] and plotted in R using Friedman's 'super smoother' with a span of 1/50[74,75]·Methods S1.

Comparison of ancient and modern genotypes

In order to investigate the marine and/or freshwater origin of adaptive alleles carried by the ancient genome we assigned the ancestral state of each allele, conditioning on allele frequencies in either geographically-proximate present-day marine and freshwater populations (Lake2, freshwater; Altafjord, marine; ''locally divergent''), or on parallel-divergent marine and freshwater populations from throughout the Northern Hemisphere species range[4] (''globally divergent''). We estimated genotype probabilities only for the Lake 2 ancient sample for which there was sufficient coverage in known freshwater-marine adaptive regions. To maximize the available data we rescaled base quality scores using mapDamage 2.0,[41] which penalizes the quality score of bases likely to be impacted by post-mortem damage based on the posterior distribution of damage-associated parameters, thereby allowing the inclusion of transitions (with a measure of confidence in the inference drawn from individual bases).

We started by identifying regions of the genome that show signs of elevated divergence among marine and freshwater populations, that therefore contain ancestry-informative variants. First, we focused on regions that were differentiated between the local Lake 2 and Altafjord present- day populations by down-sampling the five highest coverage modern samples from each of Lake 2 and Altafjord to equivalent coverage to the ancient sample. We then identified variants with fixed differences between these two modern populations. 813 genomic windows that contained data from the ancient genome sample and had variants with strong divergence in allele frequency among local geographic populations were identified first by taking 100kb of flanking sequence either side of SNPs with fixed frequency differences and then merging overlapping windows. A similar approach was used to identify 34 windows of global parallel marine-freshwater divergence. Within each window, for each of the locally-divergent and globally-divergent analyses, we used ancestry-informative SNPs (those with allele frequency differences > = 0.8 between the geographically-local or global marine and freshwater populations) to calculate the probability of the ancient genome having either 'freshwater' or 'marine' ancestry, and averaged the value for each window.

Specifically, we defined two possible ancestries for the ancient genome A_{marine} and A_{fresh} Where the allele carried by the ancient genome shares most recent common ancestry with other marine and freshwater fish respectively. The probability of observing an allele (0,1) in the ancient genome given a specific ancestral state can therefore be calculated as:

$$\Pr(allele|Ancestry) :$$

$$P(0|A_{marine}) = f_{marine}(0)$$

$$P(1|A_{marine}) = f_{marine}(1)$$

$$P(0|A_{fresh}) = f_{fresh}(0)$$

$$P(1|A_{fresh}) = f_{fresh}(1)$$

We assume uniform prior probabilities $P(A_{marine})$ and $P(A_{fresh})$, and calculate the posterior probability of each possible ancestral state as:

$$P(Ancestry|allele) = \frac{P(allele|Ancestry)P(A)}{\sum !P(Ancestry_!|allele)}$$

Current Biology, Volume *31*

Supplemental Information

Ancient and modern stickleback genomes reveal the demographic constraints on adaptation

Melanie Kirch, Anders Romundset, M. Thomas P. Gilbert, Felicity C. Jones, and Andrew D. Foote

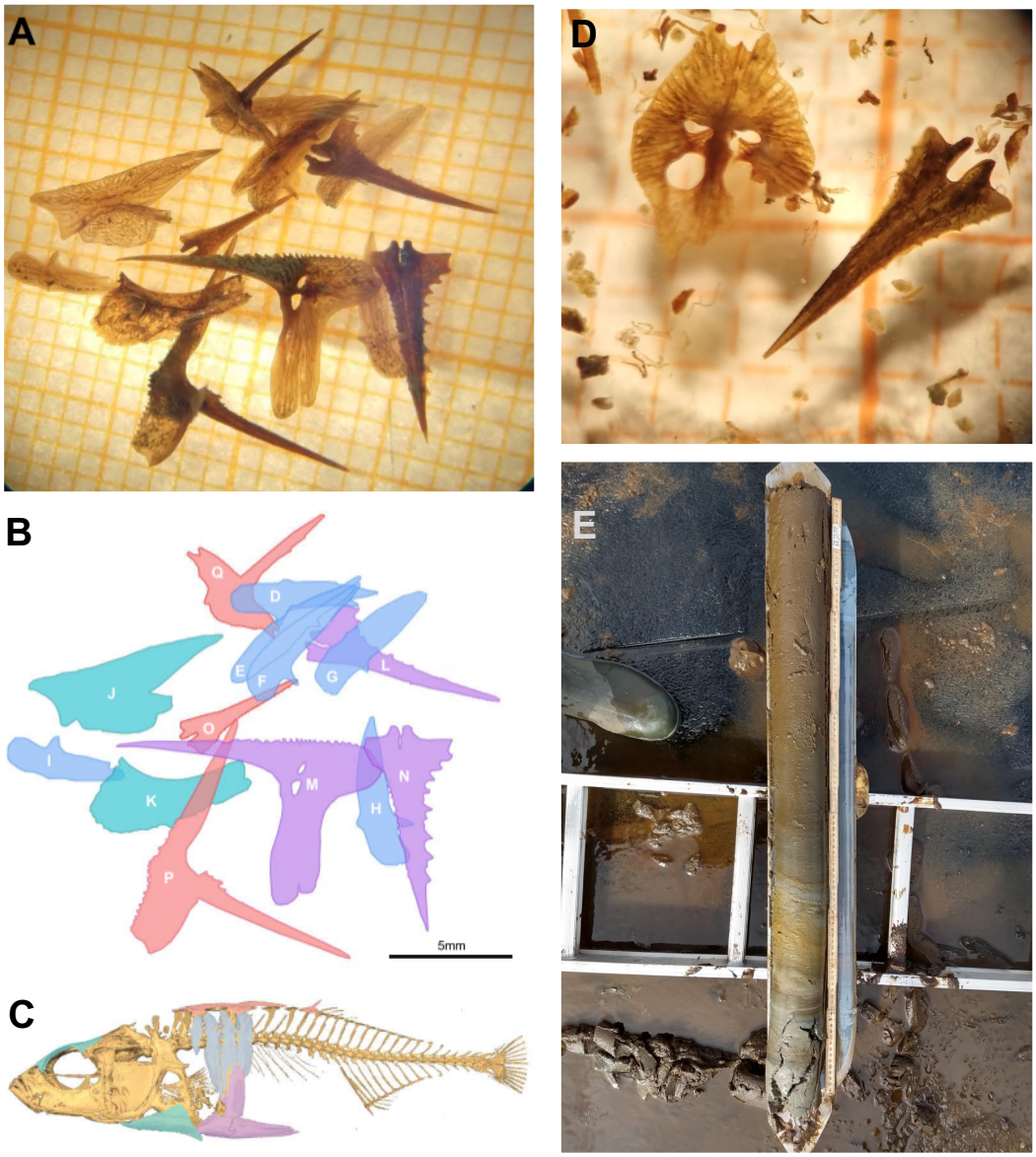


Figure S1. Ancient stickleback bones, plates and spines, Related to Figure 1. (A) Photograph of bones from Lake 1 (Klubbvatnet) near Neverfjord, Finnmark, Norway. Small squares on the backing paper are 2mm^2. (B) Illustration of bone identities from (A). D-I (blue) are lateral plates. J and K (turquoise) are from the head area. J is the plate-like dermal bone between the gills and the pelvis (ectocoracoid). K is the part of the skull plate that is above the eye (frontal). L, M and N (violet) are the pelvic bones: L and N are the pelvic spines, M is the spine-like element between them and part of the lateral protrusion. O, P and Q (red) are dorsal spines; O has lost its basal plate, while P and Q retain theirs. Scale bar in lower right corner is 5mm. (C) Positions of bones from (A & B) illustrated on an X-Ray scan of a modern freshwater fish. (D) Photograph of bones from Lake 2 (Jossavannet) near Hammerfest, Finnmark, Norway. Small squares on the backing paper are 2mm^2. (E) Sediment core from Lake 2. The marine phase of the core is on the right-side of the photograph, which is characterized by marine clay, silt and sand, which transitions to clay gyjttja with laminations during the phase where the lake is partially isolated with occasional marine contact, and then to freshwater gyttja as the lake became fully isolated.

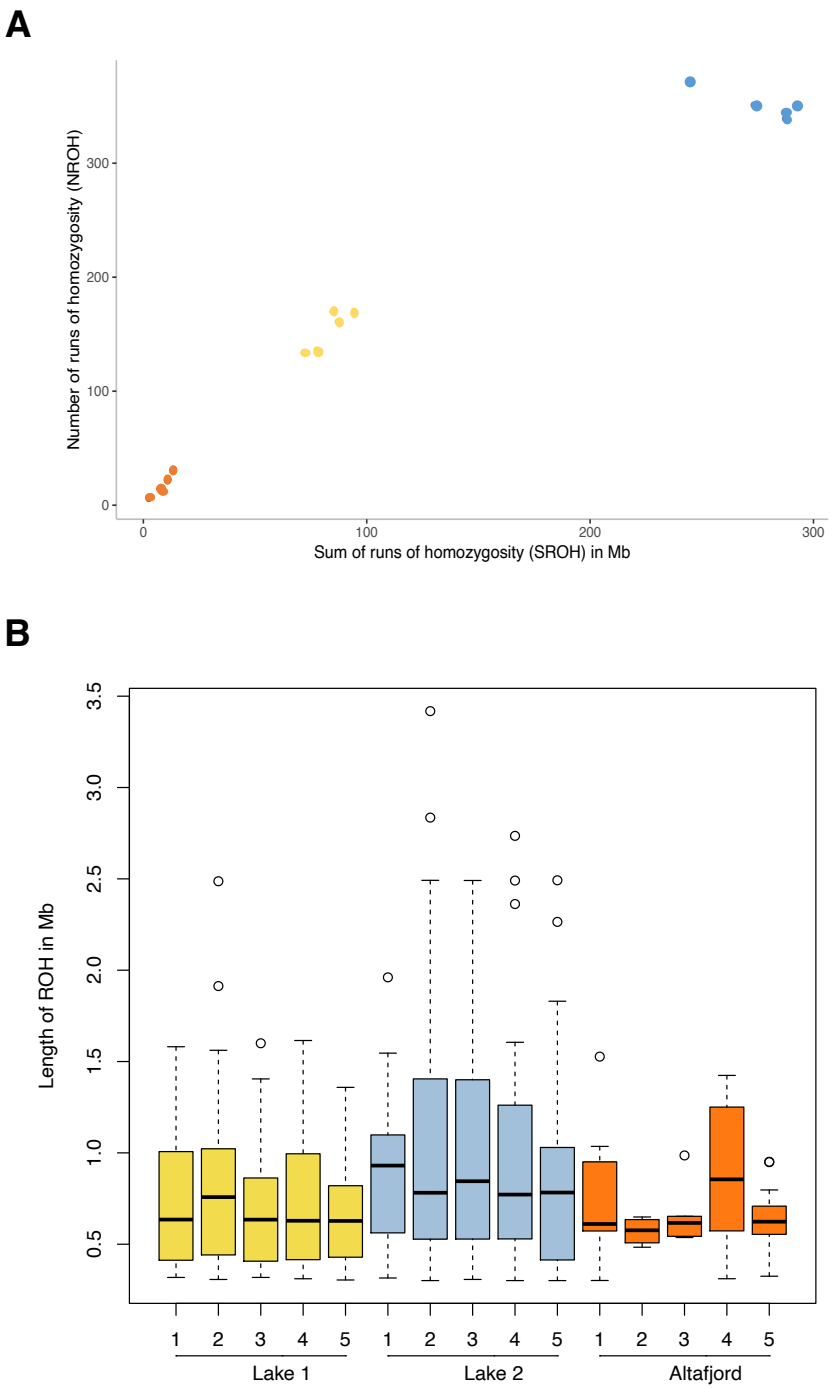


Figure S2. **Runs of homozygosity (ROH) estimates, Related to Figure 3.** (A) The sum and number of ROH for the five highest coverage genomes from Lake 1 (yellow), Lake 2 (blue) and Altafjord (orange). (B) Distribution of the length of ROH greater than 0.3 Mb in the 242 regions of the genome associated with marine-freshwater divergence in five genomes each from Lake 1 (yellow), Lake 2 (blue) and Altafjord (orange). The thick black line shows the median. The bottom and top of the box represent the 1st (Q1) and 3rd (Q3) quartile. The upper whisker corresponds to the smaller value of the maximum length of ROH or the sum of Q3 and 1.5 times the size of the box (Q3-Q1). All values above the upper whisker are shown as black circles. The lower whisker shows the smallest length of ROH for the corresponding individual.

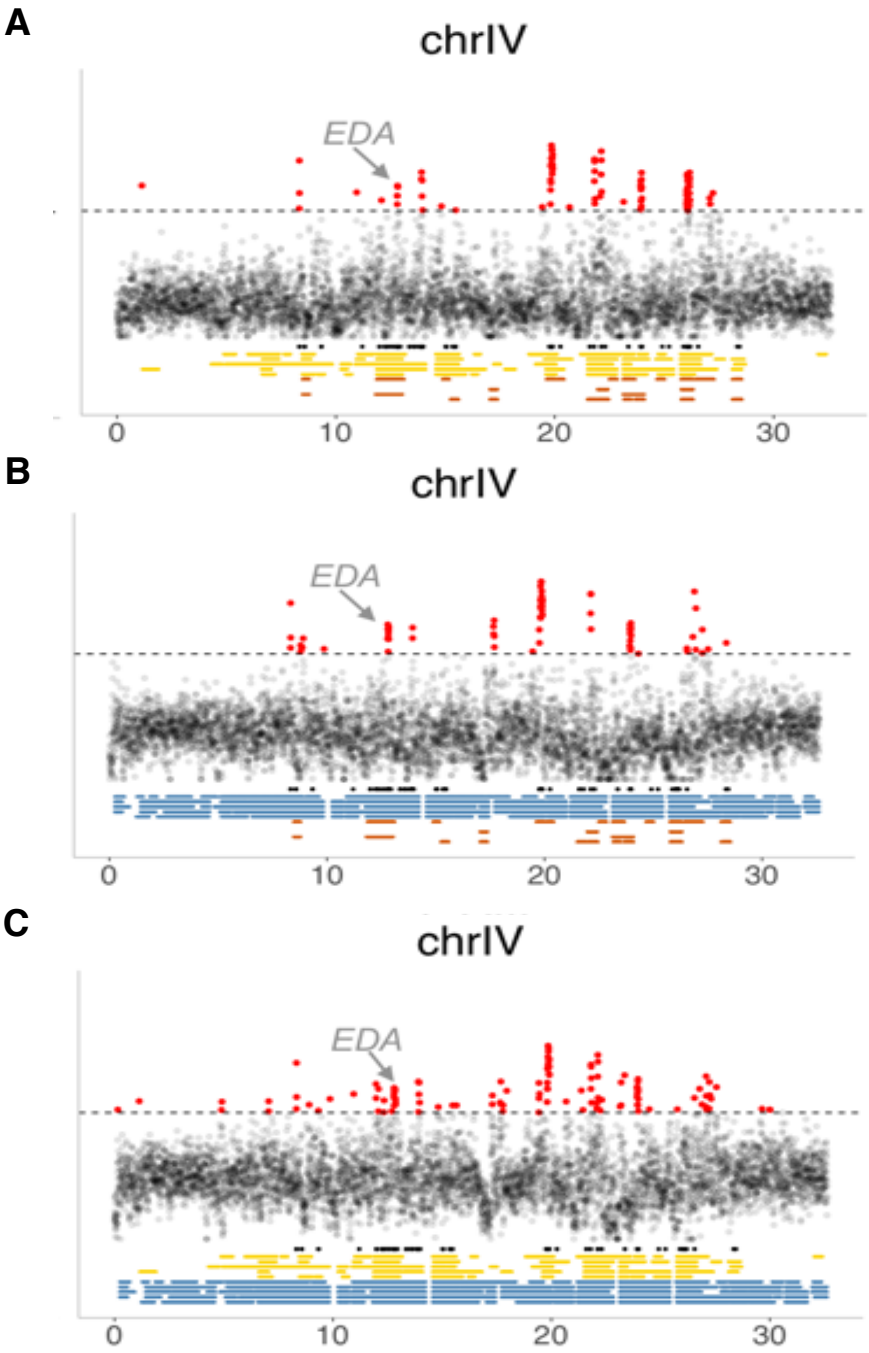


Figure S3. Manhattan plots of F_{ST} estimates for chromosome IV, Related to Figure 3 & 4. between (A) Lake 1 and Altafjord, (B) Lake 2 and Altafjord, and (C) Lake 1 and Lake 2. Manhattan plots of F_{ST} values (*y*-axis) estimated for 10kb sliding windows at sliding intervals of 5 kb. The steps on the x-axis represent Mb along the chromosome. Values above F_{ST} = 0.5 are coloured red and a dotted marker line is inserted at F_{ST} = 0.5. The black bars underneath the Manhattan plot show the location of the divergent regions from Jones *et al.*[S1] (FDR 0.05), whereas yellow, blue and orange bars represent the locations of the ROHs > 0.3 Mb of five individuals from Lake 1, Lake 2 and Altafjord, respectively.

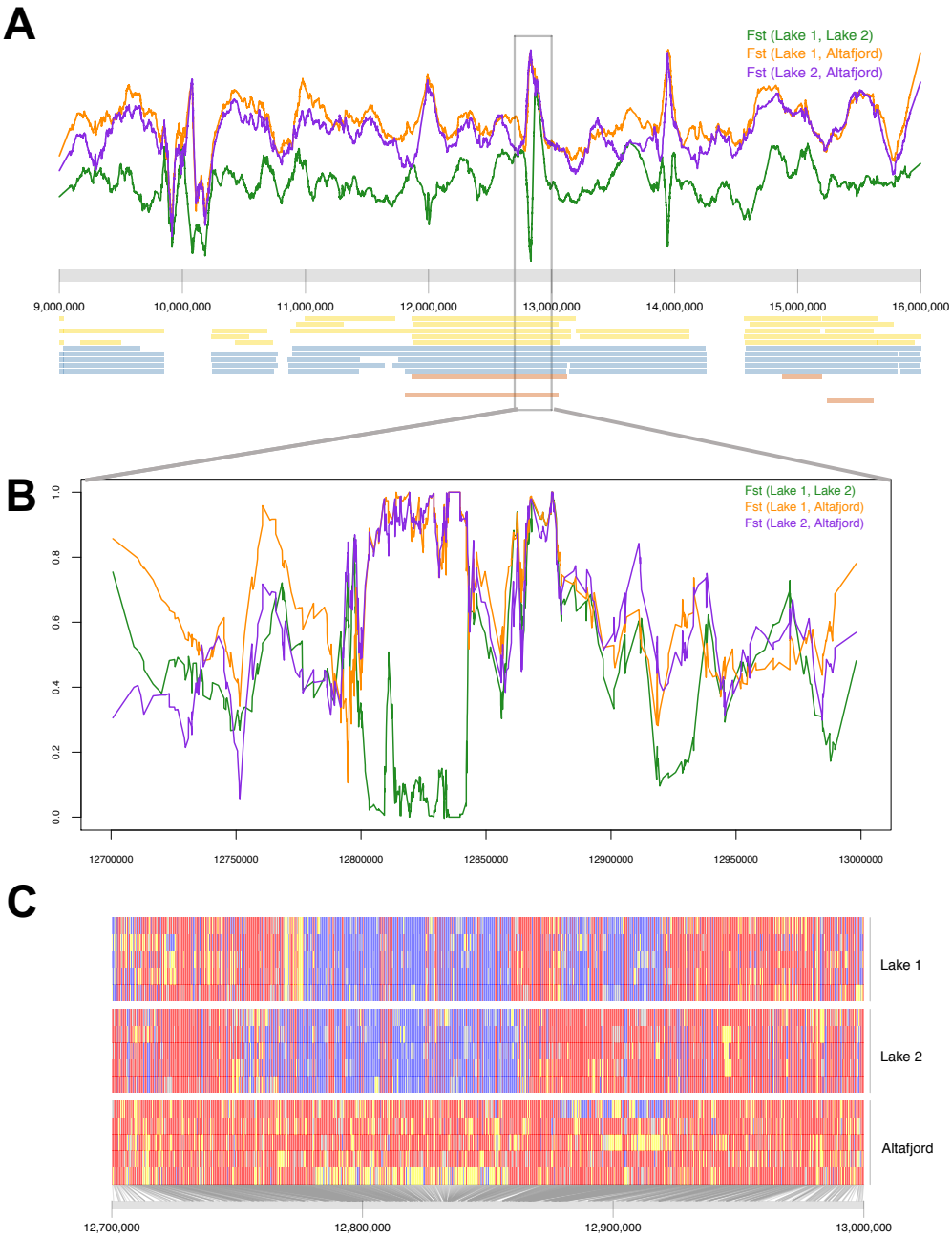


Figure S4. Freshwater populations carry different EDA haplotype, Related to Figure 4. (A) F_{ST} analyses and ROH for Lake 1, Lake 2 and Altafjord on chrIV:9,000,000-16,000,000 around the EDA region. The upper part shows curves for the F_{ST} values between each two populations, whereas the yellow, blue and red bars underneath the plot represent the locations of the ROHs > 0.3 Mb of five individuals from Lake 1, Lake 2 and Altafjord, respectively. (B) F_{ST} analyses for Lake 1, Lake 2 and Altafjord on chrIV:12,700,000-13,000,000 around the focal EDA region. The curves show F_{ST} values between each two populations. (C) Underlying genotypes at 1200 randomly picked single nucleotide polymorphisms within chrIV:12,700,000-13,000,000. Rows represent individual fish; columns represent individual single nucleotide polymorphisms; red boxes indicate alleles most common in the marine population; blue boxes indicate alleles less common in the marine population; grey boxes are missing data.

Sample Name	Location
ABW_19_Fresh	River Tyne, Scotland
ANTL_07_Marine	Antigonish Landing, Nova Scotia, Canada
BDGB_04_Marine	Bodega Bay, CA, USA
BIGL_12_Fresh	Big River Lagoon, CA, USA
BIGR_05_Marine	Big River, CA, USA
FTC_14_Fresh	Fish Trap Creek, WA, USA
GJOG_06_Marine	Gjögur, Iceland
GORT_08_Marine	Gorten Sands, Scotland
HUTU_13_Fresh	Humptulips, WA, USA
JAMA_10_Marine	Japan Marine
JMRP_09_Marine	River Tyne, Scotland
MATA_11_Fresh	Matadero Creek, CA, USA
MUDL_15_Fresh	Mud Lake, AK, USA
Neu_Marine	Neustadt, Germany
NOST_Stream	Stream, Norway
PAXB_20_Fresh	Paxton Lake, B.C. Canada
RABS_01_Marine	Rabbit Slough, AK, USA
SALR_02_MAR	Salmon River, B.C. Canada
SCX_17_Fresh	Schwalle River, Germany
SHEL_18_Fresh	River Shiel, Scotland
Lake 1	Klubbvatnet Lake, Finnmark, Norway
Lake 2	Jossavannet Lake, Finnmark, Norway
Altafjord	Altafjord, Finnmark, Norway

Table S1. The global dataset used in two PCAs and admixture analysis consists of twenty samples from Jones *et al.*[S1] and one sample each from Lake 1, Lake 2 and Altafjord, Related to STAR Methods.

Sample Name	Location	Coverage (sequenced)	Accession
Ancient_01	Lake 1	<1x	SAMN17514712
Ancient_02	Lake 2	<1x	SAMN17514713
Fish_01	Lake 1	2.88	SRR13517313
Fish_02	Lake 1	9.35	SRR13517312
Fish_03	Lake 1	3.03	SRR13517301
Fish_04	Lake 1	11.42	SRR13517290
Fish_05	Lake 1	13.57	SRR13517289
Fish_06	Lake 1	3.31	SRR13517288
Fish_07	Lake 1	4.67	SRR13517287
Fish_08	Lake 1	3.64	SRR13517286
Fish_09	Lake 1	2.91	SRR13517285
Fish_10	Lake 1	2.81	SRR13517284
Fish_11	Lake 1	11.64	SRR13517311
Fish_12	Lake 1	9.67	SRR13517310
Fish_13	Lake 1	2.89	SRR13517309
Fish_14	Lake 1	3.04	SRR13517308
Fish_15	Lake 1	3.54	SRR13517307
Fish_16	Lake 2	7.53	SRR13517306
Fish_17	Lake 2	3.01	SRR13517305
Fish_18	Lake 2	10.56	SRR13517304
Fish_19	Lake 2	2.66	SRR13517303
Fish_20	Lake 2	10.10	SRR13517302
Fish_21	Lake 2	2.16	SRR13517300
Fish_22	Lake 2	2.80	SRR13517299
Fish_23	Lake 2	11.45	SRR13517298
Fish_24	Lake 2	5.97	SRR13517297
Fish_25	Lake 2	16.75	SRR13517296
Fish_26	Altafjord	7.82	SRR13517295
Fish_27	Altafjord	14.70	SRR13517294
Fish_28	Altafjord	11.56	SRR13517293
Fish_29	Altafjord	12.05	SRR13517292
Fish_30	Altafjord	9.02	SRR13517291

Table S2. Mean depth of coverage per genome for each of the local modern stickleback samples sequenced for this study referred to as 'local' population dataset, Related to STAR Methods.

Supplemental References

S1. Jones, F.C., Grabherr, M.G., Chan, Y.F., Russell, P., Mauceli, E., Johnson, J., Swofford, R., Pirun, M., Zody, M.C., White, S., *et al.* (2012). The genomic basis of adaptive evolution in threespine sticklebacks. Nature *484*, 55–61.

S2. Feulner, P.G.D., Chain, F.J.J., Panchal, M., Huang, Y., Eizaguirre, C., Kalbe, M., Lenz, T.L., Samonte, I.E., Stoll, M., Bornberg-Bauer, E., *et al.* (2015). Genomics of Divergence along a Continuum of Parapatric Population Differentiation. PLoS Genet *11*, e1004966.

S3. Haller, B.C., and Messer, P.W. (2019). SLiM 3: Forward genetic simulations beyond the Wright–Fisher model. Molecular Biology and Evolution *36*, 632–637.

S4. Galloway, J., Cresko, W.A., and Ralph, P. (2019). A few stickleback suffice for the transport of alleles to new lakes. G3: Genes, Genomes, Genetics *10*, 505–514.

S5. Liu, S., Hansen, M.M., and Jacobsen, M.W. (2016). Region-wide and ecotype-specific differences in demographic histories of threespine stickleback populations, estimated from whole genome sequences. Mol. Ecol. *25*, 5187–5202.

EVOLUTIONARY BIOLOGY

Predicting future from past: The genomic basis of recurrent and rapid stickleback evolution

Garrett A. Roberts Kingman[1], Deven N. Vyas[2], Felicity C. Jones[3], Shannon D. Brady[1], Heidi I. Chen[1], Kerry Reid[2], Mark Milhaven[2,4], Thomas S. Bertino[2], Windsor E. Aguirre[5], David C. Heins[6], Frank A. von Hippel[7], Peter J. Park[8], Melanie Kirch[3], Devin M. Absher[9], Richard M. Myers[9], Federica Di Palma[10], Michael A. Bell[11]*, David M. Kingsley[1,12]*, Krishna R. Veeramah[2]*

Similar forms often evolve repeatedly in nature, raising long-standing questions about the underlying mechanisms. Here, we use repeated evolution in stickleback to identify a large set of genomic loci that change recurrently during colonization of freshwater habitats by marine fish. The same loci used repeatedly in extant populations also show rapid allele frequency changes when new freshwater populations are experimentally established from marine ancestors. Marked genotypic and phenotypic changes arise within 5 years, facilitated by standing genetic variation and linkage between adaptive regions. Both the speed and location of changes can be predicted using empirical observations of recurrence in natural populations or fundamental genomic features like allelic age, recombination rates, density of divergent loci, and overlap with mapped traits. A composite model trained on these stickleback features can also predict the location of key evolutionary loci in Darwin's finches, suggesting that similar features are important for evolution across diverse taxa.

INTRODUCTION

Can evolutionary outcomes be predicted? Biologists have long been fascinated with this question, including Darwin and Wallace's anticipation of the existence of Morgan's sphinx moth based on orchid morphology (*1*, *2*), Vavilov's prediction of the types of morphological variants likely to occur in plants (*3*), and Gould's gedankenexperiment about replaying the tape of life (*4*). Natural examples of recurrent evolution provide a particularly favorable opportunity to study the mechanisms that influence evolutionary predictability, including molecular patterns (*5*, *6*).

Although the predictability of evolution may appear to be in conflict with the unpredictability of historical contingency, understanding the past can yield important insights into future evolution. For example, vertebrate populations frequently harbor large reservoirs of standing genetic variation (SGV) (*7*) that give independent populations access to similar raw genetic material to respond to environmental challenges, as observed in diverse species including songbirds, cichlid fishes, and the threespine stickleback (*Gasterosteus aculeatus*) (*8*–*11*). SGV is often apparent in divergent species or populations where it is pretested by natural selection and then distributed by hybridization to related populations. Thus filtered and capable of leaping up fitness landscapes, SGV can also drive rapid evolution (*12*), helping address a very real practical challenge to testing evolutionary predictions: time.

Longitudinal studies of evolving populations have been used to estimate the tempo and strength of selection on a variety of traits in different species (*13*–*18*). Rapid phenotypic evolution over contemporary time scales has enabled hypothesis testing against detailed observations at every step in the process. There is an increasing and impressive body of research examining the genomic consequences of these phenotypic changes in microbial, invertebrate, and vertebrate systems (*19*–*26*).

Stickleback fish provide an outstanding system for further study of the genomic basis of recurrent evolution. At the end of the last Ice Age, threespine stickleback, including anadromous populations that migrate from the ocean to freshwater environments to breed, colonized and adapted to countless newly exposed freshwater environments created in the wake of retreating glaciers around the northern hemisphere (*27*, *28*). This massively parallel adaptive radiation was facilitated by natural selection acting on extensive ancient SGV (*8*, *11*). Under the "transporter" hypothesis, these variants are maintained at low frequencies in the marine populations by low levels of gene flow from freshwater populations (*29*). Reuse of ancient standing variants has enabled identification of genomewide sets of loci that are repeatedly differentiated among long-established stickleback populations (*8*, *30*–*35*). In addition, SGV enables new freshwater stickleback populations to evolve markedly within decades (*17*, *36*–*38*), including conspicuous phenotypic changes in armor plates (*17*) and body shape (*39*).

The rapidity of stickleback evolution has made it possible to begin characterizing genomic and allele frequency changes seen in very young or newly established populations under intense directional selection on multiple traits (*18*, *36*–*38*, *40*–*43*). Here, we identify key molecular features that underlie repeated and rapid evolution of freshwater stickleback by comparing genomes from diverse extant populations with the earliest generation-by-generation changes in a detailed genomic time series from three newly founded populations. We identify several basic genomic and genetic features that

[1]Department of Developmental Biology, Stanford University School of Medicine, Stanford, CA 94305-5329, USA. [2]Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY 11794-5245, USA. [3]Friedrich Miescher Laboratory of the Max Planck Society, Max-Planck-Ring, Tübingen, Germany. [4]School of Life Sciences, Arizona State University, Tempe, AZ 85281, USA. [5]Department of Biological Sciences, DePaul University, Chicago, IL 60614-3207, USA. [6]Department of Ecology and Evolutionary Biology, Tulane University, New Orleans, LA 70118, USA. [7]Department of Community, Environment and Policy, Mel & Enid Zuckerman College of Public Health, University of Arizona, Tucson, AZ 85724, USA. [8]Department of Biology, Farmingdale State College, Farmingdale, NY 11735-1021, USA. [9]HudsonAlpha Institute for Biotechnology, 601 Genome Way, Huntsville, AL 35806, USA. [10]Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, MA 02142, USA. [11]University of California Museum of Paleontology, University of California, Berkeley, Berkeley, CA 94720, USA. [12]Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA. *Corresponding author. Email: krishna.veeramah@stonybrook.edu (K.R.V.); sticklemack@gmail.com (M.A.B.); kingsley@stanford.edu (D.M.K.)

can be used to predict evolutionary outcomes in stickleback and show that they can predict genomic responses to selection in distantly related cichlids and Darwin's finches.

RESULTS

Global resequencing and EcoPeak identification

Previous whole-genome sequencing (WGS) of threespine stickleback identified 174 loci covering 1.2 Mb with alleles shared by common descent repeatedly selected in freshwater populations around the world (*8*). Just as human genetic diversity is greatest in Africa, where *Homo sapiens* arose (*44*), we hypothesized that the north Pacific region where stickleback originated (*27*) may contain a particularly rich pool of ancient adaptive alleles. To test this hypothesis, we generated whole-genome sequence data with 76–base pair (bp) paired-end Illumina reads for 38 new marine and 110 new freshwater stickleback, respectively (mean coverage of 5.5×) (sections S2, S4, S6, and S7). Combined with previous stickleback sequencing (*8*, *41*), our dataset includes 227 individual genomes: 135 genomes from 70 northeast Pacific populations in Alaska, Haida Gwaii, British Columbia, and Washington and 92 genomes from 62 populations in California, Japan, and the Atlantic coasts of North America, Iceland, and northern Europe (Fig. 1A and section S8).

We used two methods to identify loci repeatedly differentiated in freshwater populations, both based on the expectation that variants recurrently selected from SGV will be more similar among geographically separated freshwater populations than neutral loci (section S9). First, we used a genetic distance–based approach within overlapping 2500-bp windows tiled across the genome [as in the study by Jones *et al.* (*8*)]. While statistically powerful, this approach may miss younger loci with few differences between alleles and exhibits spatial resolution dependent on window size. Second, we analyzed the distribution of variants at individual bases across the genome, which has base pair–level resolution and less bias against younger loci, though at the cost of statistical power. After calling *P* value–based peaks of ecotypic (freshwater- or marine-associated) differentiation using both methods, we accepted calls at two stringency levels, either requiring agreement between the two analyses at 1% false discovery rate (FDR) (specific) or support from either at 5% FDR (sensitive). We refer to these peaks of ecotypic differentiation as EcoPeaks. We called EcoPeaks for different geographic sets of samples to find alleles that were either shared globally, within the northeast Pacific, or within other geographic regions.

Although results of the global analysis largely matched a previous report [79 of 81 most stringent calls from Jones *et al.* (*8*) in sensitive EcoPeaks ($P = 4.2 \times 10^{-21}$; table S3)], both the sensitive and specific call sets identified approximately five times as many Pacific EcoPeaks as global EcoPeaks, spanning sevenfold more of the genome (Fig. 1, E and F, and Table 1). In addition, many northeast Pacific EcoPeaks not overlapping the globally shared regions identified by Jones *et al.* (*8*) exhibit even more consistent ecotypic differentiation (assessed by *P* values) than others shared around the world (Fig. 1, B and C). Much smaller sets of non-global EcoPeaks were identified in the North Atlantic, subglacial Pacific, and supraglacial geographic regions (fig. S5), consistent with other reports (*8*, *35*).

As theoretical studies indicate that SGV is immediately available for evolution and may show an increased likelihood of large-effect alleles being advantageous compared to de novo mutations (*12*, *45*), the rich genetic reservoir observed in the northeast Pacific provides a favorable system for studying the dynamics and predictability of rapid evolutionary change (section S10). Previous studies suggest that stickleback in the northeast Pacific can adapt to freshwater environments within decades (*36*). However, thus far, studies have lacked temporal resolution of genome evolution in the critical early years of adaptation.

Rapid contemporary evolution and TempoPeak identification

To characterize the earliest stages of evolution after the establishment of new freshwater populations, we analyzed annual samples from populations that were recently founded by anadromous stickleback in three lakes in Alaska (Fig. 2A and section S1). In 1982, stickleback in Loberg Lake (LB) were exterminated to improve recreational fishing (*17*). Sometime between 1983 and 1988, LB was invaded by completely plated (~33 plates per side) anadromous stickleback [most likely from neighboring Rabbit Slough (RS)]. The characteristic freshwater, armor-reduced phenotype increased rapidly from ~16% in 1991 to ~50% by 1995 and to ~95% by 2017 (Fig. 2B) (*17*), with similarly rapid changes in overall body shape (*39*) and reproductive patterns (*46*). So as to more systematically examine even earlier generations of freshwater adaptation, Bell *et al.* (*47*) introduced ~3000 anadromous RS fish into each of two other Cook Inlet lakes without outlets that had been similarly treated to exterminate fish: Cheney Lake (CH) in 2009 and Scout Lake (SC) in 2011. Low–armor-plated (~5 to 7 plates per side) stickleback began to appear in the second and third generation after founding in CH and SC respectively, and, by 2017, they had increased to 20 to 30% (Fig. 2B).

To obtain genomewide allele frequencies across our time series, we performed pooled WGS (pool-seq) on all seven available annual samples from CH and SC since founding and eight from LB distributed between 1999 and 2017 (Fig. 2A and sections S3, S4, S7, and S13). Each freshwater pool-seq experiment consisted of 100 individuals (with three exceptions), with mean coverage of 223× per pool. In addition, we resequenced a pool of 200 anadromous RS individuals used to found the CH population in 2009 (RS2009) to 585×.

We identified single-nucleotide polymorphisms (SNPs) with significant allele frequency changes, indicating directional selection, using a modified Cochran-Mantel-Haenszel (CMH) test optimized for pool-seq data (*48*), followed by an approach analogous to our EcoPeak analysis to define both a permissive "sensitive" and a stringent "specific" set of loci that we term TempoPeaks (sections S16 to S18). Combining all three populations into a single CMH analysis (CH + SC + LB) and using RS2009 as a proxy for the founders of LB, we identified 524 sensitive and 344 specific TempoPeaks. Despite operating over very different time spans, the visual correspondence between the Pacific EcoPeaks in long-established populations and the TempoPeaks in recently established populations is notable, particularly for the specific TempoPeaks, of which 323 of 344 (94%) overlap with the sensitive Pacific EcoPeaks (Fig. 2D and section S18). In contrast, even the most lenient set of global EcoPeaks and regions from Jones *et al.* (*8*) overlap only 96 of 344 (28%) and 47 of 344 (14%) specific TempoPeaks, respectively (tables S9 and S10), emphasizing the importance of understanding the locally available SGV. Even analyzing only CH + SC (thus focusing on <10 years of freshwater adaptation), we identified 271 sensitive and 86 specific TempoPeaks, 73% and 99% of which, respectively, overlap the sensitive Pacific EcoPeaks. This marked congruity strongly suggests that the ancient SGV represented by Pacific EcoPeaks is the primary

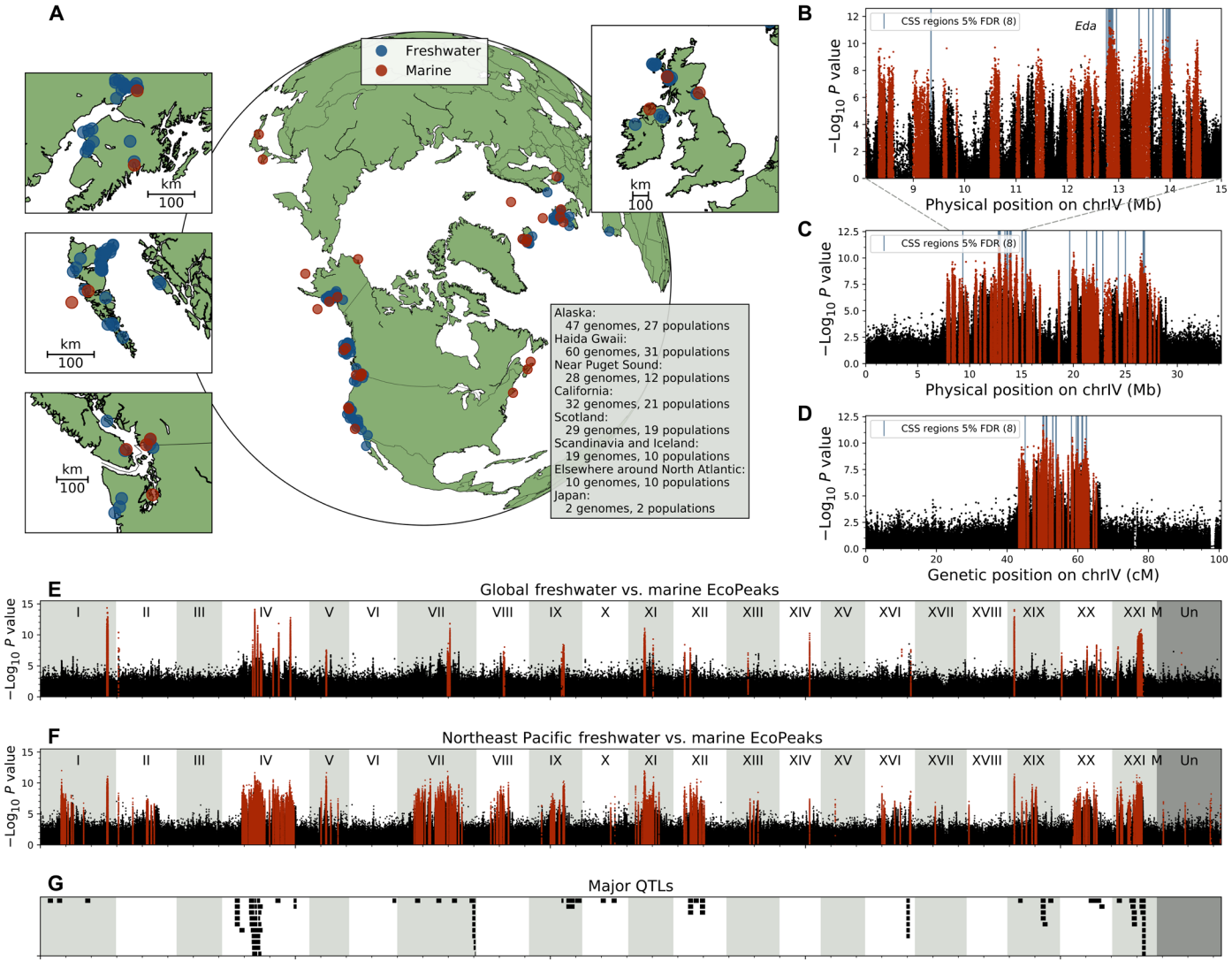


Fig. 1. Recurrent peaks of ecological sequence differentiation between marine and freshwater stickleback from different regions of world. (**A**) Marine (red) and freshwater (blue) stickleback from the locations shown were used for various analyses (table S2). (**B**) Detail of part of chrIV for single-nucleotide polymorphism (SNP)–based analysis of differential allele distribution between marine and freshwater ecotypes in the northeast Pacific basin. SNPs within specific-threshold EcoPeaks are red. A subset of regions overlap the globally shared peaks of marine-freshwater differentiation indicated by blue-colored bars [cluster separation score (CSS), 5% false discovery rate (FDR) identified by Jones *et al*. (*8*)]. (**C**) As in (B), but for the whole chromosome [dashed lines from (B) to (C)]. (**D**) Same whole chromosome as in (C), but with genetic (not physical) distance along the *x* axis. (**E** and **F**) Genomewide SNP divergence between marine and freshwater ecotypes globally and in the northeastern Pacific basin, with specific-threshold EcoPeaks in red. (**G**) Many differentiated regions overlap the location of major quantitative trait loci (QTLs) controlling various morphological, physiological, and behavioral traits in previous genetic crosses [percent variance explained (PVE) > 20, interval < 5 Mb from Peichel and Marques (*53*)].

Table 1. Overview of EcoPeaks and TempoPeaks. The comparisons by Jones *et al*. (*8*) are with the cluster separation score 5% FDR set (*8*).

	Global EcoPeaks (specific)	Pacific EcoPeaks (specific)	Pacific EcoPeaks (sensitive)	Cheney + Scout	All young
No. of regions	39	209	212	86	344
Total bases (%)	3.7 Mb (0.78%)	27.4 Mb (5.82%)	91.9 Mb (19.53%)	3.3 Mb (6.95%)	17.57 Mb (3.73%)
Median size	21.4 kb	80.2 kb	122.9 kb	21.7 kb	27.3 kb
Recovery of regions identified by Jones *et al.* (*8*)	86/174 (49.4%)	112/174 (64.3%)	158/174 (90.8%)	47/174 (27.0%)	98/174 (56.3%)
Fraction in regions identified by Jones *et al.* (*8*)	18/39 (46.2%)	29/209 (13.9%)	33/212 (15.6%)	10/86 (11.6%)	24/344 (7.0%)

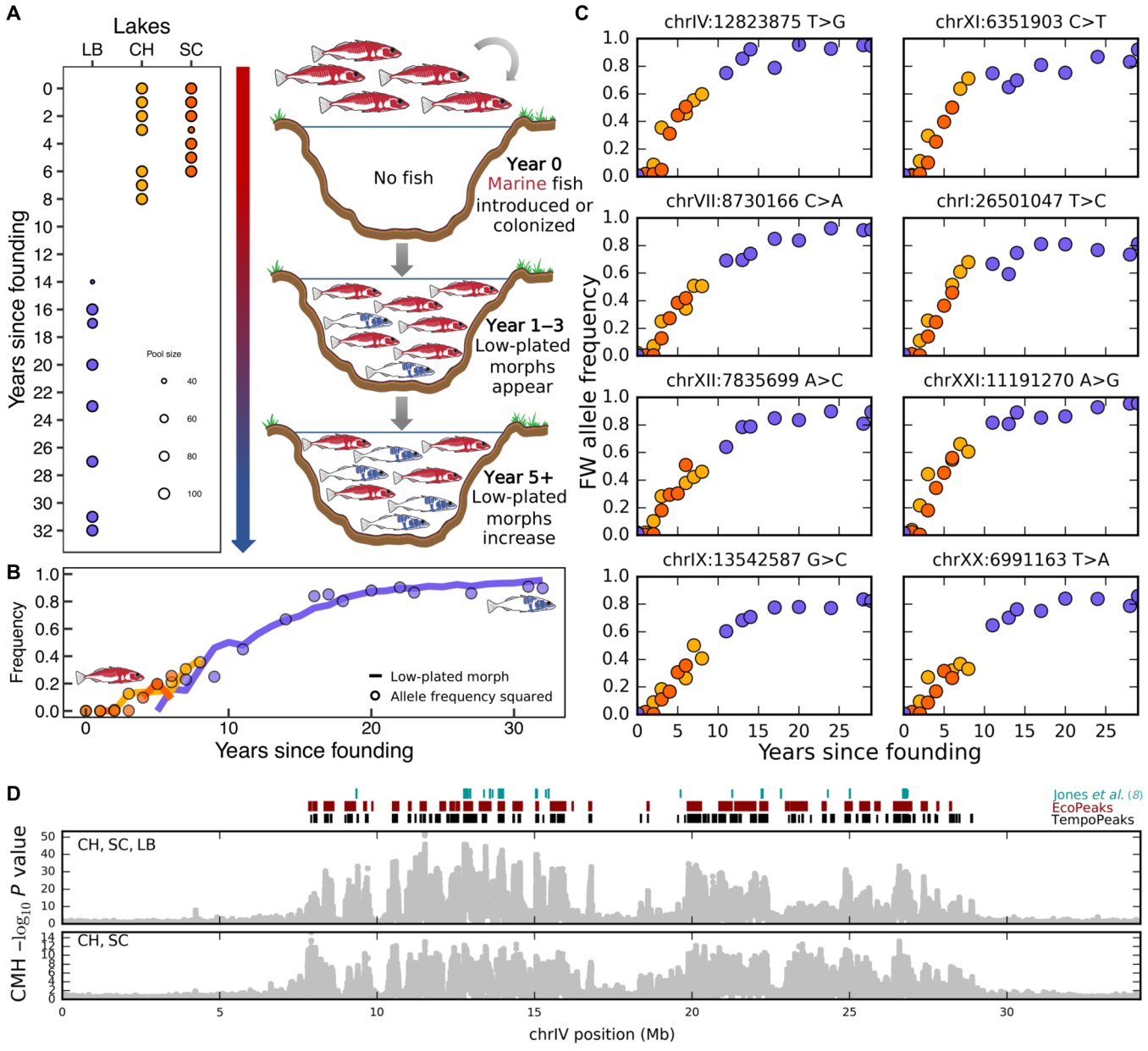


Fig. 2. Contemporary evolution occurring in freshwater transplants in Cook Inlet, Alaska. (**A**) The timing (years since founding) and approximate size of subsequent sequencing sample pools from lake populations [Loberg Lake (LB), Cheney Lake (CH), and Scout Lake (SC)] founded recently by anadromous stickleback (left) and the scenario for divergence of anadromous populations after colonizing the lakes (right). Red and blue fish represent the complete armor-plated and armor-reduced phenotypes, respectively. (**B**) Frequency of armor-reduced morphological phenotype across our CH, SC, and LB time series overlaid with the frequency squared for the freshwater (FW) *Eda* allele. LB data are based on a combination of individual genotypes and pool-seq frequencies, while CH and SC are based only on pool-seq frequencies. (**C**) Allele frequency trajectories for eight SNPs found within TempoPeaks on distinct chromosomes with the highest Cochran-Mantel-Haenszel (CMH) scores (except for chrIV:12823875, the *Eda*-plate regulatory region SNP). (**D**) Genomewide distribution of window-based CMH scores across chrIV for different combinations of transplant lakes discussed in the main text. Black, dark red, and teal bars above figure represent specific CH + SC + LB TempoPeaks, northeast Pacific EcoPeaks, and significant loci from Jones *et al.* (*8*) identified using CSS [5% FDR (*8*)], respectively.

genomic feature enabling extremely fast evolution of freshwater phenotypes in stickleback from the northeast Pacific basin.

The *Eda* SNP associated with armor plate variability (chrIV: 12,823,875 T>G (*49*)) is within the second most significant specific TempoPeak on chrIV. In both CH and SC, the G allele increases rapidly from an initial frequency of <1% to over 50% within 8 years, while approaching fixation in LB by 15 years. Notably, the square of G-allele frequencies (i.e., the expected number of GG homozygotes) tracks closely with frequencies of the low–armor plate phenotype, consistent with almost complete recessiveness (h = 0.0) for the G allele for this phenotype (Fig. 2B). Nonetheless, to fit the allele frequency trajectory of this SNP, and, in particular, the extremely rapid

increase in CH and SC, it was necessary to impose a dominance coefficient (h) of 1.0 along with a very large selection coefficient (s) of 0.55, as in a recent paper focusing on this locus (*18*).

Like *Eda*, most TempoPeaks display similarly sharp left-shifted sigmoidal allele frequency trajectories, indicating very strong and dominant-positive selection (Fig. 2C and section S20). When modeling each peak SNP as independent, we find an extremely high mean s of 0.30 (5th, 95th percentile 0.08 to 0.53) and h of 0.98 (5th, 95th percentile 0.95 to 1.0) for the 344 specific TempoPeaks found in CH + SC + LB. The estimated s values for chrIV, where there are 69 TempoPeaks, are particularly high (mean s = 0.38), consistent with the accelerated evolution of this whole chromosome observed via a chromosome-wide F_{ST} analysis comparing the founding generation of CH, SC, and LB to all subsequent years (section S15).

Features associated with EcoPeak evolution

The remarkable speed at which northeast Pacific stickleback adapt to new freshwater environments suggests that analysis of EcoPeaks may provide unique insights into optimal genomic properties for evolution. Using *Gasterosteus nipponicus*, *Gasterosteus wheatlandi*, and *Pungitius pungitius* for calibration, we estimated molecular divergence time between a pair of freshwater (Little Campbell upstream) and marine (Little Campbell downstream) stickleback in windows tiled across the genome (section S11). We find that EcoPeaks as a whole are significantly older than the rest of the genome [1600 thousand years (ka) versus 700 ka, $P < 5 \times 10^{-324}$]. Although peaks shared globally trend older than those found just within the northeast Pacific (1800 ka versus 1600 ka, $P = 0.18$), the imputed ages overlap considerably (Fig. 3A). We estimate that the majority (161 of 209) are over a million years old and have cycled between freshwater and marine environments many times during this long history, likely persisting at high frequency in freshwater habitats south of the zone of glaciation during the Ice Ages and at more northerly latitudes during previous interglacials and the Holocene.

Contrary to our expectations that recombination would disassemble regions over time, we found that older EcoPeaks are larger than younger ones (Fig. 3B). This signature is strongest at the most significant markers within each EcoPeak, which are typically older than more distal sequences (Fig. 3C). This suggests that individual regions may grow over time, with alleles originally based on an initial

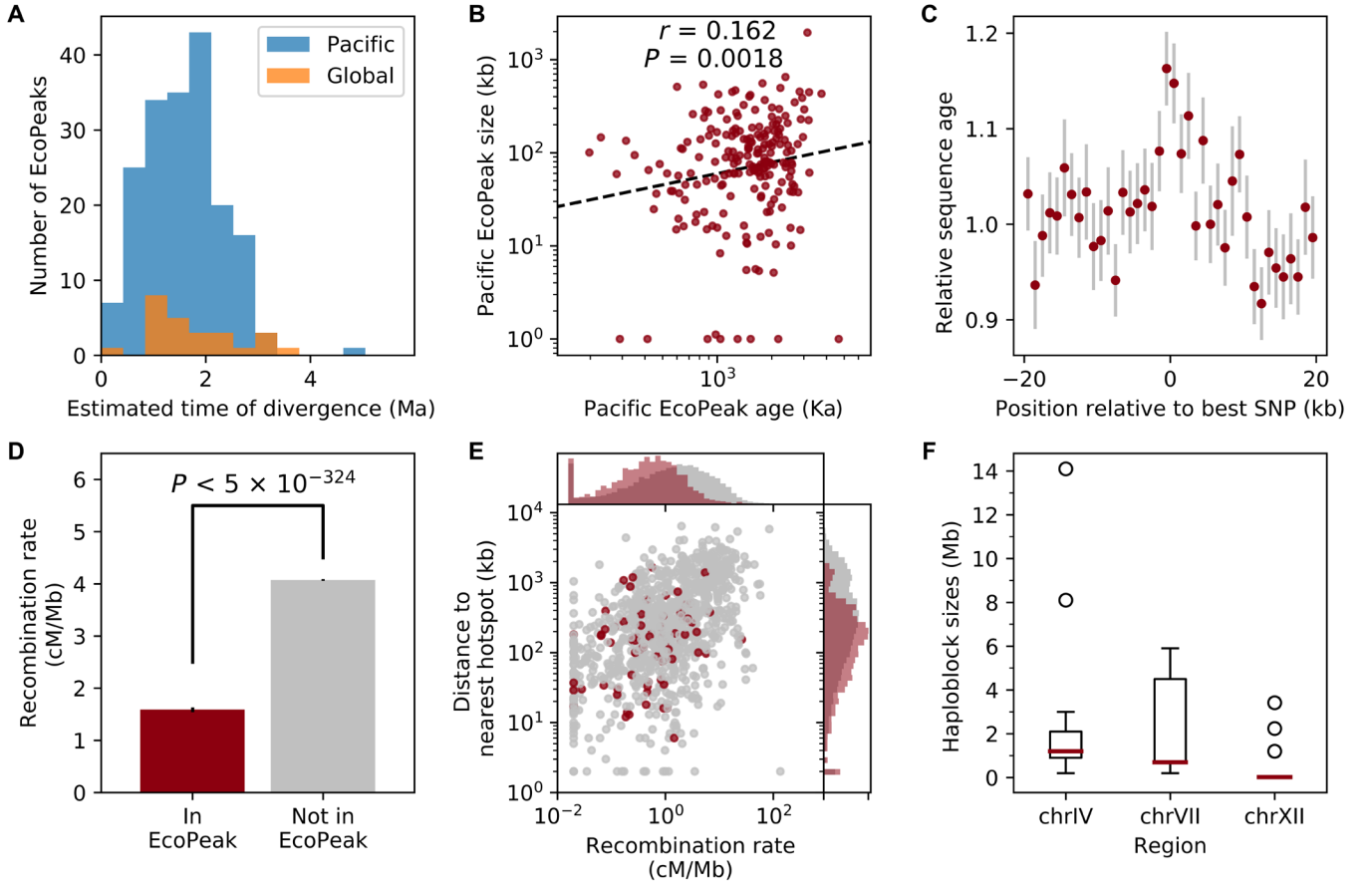


Fig. 3. EcoPeak associations with age, region size, and recombination rate. (**A**) Distribution of estimated molecular age for those EcoPeaks either shared worldwide (orange) or within the northeast Pacific (blue). Ma, million years. (**B**) EcoPeaks with older estimated molecular ages tend to be larger. (**C**) Estimated ages decline with distance on either side of EcoPeaks. Each dot represents mean age in 1-kb windows flanking the EcoPeak centers (gray bars, 1 SE). (**D**) Recombination rates tend to be lower within EcoPeaks compared to the rest of the genome, ±1 SE. (**E**) Recombination rates and distances to nearest 20× recombination hotspots, plotted for randomly subsampled 1-kb windows tiled across the genome, with marginal histograms of all windows. Locations overlapping EcoPeaks (red) are shifted to both smaller hotspot distances and lower recombination rates compared to other genomic regions (gray). (**F**) Observed haploblock size in marine fish carrying freshwater EcoPeaks on the indicated chromosomes across three marine populations. For all, specific northeast Pacific EcoPeaks are used.

beneficial mutation accumulating additional linked favorable mutations, snowballing over time to form a finely tuned haplotype with multiple adaptive changes. This is consistent with work in other species identifying examples of evolution through multiple linked mutations that together modify function of a gene (*50*–*52*) and implies that progressive allelic improvement may be common.

We also observed that EcoPeaks frequently overlap major quantitative trait loci (QTLs) in stickleback [73 of 209 overlaps observed versus 32 of 209 expected, $P < 1 \times 10^{-15}$; Fig. 1G (*53*)], suggesting that these variants underlie many mapped phenotypic traits. Just as the QTLs cluster in "supergene" complexes (*54*), so too do EcoPeaks (median observed interpeak distance 192 kb versus 795 kb expected, $P = 4.88 \times 10^{-10}$). One particularly large complex (chrIV: 8 to 17 Mb) contains 22 EcoPeaks and the major QTLs controlling many aspects of both defensive armor and trophic morphology (e.g., the length of dorsal and pelvic spines, the number of armor plates through *Eda*, gill rakers, and teeth). Thus, clustering may have important functional effects by allowing multiple traits and underlying EcoPeaks to be selected and inherited as a single unit, especially when in tight linkage. A fine-scale recombination map of RS stickleback (generated with LDhelmet (*55*)) shows that EcoPeaks are highly enriched in regions of low average recombination, forming tightly linked haploblocks (Fig. 3D, compare Fig. 1, C and D; section S14). EcoPeaks are also enriched near local recombination hotspots within their neighborhood (Fig. 3E), potentially facilitating reassembly of larger haplotype blocks upon freshwater colonization (also see section S19).

To further examine the frequency and size of haploblocks in individual fish, we surveyed 1643 stickleback from three Alaskan marine populations by SNP array genotyping (sections S5 and S12). While most marine fish heterozygous for freshwater alleles carry a relatively small haploblock, some carry multi-megabase haploblocks containing multiple EcoPeaks (Fig. 3F). Thus, a proper treatment of rapid stickleback evolution needs to account for the complex linkage of EcoPeaks rather than treating them independently.

Modeling the genomic landscape of contemporary evolution

To estimate a more realistic distribution of fitness effects (DFE) that incorporates the genome's recombination landscape, we developed a deep neural network (DNN) approach that uses forward simulations (section S21). Our simulations, which are conceptually similar to those of Galloway *et al.* (*56*), attempted to replicate the dynamics of the "transporter model" (*29*), with one large (N_e = 10,000) anadromous population connected independently by gene flow to 10 smaller (N_e = 1000) established freshwater populations. After 1000 generations, we founded three new freshwater populations from the anadromous population, thus generating simulated allele frequency trajectories that reflect our annual LB, CH, and SC samples (Fig. 4A).

Focusing our DNN analysis on a subset of 19 specific TempoPeak SNPs separated by ≥0.4 cM (~100 kb) along chrIV, we closely replicated observed allele trajectories of positively selected freshwater alleles across all SNPs simultaneously using a beta distribution–shaped DFE, for which the mean *s* across the 19 TempoPeaks was 0.063 and the standard deviation was 0.030, with reciprocal fitness costs implemented in the marine population (Fig. 4C). The estimated *s* from our DNN was thus substantially smaller than the mean of 0.48 when each SNP was considered independently. In addition, 18 of 19 SNPs were predicted to be fully dominant and none fully recessive under the best model.

We validated our best-fit DNN model by simulating the 19 selected TempoPeaks SNPs with the estimated DFE along with ~400k neutral SNPs distributed randomly along chrIV. Despite the neutral SNPs not being used in training the DNN, we were able to mimic the overall topology of the CMH scores across the entire genome, suggesting that our model was capturing the overall genomic architecture of freshwater adaptation (Fig. 4D). Our best-fit DNN model also appeared to recapitulate much of the haplotype structure of the array data from individuals from RS, LB1999, and LB2013 (Fig. 4B). Notably, the transition to freshwater alleles appears to be somewhat slower on the right half of chrIV, where there are fewer EcoPeaks, TempoPeaks, and QTLs, and this difference was observable in both the empirical and simulated data.

Overall, our model suggests that extremely rapid and replicable allele frequency increases on chrIV in LB, CH, and SC are mostly driven by multiple linked (primarily) dominant alleles, each with relatively smaller *s* values that act in concert, with recombination hotspots between them (section S19) allowing rapid reassembly of optimum freshwater haplotypes, consistent with the transporter hypothesis. The lower individual *s* values may allow these dominant alleles to persist in the marine environment at low frequency after being disassembled by recombination, especially if some act in epistasis.

Biological features with predictive power

Given the genomewide dynamism of the earliest stages of freshwater adaptation, we attempted to identify genomic features that predict the speed of evolution at TempoPeaks and understand why some peaks are consistently selected more rapidly than others (section S22). We used CMH scores as a proxy of evolutionary speed for each TempoPeak in CH + SC + LB and regressed these against a variety of sequence features.

The best predictor for the speed of evolution is the degree of ecotypic differentiation between marine and long-established freshwater populations (Pacific EcoPeak *P* value), with variants more commonly differentiated in the northeast Pacific being selected more quickly (Fig. 5A and fig. S81). Fisher's geometric model indicates that alleles with large effects are usually disfavored; however, the "prefiltering" of ancient SGV that counters this tendency (*12*) largely benefits alleles that are broadly positively selected, possibly explaining this result.

We also found that larger TempoPeaks are typically selected more rapidly. Similarly, greater TempoPeak density predicts more rapid divergence, suggesting that our simulation accurately reflects how nearby loci mutually reinforce their collective selection. Overlap with major QTLs also has a strong association with rapid evolution, while other variables such as increased sequence divergence, decreased recombination rate, increased gene overlap, increased sequence conservation, increased Ka/Ks, and decreased ancestral marine frequency have smaller contributions to predictive power for speed of selection (Fig. 5A).

We also tested whether underlying sequence characteristics could predict not only the speed of selection in CH + SC + LB but also the location of the selected regions themselves (section S23). Recombination rate, QTL overlap, allelic age, and an integrated genomic context score (section S23) that incorporate the previous features are all useful predictors (Fig. 5B). By combining these fundamental features into a logistic model trained on the survey of extant populations, the most confident predictions of selected regions in the rest of the genome achieve 85% precision. This model performs 67% as well as predictions based only on empirical repeatability in extant populations in the northeast Pacific (Fig. 5B). Thus, our understanding of underlying principles reflects an incomplete yet substantial proportion of evolutionary repeatability.

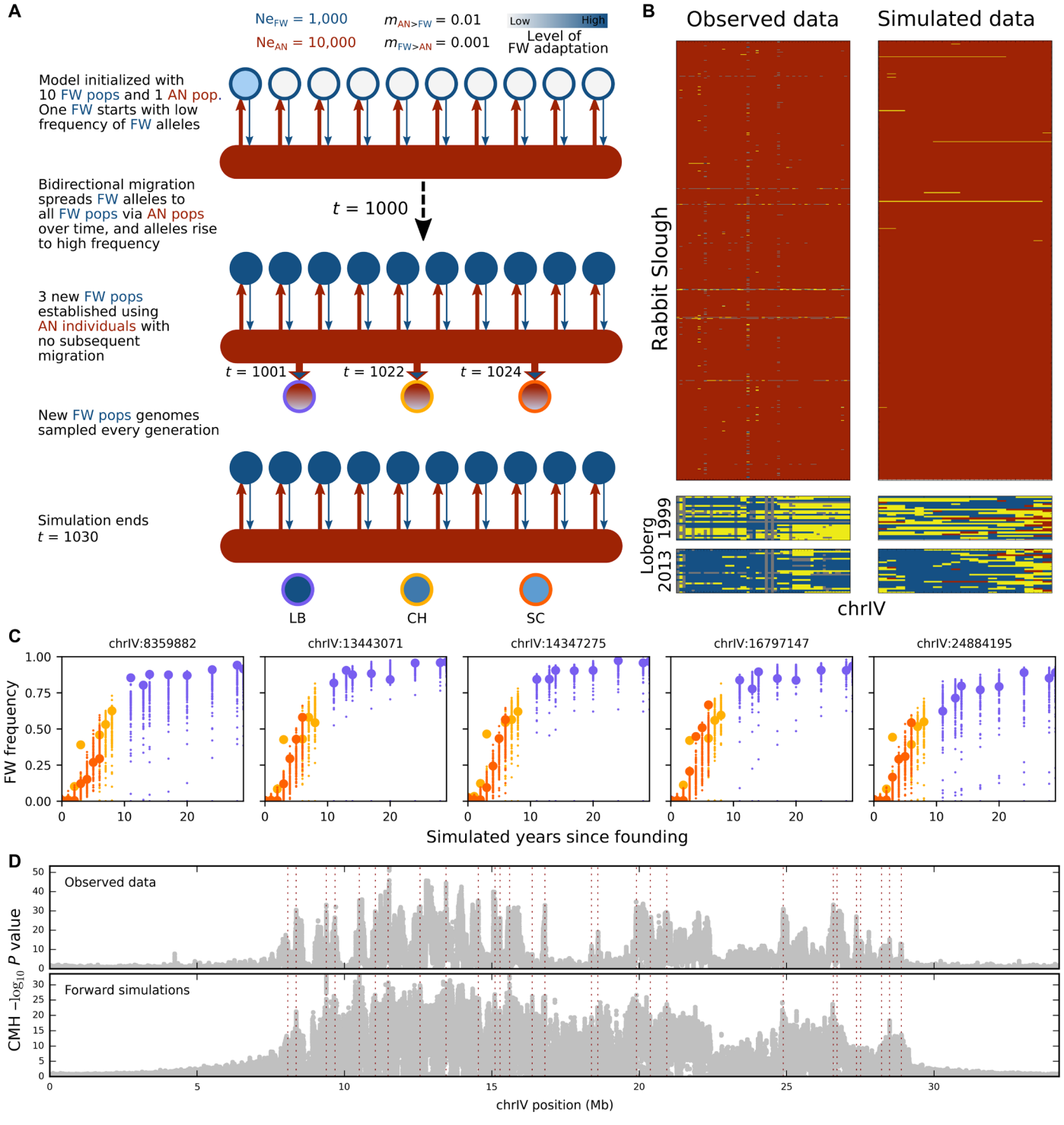


Fig. 4. DNN simulation–based modeling of rapid and repeated stickleback evolution. (**A**) Schematic showing evolutionary model of forward simulations under the transporter hypothesis. Red horizontal bars, anadromous (AN) ancestor; blue circles, descendant freshwater isolates; red to blue shaded circles, three adapting freshwater populations (i.e., LB, CH, and SC) founded recently by anadromous stickleback; and arrows, gene flow or founding events. (**B**) Genotypes across chrIV for freshwater-associated SNPs in RS ($n = 750$), LB in 1999 ($n = 25$), and LB in 2013 for (left) observed and (right) simulated data under best-fit DNN model. anadromous homozygous, red; heterozygous, yellow; and freshwater homozygous genotypes, blue; respectively. (**C**). Allele frequency trajectories for LB, CH, and SC in 100 simulations under the best-fit DNN model for five randomly selected SNPs. Larger points, observed data. (**D**) Distribution of average CMH scores in windows of 2500 bp across chrIV for (top) observed and (bottom) simulated data under best-fit DNN model. Red dotted lines, locations of SNPs under selection and used to fit DNN.

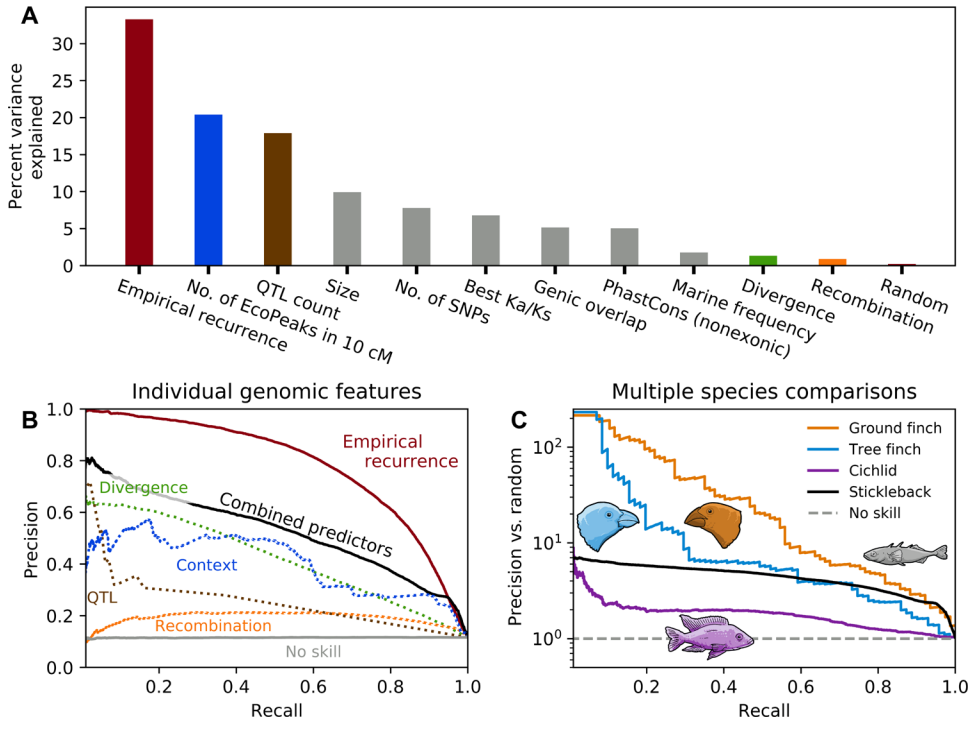


Fig. 5. Properties underlying speed and locus of selection in stickleback, cichlids, and Darwin's finches. (**A**) Variance in the speed of TempoPeak selection explained by different underlying genomic features, including colored bars: empirical recurrence of marine-freshwater differentiation (peak Pacific ecotypic *P* value), number of additional Pacific EcoPeaks within 10 cM, number of major QTLs overlapped, sequence divergence, and recombination rate; gray bars: genomic size of EcoPeak, total number of variable nucleotides, elevated Ka/Ks in coding regions, overlap with genic sequences, overlap with conserved noncoding sequence (PhastCons nonexonic), and carrier frequency of freshwater alleles in marine populations. (**B**) Precision-recall curve for predicting the locations of selected loci in CH + SC + LB lakes by either individual genomic features (dotted lines), a composite model trained with these basic predictors, or the empirical expectation of recurrence based on many extant populations. Precision is the fraction of predictions that are accurate, while recall is the fraction of true positives that are correctly predicted. "No skill" refers to the performance expected by random chance. (**C**) Performance above chance of the composite model applied to stickleback, cichlids, and two representative pairs of species of Darwin's finches (ground finches: *Geospiza magnirostris* versus *Geospiza propinqua*; tree finches: *Camarhynchus pauper* versus *Camarhynchus psittacula*).

Parallels in distant species

To test the generality of these predictive factors, we applied the stickleback-trained model to a dataset of 12 pairs of species of Darwin's finches (section S23) (*57*). Darwin's finches have undergone adaptive radiation in the Galápagos Islands over the last several hundred thousand years, are ~435 million years divergent from stickleback, and face very different selective pressures. As in stickleback, however, the "islands of divergence" of all 12 analyzed pairs of species of Darwin's finches (sensu Han *et al.*) are enriched for ancient alleles overlapping mapped QTLs with low recombination rates. The top 100 windows predicted by the stickleback model recover a median of 28-fold more previously identified islands of divergence than expected by chance ($P < 1 \times 10^{-10}$; Fig. 5C), including the *Alx1* and *Hmga2* loci implicated in beak morphology in multiple species pairs (even without QTL input). The model also recovers a substantial proportion of differentiated loci in a recent case of cichlid speciation (*58*). Thus, a handful of basic genomic properties allow strong quantitative predictions of the location of key evolutionary loci, even across widely separated branches of life.

DISCUSSION

The importance of SGV for evolution is becoming increasingly apparent, especially in species with large genome sizes (*59*), including humans (*60*). At first glance, the dependence of threespine stickleback on SGV for freshwater adaptation may appear to be a peculiarity in terms of repeatability and speed and their particular natural history. However, by more comprehensively understanding the dynamics of this highly optimized process, we have extracted general features of genome architecture and evolution that successfully translate to species on distant branches of the tree of life, thus demonstrating the tremendous power of the stickleback system to identify unifying principles that underlie evolutionary change.

MATERIALS AND METHODS

Sample collection and DNA preparation

Fish for all downstream genomic analyses were trapped following Institutional Animal Care and Use Committee (IACUC) guidelines using unbaited minnow traps set near shore, immediately euthanized with MS-222 (tricaine methanesulfonate), and then preserved in 70 or 95% ethanol (section S1). DNA extraction was performed using either phenol:chloroform isolation (*61*) and quantified using a NanoDrop spectrophotometer or using the DNeasy 96 Blood & Tissue Kit following the standard "animal tissue" protocol and quantified using the Qubit High-sensitivity DNA Assay (section S2). Equimolar pooling of samples from RS, CH, SC, and LB was performed using an Opentrons OT-2 robot (section S3).

Genome sequencing and genotyping

Samples from long-established populations underwent WGS sequencing on a HiSeq 2000 using 2 × 76-bp paired-end sequencing libraries. Contemporary pools were sequenced on either a NovaSeq 6000 or Illumina HiSeq 2500 using 2 × 150-bp paired-end sequencing libraries. Contemporary WGS was performed by Beijing Genomics Institute using their proprietary DNBseq technology using 2 × 100-bp paired-end libraries (section S4). A custom Illumina 384 GoldenGate array was designed for SNP genotyping (section S5).

Bioinformatic processing

We constructed a slightly modified reference genome based on the recent Hi-C–guided improvement of the stickleback genome (*62*), to which we refer as *gasAcu1-4*, that includes a new chrP and a new mitochondrial genome (section S6). Reads were mapped to *gasAcu1-4* using bwa mem (*63*) and Picard was used to add read groups and mark duplicate reads. Indel realignment and base quality recalibration were performed using Genome Analysis Toolkit (GATK) (*64*). HaplotypeCaller and GenotypeGVCFs were used for variant calling for WGS data. Allele frequencies in pool-seq populations were calculated using the maximum-likelihood method of Lynch *et al.* (*65*) and PoPoolation2 (v1201) (section S7) (*66*).

Analysis

EcoPeaks were identified using two approaches, the first following the same genetic distance–based approach as Jones *et al.* (*8*) based on 2500-bp windows sliding every 500 bp, and the second analyzing the distribution of allele counts between marine and freshwater populations at every base position in the genome with two alleles present at >10% frequency in the combined analysis metapopulation. For both the SNP-based and 2500-bp window–based analyses, nearby significant values were grouped into the EcoPeaks that behaved as a single unit using a greedy algorithm. Peaks were filtered at either a 1% FDR for the specific calls or at 5% for the sensitive calls. The single base and window peaks were then intersected for the final specific calls or unioned for the final sensitive calls (section S9). Allelic divergence and age were computed from five upstream (freshwater) and five downstream (marine) fish from Little Campbell River. Nonoverlapping 1-kb windows were tiled across the genome, and variants homozygous for different alleles were counted and used to compute marine-freshwater sequence divergence d (section S11). A rho-based recombination map was constructed on the basis of 20 RS genomes using LDhelmet (*55*) following a similar methodology to that of Shanfelter *et al.* (*67*), though with some minor modifications, and converted to genetic distance based on the pedigree-based linkage map generated by Glazer *et al.* (*68*) (section S14). F_{ST} was calculated for each pool-seq population against its youngest counterpart using the ratio of averages method implemented by Bhatia *et al.* (*69*) (section S18). We used a modified CMH test (*48*) to identify SNPs that had shown a significant change in allele frequency in our contemporary time-series data (section S17). We followed the same general EcoPeak identification methodology to define TempoPeaks from our CMH P values. Sensitive TempoPeaks were based on a 5% Bonferroni-corrected P value threshold merging SNP and window-defined peaks. Specific TempoPeaks were based on a 1% P value threshold only considering window-defined peaks (section S18). We applied the deterministic method described by Taus *et al.* (*70*) to estimate s for the SNP with the largest CMH score in each significant TempoPeak. We estimated s and $p0$ (initial allele frequency), both assuming that the dominance coefficient h is 0.5, as well as simultaneously estimating s, $p0$, and h (section S20). We additionally developed a DNN approach to estimate the DFE of multiple linked TempPeaks on chrIV. This analysis includes three main stages: (i) simulating linked loci with positive selection parameterized with various randomly drawn DFEs within the context of the demographic and evolutionary model of the transporter hypothesis, (ii) using the DNN framework to estimate the best DFE given the observed allele frequency trajectory data from our pool-seq experiments, and (iii) comparing the transporter hypothesis under the best-fit DFE to various features of the observed genomic data (section S21). Genomic features of interest were then used to predict speed of TempoPeak selection in the contemporary evolution populations using a linear regression model (section S22). We also applied genomic features of interest to predict the genomic loci of selection in the contemporary evolution populations via multivariate logistic regression. Last, we applied this model with the same terms and weights to datasets from Darwin's finches (*57*) and Lake Victoria cichlids (*58*) using the previously published analyses as our truth sets (section S23).

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/7/25/eabg5285/DC1

REFERENCES AND NOTES

1. C. Darwin, *Fertilisation of Orchids* (Murray, 1862).
2. A. R. Wallace, Creation by Law. *Q. J. Sci.* **4**, 470–488 (1867).
3. N. I. Vavilov, The law of homologous series in variation. *J. Genet.* **12**, 47–89 (1922).
4. S. J. Gould, *Wonderful Life* (Norton, 1989).
5. B. H. Good, M. J. McDonald, J. E. Barrick, R. E. Lenski, M. M. Desai, The dynamics of molecular evolution over 60,000 generations. *Nature* **551**, 45–50 (2017).
6. D. Blank, L. Wolf, M. Ackermann, O. K. Silander, The predictability of molecular evolution during functional innovation. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 3044–3049 (2014).
7. Z. D. Blount, R. E. Lenski, J. B. Losos, Contingency and determinism in evolution: Replaying life's tape. *Science* **362**, eaam5979 (2018).
8. F. C. Jones, M. G. Grabherr, Y. F. Chan, P. Russell, E. Mauceli, J. Johnson, R. Swofford, M. Pirun, M. C. Zody, S. White, E. Birney, S. Searle, J. Schmutz, J. Grimwood, M. C. Dickson, R. M. Myers, C. T. Miller, B. R. Summers, A. K. Knecht, S. D. Brady, H. Zhang, A. A. Pollen, T. Howes, C. Amemiya; Broad Institute Genome Sequencing Platform & Whole Genome Assembly Team, E. S. Lander, F. D. Palma, K. Lindblad-Toh, D. M. Kingsley, The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**, 55–61 (2012).
9. Y.-T. Lai, C. K. L. Yeung, K. E. Omland, E.-L. Pang, Y. Hao, B.-Y. Liao, H.-F. Cao, B.-W. Zhang, C.-F. Yeh, C.-M. Hung, H.-Y. Hung, M.-Y. Yang, W. Liang, Y.-C. Hsu, C.-T. Yao, L. Dong, K. Lin, S.-H. Li, Standing genetic variation as the predominant source for adaptation of a songbird. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 2152–2157 (2019).
10. Y.-H. E. Loh, E. Bezault, F. M. Muenzel, R. B. Roberts, R. Swofford, M. Barluenga, C. E. Kidd, A. E. Howe, F. Di Palma, K. Lindblad-Toh, J. Hey, O. Seehausen, W. Salzburger, T. D. Kocher, J. T. Streelman, Origins of shared genetic variation in African cichlids. *Mol. Biol. Evol.* **30**, 906–917 (2013).
11. P. F. Colosimo, K. E. Hosemann, S. Balabhadra, G. Villarreal Jr., M. Dickson, J. Grimwood, J. Schmutz, R. Myers, D. Schluter, D. M. Kingsley, Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. *Science* **307**, 1928–1933 (2005).
12. R. D. H. Barrett, D. Schluter, Adaptation from standing genetic variation. *Trends Ecol. Evol.* **23**, 38–44 (2008).
13. J. A. Endler, *Natural Selection in the Wild* (Princeton Univ. Press, 1986).
14. J. G. Kingsolver, H. E. Hoekstra, J. M. Hoekstra, D. Berrigan, S. N. Vignieri, C. E. Hill, A. Hoang, P. Gibert, P. Beerli, The strength of phenotypic selection in natural populations. *Am. Nat.* **157**, 245–261 (2001).
15. D. A. Reznick, H. Bryga, J. A. Endler, Experimentally induced life-history evolution in a natural population. *Nature* **346**, 357–359 (1990).
16. P. R. Grant, B. R. Grant, Unpredictable evolution in a 30-year study of Darwin's finches. *Science* **296**, 707–711 (2002).
17. M. A. Bell, W. E. Aguirre, N. J. Buck, Twelve years of contemporary armor evolution in a threespine stickleback population. *Evolution* **58**, 814–824 (2004).

18. D. Schluter, K. B. Marchinko, M. E. Arnegard, H. Zhang, S. D. Brady, F. C. Jones, M. A. Bell, D. M. Kingsley, Fitness maps to a large-effect locus in introduced stickleback populations. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e1914889118 (2021).
19. A. O. Bergland, E. L. Behrman, K. R. O'Brien, P. S. Schmidt, D. A. Petrov, Genomic evidence of rapid and stable adaptive oscillations over seasonal time scales in *Drosophila*. *PLOS Genet.* **10**, e1004775 (2014).
20. S. F. Levy, J. R. Blundell, S. Venkataram, D. A. Petrov, D. S. Fisher, G. Sherlock, Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature* **519**, 181–186 (2015).
21. P. Nosil, R. Villoutreix, C. F. de Carvalho, T. E. Farkas, V. Soria-Carrasco, J. L. Feder, B. J. Crespi, Z. Gompert, Natural selection and the predictability of evolution in *Timema* stick insects. *Science* **359**, 765–770 (2018).
22. J. T. Anderson, C.-R. Lee, C. A. Rushworth, R. I. Colautti, T. Mitchell-Olds, Genetic trade-offs and conditional neutrality contribute to local adaptation. *Mol. Ecol.* **22**, 699–708 (2013).
23. N. O. Therkildsen, A. P. Wilder, D. O. Conover, S. B. Munch, H. Baumann, S. R. Palumbi, Contrasting genomic shifts underlie parallel phenotypic evolution in response to fishing. *Science* **365**, 487–490 (2019).
24. R. D. H. Barrett, S. Laurent, R. Mallarino, S. P. Pfeifer, C. C. Y. Xu, M. Foll, K. Wakamatsu, J. S. Duke-Cohan, J. D. Jensen, H. E. Hoekstra, Linking a mutation to survival in wild mice. *Science* **363**, 499–504 (2019).
25. T. J. Thurman, R. D. H. Barrett, The genetic consequences of selection in natural populations. *Mol. Ecol.* **25**, 1429–1448 (2016).
26. M. A. Bell, W. E. Aguirre, Contemporary evolution, allelic recycling, & adaptive radiation of the threespine stickleback. *Evol. Ecol. Res.* **15**, 377–411 (2013).
27. M. A. Bell, S.A. Foster, Ed., *The Evolutionary Biology of the Threespine Stickleback* (Oxford Univ. Press, 1994).
28. B. Fang, J. Merilä, M. Matschiner, P. Momigliano, Estimating uncertainty in divergence times among three-spined stickleback clades using the multispecies coalescent. *Mol. Phylogenet. Evol.* **142**, 106646 (2020).
29. D. Schluter, G. L. Conte, Genetics and ecological speciation. *Proc. Natl. Acad. Sci. U.S.A.* **106** (Suppl. 1), 9955–9962 (2009).
30. P. A. Hohenlohe, S. Bassham, P. D. Etter, N. Stiffler, E. A. Johnson, W. A. Cresko, Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLOS Genet.* **6**, e1000862 (2010).
31. F. C. Jones, Y. F. Chan, J. Schmutz, J. Grimwood, S. D. Brady, A. M. Southwick, D. M. Absher, R. M. Myers, T. E. Reimchen, B. E. Deagle, D. Schluter, D. M. Kingsley, A genome-wide SNP genotyping array reveals patterns of global and repeated species-pair divergence in sticklebacks. *Curr. Biol.* **22**, 83–90 (2012).
32. B. E. Deagle, F. C. Jones, D. M. Absher, D. M. Kingsley, T. E. Reimchen, Phylogeography and adaptation genetics of stickleback from the Haida Gwaii archipelago revealed using genome-wide single nucleotide polymorphism genotyping. *Mol. Ecol.* **22**, 1917–1932 (2013).
33. M. Roesti, S. Gavrilets, A. P. Hendry, W. Salzburger, D. Berner, The genomic signature of parallel adaptation from shared genetic variation. *Mol. Ecol.* **23**, 3944–3956 (2014).
34. T. C. Nelson, W. A. Cresko, Ancient genomic variation underlies repeated ecological adaptation in young stickleback populations. *Evol. Lett.* **2**, 9–21 (2018).
35. B. Fang, P. Kemppainen, P. Momigliano, X. Feng, J. Merilä, On the causes of geographically heterogeneous parallel evolution in sticklebacks. *Nat. Ecol. Evol.* **4**, 1105–1115 (2020).
36. E. A. Lescak, S. L. Bassham, J. Catchen, O. Gelmond, M. L. Sherbick, F. A. von Hippel, W. A. Cresko, Evolution of stickleback in 50 years on earthquake-uplifted islands. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E7204–E7212 (2015).
37. N. V. Terekhanova, M. D. Logacheva, A. A. Penin, T. V. Neretina, A. E. Barmintseva, G. A. Bazykin, A. S. Kondrashov, N. S. Mugue, Fast evolution from precast bricks: genomics of young freshwater populations of threespine stickleback *Gasterosteus aculeatus*. *PLOS Genet.* **10**, e1004696 (2014).
38. S. Bassham, J. Catchen, E. Lescak, F. A. von Hippel, W. A. Cresko, Repeated selection of alternatively adapted haplotypes creates sweeping genomic remodeling in stickleback. *Genetics* **209**, 921–939 (2018).
39. W. E. Aguirre, M. A. Bell, Twenty years of body shape evolution in a threespine stickleback population adapting to a lake environment. *Biol. J. Linn. Soc. Lond.* **105**, 817–831 (2012).
40. A. Garcia-Elfring, A. Paccard, T. J. Thurman, B. A. Wasserman, E. P. Palkovacs, A. P. Hendry, R. D. H. Barrett, Using seasonal genomic changes to understand historical adaptation to new environments: Parallel selection on stickleback in highly-variable estuaries. *Mol. Ecol.* **30**, 2054–2064 (2021).
41. D. A. Marques, F. C. Jones, F. Di Palma, D. M. Kingsley, T. E. Reimchen, Experimental evidence for rapid genomic adaptation to a new niche in an adaptive radiation. *Nat. Ecol. Evol.* **2**, 1128–1138 (2018).
42. R. D. H. Barrett, S. M. Rogers, D. Schluter, Natural selection on a major armor gene in threespine stickleback. *Science* **322**, 255–257 (2008).
43. T. G. Laurentino, D. Moser, M. Roesti, M. Ammann, A. Frey, F. Ronco, B. Kueng, D. Berner, Genomic release-recapture experiment in the wild reveals within-generation polygenic selection in stickleback fish. *Nat. Commun.* **11**, 1928 (2020).
44. N. Yu, F.-C. Chen, S. Ota, L. B. Jorde, P. Pamilo, L. Patthy, M. Ramsay, T. Jenkins, S.-K. Shyue, W.-H. Li, Larger genetic differences within Africans than between Africans and Eurasians. *Genetics* **161**, 269–274 (2002).
45. D. Schluter, E. A. Clifford, M. Nemethy, J. S. McKinnon, Parallel evolution and inheritance of quantitative traits. *Am. Nat.* **163**, 809–822 (2004).
46. J. A. Baker, D. C. Heins, J. E. Baum, Trajectory and rate of change in female life-history traits following colonization of a freshwater, lacustrine environment by oceanic threespine stickleback. *Evol. Ecol. Res.* **20**, 247–263 (2019).
47. M. A. Bell, D. C. Heins, M. A. Wund, F. A. von Hippel, R. Massengill, K. Dunker, G. A. Bristow, W. E. Aguirre, Reintroduction of threespine stickleback into Cheney and Scout Lakes, Alaska. *Evol. Ecol. Res.* **17**, 157–178 (2016).
48. K. Spitzer, M. Pelizzola, A. Futschik, Modifying the Chi-square and the CMH test for population genetic inference: Adapting to overdispersion. *Ann. Apl. Stats.* **14**, 202–220 (2020).
49. N. M. O'Brown, B. R. Summers, F. C. Jones, S. D. Brady, D. M. Kingsley, A recurrent regulatory change underlying altered expression and Wnt response of the stickleback armor plates gene EDA. *eLife* **4**, e05290 (2015).
50. A. P. McGregor, V. Orgogozo, I. Delon, J. Zanet, D. G. Srinivasan, F. Payre, D. L. Stern, Morphological evolution through multiple *cis*-regulatory mutations at a single gene. *Nature* **448**, 587–590 (2007).
51. S. Koshikawa, M. W. Giorgianni, K. Vaccaro, V. A. Kassner, J. H. Yoder, T. Werner, S. B. Carroll, Gain of *cis*-regulatory activities underlies novel domains of wingless gene expression in *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 7524–7529 (2015).
52. L. F. Stam, C. C. Laurie, Molecular dissection of a major gene effect on a quantitative trait: The level of alcohol dehydrogenase expression in *Drosophila melanogaster*. *Genetics* **144**, 1559–1564 (1996).
53. C. L. Peichel, D. A. Marques, The genetic and molecular architecture of phenotypic diversity in sticklebacks. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **372**, 20150486 (2017).
54. C. T. Miller, A. M. Glazer, B. R. Summers, B. K. Blackman, A. R. Norman, M. D. Shapiro, B. L. Cole, C. L. Peichel, D. Schluter, D. M. Kingsley, Modular skeletal evolution in sticklebacks is controlled by additive and clustered quantitative trait loci. *Genetics* **197**, 405–420 (2014).
55. A. H. Chan, P. A. Jenkins, Y. S. Song, Genome-wide fine-scale recombination rate variation in *Drosophila melanogaster*. *PLOS Genet.* **8**, e1003090 (2012).
56. J. Galloway, W. A. Cresko, P. Ralph, A few stickleback suffice for the transport of alleles to new lakes. *G3* **10**, 505–514 (2020).
57. F. Han, S. Lamichhaney, B. R. Grant, P. R. Grant, L. Andersson, M. T. Webster, Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among Darwin's finches. *Genome Res.* **27**, 1004–1015 (2017).
58. J. I. Meier, D. A. Marques, C. E. Wagner, L. Excoffier, O. Seehausen, Genomics of parallel ecological speciation in Lake Victoria cichlids. *Mol. Biol. Evol.* **35**, 1489–1506 (2018).
59. W. Mei, M. G. Stetter, D. J. Gates, M. C. Stitzer, J. Ross-Ibarra, Adaptation in plant genomes: Bigger is different. *Am. J. Bot.* **105**, 16–19 (2018).
60. R. D. Hernandez, J. L. Kelley, E. Elyashiv, S. C. Melton, A. Auton, G. McVean; 1000 Genomes Project, G. Sella, M. Przeworski, Classic selective sweeps were rare in recent human evolution. *Science* **331**, 920–924 (2011).
61. C. L. Peichel, K. S. Nereng, K. A. Ohgi, B. L. E. Cole, P. F. Colosimo, C. A. Buerkle, D. Schluter, D. M. Kingsley, The genetic architecture of divergence between threespine stickleback species. *Nature* **414**, 901–905 (2001).
62. C. L. Peichel, S. T. Sullivan, I. Liachko, M. A. White, Improvement of the threespine stickleback genome using a Hi-C-based proximity-guided assembly. *J. Hered.* **108**, 693–700 (2017).
63. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
64. M. A. DePristo, E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, G. del Angel, M. A. Rivas, M. Hanna, A. McKenna, T. J. Fennell, A. M. Kernytsky, A. Y. Sivachenko, K. Cibulskis, S. B. Gabriel, D. Altshuler, M. J. Daly, A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
65. M. Lynch, D. Bost, S. Wilson, T. Maruki, S. Harrison, Population-genetic inference from pooled-sequencing data. *Genome Biol. Evol.* **6**, 1210–1218 (2014).
66. R. Kofler, R. V. Pandey, C. Schlötterer, PoPoolation2: Identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics* **27**, 3435–3436 (2011).

67. A. F. Shanfelter, S. Archambeault, M. A. White, Divergent fine-scale recombination landscapes between a freshwater and marine population of threespine stickleback fish. *Genome Biol. Evol.* **11**, 1573–1585 (2019).
68. A. M. Glazer, E. E. Killingbeck, T. Mitros, D. S. Rokhsar, C. T. Miller, Genome assembly improvement and mapping convergently evolved skeletal traits in sticklebacks with genotyping-by-sequencing. *G3* **5**, 1463–1472 (2015).
69. G. Bhatia, N. Patterson, S. Sankararaman, A. L. Price, Estimating and interpreting FST: The impact of rare variants. *Genome Res.* **23**, 1514–1521 (2013).
70. T. Taus, A. Futschik, C. Schlötterer, Quantifying selection with pool-seq time series data. *Mol. Biol. Evol.* **34**, 3023–3034 (2017).
71. W. Aguirre, P. Doherty, M. Bell, Genetics of lateral plate and Gillraker phenotypes in a rapidly evolving population of threespine stickleback. *Behaviour* **141**, 1465–1483 (2004).
72. A. Y. K. Albert, S. Sawaya, T. H. Vines, A. K. Knecht, C. T. Miller, B. R. Summers, S. Balabhadra, D. M. Kingsley, D. Schluter, The genetics of adaptive shape shift in stickleback: Pleiotropy and effect size. *Evolution* **62**, 76–85 (2008).
73. S. Arif, W. E. Aguirre, M. A. Bell, Evolutionary diversification of opercle shape in Cook Inlet threespine stickleback. *Biol. J. Linn. Soc. Lond.* **97**, 832–844 (2009).
74. W. E. Aguirre, K. E. Ellis, M. Kusenda, M. A. Bell, Phenotypic variation and sexual dimorphism in anadromous threespine stickleback: Implications for postglacial adaptive radiation. *Biol. J. Linn. Soc. Lond.* **95**, 465–478 (2008).
75. M. D. Shapiro, M. E. Marks, C. L. Peichel, B. K. Blackman, K. S. Nereng, B. Jónsson, D. Schluter, D. M. Kingsley, Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* **428**, 717–723 (2004).
76. Y. F. Chan, M. E. Marks, F. C. Jones, G. Villarreal Jr., M. D. Shapiro, S. D. Brady, A. M. Southwick, D. M. Absher, J. Grimwood, J. Schmutz, R. M. Myers, D. Petrov, B. Jónsson, D. Schluter, M. A. Bell, D. M. Kingsley, Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. *Science* **327**, 302–305 (2010).
77. A. S. Hinrichs, D. Karolchik, R. Baertsch, G. P. Barber, G. Bejerano, H. Clawson, M. Diekhans, T. S. Furey, R. A. Harte, F. Hsu, J. Hillman-Jackson, R. M. Kuhn, J. S. Pedersen, A. Pohl, B. J. Raney, K. R. Rosenbloom, A. Siepel, K. E. Smith, C. W. Sugnet, A. Sultan-Qurraie, D. J. Thomas, H. Trumbower, R. J. Weber, M. Weirauch, A. S. Zweig, D. Haussler, W. J. Kent, The UCSC Genome Browser Database: Update 2006. *Nucleic Acids Res.* **34**, D590–D598 (2006).
78. W. J. Kent, R. Baertsch, A. Hinrichs, W. Miller, D. Haussler, Evolution's cauldron: Duplication, deletion, and rearrangement in the mouse and human genomes. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 11484–11489 (2003).
79. W. J. Kent, BLAT--the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
80. B. L. Aken, S. Ayling, D. Barrell, L. Clarke, V. Curwen, S. Fairley, J. Fernandez Banet, K. Billis, C. García Girón, T. Hourlier, K. Howe, A. Kähäri, F. Kokocinski, F. J. Martin, D. N. Murphy, R. Nag, M. Ruffier, M. Schuster, Y. A. Tang, J.-H. Vogel, S. White, A. Zadissa, P. Flicek, S. M. J. Searle, The Ensembl gene annotation system. *Database*. 2016, baw093 (2016).
81. M. Schubert, S. Lindgreen, L. Orlando, AdapterRemoval v2: Rapid adapter trimming, identification, and read merging. *BMC. Res. Notes* **9**, 88 (2016).
82. D. W. Hagen, Isolating mechanisms in threespine sticklebacks (*Gasterosteus*). *J. Fish. Res. Board Can.* **24**, 1637–1692 (1967).
83. K. Yoshida, T. Makino, K. Yamaguchi, S. Shigenobu, M. Hasebe, M. Kawata, M. Kume, S. Mori, C. L. Peichel, A. Toyoda, A. Fujiyama, J. Kitano, Sex chromosome turnover contributes to genomic divergence between incipient stickleback species. *PLOS Genet.* **10**, e1004223 (2014).
84. M. Higuchi, A. Goto, Genetic evidence supporting the existence of two distinct species in the genus *Gasterosteus* around Japan. *Environ. Biol. Fishes* **47**, 1–16 (1996).
85. S. Varadharajan, P. Rastas, A. Löytynoja, M. Matschiner, F. C. F. Calboli, B. Guo, A. J. Nederbragt, K. S. Jakobsen, J. Merilä, A high-quality assembly of the nine-spined stickleback (*Pungitius pungitius*) genome. *Genome Biol. Evol.* **11**, 3291–3308 (2019).
86. M. Kirkpatrick, N. Barton, Chromosome inversions, local adaptation and speciation. *Genetics* **173**, 419–434 (2006).
87. S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, P. C. Sham, PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
88. B. S. Gaut, A. D. Long, The lowdown on linkage disequilibrium. *Plant Cell* **15**, 1502–1506 (2003).
89. R. C. Lewontin, The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics* **49**, 49–67 (1964).
90. K. Osoegawa, K. C. Mallempati, S. Gangavarapu, A. Oki, K. Gendzekhadze, S. R. Marino, N. K. Brown, M. P. Bettinotti, E. T. Weimer, G. Montero-Martín, L. E. Creary, T. A. Vayntrub, C.-J. Chang, M. Askar, S. J. Mack, M. A. Fernández-Viña, HLA alleles and haplotypes observed in 263 US families. *Hum. Immunol.* **80**, 644–660 (2019).
91. P. A. Hohenlohe, S. Bassham, M. Currey, W. A. Cresko, Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **367**, 395–408 (2012).
92. C. C. Chang, C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, J. J. Lee, Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
93. S. Purcell, C. Chang, *PLINK*; www.cog-genomics.org/plink/1.9/.
94. N. Patterson, A. L. Price, D. Reich, Population structure and eigenanalysis. *PLOS Genet.* **2**, e190 (2006).
95. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
96. O. Delaneau, B. Howie, A. J. Cox, J.-F. Zagury, J. Marchini, Haplotype estimation using sequencing reads. *Am. J. Hum. Genet.* **93**, 687–696 (2013).
97. M. A. White, J. Kitano, C. L. Peichel, Purifying selection maintains dosage-sensitive genes during degeneration of the threespine stickleback Y chromosome. *Mol. Biol. Evol.* **32**, 1981–1995 (2015).
98. G. Lunter, M. Goodson, Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res.* **21**, 936–939 (2011).
99. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
100. M. A. Bell, J. D. Stewart, P. J. Park, The world's oldest fossil threespine stickleback fish. *Copeia* **2009**, 256–265 (2009).
101. B. Guo, F. J. J. Chain, E. Bornberg-Bauer, E. H. Leder, J. Merilä, Genomic divergence between nine- and three-spined sticklebacks. *BMC Genomics* **14**, 756 (2013).
102. R. Kawahara, M. Miya, K. Mabuchi, T. J. Near, M. Nishida, Stickleback phylogenies resolved: Evidence from mitochondrial genomes and 11 nuclear genes. *Mol. Phylogenet. Evol.* **50**, 401–404 (2009).
103. T. S. Korneliussen, A. Albrechtsen, R. Nielsen, ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics* **15**, 356 (2014).
104. C. S. Smukowski Heil, C. Ellison, M. Dubin, M. Noor, Recombining without hotspots: A comprehensive evolutionary portrait of recombination in two closely related species of *Drosophila*. *Genome Biol. Evol.* **7**, 2829–2842 (2015).
105. R. R. Hudson, M. Slatkin, W. P. Maddison, Estimation of levels of gene flow from DNA sequence data. *Genetics* **132**, 583–589 (1992).
106. Á. Jónás, T. Taus, C. Kosiol, C. Schlötterer, A. Futschik, Estimating the effective population size from temporal allele frequency changes in experimental evolution. *Genetics* **204**, 723–735 (2016).
107. N. O. Rode, Y. Holtz, K. Loridon, S. Santoni, J. Ronfort, L. Gay, How to optimize the precision of allele and haplotype frequency estimates using pooled-sequencing data. *Mol. Ecol. Resour.* **18**, 194–203 (2018).
108. C. Vlachos, C. Burny, M. Pelizzola, R. Borges, A. Futschik, R. Kofler, C. Schlötterer, Benchmarking software tools for detecting and quantifying selection in evolve and resequencing studies. *Genome Biol.* **20**, 169 (2019).
109. W. G. Hill, A. Robertson, The effect of linkage on limits to artificial selection. *Genet. Res.* **8**, 269–294 (1966).
110. J. H. Gillespie, *Population Genetics: A Concise Guide* (Johns Hopkins University Press, 1998).
111. Z. He, X. Dai, M. Beaumont, F. Yu, Detecting and quantifying natural selection at two linked loci from time series data of allele frequencies with forward-in-time simulations. *Genetics* **216**, 521–541 (2020).
112. M. Kimura, A model of a genetic system which leads to closer linkage by natural selection. *Evolution* **10**, 278–287 (1956).
113. J. Terhorst, C. Schlötterer, Y. S. Song, Multi-locus analysis of genomic time series data from experimental evolution. *PLOS Genet.* **11**, e1005069 (2015).
114. C. J. Battey, P. L. Ralph, A. D. Kern, Predicting geographic location from genetic variation with deep neural networks. *eLife* **9**, e54507 (2020).
115. A. D. Kern, D. R. Schrider, diploS/HIC: An updated approach to classifying selective sweeps. *G3* **8**, 1959–1970 (2018).
116. D. R. Schrider, A. D. Kern, Supervised machine learning for population genetics: A new paradigm. *Trends Genet.* **34**, 301–312 (2018).
117. B. C. Haller, P. W. Messer, SLiM 3: Forward genetic simulations beyond the Wright–Fisher model. *Mol. Biol. Evol.* **36**, 632–637 (2019).
118. C. D. Huber, A. Durvasula, A. M. Hancock, K. E. Lohmueller, Gene expression drives the evolution of dominance. *Nat. Commun.* **9**, 2750 (2018).
119. A. F. Agrawal, M. C. Whitlock, Inferences about the distribution of dominance drawn from yeast gene knockout data. *Genetics* **187**, 553–566 (2011).
120. A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems* (O'Reilly Media Inc., 2019).
121. T. A. O'Brien, K. Kashinath, N. R. Cavanaugh, W. D. Collins, J. P. O'Brien, A fast and objective multidimensional kernel density estimation method: fastKDE. *Comput. Stat. Data Anal.* **101**, 148–160 (2016).
122. T. A. O'Brien, W. D. Collins, S. A. Rauscher, T. D. Ringler, Reducing the computational cost of the ECF using a nuFFT: A fast and objective probability density estimation method. *Comput. Stat. Data Anal.* **79**, 222–234 (2014).

123. J. A. Chaves, E. A. Cooper, A. P. Hendry, J. Podos, L. F. De León, J. A. M. Raeymaekers, W. O. MacMillan, J. A. C. Uy, Genomic variation at the tips of the adaptive radiation of Darwin's finches. *Mol. Ecol.* **25**, 5282–5295 (2016).
124. S. Singhal, E. M. Leffler, K. Sannareddy, I. Turner, O. Venn, D. M. Hooper, A. I. Strand, Q. Li, B. Raney, C. N. Balakrishnan, S. C. Griffith, G. McVean, M. Przeworski, Stable recombination hotspots in birds. *Science* **350**, 928–932 (2015).
125. S. Lamichhaney, F. Han, J. Berglund, C. Wang, M. S. Almén, M. T. Webster, B. R. Grant, P. R. Grant, L. Andersson, A beak size locus in Darwin's finches facilitated character displacement during a drought. *Science* **352**, 470–474 (2016).
126. S. Lamichhaney, J. Berglund, M. S. Almén, K. Maqbool, M. Grabherr, A. Martinez-Barrio, M. Promerová, C.-J. Rubin, C. Wang, N. Zamani, B. R. Grant, P. R. Grant, M. T. Webster, L. Andersson, Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* **518**, 371–375 (2015).

Acknowledgments: We thank the many individuals who contributed to this study. Fish samples used for geographic surveys: S. Arnott, B. Blackman, F. Chan, P. Colosimo, A. Dalziel, B. Deagle, D. P. Højgaard, J. A. Jacobsen, B. Jonsson, R. King, D. Kueltz, A. Maccoll, J. McKinnon, C. Miller, S. Mori, K. O'Brien, C. Peichel, M. Ravinet, M. Rhodes-Reese and NOAA, T. Reimchen, J. Richmond, D. Schluter, M. Shapiro, B. Summers, and T. Vines. Undergraduate laboratory assistants: J. Ancona, H. Babalola, P. Chohan, J. F. Gaige, Z. Khan, and J. Mallozzi. Sampling access: M. Tauriainen and the staff of T & J Gravel Products permitted us to collect stickleback on their property from Scout and Loberg lakes, respectively. Field assistants, Alaska: S. Abrams, D. Arciari, D. B. Bell, M. R. Bell, S. R. Bell, B. Berland, M. Bobb, G. A. Bristow, K. T. Ellis, V. Ely, J. Fitzgerald, A. K. Gangavelli, M. A. Hahn, A. C. Havens, A. Hernandez, L. Hitt, J. Johnson, E. Kalabacas, A. Karve, H. Knoper, F. Kreier, M. Kurz, R. Lucas, A. McGarry, M. McGee, B. K. Lohman, R. Paitz, A. Plaunova, J. L. Rollins, H. Schultz, M. Sekiya, D. L. Soltz, L. Stein, A. C. Thompson, M. P. Travis Heide Viitaniemi, J. I. Wucherpfennig, and K. T. Xie. We thank A. Hinrichs, H. Clawson, K. Smith, and D. Karolchik for contribution to the UCSC Stickleback Genome Browser annotations for *gasAcu1*, which were lifted to *gasAcu1-4* for this study. IACUC approvals: K.R.V.: 1446584, Stony Brook University. M.A.B.: 237429, Stony Brook University. D.M.K: 13834, Stanford University. F.C.J.: 35/9185.82-5 EB01/09 A, Baden-Württemberg Regierungspräsidium, Germany. Friedrich Miescher Laboratory of Max Planck Society, Tübingen, Germany. D.C.H.: 0304R-UT-C, 0304R2, 0304R3, and 0304R4, Tulane University. **Funding:** This work was supported by NSF BSR8905758, BSR9046191, DEB0211391, DEB0322818, DEB0509070, and DEB0919184 to M.A.B.; NIH R01GM124330 to K.R.V.; DFG SPP1819 and ERC 617219 to F.C.J.; by Newcomb Institute grants to D.C.H.; NSF GRFP to G.A.R.K.; NSF GRFP 1656518 and NIH 5T32GM007790 and Stanford CEHG Fellowship to H.I.C.; and NIH 3P50HG002568 and 3P50HG002568-09S1 to D.M.K. D.M.K. is an investigator of HHMI.

Author contributions: M.A.B: experimental design, Alaskan sampling and population founding, tissue sampling, and morphological data collection. P.J.P.: Alaskan sampling and population founding. F.A.v.H.: Alaskan sampling and logistical support. W.E.A.: Alaskan sampling and morphological data collection. D.C.H.: Alaskan sampling. G.A.R.K.: analysis of geographic populations and genomic properties predictive of evolution. D.N.V., M.M., and T.S.B.: DNA extraction, quantitation, selection, and pooling of Alaskan samples. F.C.J.: design of geographic sampling and SNP array and SNP array analysis. S.D.B.: sample curation, DNA preparation, and SNP array calling. D.M.A. and R.M.M.: SNP array genotyping. M.K.: SNP array analysis. H.I.C.: transfer and visualization of genome annotations. F.D.P.: geographic population sequencing. D.M.K.: experimental design and conceptual guidance. K.R.V. and K.R.: analysis and modeling of contemporary populations. G.A.R.K. and K.R.V. wrote the manuscript with input from all authors. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. All WGS Illumina data of extant populations have been deposited in the Sequence Read Archive (www.ncbi.nlm.nih.gov/sra) at accession PRJNA247503. All WGS Illumina data for contemporary pool-seq experiments have been deposited in the Sequence Read Archive at accession PRJNA671824. All WGS BGI data for RS 2009 genomes used to construct the recombination map have been deposited in the Sequence Read Archive at accession PRJNA671690. SNP genotyping array data and ancestral reference sequence have been deposited on Data Dryad (https://doi.org/10.5061/dryad.pvmcvdnjm). EcoPeak, TempoPeak, and RS recombination rate data can be visualized and downloaded [via the Table Browser (*71*) of the UCSC Genome Browser (http://genome.ucsc.edu/) (*72*)] by copying the following track hub (*73*) URL into the "My Hubs" tab at https://genome.ucsc.edu/cgi-bin/hgHubConnect: https://sbwdev.stanford.edu/kingsleyAssemblyHub/hub.txt. The assembly hub must be opened through the UCSC Genome Browser, not directly. We provide the *gasAcu1-4* reference genome as well as liftOver chains for converting to and from the original Broad S1 stickleback reference genome (*gasAcu1*) via Data Dryad at https://doi.org/10.5061/dryad.547d7wm6t.

Submitted 11 January 2021
Accepted 5 May 2021
Published 18 June 2021
10.1126/sciadv.abg5285

Citation: G. A. Roberts Kingman, D. N. Vyas, F. C. Jones, S. D. Brady, H. I. Chen, K. Reid, M. Milhaven, T. S. Bertino, W. E. Aguirre, D. C. Heins, F. A. von Hippel, P. J. Park, M. Kirch, D. M. Absher, R. M. Myers, F. Di Palma, M. A. Bell, D. M. Kingsley, K. R. Veeramah, Predicting future from past: The genomic basis of recurrent and rapid stickleback evolution. *Sci. Adv.* **7**, eabg5285 (2021).

Extract of Supplementary Materials for

Predicting future from past: The genomic basis of recurrent and rapid stickleback evolution

Garrett A. Roberts Kingman, Deven N. Vyas, Felicity C. Jones, Shannon D. Brady, Heidi I. Chen, Kerry Reid,Mark Milhaven, Thomas S. Bertino, Windsor E. Aguirre, David C. Heins, Frank A. von Hippel, Peter J. Park, Melanie Kirch, Devin M. Absher, Richard M. Myers, Federica Di Palma, Michael A. Bell*, David M. Kingsley*, Krishna R. Veeramah*

*Corresponding author. Email: krishna.veeramah@stonybrook.edu (K.R.V.); sticklemack@gmail.com (M.A.B.); kingsley@stanford.edu (D.M.K.)

Published 18 June 2021, *Sci. Adv.* **7**, eabg5285 (2021) DOI: 10.1126/sciadv.abg5285

The PDF file includes:

Sections S1, S2, S5 and S12
Figs. S1, S10 to S16
References

Other Supplementary Material for this manuscript includes the following:

(available at advances.sciencemag.org/cgi/content/full/7/25/eabg5285/DC1)

Sections S1 to S23
Figs. S1 to S89
Tables S1, S2, S11 to S14 Movies S1 to S4
Legends for tables S1, S2, S11 to S14 Tables S3 to S10 and S15 to S19 Legends for movies S1 to S4
References

1. SAMPLE COLLECTION

General sampling methods for established populations. Fish were generally trapped using unbaited minnow traps set near the shore, immediately euthanized with MS 222 (tricaine methanesulfonate), then preserved in 70% ethanol for transport and storage.

Sample collections for SNP Genotyping array. Marine threespine stickleback were collected from Rabbit slough (751 specimens), Resurrection Bay (655 specimens) and Glacier Spit (237 specimens) in Alaska and genotyped using a custom SNP genotyping array detailed in section 8.

General sampling methods and background for contemporary evolution lakes. Annual samples were made from populations in three Alaskan lakes (Loberg, Cheney and Scout) to provide an evolutionary time series from founding to study rapid genomic evolution (https://www.adfg.alaska.gov/index.cfm?adfg=fishingSportLakeData.main&StockingAreaID=2 for lake information, **Figure S1**). These populations had been founded recently by anadromous (sea-run) threespine stickleback, and genetically determined phenotypic traits (*71*, *72*) were evolving rapidly in all of them (*17*, *26*, *39*, *47*, *73*). Stickleback were sampled under collecting permits from the Alaska Department of Fish and Game and sacrificed under protocols approved by the Institutional Animal Care and Use Committee of Stony Brook University to MAB. Each sample was trapped using unbaited Gees® Galvanized Minnow Traps (G-40), which were set for about 24 h near shore, on the bottom, at about 1 m depth. Fish were sacrificed using an overdose of MS 222 (Tricaine methanesulfonate) until they were unresponsive to a tap on the side of the bucket. They were rinsed in lake water and immediately immersed in 100 to 70% ethanol. The ethanol was replaced with fresh ethanol within 24 hours, and the sample was stored at room temperature. Sample details are in **Table S1**. The anadromous threespine stickleback population that breeds in Rabbit Slough (*74*), about 1.25 km from Loberg Lake, or a closely related population from an adjacent tributary in the same drainage probably founded the Loberg Lake population naturally (*17*). MAB introduced anadromous Rabbit Slough adults into Cheney and Scout lakes in 2009 and 2011, respectively (*47*). Thus, all three lake populations that we used to study contemporary genomic evolution were derived from the same or closely related anadromous threespine stickleback populations.

Rabbit Slough samples. Two thousand nine hundred sixty-four (2964) adult anadromous (sea-run) threespine stickleback were trapped in Rabbit Slough, Matanuska-Susitna Borough, Alaska as they migrated upstream to breed in 2009 (*47*). They were captured using nine one quarter-inch mesh traps, set entirely across the outlet of a culvert that discharges from under the west side of the Parks Highway. The first dorsal spine was clipped from each specimen before it was released into Cheney Lake, and the spines were immediately put into 95% ethanol.

Loberg Lake samples. Loberg Lake is a 4.5 ha lake near Palmer, in the Matanuska-Susitna Borough, Alaska at about 61.56º N latitude and 149.26º W longitude. The population was founded between 1983 and 1988, and we treat 1985 as the year it originated (*17*). It was apparently founded by a substantial number of anadromous threespine stickleback that entered the lake through a spring that discharges into Spring Creek, which is in the same drainage and close to Rabbit Slough (*17*). Annual samples from Loberg Lake were first made in 1990, and subsamples were preserved in ethanol for DNA sequencing starting in 1999 (*17*, *26*). Most of these samples came from a site about halfway between sites A and C of Bell et al. (*17*), but some came from the five other sampling sites around the lake. Samples were made with six to 20 one-eighth and one-quarter-inch mesh traps.

Cheney Lake introduction and samples. Cheney Lake is a 9.9 ha lake in Anchorage City-Borough, Alaska at 61.2° N. latitude and 149.76° W longitude (*47*). It was treated with rotenone to exterminate an exotic species in the fall of 2008 and between 29 May and 3 June 2009, 2964 adult anadromous stickleback were transported in their native water in aerated coolers (i.e., ice chests) from Rabbit Slough to the University of Alaska Anchorage (UAA), 55 km away. Their first dorsal spine was clipped for DNA extraction. They were held in aerated, aged tap water, and transported within 24 hours to Cheney Lake, about 5 km away, in aerated, aged tap water in a cooler, and released at the south end of the lake. The water in the cooler was exchanged with Cheney Lake water in a series of dilutions to reduce the temperature difference to <4 C. The stickleback were caught with large aquarium nets and released into the lake. They formed a linear school and swam rapidly into deeper water.

Annual sampling started in Cheney Lake in the late summer of 2009, when juvenile stickleback were abundant. Capture efficiency declined during the next two years, and virtually no stickleback were caught in 2013 and 2014. In 2015, they were caught in small numbers, which have increased progressively since then (*47*). Since 2015, traps have been set only at the tip of a peninsula on the north shore, where capture was most efficient. Samples were available for DNA extraction each year besides 2013 and 2014.

Scout Lake introduction and samples. Scout Lake is a 38.5 ha lake in Stirling, Kenai Peninsula Borough, Alaska at 60.54° N latitude and 150.83° W longitude (*47*). Between 4 June and 4 July 2011, 3047 adult stickleback from the Rabbit Slough anadromous population were released into Scout Lake using the same methods as in Cheney Lake. However, they were not abundant in Rabbit Slough this year and had to be accumulated for several days at UAA before transport to Scout Lake, about 216 km away, and released at the east end of the lake. Annual samples were first made from this population in the late summer of 2011. Capture efficiency declined precipitously in 2013 and 2014, and their numbers have been recovering since then (*47*). Annual samples for DNA extraction have been made each year since.

Tissue sampling. Each sample was given code letters and a descriptive name, and each fish was individually numbered. Eight stickleback at a time were removed from the storage bottle and a pectoral or caudal fin (occasionally two fins from small fish) was grasped with

needle-nose forceps and cut off with iridectomy scissors, which were washed in alcohol between specimens and in bleach between samples from different populations or dates. Fin clips were stored in 70% ethanol in 1.5 ml microcentrifuge tubes or 48-well plates. The fish from which fins were cut were stored in labeled 10 ml screw-cap tubes with two fish per tube (head up, head down). Fish from which a fin was not clipped were left in the storage bottle. The plates of fin clips were sealed in Saran Plastic Wrap® and taped shut for storage and shipment. Fin clips were made in DMK's laboratory and shipped in a cooler to KRV's laboratory by MAB. All ethanol-preserved stickleback specimens are stored in DMK's laboratory in a cold room.

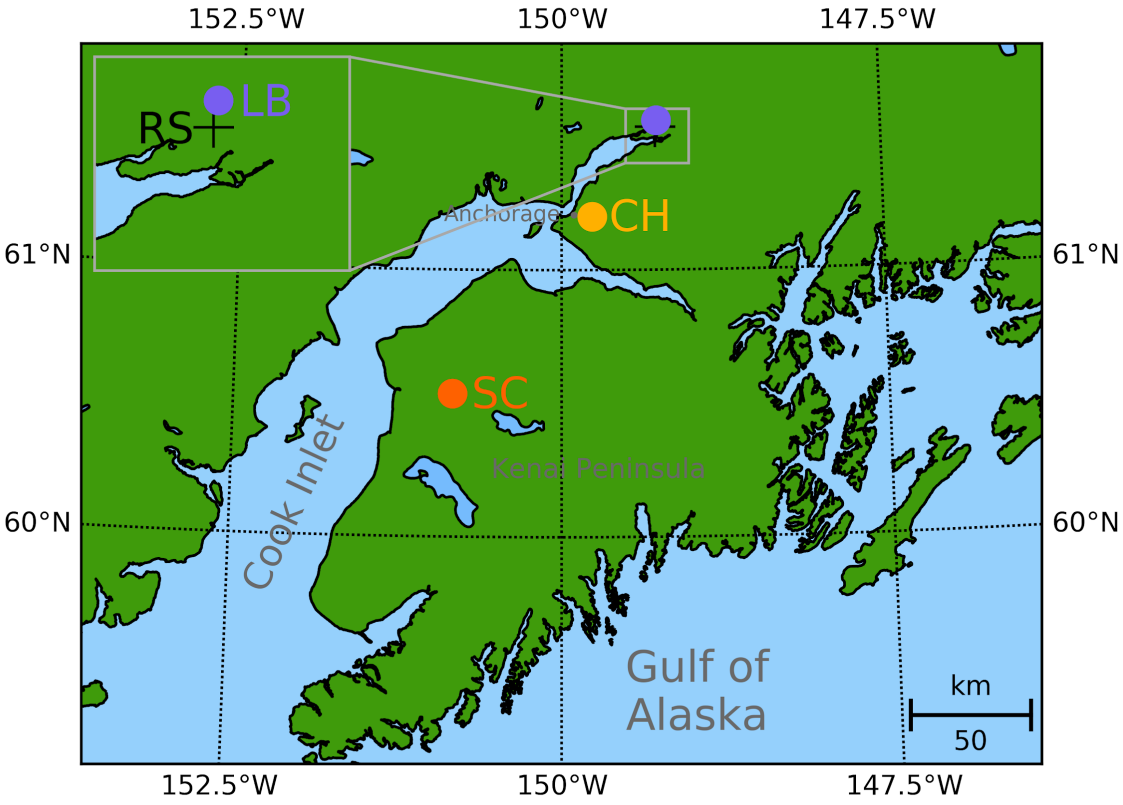


Figure S1: Map of Alaska showing location of three contemporary freshwater lakes as well as Rabbit Slough.

2. DNA EXTRACTION AND QUANTIFICATION

Geographic population survey samples. DNA from caudal and/or pectoral fin tissues was extracted using overnight proteinase K digestion followed by phenol:chloroform isolation with PhaseLock tubes as previously described (*61*). DNA concentration was quantified using a NanoDrop spectrophotometer.

SNP genotyping array samples. DNA from SNP genotyping array samples were extracted from clipped spines or fins using overnight proteinase K digestion followed by phenol:chloroform isolation as previously described (*61*). DNA concentration was quantified using a NanoDrop spectrophotometer.

Contemporary evolution samples. DNA from contemporary samples was extracted from clipped fins or spines using the DNeasy 96 Blood & Tissue Kit following the standard “animal tissue” protocol. Tissue was digested with proteinase K for 24-48 hours, depending on the observed state of the digest. In general, only one lake time point was used per 96-well plate to avoid cross-contamination between lake time-points. RNase A was either applied to individual samples during these extractions or to samples after pooling (see **Table S1**). DNA for each individual sample was quantified using the Qubit High-sensitivity DNA assay.

Table S1: Sample information and sequencing statistics for contemporary evolution lakes undergoing PoolSeq.
See separate excel file

6. REFERENCE GENOME CONSTRUCTION

gasAcu1-4 creation. We use as our reference genome a slight modification of the recent Hi-C guided improvement of the stickleback genome (*62*), which we term *gasAcu1-4*. The Peichel et al. (*62*) Hi-C genome was not modified in any way except to address two long-standing issues that have persisted throughout many versions of the stickleback genome.

First, the subtelomeric region of chrVII is of great biological interest due to its well-documented role in controlling pelvic spine development (*75*). However, due to its highly repetitive nature, it is extremely difficult to assemble and many important sequences, including the key gene *Pitx1*, are missing entirely from existing genome assemblies, while other sequences in the region are scattered in small unassembled scaffolds. We address this issue by including the sequence from Salmon River BAC clones (Genbank GU130435) (*76*) as chrP and removing overlapping fragmented sequences from chrUn and the end of chrVII. We note that chrP is derived from a marine population, while the rest of the genome is from a freshwater population (Bear Paw Lake), so all analyses concerning this chromosome must be interpreted as such.

Second, the mitochondrial genome was previously split into two fragments buried within chrUn, while a separate mitochondrial genome sequence from Northern Japan was added as chrM. We corrected these issues by removing the duplicated Bear Paw Lake mitochondrial genome from chrUn and using it to replace the exogenous chrM sequence, resulting in a single copy of the mitochondrial genome derived entirely from Bear Paw Lake.

Liftover chains. LiftOver (*77*) chains (*78*) between gasAcu1 and gasAcu1-4 were generated from BLAT (*79*) alignments using the script https://github.com/ENCODE-DCC/kentUtils/blob/master/src/hg/utils/automation/doSameSpeciesLiftOver.pl as well as Kent Utilities (https://github.com/ucscGenomeBrowser/kent) following the procedure described here: http://genomewiki.ucsc.edu/index.php/DoSameSpeciesLiftOver.pl

Annotation. Ensembl (*80*) release 94 gene annotations for *gasAcu1* were downloaded via the UCSC Genome Browser and lifted to *gasAcu1-4* using all default liftOver parameters.

Major Quantitative Trait Loci. In order to annotate our new genome build with quantitative trait loci (QTL), we began with all QTLs from the comprehensive 2017 review of all reported stickleback QTLs by Marques and Peichel (*53*). We first filtered out results from stickleback species other than *Gasterosteus aculeatus,* and then for each reported interval lifted the coordinates from *gasAcu1* to *gasAcu1-4*. Finally, to create a set of well-bounded major QTL, we required each QTL to have PVE > 20 and have a reported confidence interval < 5Mb. All markers on chrP were assigned to the end of chrVII, where *Pitx1* physically resides. No filtering was performed on account of trait similarity within a cross or multiple crosses analyzing the same trait. This yielded 108 major QTLs. Some loci contain more overlapping QTL reports than visualized in Fig. 1F.

12. FREQUENCY AND HAPLOTYPE STRUCTURE FROM ARRAY GENOTYPES

To characterise the frequency, extent of haplotype tracts, and linkage disequilibrium of freshwater-adaptive standing genetic variation among marine stickleback, we genotyped 751 Rabbit Slough (RABS), 655 Resurrection Bay (RSBY) and 237 Glacier Spit (GLSP) wild caught marine stickleback using a custom Illumina GoldenGate array.

Data filtering. The genotyping data was filtered to remove individuals with excessive missing genotype calls (mind 0.3), and SNPs with excessive failure rate or heterozygosity using Plink (*87*) (filter parameters `--mind 0.33`; `--geno 0.33`; `--het 0.8`) to retain a dataset comprising 302 snps and 742, 628, 199 individuals for RABS, RSBY and GLSP respectively.

Estimating Frequency of Freshwater-adaptive alleles. Diploid genotypes were recoded to represent the number of derived freshwater alleles [0,1 or 2] where the derived allele was defined to be the minor allele across all marine fish. To estimate the frequency of freshwater adaptive alleles in the marine populations we first extracted SNPs falling within 'sensitive' EcoPeaks and/or adaptive loci identified by Jones et al. (*8*). To ensure the distribution of allele frequencies was not biased towards genomic regions containing several closely linked adaptive loci or by adaptive regions tagged by multiple array SNPs, we merged the bed intervals of adaptive loci within 50kb of each other and estimated allele frequency using a single SNP from each merged region (44 adaptive regions total).

The frequency of freshwater adaptive alleles ranged from 0-0.07 (RABS), 0-0.09 (GLSP) and 0-0.14 (RSBY), with median values less than 0.01 for all RSBY and GLSP and less than 0.001 for RABS (**Figure S10**). Rabbit Slough fish are less likely to carry freshwater adaptive alleles than fish from Glacier Split and Resurrection Bay suggesting they are less affected by introgression with local freshwater populations. We note that the frequency of alleles at each locus is highly correlated among populations with the exception of two loci at comparatively high frequency in RSBY and GLSP, but low frequency in RABS (the chrI inversion at chrI:26.5Mb, and chrV:8.35Mb). This hints at some fine grained population substructure in marine populations - Rabbit Slough is located in the Knik Arm of the Cook Inlet, and is more than 340km away as the fish swims from Glacier Spit and Resurrection Bay (both of which are more proximal to the North Pacific Ocean).

Identifying Carriers of Freshwater Alleles
For each individual, we summed the number of derived alleles at SNPs tagging adaptive loci and classified fish carrying one or more alleles to be putative "carriers" of freshwater-adaptive standing genetic variation. Direct migration of freshwater fish or F1 hybrids through the marine environment to neighbouring freshwater populations has been hypothesized to be an efficient mechanism for the spread of freshwater adaptive alleles (*56*), assuming the migrants could survive strong negative selection pressures while in the marine environment. In this survey of

1569 marine stickleback, we did not find any evidence of F1 hybrids or direct freshwater migrants. In contrast, we observed considerable diversity in the standing genetic variation carried among individuals within and between marine populations (**Figure S11-S13**). As expected from the low allele frequency within each population freshwater-adaptive alleles were most commonly carried in heterozygous form at typically one locus. However, rare individuals carrying freshwater adaptive haplotypes at multiple loci across the genome were also observed (range 1-8 chromosomes of 13 studied). For example, two individuals in RSBY carry homozygous tracts of freshwater-adaptive alleles at multiple, but not all loci targeted by the genotyping array (**Figure S12**). Notably these individuals are a mosaic of freshwater homozygous, heterozygous and marine homozygous tracts across their genome indicating they arise from recent admixture and backcrossing among marine and freshwater ecotypes. This highlights the importance of hybridization and admixture as a source of standing genetic variation that can be used as a substrate for rapid adaptation in the future.

Extent of Freshwater Haplotype Blocks Carried by Marine Fish

The size of freshwater haplotype blocks carried by marine individuals and the potential for the blocks to include freshwater adaptive variation at multiple loci across a chromosome) has direct implications for the rate of rapid adaptation in subsequent freshwater colonisations. Since the size of freshwater haplotype blocks will decay at a rate proportional to 1-recombination rate, the extent of the haplotype block along a chromosome, to the degree that it can be determined from the SNPs tagged by this array, is an indicator of the amount of time the freshwater adaptive haplotype has been in the marine environment. F1 hybrids are expected to be heterozygous at adaptive loci across the entire length of each chromosome, while subsequent generations will carry smaller tracts of heterozygosity. The SNP genotyping array tagged multiple adaptive loci on chromosome IV spanning 14Mb enabling us to explore the extent of freshwater haplotype blocks carried by marine fish.

Of 1484 chromosomes IV sampled from RABS marine fish, only one fish showed a contiguous run of heterozygosity that spanned SNPs in a small 20kb region at the EDA locus chrIV:12821275-12841217 (**Figure S11**). This individual is homozygous for marine alleles at other adaptive loci on chrIV indicating the freshwater haplotype block observed has been present in the marine population for many generations. The paucity of any other marine fish carrying extensive freshwater adaptive haplotype tracts on chrIV in RABS contrasts with our samples from RSBY and GLSP, and may indicate lower geographic proximity to marine-freshwater hybrid zones and/or stronger selective purging of freshwater alleles in RABS marine fish. In RSBY, we observed extensive haplotype blocks spanning at least 4 distinct adaptive loci tagged on the array over as much as 11.8Mb in four of 1256 chromosomes IV (0.0032%; **Figure S12**). Further, in GLSP carriers, a single individual out of 398 chromosomes sampled (0.0025%) was found to be heterozygous at 96% of chrIV SNPs tagged indicating it carries freshwater adaptive alleles at as many as six adaptive loci on this chromosome (**Figure S13**). Combined, these results indicate that large freshwater-adaptive haplotype blocks spanning multiple adaptive loci can be found in the marine population at very low frequency. As a source of partially assembled (linked) adaptive cassettes, these rare large haplotype blocks

have the potential to dramatically increase the rate of adaptation in new freshwater habitats. We also note that these haplotype blocks found in marine carriers are several orders of magnitude larger than individual EcoPeaks, as expected for freshwater adaptive variants that only transiently pass through marine populations.

Linkage Disequilibrium Among Adaptive Loci

Since divergent adaptation to marine and freshwater environments is highly polygenic involving loci on almost all chromosomes, we explored the extent to which marine stickleback carry freshwater adaptive alleles on multiple chromosomes and looked for evidence of inter-chromosomal linkage disequilibrium (LD). The frequency dependency of some LD estimators (eg r^2) can make comparisons across loci challenging - especially for paucimorphisms (rare alleles) whose frequencies may differ by an order of magnitude across loci (*88*). In addition to r^2, we therefore also calculated D' (*89*) in which the two-loci linkage disequilibrium is standardized using the allele frequencies at the loci in question. Calculations were performed in `R` with unphased genotype data using the library `Pould` (*90*).

Previous studies reported evidence for considerable linkage disequilibrium among loci in Alaskan marine stickleback from Rabbit Slough and Resurrection Bay (*91*). However, the low frequency of freshwater adaptive alleles in these populations (<1%), makes the conclusions of the previous study questionable given the modest sample size used (18 individuals from Resurrection Bay, and 14 individuals from Rabbit Slough). Further, the findings are most likely false positives caused by artificial population substructure: the authors analysed fish from Rabbit Slough and Resurrection Bay together as a single population, assuming no geographic substructure exists among their samples, yet our analysis of allele frequencies at adaptive loci (above) show such structure very likely exists.

Here, with considerably larger samples sizes of 742, 628, and 199 individuals for Rabbit Slough (**Figure S14**), Resurrection Bay (**Figure S15**) and Glacier Spit (**Figure S16**), respectively, we find evidence for only modest linkage disequilibrium among adaptive loci across the genome. Our analyses are still relatively underpowered given that most interchromosomal allelic associations are driven by very rare carriers (e.g., in Rabbit Slough, the heatmap in **Figure S14** suggests strong LD among chrXI and chrXIX (r^2), that is driven by a single carrier individual who is heterozygous for alleles on both chrXI and chrXIX (**Figure S11**). Further, the D' standardized measure of LD suffers from the problem of finding complete LD=1 when a rare allele is only ever sampled in association with one of the alleles at the second locus. Combined, we find little evidence for inter-chromosomal linkage disequilibrium among adaptive loci and stress the importance of very large sample sizes required (perhaps 10 fold larger than our current sample sizes) in order to explore this phenomena among paucimorphisms with sufficient power.

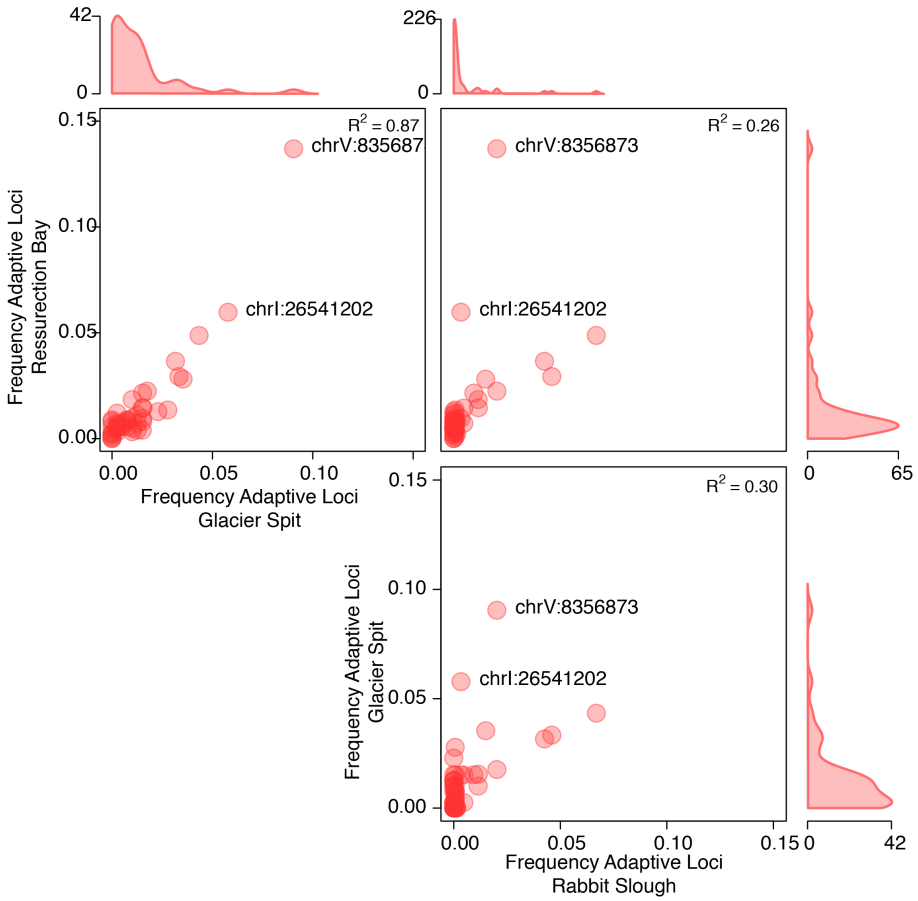


Figure S10: Frequency of freshwater derived alleles in three Alaskan marine populations. Each point represents an adaptive locus with frequency estimated from a single SNP per locus tagging the divergent marine and freshwater haplotypes. Density plots show the distribution of frequencies at the 44 adaptive loci analyzed.

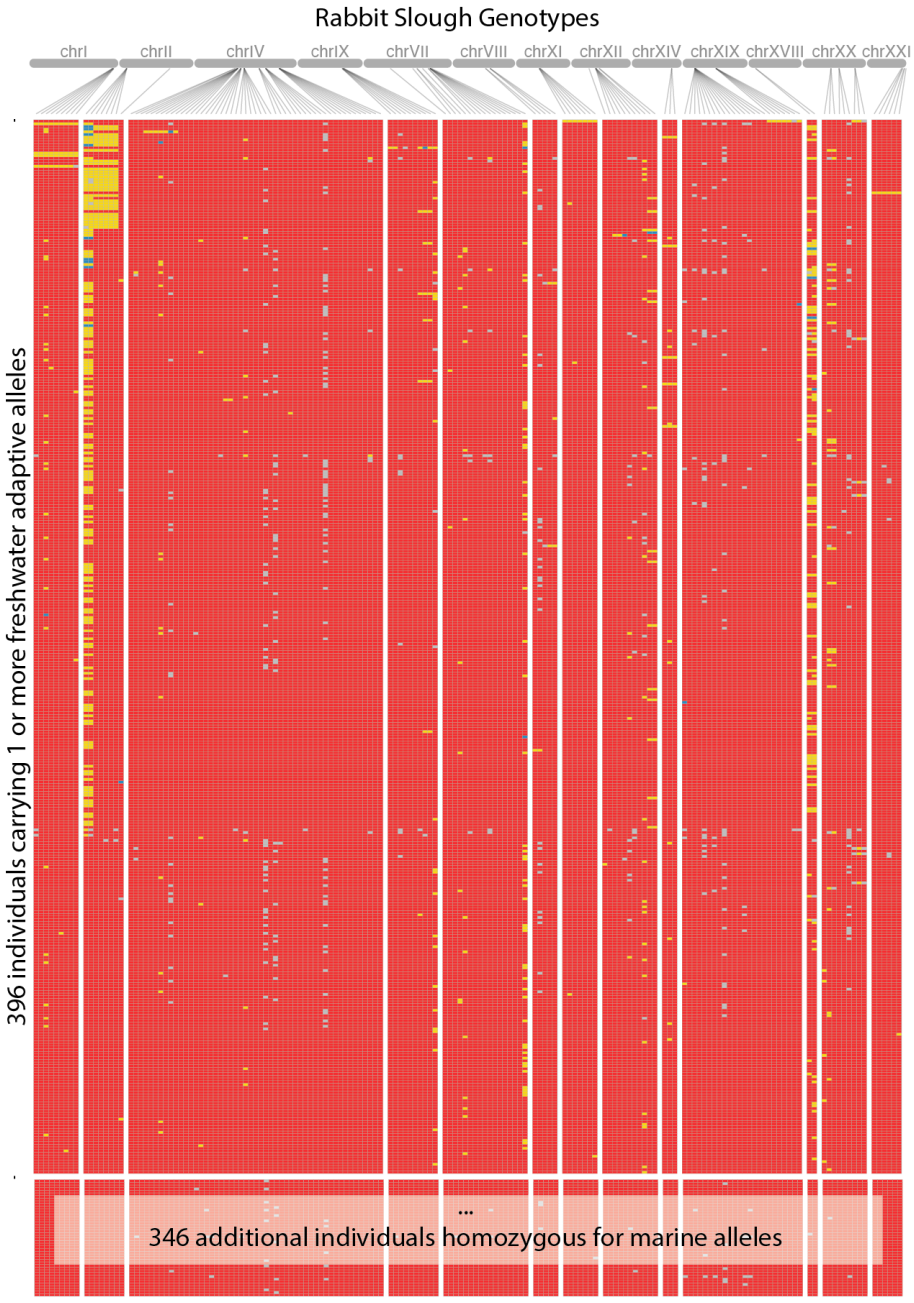


Figure S11: A visual genotype showing 742 Rabbit Slough marine fish at tagged adaptive loci across the genome. Rows represent individuals (sorted from top to bottom in descending order based on the total number of derived alleles each individual carries); columns represent SNPs. Red = homozygous for marine (major) allele; Blue = homozygous for freshwater (derived/minor) allele; Yellow = heterozygous; Grey = missing data.

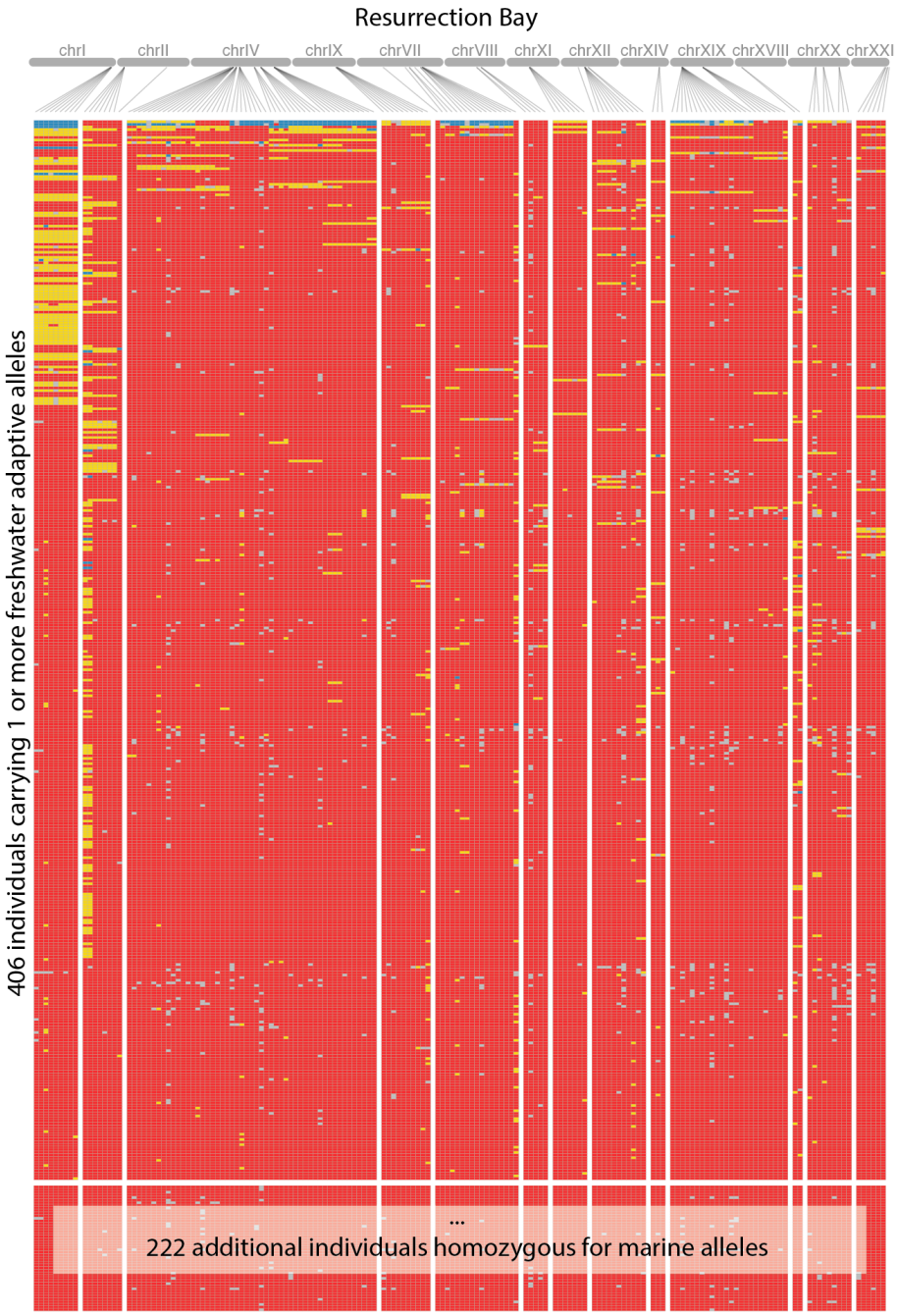


Figure S12: A visual genotype showing 628 Resurrection Bay marine fish at tagged adaptive loci across the genome. Rows represent individuals (sorted from top to bottom in descending order based on the total number of derived alleles each individual carries); columns represent SNPs. Red = homozygous for marine (major) allele; Blue = homozygous for freshwater (derived/minor) allele; Yellow = heterozygous; Grey = missing data.

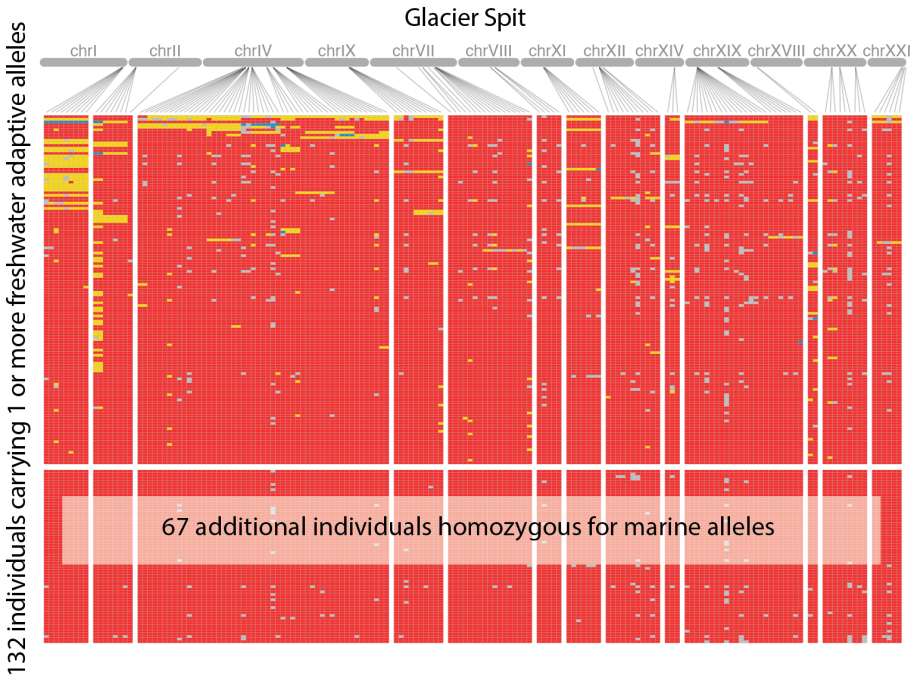


Figure S13: A visual genotype showing 742 Glacier Spit marine fish at tagged adaptive loci across the genome. Rows represent individuals (sorted from top to bottom in descending order based on the total number of derived alleles each individual carries); columns represent SNPs. Red = homozygous for marine (major) allele; Blue = homozygous for freshwater (derived/minor) allele; Yellow = heterozygous; Grey = missing data.

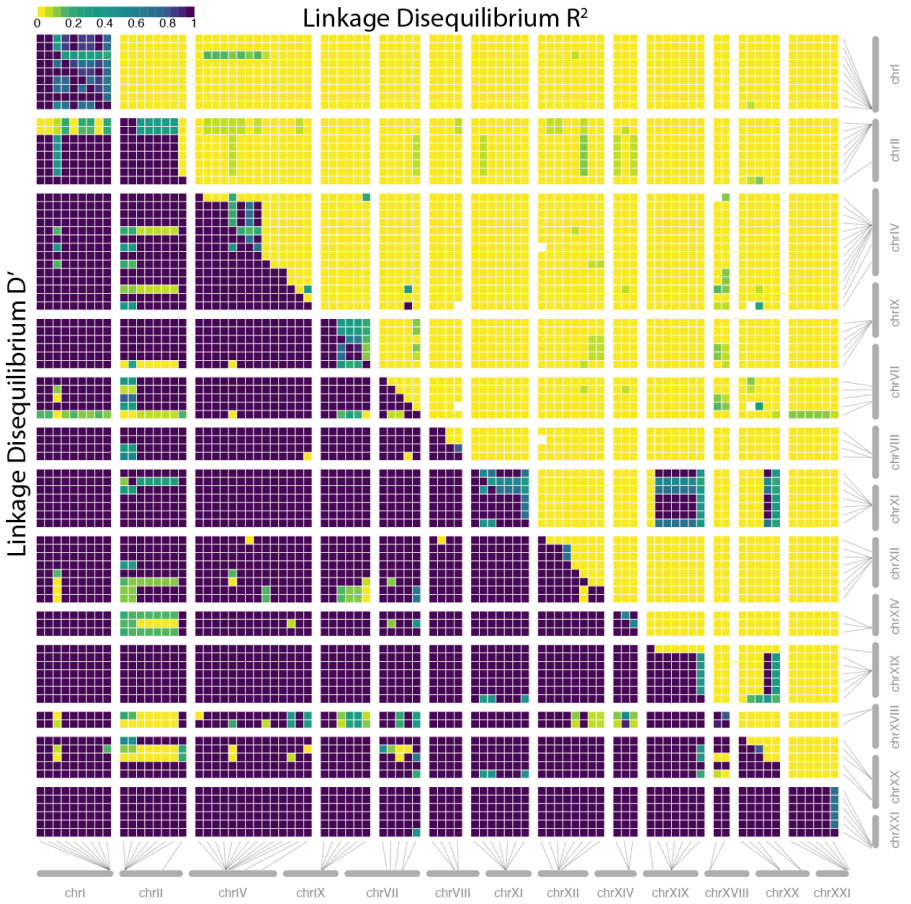


Figure S14: Estimates of linkage disequilibrium among adaptive alleles across the genome of Rabbit Slough fish. Upper triangle shows r^2, while lower triangle shows D'. SNPs that were fixed for one allele in the focal population were excluded from the analysis. Darker colors indicate strong LD.

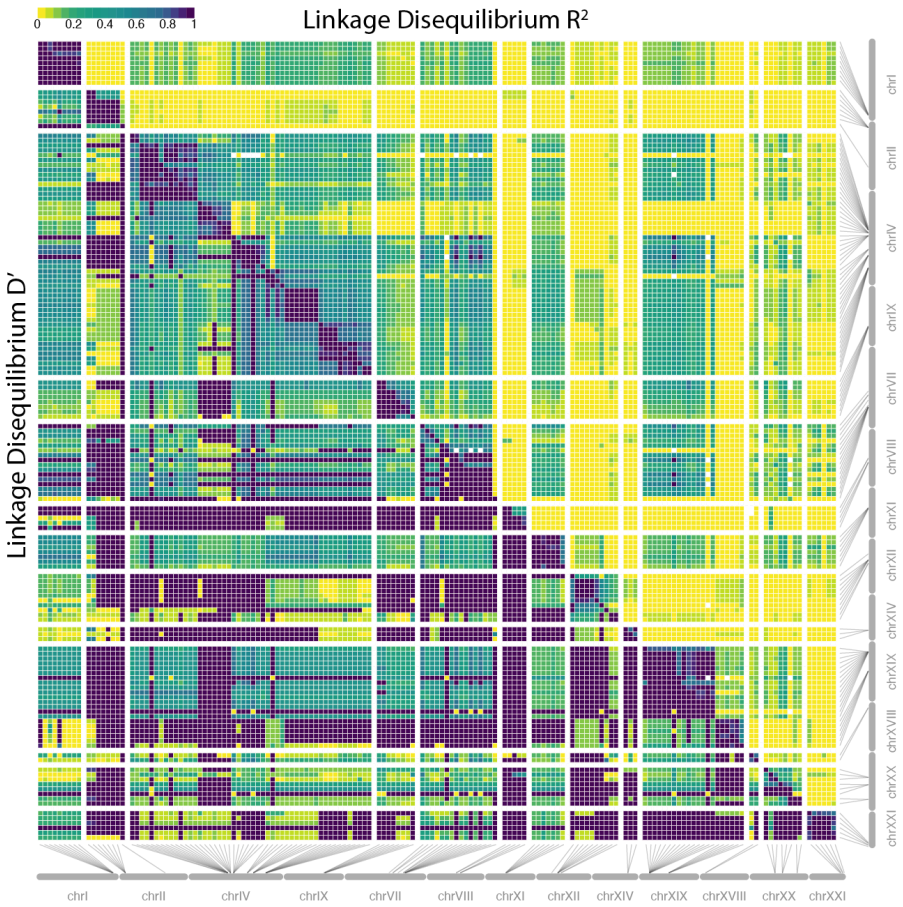


Figure S15: Estimates of linkage disequilibrium among adaptive alleles across the genome of Resurrection Bay fish. Upper triangle shows r^2, while lower triangle shows D'. SNPs that were fixed for one allele in the focal population were excluded from the analysis. Darker colors indicate strong LD.

Figure S16: Estimates of linkage disequilibrium among adaptive alleles across the genome of Glacier Spit fish. Upper triangle shows r^2, while lower triangle shows D'. SNPs that were fixed for one allele in the focal population were excluded from the analysis. Darker colors indicate strong LD.

REFERENCES AND NOTES

1. C. Darwin, *Fertilisation of Orchids* (Murray, 1862).

2. A. R. Wallace, Creation by Law. *Q. J. Sci.* **4**, 470–488 (1867).

3. N. I. Vavilov, The law of homologous series in variation. *J. Genet.* **12**, 47–89 (1922).

4. S. J. Gould, *Wonderful Life* (Norton, 1989).

5. B. H. Good, M. J. McDonald, J. E. Barrick, R. E. Lenski, M. M. Desai, The dynamics of molecular evolution over 60,000 generations. *Nature* **551**, 45–50 (2017).

6. D. Blank, L. Wolf, M. Ackermann, O. K. Silander, The predictability of molecular evolution during functional innovation. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 3044–3049 (2014).

7. Z. D. Blount, R. E. Lenski, J. B. Losos, Contingency and determinism in evolution: Replaying life's tape. *Science* **362**, eaam5979 (2018).

8. F. C. Jones, M. G. Grabherr, Y. F. Chan, P. Russell, E. Mauceli, J. Johnson, R. Swofford, M. Pirun, M. C. Zody, S. White, E. Birney, S. Searle, J. Schmutz, J. Grimwood, M. C. Dickson, R. M. Myers, C. T. Miller, B. R. Summers, A. K. Knecht, S. D. Brady, H. Zhang, A. A. Pollen, T. Howes, C. Amemiya; Broad Institute Genome Sequencing Platform & Whole Genome Assembly Team, E. S. Lander, F. D. Palma, K. Lindblad-Toh, D. M. Kingsley, The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**, 55–61 (2012).

9. Y.-T. Lai, C. K. L. Yeung, K. E. Omland, E.-L. Pang, Y. Hao, B.-Y. Liao, H.-F. Cao, B.-W. Zhang, C.-F. Yeh, C.-M. Hung, H.-Y. Hung, M.-Y. Yang, W. Liang, Y.-C. Hsu, C.-T. Yao, L. Dong, K. Lin, S.-H. Li, Standing genetic variation as the predominant source for adaptation of a songbird. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 2152–2157 (2019).

10. Y.-H. E. Loh, E. Bezault, F. M. Muenzel, R. B. Roberts, R. Swofford, M. Barluenga, C. E. Kidd, A. E. Howe, F. Di Palma, K. Lindblad-Toh, J. Hey, O. Seehausen, W. Salzburger, T. D. Kocher, J. T. Streelman, Origins of shared genetic variation in African cichlids. *Mol. Biol. Evol.* **30**, 906–917 (2013).

11. P. F. Colosimo, K. E. Hosemann, S. Balabhadra, G. Villarreal Jr., M. Dickson, J. Grimwood, J. Schmutz, R. Myers, D. Schluter, D. M. Kingsley, Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. *Science* **307**, 1928–1933 (2005).

12. R. D. H. Barrett, D. Schluter, Adaptation from standing genetic variation. *Trends Ecol. Evol.* **23**, 38–44 (2008).

13. J. A. Endler, *Natural Selection in the Wild* (Princeton Univ. Press, 1986).

14. J. G. Kingsolver, H. E. Hoekstra, J. M. Hoekstra, D. Berrigan, S. N. Vignieri, C. E. Hill, A. Hoang, P. Gibert, P. Beerli, The strength of phenotypic selection in natural populations. *Am. Nat.* **157**, 245–261 (2001).

15. D. A. Reznick, H. Bryga, J. A. Endler, Experimentally induced life-history evolution in a natural population. *Nature* **346**, 357–359 (1990).

16. P. R. Grant, B. R. Grant, Unpredictable evolution in a 30-year study of Darwin's finches. *Science* **296**, 707–711 (2002).

17. M. A. Bell, W. E. Aguirre, N. J. Buck, Twelve years of contemporary armor evolution in a threespine stickleback population. *Evolution* **58**, 814–824 (2004).

18. D. Schluter, K. B. Marchinko, M. E. Arnegard, H. Zhang, S. D. Brady, F. C. Jones, M. A. Bell, D. M. Kingsley, Fitness maps to a large-effect locus in introduced stickleback populations. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e1914889118 (2021).

19. A. O. Bergland, E. L. Behrman, K. R. O'Brien, P. S. Schmidt, D. A. Petrov, Genomic evidence of rapid and stable adaptive oscillations over seasonal time scales in *Drosophila*. *PLOS Genet.* **10**, e1004775 (2014).

20. S. F. Levy, J. R. Blundell, S. Venkataram, D. A. Petrov, D. S. Fisher, G. Sherlock, Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature* **519**, 181–186 (2015).

21. P. Nosil, R. Villoutreix, C. F. de Carvalho, T. E. Farkas, V. Soria-Carrasco, J. L. Feder, B. J. Crespi, Z. Gompert, Natural selection and the predictability of evolution in *Timema* stick insects. *Science* **359**, 765–770 (2018).

22. J. T. Anderson, C.-R. Lee, C. A. Rushworth, R. I. Colautti, T. Mitchell-Olds, Genetic trade-offs and conditional neutrality contribute to local adaptation. *Mol. Ecol.* **22**, 699–708 (2013).

23. N. O. Therkildsen, A. P. Wilder, D. O. Conover, S. B. Munch, H. Baumann, S. R. Palumbi, Contrasting genomic shifts underlie parallel phenotypic evolution in response to fishing. *Science* **365**, 487–490 (2019).

24. R. D. H. Barrett, S. Laurent, R. Mallarino, S. P. Pfeifer, C. C. Y. Xu, M. Foll, K. Wakamatsu, J. S. Duke-Cohan, J. D. Jensen, H. E. Hoekstra, Linking a mutation to survival in wild mice. *Science* **363**, 499–504 (2019).

25. T. J. Thurman, R. D. H. Barrett, The genetic consequences of selection in natural populations. *Mol. Ecol.* **25**, 1429–1448 (2016).

26. M. A. Bell, W. E. Aguirre, Contemporary evolution, allelic recycling, & adaptive radiation of the threespine stickleback. *Evol. Ecol. Res.* **15**, 377–411 (2013).

27. M. A. Bell, S.A. Foster, Ed., *The Evolutionary Biology of the Threespine Stickleback* (Oxford Univ. Press, 1994).

28. B. Fang, J. Merilä, M. Matschiner, P. Momigliano, Estimating uncertainty in divergence times among three-spined stickleback clades using the multispecies coalescent. *Mol. Phylogenet. Evol.* **142**, 106646 (2020).

29. D. Schluter, G. L. Conte, Genetics and ecological speciation. *Proc. Natl. Acad. Sci. U.S.A.* **106** Suppl. 1, 9955–9962 (2009).

30. P. A. Hohenlohe, S. Bassham, P. D. Etter, N. Stiffler, E. A. Johnson, W. A. Cresko, Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLOS Genet.* **6**, e1000862 (2010).

31. F. C. Jones, Y. F. Chan, J. Schmutz, J. Grimwood, S. D. Brady, A. M. Southwick, D. M. Absher, R. M. Myers, T. E. Reimchen, B. E. Deagle, D. Schluter, D. M. Kingsley, A genome-wide SNP genotyping array reveals patterns of global and repeated species-pair divergence in sticklebacks. *Curr. Biol.* **22**, 83–90 (2012).

32. B. E. Deagle, F. C. Jones, D. M. Absher, D. M. Kingsley, T. E. Reimchen, Phylogeography and adaptation genetics of stickleback from the Haida Gwaii archipelago revealed using genome-wide single nucleotide polymorphism genotyping. *Mol. Ecol.* **22**, 1917–1932 (2013).

33. M. Roesti, S. Gavrilets, A. P. Hendry, W. Salzburger, D. Berner, The genomic signature of parallel adaptation from shared genetic variation. *Mol. Ecol.* **23**, 3944–3956 (2014).

34. T. C. Nelson, W. A. Cresko, Ancient genomic variation underlies repeated ecological adaptation in young stickleback populations. *Evol. Lett.* **2**, 9–21 (2018).

35. B. Fang, P. Kemppainen, P. Momigliano, X. Feng, J. Merilä, On the causes of geographically heterogeneous parallel evolution in sticklebacks. *Nat. Ecol. Evol.* **4**, 1105–1115 (2020).

36. E. A. Lescak, S. L. Bassham, J. Catchen, O. Gelmond, M. L. Sherbick, F. A. von Hippel, W. A. Cresko, Evolution of stickleback in 50 years on earthquake-uplifted islands. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E7204–E7212 (2015).

37. N. V. Terekhanova, M. D. Logacheva, A. A. Penin, T. V. Neretina, A. E. Barmintseva, G. A. Bazykin, A. S. Kondrashov, N. S. Mugue, Fast evolution from precast bricks: genomics of young freshwater populations of threespine stickleback *Gasterosteus aculeatus*. *PLOS Genet.* **10**, e1004696 (2014).

38. S. Bassham, J. Catchen, E. Lescak, F. A. von Hippel, W. A. Cresko, Repeated selection of alternatively adapted haplotypes creates sweeping genomic remodeling in stickleback. *Genetics* **209**, 921–939 (2018).

39. W. E. Aguirre, M. A. Bell, Twenty years of body shape evolution in a threespine stickleback population adapting to a lake environment. *Biol. J. Linn. Soc. Lond.* **105**, 817–831 (2012).

40. A. Garcia-Elfring, A. Paccard, T. J. Thurman, B. A. Wasserman, E. P. Palkovacs, A. P. Hendry, R. D. H. Barrett, Using seasonal genomic changes to understand historical adaptation to new environments: Parallel selection on stickleback in highly-variable estuaries. *Mol. Ecol.* **30**, 2054–2064 (2021).

41. D. A. Marques, F. C. Jones, F. Di Palma, D. M. Kingsley, T. E. Reimchen, Experimental evidence for rapid genomic adaptation to a new niche in an adaptive radiation. *Nat. Ecol. Evol.* **2**, 1128–1138 (2018).

42. R. D. H. Barrett, S. M. Rogers, D. Schluter, Natural selection on a major armor gene in threespine stickleback. *Science* **322**, 255–257 (2008).

43. T. G. Laurentino, D. Moser, M. Roesti, M. Ammann, A. Frey, F. Ronco, B. Kueng, D. Berner, Genomic release-recapture experiment in the wild reveals within-generation polygenic selection in stickleback fish. *Nat. Commun.* **11**, 1928 (2020).

44. N. Yu, F.-C. Chen, S. Ota, L. B. Jorde, P. Pamilo, L. Patthy, M. Ramsay, T. Jenkins, S.-K. Shyue, W.-H. Li, Larger genetic differences within Africans than between Africans and Eurasians. *Genetics* **161**, 269–274 (2002).

45. D. Schluter, E. A. Clifford, M. Nemethy, J. S. McKinnon, Parallel evolution and inheritance of quantitative traits. *Am. Nat.* **163**, 809–822 (2004).

46. J. A. Baker, D. C. Heins, J. E. Baum, Trajectory and rate of change in female life-history traits following colonization of a freshwater, lacustrine environment by oceanic threespine stickleback. *Evol. Ecol. Res.* **20**, 247–263 (2019).

47. M. A. Bell, D. C. Heins, M. A. Wund, F. A. von Hippel, R. Massengill, K. Dunker, G. A. Bristow, W. E. Aguirre, Reintroduction of threespine stickleback into Cheney and Scout Lakes, Alaska. *Evol. Ecol. Res.* **17**, 157–178 (2016).

48. K. Spitzer, M. Pelizzola, A. Futschik, Modifying the Chi-square and the CMH test for population genetic inference: Adapting to overdispersion. *Ann. Apl. Stats.* **14**, 202–220 (2020).

49. N. M. O'Brown, B. R. Summers, F. C. Jones, S. D. Brady, D. M. Kingsley, A recurrent regulatory change underlying altered expression and Wnt response of the stickleback armor plates gene EDA. *eLife* **4**, e05290 (2015).

50. A. P. McGregor, V. Orgogozo, I. Delon, J. Zanet, D. G. Srinivasan, F. Payre, D. L. Stern, Morphological evolution through multiple *cis*-regulatory mutations at a single gene. *Nature* **448**, 587–590 (2007).

51. S. Koshikawa, M. W. Giorgianni, K. Vaccaro, V. A. Kassner, J. H. Yoder, T. Werner, S. B. Carroll, Gain of *cis*-regulatory activities underlies novel domains of wingless gene expression in *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 7524–7529 (2015).

52. L. F. Stam, C. C. Laurie, Molecular dissection of a major gene effect on a quantitative trait: The level of alcohol dehydrogenase expression in *Drosophila melanogaster*. *Genetics* **144**, 1559–1564 (1996).

53. C. L. Peichel, D. A. Marques, The genetic and molecular architecture of phenotypic diversity in sticklebacks. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **372**, 20150486 (2017).

54. C. T. Miller, A. M. Glazer, B. R. Summers, B. K. Blackman, A. R. Norman, M. D. Shapiro, B. L. Cole, C. L. Peichel, D. Schluter, D. M. Kingsley, Modular skeletal evolution in sticklebacks is controlled by additive and clustered quantitative trait loci. *Genetics* **197**, 405–420 (2014).

55. A. H. Chan, P. A. Jenkins, Y. S. Song, Genome-wide fine-scale recombination rate variation in *Drosophila melanogaster*. *PLOS Genet.* **8**, e1003090 (2012).

56. J. Galloway, W. A. Cresko, P. Ralph, A few stickleback suffice for the transport of alleles to new lakes. *G3* **10**, 505–514 (2020).

57. F. Han, S. Lamichhaney, B. R. Grant, P. R. Grant, L. Andersson, M. T. Webster, Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among Darwin's finches. *Genome Res.* **27**, 1004–1015 (2017).

58. J. I. Meier, D. A. Marques, C. E. Wagner, L. Excoffier, O. Seehausen, Genomics of parallel ecological speciation in Lake Victoria cichlids. *Mol. Biol. Evol.* **35**, 1489–1506 (2018).

59. W. Mei, M. G. Stetter, D. J. Gates, M. C. Stitzer, J. Ross-Ibarra, Adaptation in plant genomes: Bigger is different. *Am. J. Bot.* **105**, 16–19 (2018).

60. R. D. Hernandez, J. L. Kelley, E. Elyashiv, S. C. Melton, A. Auton, G. McVean; 1000 Genomes Project, G. Sella, M. Przeworski, Classic selective sweeps were rare in recent human evolution. *Science* **331**, 920–924 (2011).

61. C. L. Peichel, K. S. Nereng, K. A. Ohgi, B. L. E. Cole, P. F. Colosimo, C. A. Buerkle, D. Schluter, D. M. Kingsley, The genetic architecture of divergence between threespine stickleback species. *Nature* **414**, 901–905 (2001).

62. C. L. Peichel, S. T. Sullivan, I. Liachko, M. A. White, Improvement of the threespine stickleback genome using a Hi-C-based proximity-guided assembly. *J. Hered.* **108**, 693–700 (2017).

63. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

64. M. A. DePristo, E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, G. del Angel, M. A. Rivas, M. Hanna, A. McKenna, T. J. Fennell, A. M. Kernytsky, A. Y. Sivachenko, K. Cibulskis, S. B. Gabriel, D. Altshuler, M. J. Daly, A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).

65. M. Lynch, D. Bost, S. Wilson, T. Maruki, S. Harrison, Population-genetic inference from pooled-sequencing data. *Genome Biol. Evol.* **6**, 1210–1218 (2014).

66. R. Kofler, R. V. Pandey, C. Schlötterer, PoPoolation2: Identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics* **27**, 3435–3436 (2011).

67. A. F. Shanfelter, S. Archambeault, M. A. White, Divergent fine-scale recombination landscapes between a freshwater and marine population of threespine stickleback fish. *Genome Biol. Evol.* **11**, 1573–1585 (2019).

68. A. M. Glazer, E. E. Killingbeck, T. Mitros, D. S. Rokhsar, C. T. Miller, Genome assembly improvement and mapping convergently evolved skeletal traits in sticklebacks with genotyping-by-sequencing. *G3* **5**, 1463–1472 (2015).

69. G. Bhatia, N. Patterson, S. Sankararaman, A. L. Price, Estimating and interpreting FST: The impact of rare variants. *Genome Res.* **23**, 1514–1521 (2013).

70. T. Taus, A. Futschik, C. Schlötterer, Quantifying selection with pool-seq time series data. *Mol. Biol. Evol.* **34**, 3023–3034 (2017).

71. W. Aguirre, P. Doherty, M. Bell, Genetics of lateral plate and Gillraker phenotypes in a rapidly evolving population of threespine stickleback. *Behaviour* **141**, 1465–1483 (2004).

72. A. Y. K. Albert, S. Sawaya, T. H. Vines, A. K. Knecht, C. T. Miller, B. R. Summers, S. Balabhadra, D. M. Kingsley, D. Schluter, The genetics of adaptive shape shift in stickleback: Pleiotropy and effect size. *Evolution* **62**, 76–85 (2008).

73. S. Arif, W. E. Aguirre, M. A. Bell, Evolutionary diversification of opercle shape in Cook Inlet threespine stickleback. *Biol. J. Linn. Soc. Lond.* **97**, 832–844 (2009).

74. W. E. Aguirre, K. E. Ellis, M. Kusenda, M. A. Bell, Phenotypic variation and sexual dimorphism in anadromous threespine stickleback: Implications for postglacial adaptive radiation. *Biol. J. Linn. Soc. Lond.* **95**, 465–478 (2008).

75. M. D. Shapiro, M. E. Marks, C. L. Peichel, B. K. Blackman, K. S. Nereng, B. Jónsson, D. Schluter, D. M. Kingsley, Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* **428**, 717–723 (2004).

76. Y. F. Chan, M. E. Marks, F. C. Jones, G. Villarreal Jr., M. D. Shapiro, S. D. Brady, A. M. Southwick, D. M. Absher, J. Grimwood, J. Schmutz, R. M. Myers, D. Petrov, B. Jónsson, D.

Schluter, M. A. Bell, D. M. Kingsley, Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. *Science* **327**, 302–305 (2010).

77. A. S. Hinrichs, D. Karolchik, R. Baertsch, G. P. Barber, G. Bejerano, H. Clawson, M. Diekhans, T. S. Furey, R. A. Harte, F. Hsu, J. Hillman-Jackson, R. M. Kuhn, J. S. Pedersen, A. Pohl, B. J. Raney, K. R. Rosenbloom, A. Siepel, K. E. Smith, C. W. Sugnet, A. Sultan-Qurraie, D. J. Thomas, H. Trumbower, R. J. Weber, M. Weirauch, A. S. Zweig, D. Haussler, W. J. Kent, The UCSC Genome Browser Database: Update 2006. *Nucleic Acids Res.* **34**, D590–D598 (2006).

78. W. J. Kent, R. Baertsch, A. Hinrichs, W. Miller, D. Haussler, Evolution's cauldron: Duplication, deletion, and rearrangement in the mouse and human genomes. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 11484–11489 (2003).

79. W. J. Kent, BLAT--the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).

80. B. L. Aken, S. Ayling, D. Barrell, L. Clarke, V. Curwen, S. Fairley, J. Fernandez Banet, K. Billis, C. García Girón, T. Hourlier, K. Howe, A. Kähäri, F. Kokocinski, F. J. Martin, D. N. Murphy, R. Nag, M. Ruffier, M. Schuster, Y. A. Tang, J.-H. Vogel, S. White, A. Zadissa, P. Flicek, S. M. J. Searle, The Ensembl gene annotation system. *Database*. **2016**, baw093 (2016).

81. M. Schubert, S. Lindgreen, L. Orlando, AdapterRemoval v2: Rapid adapter trimming, identification, and read merging. *BMC. Res. Notes* **9**, 88 (2016).

82. D. W. Hagen, Isolating mechanisms in threespine sticklebacks (*Gasterosteus*). *J. Fish. Res. Board Can.* **24**, 1637–1692 (1967).

83. K. Yoshida, T. Makino, K. Yamaguchi, S. Shigenobu, M. Hasebe, M. Kawata, M. Kume, S. Mori, C. L. Peichel, A. Toyoda, A. Fujiyama, J. Kitano, Sex chromosome turnover contributes to genomic divergence between incipient stickleback species. *PLOS Genet.* **10**, e1004223 (2014).

84. M. Higuchi, A. Goto, Genetic evidence supporting the existence of two distinct species in the genus *Gasterosteus* around Japan. *Environ. Biol. Fishes* **47**, 1–16 (1996).

85. S. Varadharajan, P. Rastas, A. Löytynoja, M. Matschiner, F. C. F. Calboli, B. Guo, A. J. Nederbragt, K. S. Jakobsen, J. Merilä, A high-quality assembly of the nine-spined stickleback (*Pungitius pungitius*) genome. *Genome Biol. Evol.* **11**, 3291–3308 (2019).

86. M. Kirkpatrick, N. Barton, Chromosome inversions, local adaptation and speciation. *Genetics* **173**, 419–434 (2006).

87. S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, P. C. Sham, PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

88. B. S. Gaut, A. D. Long, The lowdown on linkage disequilibrium. *Plant Cell* **15**, 1502–1506 (2003).

89. R. C. Lewontin, The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics* **49**, 49–67 (1964).

90. K. Osoegawa, K. C. Mallempati, S. Gangavarapu, A. Oki, K. Gendzekhadze, S. R. Marino, N. K. Brown, M. P. Bettinotti, E. T. Weimer, G. Montero-Martín, L. E. Creary, T. A. Vayntrub, C.-J. Chang, M. Askar, S. J. Mack, M. A. Fernández-Viña, HLA alleles and haplotypes observed in 263 US families. *Hum. Immunol.* **80**, 644–660 (2019).

91. P. A. Hohenlohe, S. Bassham, M. Currey, W. A. Cresko, Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **367**, 395–408 (2012).

92. C. C. Chang, C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, J. J. Lee, Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).

93. S. Purcell, C. Chang, *PLINK*; www.cog-genomics.org/plink/1.9/.

94. N. Patterson, A. L. Price, D. Reich, Population structure and eigenanalysis. *PLOS Genet.* **2**, e190 (2006).

Title

Adaptive standing genetic variation from disparate sources provides a substrate for rapid adaptation via hard and soft sweeps

Authors

Kirch M[1], Avdievich E[1], Venkataramani K[1], Kucka M[1], Thome R[1], Philippos G[1], Gaspar L[1], Peichel C[2], Archambeault S[2], Bell M[3], Veeramah K[4], Kingsley DM[5,6], Jones FC[1]

Affiliations

1. Friedrich Miescher Laboratory of the Max Planck Society, Tübingen, Germany.
2. Institute of Evolutionary Ecology, University of Bern, University of Bern, Switzerland.
3. Museum of Paleontology, University of California, Berkeley, CA, USA
4. Department of Ecology and Evolution, Stony Brook University, New York USA
5. Department of Developmental Biology, School of Medicine, Stanford University, CA USA
6. Howard Hughes Medical Institute, Chevy Chase, MD, USA

Abstract

A key goal in evolutionary biology is to better understand the mechanisms by which a population rapidly adapts to a new or changing environment. Compared to *de novo* mutations, standing genetic variation (SGV) provides a ready-available substrate for rapid adaptation and population genetics theory predicts that adaptation via SGV outrivals newly arising mutations in most cases. Despite its crucial role in rapid adaptation, the availability and maintenance of SGV is largely understudied. Several obstacles have to be overcome for characterising SGV: the polygenic basis of adaptation to a new environment needs to be known in order to assess the SGV available at these loci in the founder population; in addition, large sample sizes are necessitated to detect SGV present at low frequencies. Threespine stickleback fish (*Gasterosteus aculeatus*) offer an ideal system to tackle these challenges as the genomic basis of adaptation in sticklebacks was previously identified and large sample sizes are accessible (Jones *et al.*, 2012). Sticklebacks have repeatedly adapted to divergent marine and freshwater environments in the face of ongoing gene-flow and parallel adaptation to freshwater environments happens via reuse of freshwater-adaptive SGV available in marine populations at low frequencies. Targeted enrichment sequencing at individual resolution was performed for more than 9 000 marine threespine sticklebacks to characterise SGV in a large empirical data set. This individual sequencing data shows that freshwater-adaptive variants are present at low frequencies in two marine populations with frequency distributions differing significantly between the two sampled stickleback populations from Alaska and Washington. Freshwater-adaptive variants are present as multigeneration introgressed alleles and each individual marine fish “carries” multiple freshwater-adaptive alleles across the genome, potentially facilitating future rapid freshwater adaptation. By comparing our data to genomic datasets of globally-diverse sticklebacks we infer that freshwater-adaptive haplotypes cluster with North-American, European and Californian freshwater individuals suggesting that SGV is introduced by multiple, disparate sources into the marine population. Additional whole genome sequencing data revealed patterns of soft as well as hard sweeps in an Alaskan freshwater population. Combined, we present insight into the origins, availability, and maintenance of SGV and the evolutionary potential of biodiversity.

Introduction

Standing genetic variation (SGV) is likely to play a major role in the survival of species in response to current anthropogenic climate change. Aside from *de novo* mutations, adaptation from SGV represents one of the key mechanisms for rapid, adaptive evolution. Yet despite this pivotal role, we have relatively poor understanding of factors that shape the availability of SGV and thus facilitate or constrain rates of future adaptation.

Adaptation via SGV may outrival newly arising mutations for several reasons (Barrett & Schluter, 2008; Reid *et al.*, 2021). Firstly, SGV is already present as multiple copies in the population whereas *de novo* mutations are present as a single copy. This increases the likelihood of fixation for pre-existing variation (Hermisson & Pennings, 2005). Secondly, being present as multiple copies, SGV can rise to high frequencies quicker (Hermisson & Pennings, 2005). Additionally, adaptation via SGV bypasses the time waiting for a new beneficial mutation to arise (Reid *et al.*, 2021). Last but not least, SGV can be pre-tested and optimised over time (Reid *et al.*, 2021), thus providing potentially fine-tuned haplotypes for rapid adaptation. A substantial role for SGV in facilitating rapid adaptation has been demonstrated in diverse organisms (McGee *et al.*, 2020; Wooldridge *et al.*, 2022; Konečná *et al.*, 2021).

Understanding of how adaptive SGV is spread and maintained can provide insight into the evolutionary and demographic processes that promote or constrain future adaptive potential of a population. Beyond case studies of a handful of large-phenotypic effect loci like the Eda locus in stickleback fish (Colosimo *et al.*, 2005), knowledge of SGV relevant for future adaptation is sparse. For instance, the frequencies of adaptive-SGV have only been empirically quantified at a few selected loci leaving us in the dark about the frequencies of SGV for polygenic adaptation. How does the availability of polygenic adaptive SGV within and among individuals of a founding population influence rates of adaptation? Are adaptive alleles carried into a new environment separately by several individuals and re-assembled through recombination over subsequent generations, or alternatively does rapid adaptation occur primarily via selection on alleles carried in linkage by a single or small number of founders? As adaptive SGV introgresses through populations in divergent environments and subject to different selection pressures, is it deleterious or costly for an individual (and population) to carry (and maintain) divergently adaptive SGV?

Despite its likely importance in rapid adaptive evolution, SGV remains largely understudied for a number of challenging reasons: firstly, SGV underlying future adaptive evolutionary change are rarely known in advance, making it difficult to know which loci to study; secondly, SGV mediating rapid adaptive evolution can be present at very low frequencies in a population, and quantifying their abundance and linkage to other adaptive alleles necessitates very large sample sizes. In order to understand how adaptive-SGV facilitates rapid adaptation

we therefore need detailed knowledge of its availability including the frequencies of adaptive alleles and of the linkage disequilibrium among them in a system in which the loci most likely underlying future adaptation events are predictable and can be studied in large numbers of individuals (Kingman *et al.*, 2021).

With a high-resolution map of genomic loci underlying repeated and parallel adaptive divergence among marine and freshwater habitats (Jones *et al.*, 2012), and the ability to study large numbers of individuals, threespine stickleback (*Gasterosteus aculeatus*) represent a well poised system for studying the nature of SGV underlying rapid adaptive change. Since the end of the Last Glacial Maximum (~10 000 - 20 000 years ago), anadromous (hereafter called marine) sticklebacks colonised freshwater environments repeatedly and independently across the Northern Hemisphere providing powerful biological replicates of the evolutionary process (Bell & Foster, 1994). Marine fish experience high selective forces (selection coefficients up to 0.5 at some loci (Barrett & Schluter, 2008; Kingman *et al.*, 2021)) and can adapt to new freshwater environments within decades (Bell *et al.*, 2004; Lescak *et al.*, 2015; Terekhanova *et al.*, 2014; Bassham *et al.*, 2018). Jones *et al.* identified a set of 242 adaptive loci (FDR 0.05) underlying this global parallel adaptation, covering 2.4 Mb (0.5%) of the stickleback genome (Jones *et al.*, 2012). This set of loci includes the well-studied Ectodysplasin locus (Eda) underlying armour plate differences in stickleback ecotypes (Colosimo *et al.*, 2005) and subject to strong selection in the wild (Schluter *et al.*, 2021).

Previous studies have shown that adaptation via reuse of ancient SGV (e.g. up to two million years old at the Eda locus), plays a major role in adaptation of sticklebacks to freshwater habitats (Colosimo *et al.*, 2005; Colosimo *et al.*, 2004; Cresko *et al.*, 2004; Miller *et al.*, 2007; Jones *et al.*, 2012; Terekhanova *et al.*, 2014; Kingman *et al.*, 2021). At numerous loci across the genome freshwater haplotypes are used in parallel for adaptation to independent freshwater populations at disparate geographic locations (Colosimo *et al.*, 2005; Jones *et al.*, 2012). At those same loci, highly divergent marine haplotypes are found in marine sticklebacks, suggesting divergent selection pressures favour the maintenance of divergent haplotypes in the respective marine and freshwater environments. Given freshwater sticklebacks have “resident” life history and lack migratory behavior (Moyle, 1976; Wootton, 1984), these findings pose the question of how freshwater-adaptive alleles can be transported from one freshwater population to another when they are only connected through a marine environment. The most likely mechanism is through hybridization and introgression with marine ecotypes - alleles are “transported” and spread through local environments with opposing selection pressures via introgression (Jones *et al.*, 2006), termed as the “transporter hypothesis” by Schluter & Conte (2009; Figure 1A). Consistent with this, freshwater-adaptive alleles are expected to be present in marine populations at low frequencies (for example the freshwater haplotype at the Eda locus was found at a frequency of 0.038 in 109 Californian

marine sticklebacks, Colosimo *et al.*, 2005). Combined with their anadromous life history (migration from the ocean to breed in rockpools, estuaries and lower reaches of rivers) marine stickleback populations represent an important source of freshwater-adaptive SGV in newly formed freshwater habitats.

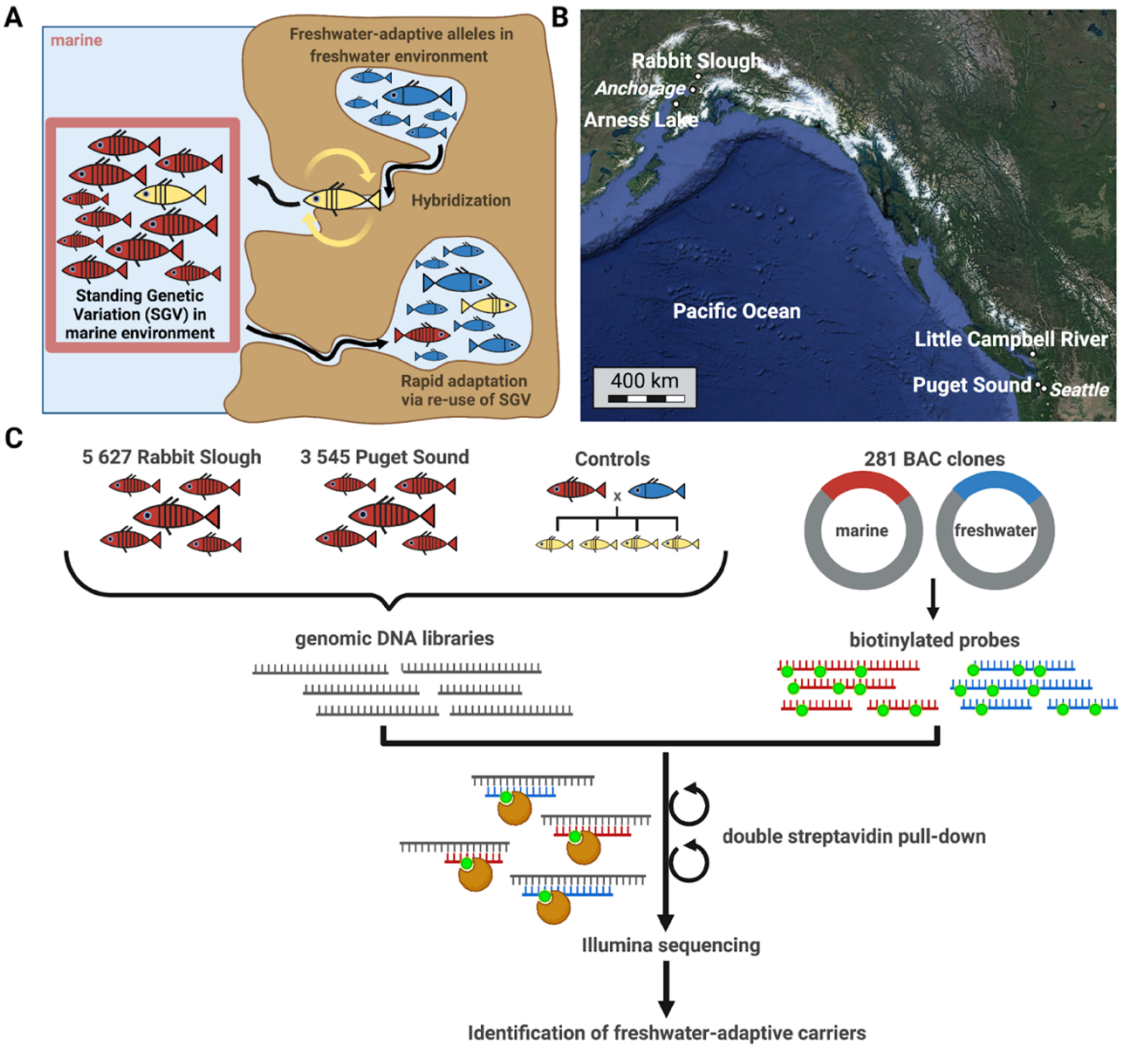


Figure 1: (A) Freshwater-adaptive standing genetic variation is present and spread by marine populations due to hybridisation and introgression in marine-freshwater hybrid zones. Despite divergence between ecotypes being maintained by a selection-migration balance at adaptive loci, freshwater-adaptive alleles are found in the marine population at low frequencies as standing genetic variation. When sticklebacks from the marine population colonise a new freshwater habitat, already present "pre-screened" adaptive variation can be re-used for rapid adaptation. (B) Marine samples were sampled at Rabbit Slough (RABS) and Puget Sound (PGTS). Parents from the control nuclear family were wild-derived strains from Little Campbell River and freshwater samples were sampled at Arness Lake. (Details see Table S1). (C) Genomic libraries were prepared for marine samples from RABS, PGTS as well as the control samples. Previously identified divergent loci were pulled down with biotinylated BAC probes and subsequently sequenced with Illumina (Jones *et al.*, 2012). (see section S1.3).

To assess the constraints on availability and maintenance of polygenic SGV used in rapid adaptation of sticklebacks to freshwater habitats, we focused on a previously identified subset of the genome involved in parallel adaptation (Jones *et al.*, 2012) and collected pulled-down sequencing data at the level of individuals for more than 9 000 marine sticklebacks.

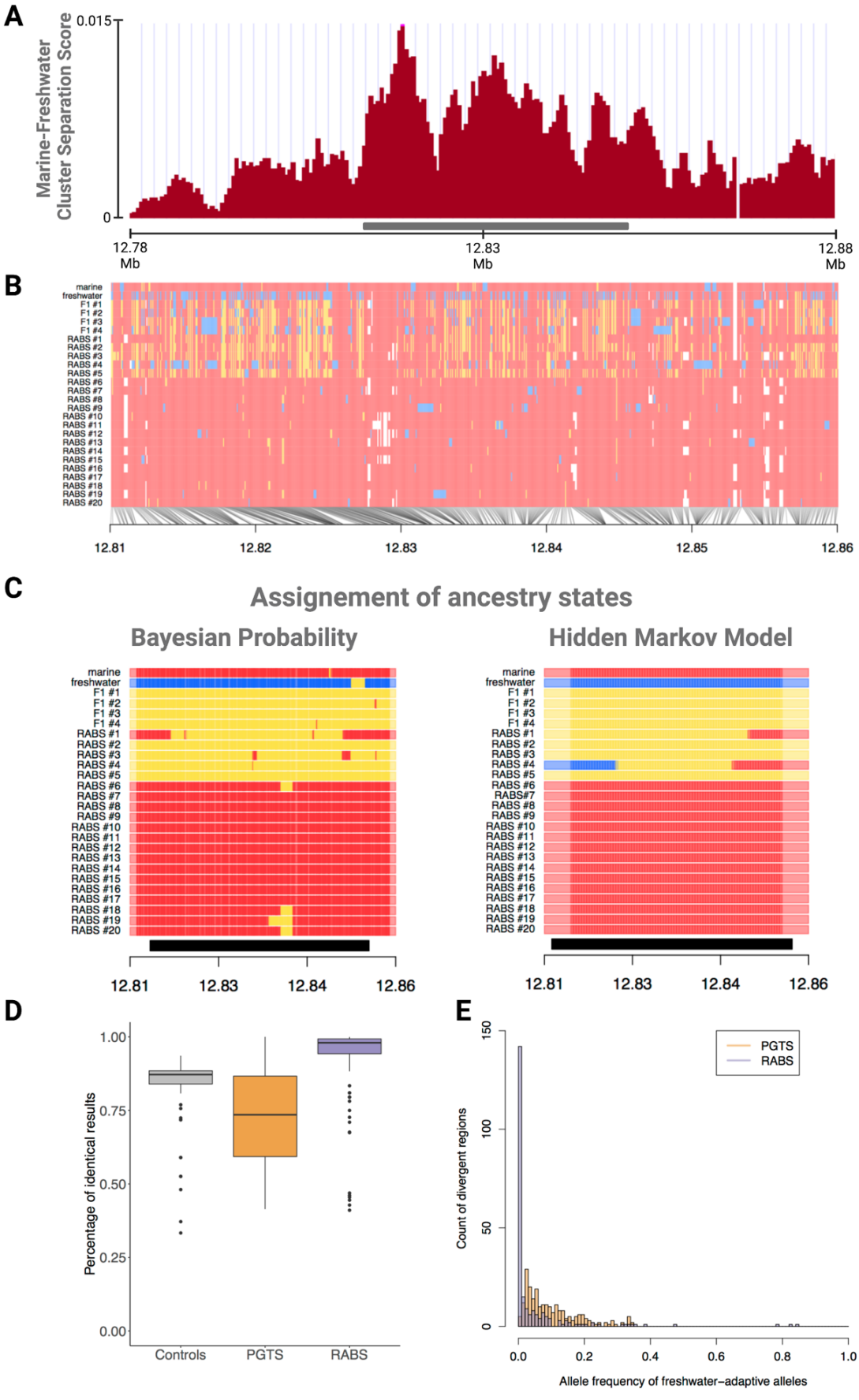
Results

More than 9 000 marine threespine stickleback fish were collected from two different sites in North America: Puget Sound (PGTS) in Washington and Rabbit Slough (RABS) in Alaska (Figure 1, Table S1). Targeted enrichment short-read sequencing was used to collect partial genomic data for 3 273 PGTS and 5 490 RABS samples at 240 previously identified adaptive regions (Jones *et al.* 2012, FDR 0.05) and flanking sequence as well as 10 randomly selected regions covering in total 26.7 Mb (5.8%) of the genome with a mean read depth coverage of 4.6+/-1.9 SD (Table S2, S3). At each locus and for each fish, the genomic data was "segmented" into freshwater, marine, and heterozygous ancestry states using two different statistical approaches (Figure 2).

Firstly, since there is deep evolutionary divergence times between marine and freshwater haplotypes at parallel adaptive loci (Colosimo *et al.*, 2005; Jones *et al.*, 2012) we used the density of heterozygous sites and derived alleles in sliding windows of 10kb as input for a Hidden Markov Model (HMM) that segmented the data into "homozygous marine", "heterozygous" or "freshwater" states (see section S1.5.2). The emission values for these states were trained using genomic data of parents and offspring from a control family (marine x freshwater), and transition probabilities were proportional to genome-wide recombination rate. This approach is agnostic of the specific alleles involved, except for their ancestral (defined as the most frequent allele in the marine population) or derived (defined as non-ancestral allele) state, and has the potential to assign homozygous-marine, -freshwater and heterozygous states to previously unstudied genomic regions. Secondly, we used *a priori* defined ecotype-specific single nucleotide polymorphism (SNPs) based on a Canadian training data set in a Bayesian Probability (BP) approach to assign ancestry state to 30-SNP sliding windows across the targeted genomic regions in each individual fish (see section 1.5.1). The results of both methods coincided by 0.85+/-0.10 SD for all samples, by 0.93+/-0.12 SD for the RABS population and by 0.72+/-0.16 SD for the PGTS population (Figure 2D). The subsequently presented results were obtained by the Bayesian Probability approach with the results of the HMM approach marked in parentheses.

SGV is present at low frequencies

Freshwater-adaptive alleles are mostly present at low frequencies in both marine populations with some freshwater adaptive alleles in the RABS population found in as few as 1 out of 10 980 sampled chromosomes (Figure 2E, Figure S1) – highlighting limited availability of some adaptive alleles for future adaptation from SGV, as well as the importance of studying large sample sizes. Mean allele frequencies were 0.05+/-0.12 SD (HMM: 0.03+/-0.08 SD) in the RABS population and 0.10+/-0.08 SD (HMM: 0.11+/-0.10 SD) in the PGTS population with ranges from 9e-5 to 0.84 (HMM: 0 to 0.5) and 0 to 0.35 (HMM: 0 to 0.48), respectively.


A
Marine-Freshwater Cluster Separation Score
0.015
0
12.78 Mb
12.83 Mb
12.88 Mb
B
marine
freshwater
F1 #1
F1 #2
F1 #3
F1 #4
RABS #1
RABS #2
RABS #3
RABS #4
RABS #5
RABS #6
RABS #7
RABS #8
RABS #9
RABS #10
RABS #11
RABS #12
RABS #13
RABS #14
RABS #15
RABS #16
RABS #17
RABS #18
RABS #19
RABS #20
12.81
12.82
12.83
12.84
12.85
12.86
C
Assignement of ancestry states
Bayesian Probability
Hidden Markov Model
D
Percentage of identical results
1.00
0.75
0.50
0.25
0.00
Controls
PGTS
RABS
E
Count of divergent regions
150
100
50
0
PGTS
RABS
0.0
0.2
0.4
0.6
0.8
1.0
Allele frequency of freshwater−adaptive alleles

175

Figure 2: Individuals carrying freshwater-adaptive standing genetic variation at divergent loci were identified using two different statistical approaches: a Bayesian Probability method and a Hidden Markov Model (here shown at the chromosome IV EDA locus as example). (A) Marine-freshwater cluster separation score (CSS) around the EDA region. Divergent loci were previously identified by CSS (Jones *et al.*, 2012). (B) Visual genotype for the control family as well as 20 individuals from the RABS population at the EDA region, a subset of SNPs is shown. (C) Carriers at each divergent locus were identified by two different approaches. The Bayesian Probability approach (on the left) focused on ecotype informative SNPs from a training data set of 26 Canadian individuals with an allele frequency difference above 0.5 between marine and freshwater sticklebacks. Ancestry states were subsequently assigned for windows of 30 SNPs. The Hidden Markov Model (on the right) was trained with the control family and assigned ancestry states based on the density of derived alleles and of heterozygosity. (D) Percentage of identical results obtained by both approaches. (E) Availability of SGV. Histogram of allele frequencies of freshwater-adaptive alleles for each of 232 divergent loci in two marine populations (PGTS and RABS) based on the results obtained with the Bayesian Probability approach (see section 1.5.1).

Frequencies different between populations

We observed significant differences in allele frequencies of SGV carried in the two populations (pearson correlation coefficient 0.26, HMM: 0.30). Furthermore, allele frequencies in PGTS are significantly higher than in the RABS population (paired one-tailed t-test p = 1.3e-10, HMM p<2.2e-16) and 85% of the loci (HMM: 91%) show higher frequencies in the PGTS population (Figure S2, S3). For instance, 335 (HMM: 230) out of 6 546 sampled PGTS chromosomes are carrying a freshwater-adaptive allele at the Eda locus whereas only 4 (HMM: 5) of 10 722 RABS chromosomes do (frequencies of 0.049 (HMM: 0.034) and 0.0004 (HMM: 0.0005), respectively). Similar results were found for another divergent region on chrIV, the “Wnt7b locus” approximately 4 cM away from Eda: 719 (HMM: 979) of 6 546 PGTS chromosomes are carrying a freshwater-adaptive allele, whereas only 3 (HMM: 3) of 10 722 RABS chromosomes do (frequencies of 0.11 (HMM: 0.14) and 0.0003 (HMM: 0.0003), respectively).

Haplotype decay through recombination at adaptive loci

When a haplotype introgresses into a population via hybridization, recombination results in a decay of the introgressed haplotype block size over time at a rate proportional to the local genomic recombination rate. The size of an introgressed haplotype block, its start and end points and variation in these features can therefore provide a great deal of information about the historical recombination events of adaptive SGV haplotypes. We stress that the “haplotype blocks”, detected by the Bayesian Probability approach and referred to in this study, capture historical recombination events defining the boundaries of an ancient derived freshwater haplotype since its introgression from the source population, and do not reflect the size of a phased haplotype with linked marine alleles in the genetic sense of the term. We therefore refer to these as “ancient haplotype blocks”. While the decay in size of an ancient haplotype block is an indicator of the age of the haplotype, it does not necessarily represent only the time the haplotype has spent in the marine environment. Parallel adaptation from SGV may cause

already broken down ancient haplotypes to be swept to high frequency (“fixed”) in another freshwater population and also introgress (“leak”) in this form back into the marine population. Thus, the size of the freshwater-adaptive ancient haplotype blocks captures a mixture of selection, introgression and decay throughout stickleback adaptation history.

Until now, it is still unclear, whether freshwater-adaptive alleles are neutral in the marine environment or whether they present a genetic load for marine individuals and are thus quickly purged from the marine population. After accounting for the local genomic recombination rate, the frequency and size of ancient blocks together may provide some clues as to the differential introgression rate across loci and the role of selection in maintaining or constraining the availability of SGV for future adaptive potential.

The freshwater-adaptive haplotype blocks present in marine “carriers” differ in their sizes and their starting and end points. For instance, each of the four Eda carriers in the RABS population shows a different freshwater-adaptive haplotype block (Figure 3A). The PGTS Eda carriers also depict different flavours of haplotypes, with freshwater-adaptive ancestry states being present for the right and/or left flank of the divergent region in different carrier individuals (Figure 3A). Since the detected freshwater-adaptive haplotype blocks are differing in size and position, the freshwater-adaptive material has apparently already undergone several recombination events at this locus.

We observed a significant positive relationship between the frequency of ancient adaptive blocks at each locus and the size of the ancient adaptive block in marine “carriers”. In both populations, larger ancient blocks are present at loci with higher frequency in the population (Figure 3) – suggesting that freshwater adaptive ancient haplotype blocks enter the population at high rates, but do not persist long enough for complete breakdown to smaller haplotype sizes. Similarly, loci with smaller ancient haplotype blocks tend to found at lower frequencies suggesting loss of alleles from the population over time (via selection or drift) while haplotype block size decays via recombination.

Assuming all loci enter the marine population at the same frequency, of particular interest to us, are the loci that deviate from this positive relationship. Loci that are at low frequency, but have relatively large ancient block size (upper left part of plot), e.g. Eda in RABS (marked in red in Figure 3B), comprise loci rapidly purged from the population before recombination can break them down. These loci thus present a theoretical large “genetic load” to the marine population, and would suggest that purging of these ancient alleles by selection has the potential to constrain rates of future adaptation from SGV by limiting the availability of adaptive alleles. In contrast, loci that are present at high frequency, but have small ancient block size (lower right part of plot) are those that present little genetic load to the marine population and can persist at high frequencies for long enough that recombination decays adaptive blocks to small size.

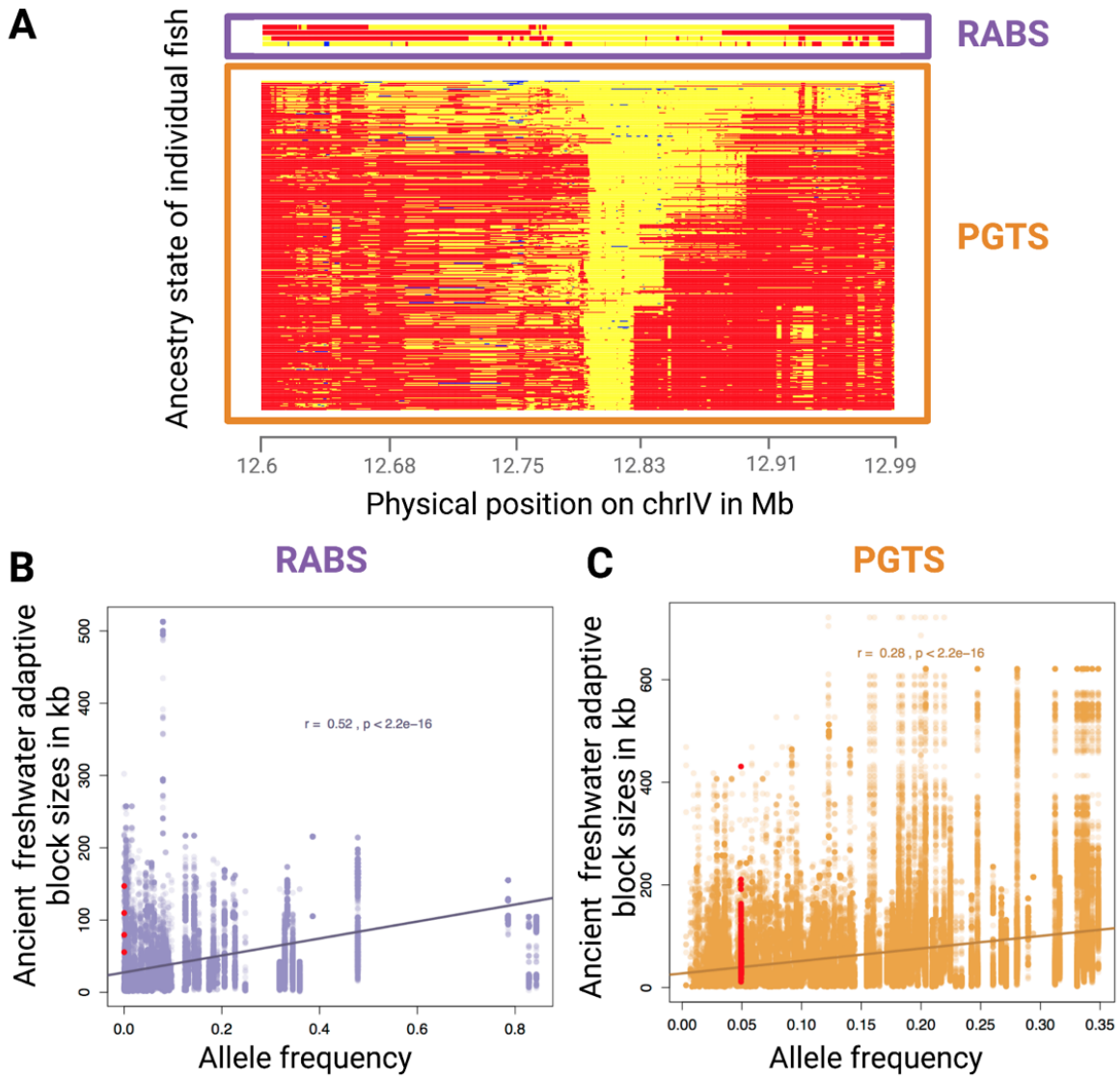


Figure 3: Freshwater-adaptive standing genetic variation in marine fish is present as ancient haplotype blocks of varying sizes. (A) Pile-up of freshwater-adaptive haplotype blocks carried at the Eda locus by individuals identified to be carrying freshwater-adaptive Eda alleles in the RABS and PGTS populations. Red - marine homozygous ancestry, yellow- heterozygous ancestry, blue - freshwater homozygous ancestry. (B,C) Ancient freshwater-adaptive block sizes carried by RABS carriers (B) or PGTS carriers (C) at all divergent regions plotted against their allele frequencies in the respective population. The ancient block size is determined as the sum of all freshwater-adaptive ancestry haplotype blocks overlapping with the respective divergent region. The results for the Eda locus are plotted in red. Results presented here were quantified based on a Bayesian Probability approach (see section S1.5.1).

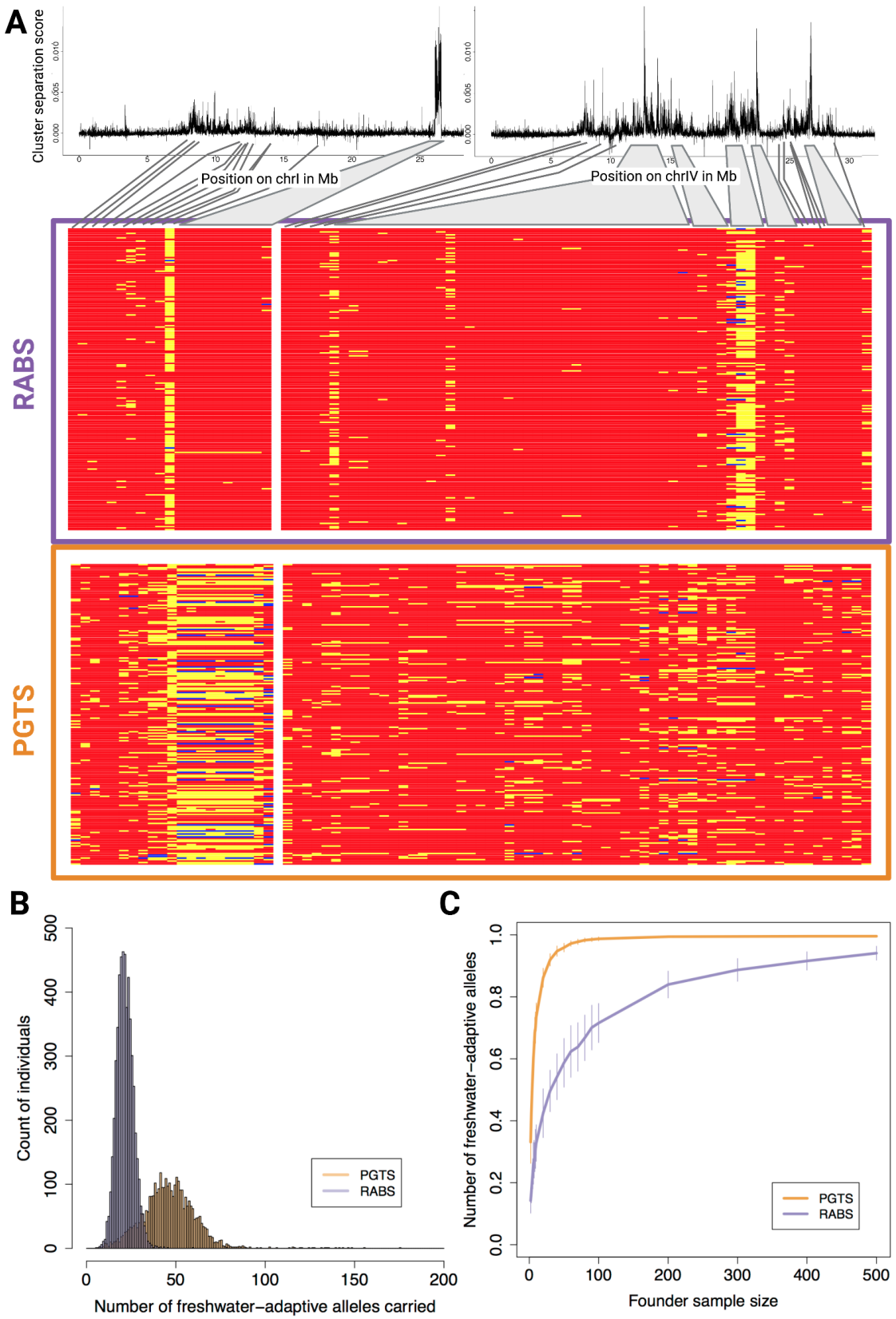
In PGTS, the twenty loci with the largest ancient block sizes (above 600kb) show medium frequencies ranging from 0.12 to 0.35. Half of them fall within the chrI inversion and the other half is located on chrXI. In RABS, the seven loci with the largest ancient block sizes (above 250 kb) are located on chrI, chrXI, chrXX and chrXXI with frequencies ranging from 0.0002 to 0.08. One of the loci falls within the chrI inversion. Inversions are genomic regions with suppressed recombination and therefore, haplotypes are broken down slower in these regions.

All in all, the ancient blocks are significantly longer for PGTS carriers compared to RABS carriers at 71% of the divergent loci with p-values < 0.05 (Figure 3). Freshwater-adaptive alleles are thus not only more frequent in the PGTS population, but also carried in bigger blocks compared to the RABS population. In general, RABS appears to be either exposed to lower introgression and/or stronger selection.

Multiple freshwater-adaptive alleles are carried by individual marine fish

Until now, the analysis was focused on each adaptive locus separately. However, adaptation is typically polygenic and in sticklebacks an ensemble of freshwater-adaptive alleles is essential for adaptation. A key question in adaptation of sticklebacks is whether SGV is provided by recently introgressed individuals, e.g. F1s between marine and freshwater individuals, or whether the freshwater allele ensemble used in adaptation is initially broken down and divided on many individuals in the marine population before being swept and reassembled during freshwater adaptation. By examining the ancestry state across linked adaptive loci we note that “carriers” in our data don’t present as F1s and predominantly don’t carry large blocks of several adaptive loci successively. RABS “carriers”, for instance, mostly carry only one or two adaptive alleles in a block, whereas PGTS “carriers” show longer freshwater-adaptive allele blocks. In both populations, individuals carry multiple allele blocks within chrI and chrIV. These allele blocks are interspersed and disconnected by marine alleles. Adaptive loci within the chrI inversion are often carried as one successive block as expected due to reduced recombination within inversions (Figure 4B). Considering all 232 adaptive loci studied across the genome, PGTS individuals carry 6 to 176 out of 232 freshwater-adaptive alleles (HMM: 1 to 127 out of 142) with a mean of 46.8+/-15.2 SD (HMM: 31.5+/-12.9 SD), whereas RABS individuals carry 7 to 79 out of 232 freshwater-adaptive alleles (HMM: 1 to 53 out of 142) with a mean of 22.1+/-4.8 SD (HMM: 8.5+/-4.0 SD) (Figure 4B, Figure S4). These results support the hypothesis that freshwater-adaptive material was introduced by several recombination events or originates from several introgressed individuals carrying different allele ensembles. Furthermore, it shows that freshwater-adaptive alleles are dispersed and that all marine fish carry some amount of freshwater-adaptive SGV.

Given these results, the question arises, how many freshwater-adaptive material is available in a marine founder population colonising a freshwater lake. In order to assess this, different amounts of marine individuals were randomly sampled from each of the marine populations. Subsequently, the percentage of freshwater-adaptive alleles present at 232 divergent loci in a given founder population was determined (Figure 4C). With SGV being prevalent in PGTS, the astonishingly low number of only 50 randomly sampled marine PGTS individuals provided already 96% (+/- 1% SD) of the freshwater-adaptive allele ensemble and 500 PGTS individuals entailed all 231 (+/- 0.1 SD) adaptive alleles that were previously


A
Cluster separation score
Position on chrI in Mb
Position on chrIV in Mb
RABS
PGTS
B
Count of individuals
Number of freshwater−adaptive alleles carried
PGTS
RABS
C
Number of freshwater−adaptive alleles
Founder sample size
PGTS
RABS

305

Figure 4: Freshwater-adaptive standing genetic variation is carried by marine fish at multiple loci cross the genome. (A) Carrier states for 21 divergent regions on chrI and 61 divergent regions on chrIV, shown with the extent of global parallel marine freshwater divergence (cluster separation score, Jones *et al.*, 2012). Rows represent 100 randomly picked individuals from RABS and PGTS, respectively, and each column represents a divergent region. Red - no carrier, yellow- heterozygous carrier, blue - freshwater homozygous carrier of freshwater-adaptive allele. (B) Individual marine fish carry multiple freshwater-adaptive alleles across the genome. A histogram shows the number of freshwater-adaptive alleles carried by the RABS and PGTS individuals genome-wide (232 divergent regions in total). (C) Sampling from the marine population at random would require relatively small sample sizes (e.g. less than 500) to sample freshwater-adaptive alleles at all global parallel adaptive loci identified in Jones *et al.*, 2012. Plots show the percentage of freshwater-adaptive alleles present at 231 (PGTS) or 232 (RABS) divergent loci for different founder population sizes randomly sampled from PGTS or RABS, respectively. Mean and standard deviation were determined based on 100 iterations. Results presented here were quantified based on a Bayesian Probability approach (see section S1.5.1).

detected to be present in the PGTS population. 50 RABS individuals only carry 58% (+/- 9% SD) of all analysed 232 adaptive alleles. Since some freshwater-adaptive alleles are only carried by very few individuals in the RABS population, even 500 founder individuals only entail 94% (+/- 3% SD) and not the complete freshwater-adaptive allele ensemble. However, population sizes of 3000 RABS individuals used for colonisation of freshwater lakes in Kingman *et al.* encompassed in this analysis 231.7 (+/- 0.5 SD) adaptive alleles (Kingman *et al.*, 2021). These results suggest that freshwater-adaptive material is highly prevalent in a founder population. Even if adaptive alleles are present at low frequencies in the marine population such as in RABS, the complete freshwater-adaptive allele ensemble is present when enough marine individuals are assembled.

Linkage Disequilibrium among adaptive alleles

Since most individuals carry several adaptive loci and stickleback adaptation is highly polygenic, we investigated whether specific allelic combinations were prevalent in the carrier individuals, and therefore determined the extent of intra- as well as inter-chromosomal linkage disequilibrium (LD) in RABS and PGTS. Especially in RABS, some alleles are very rare (so-called paucimorphisms) and complicate the reliable determination of LD. To overcome this challenge, r^2 and D' were calculated for both populations, whereat D' is standardised based on the allele frequencies at the loci of interest and thus correcting for frequencies of rare alleles differing by several magnitudes.

PGTS shows strong LD for all 10 adaptive loci within the chrI inversion (r^2: 0.63+/-0.34 SD, D': 0.83+/-0.23 SD) confirmed by Chi-square p-values below 1e-16 (Figure 5, Figure S5). LD for 61 adaptive loci on chrIV is in general weak (r^2: 0.02+/-0.08 SD, D': 0.16+/-0.20 SD) even lower than the mean of intra-chromosomal LD for all 231 adaptive loci included in the analysis (r^2: 0.05+/-0.15 SD, D': 0.19+/-0.24 SD) (Figure 5, Figure S6). Only very few close-by adaptive loci show strong linkage. As an example, the probability for a PGTS individual which carries a freshwater-adaptive allele at the Eda locus (chrIV:12,811,394-12,856,894) to also carry a

freshwater-adaptive allele at the approximately 4 cM apart Wnt7b locus (chrIV:26,703,207-26,767,873) was 0.19 and almost identical to the probability for a non-carrier (0.20) leading to very low LD values (r^2: 2e-06 SD, D': 0.002). The probability of these close-by loci to be separated by recombination during meiosis is only 0.04 and thus the empirically obtained likelihood corresponds to about 96 generations of recombination in order to break-down the linkage between the two adaptive loci. This indicates that freshwater-adaptive alleles on chrIV are already present for many generations in the marine population and/or already introgressed as incomplete freshwater-adaptive allele block into the marine population. In RABS, intra-chromosomal LD is also high for all 10 adaptive loci within the chrI inversion (r^2: 0.73+/-0.36 SD, D': 0.88+/-0.17 SD) reassured by Chi-square p-values below 1e-16 (Figure S7, S8). LD for 62 adaptive loci on chrIV (r^2: 0.02+/-0.06 SD, D': 0.27+/-0.20 SD) is moderate with similar ranges as the mean of intra-chromosomal LD for all 232 adaptive loci (r^2: 0.03+/-0.12 SD, D': 0.27+/-0.24 SD) (Figure S9). Since only four RABS individuals carried a freshwater-adaptive allele at the Eda locus and three carried adaptive alleles at the "Wnt7b" locus approximately 4cM away, analysis focusing only on these individuals was largely underpowered.

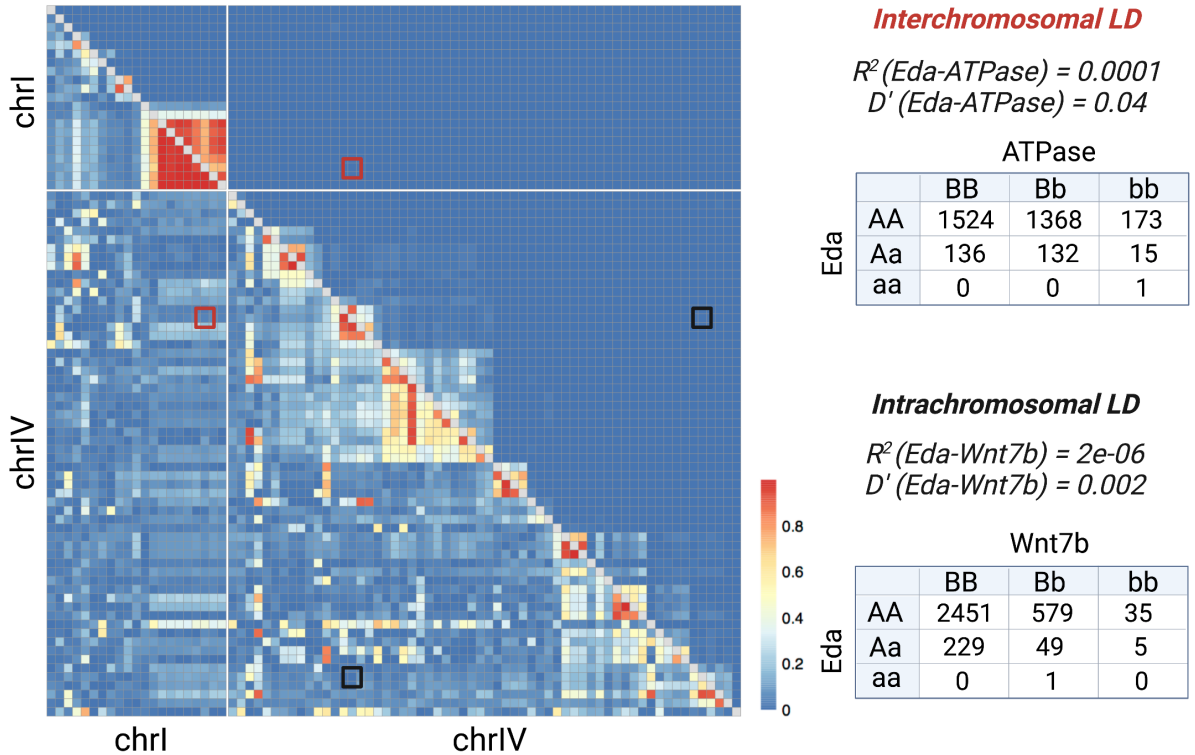


Figure 5: Within PGTS marine individuals significant inter-chromosomal association (linkage disequilibrium) was detected among freshwater-adaptive alleles carried on chromosomes I and IV, while no significant intra-chromosomal linkage disequilibrium was detected among alleles on chromosome IV. Estimates of linkage disequilibrium among adaptive alleles in chrI and chrIV for PGTS individuals based on carrier information (see section S1.6.8). The upper triangle shows r^2 and the lower triangle shows D'. Red boxes mark the inter-chromosomal LD between Eda (chrIV:12,811,394-12,856,894) and ATPase (chrIV:26,703,207-26,767,873) and black boxes the intra-chromosomal LD between Eda (chrIV:12,811,394-12,856,894) and Wnt7b (chrI:26,312,062-26,384,706).

R^2 values for inter-chromosomal LD indicated low linkage for PGTS as well as RABS (PGTS: r^2: 0.0007+/-0.0014 SD; RABS: r^2: 0.004+/-0.014 SD). After correcting for frequencies of rare alleles, however, D' values suggested stronger linkage (PGTS: D': 0.10+/-0.14 SD; RABS: D': 0.20+/-0.20 SD). The probability for loci on different chromosomes to be separated by recombination during meiosis is 0.5. Hence, LD is expected to decay quickly for them and mean D' values detected here correspond to only one to three generations of recombination. In general, our data supports the presence of epistatic effects favouring specific inter-chromosomal allelic combinations. Marine individuals carrying freshwater-adaptive inter-chromosomal allele combinations might largely contribute to a quick re-assembly of the freshwater-adaptive allele ensemble in a newly colonised freshwater habitat. However, the weak linkage between adaptive loci within chromosomes might indicate a constraint for rapid adaptation. In order to re-assemble these loci several generations of recombination might be needed.

Lack of F1 hybrids in marine populations

To investigate the breakdown of the freshwater-adaptive allele ensemble further, the breakdown was simulated during consecutive generations of introgression based on the carrier information at 232 divergent loci. In this approach, F1 hybrids between divergent ecotypes can be distinguished from multi-generational introgressed individuals by comparing their heterozygosity and freshwater ancestry fraction within the examined loci. Specifically, F1s are expected to have high heterozygosity and intermediate freshwater ancestry levels, while F2s present lower heterozygosity values and F1 backcrosses with marine samples show lower levels of both variables (Figure 6A, Figure S11A). The empirical data for all marine individuals compared to the simulated data showed that no F1s are present in either of the marine populations. Individuals from PGTS population match F1 backcrosses with marine ancestors (F1-Marine) and F1-Marine backcrosses with marine ancestors, whereas RABS population only match F1-Marine backcrosses with marine ancestors (Figure 6B, Figure S11B). These results highlight further that freshwater-adaptive alleles are not provided by single recently introgressed individuals, but rather are present in the marine population for several generations. When the alleles are randomly distributed among the individuals in a population, particularly the variation between individuals is removed (Figure S10, S12). In comparison to the observed data, this supports that different alleles and allele combinations are experiencing different selective pressures and that the degree of introgression can vary among alleles.

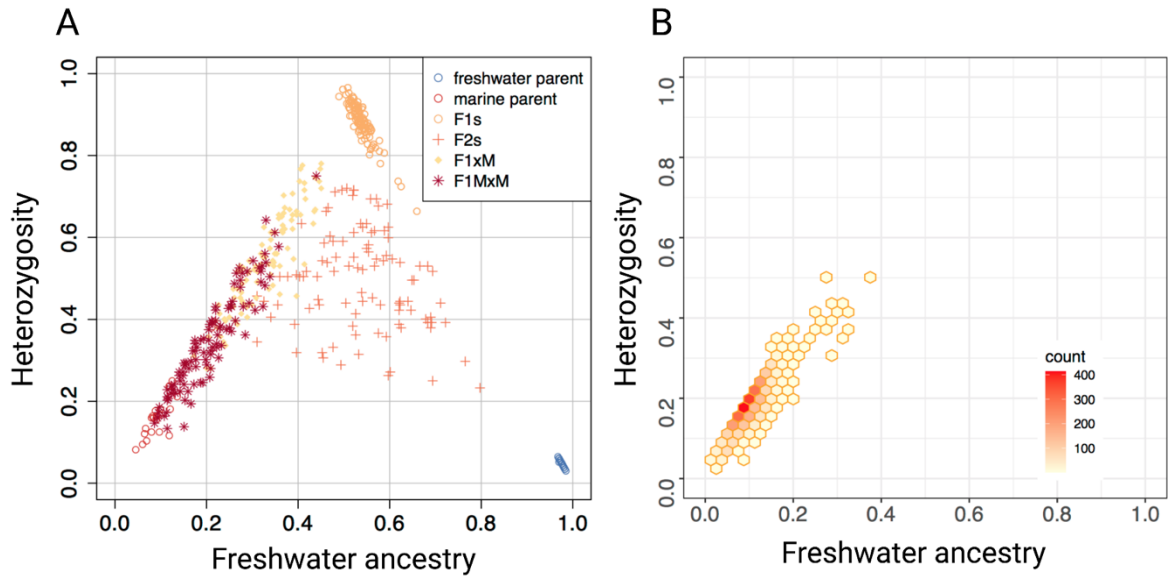


Figure 6: Marine fish carrying freshwater-adaptive alleles are products of multiple generations of introgression into an ancestral marine genomic background. (A) Simulations for the breakdown of the freshwater-adaptive allele ensemble during multiple generations of introgression. Simulations were performed based on carrier information at 232 divergent loci. (B) Comparison of empirical data of PGTS population to simulated data in panel A suggest PGTS marine individuals represent predominantly F1-Marine backcrosses with marine ancestors (F1MxM) and a smaller number of backcrosses of hybrids with marine ancestors (F1xM). Results presented here were quantified based on a Bayesian Probability approach (see section S1.5.1).

Geographically disparate sources of freshwater-adaptive standing genetic variation in marine fish

Introgression plays a major role for the availability of SGV. Until now, it is unknown whether this variation originates from a single hybrid zone or whether it is introduced into the marine population from multiple geographically disparate sources. To investigate the origin of freshwater-adaptive alleles in the marine population, we projected phased haplotypes of carrier individuals onto the principal components of variation among 103 global freshwater individuals from Alaska, Canada, Europe and California (Kingman *et al.*, 2021, Table S10). For some loci, the global freshwater individuals were separated by geography.

Major axes of variation separate populations by longitude (e.g. Atlantic freshwater (European) and Californian populations cluster separately from North American/Alaskan populations on principal component one), and latitude (e.g. Northern Pacific (Alaskan and Canadian) populations separate from Southern Pacific (Californian) populations along principal component two). At the central 6 kb block of the Eda region (chrIV:12,818,000 – 12,824,000) the freshwater individuals were mostly divided into three clusters, grouping individuals from Alaska and Canada, Canada and Europe, and California together. For a 35 kb block at the ATPase locus within chrI inversion (chrI: 26,311,706 – 26,346,706), the samples were divided into a Californian, a European and several North-American individual groups (Figure 7).

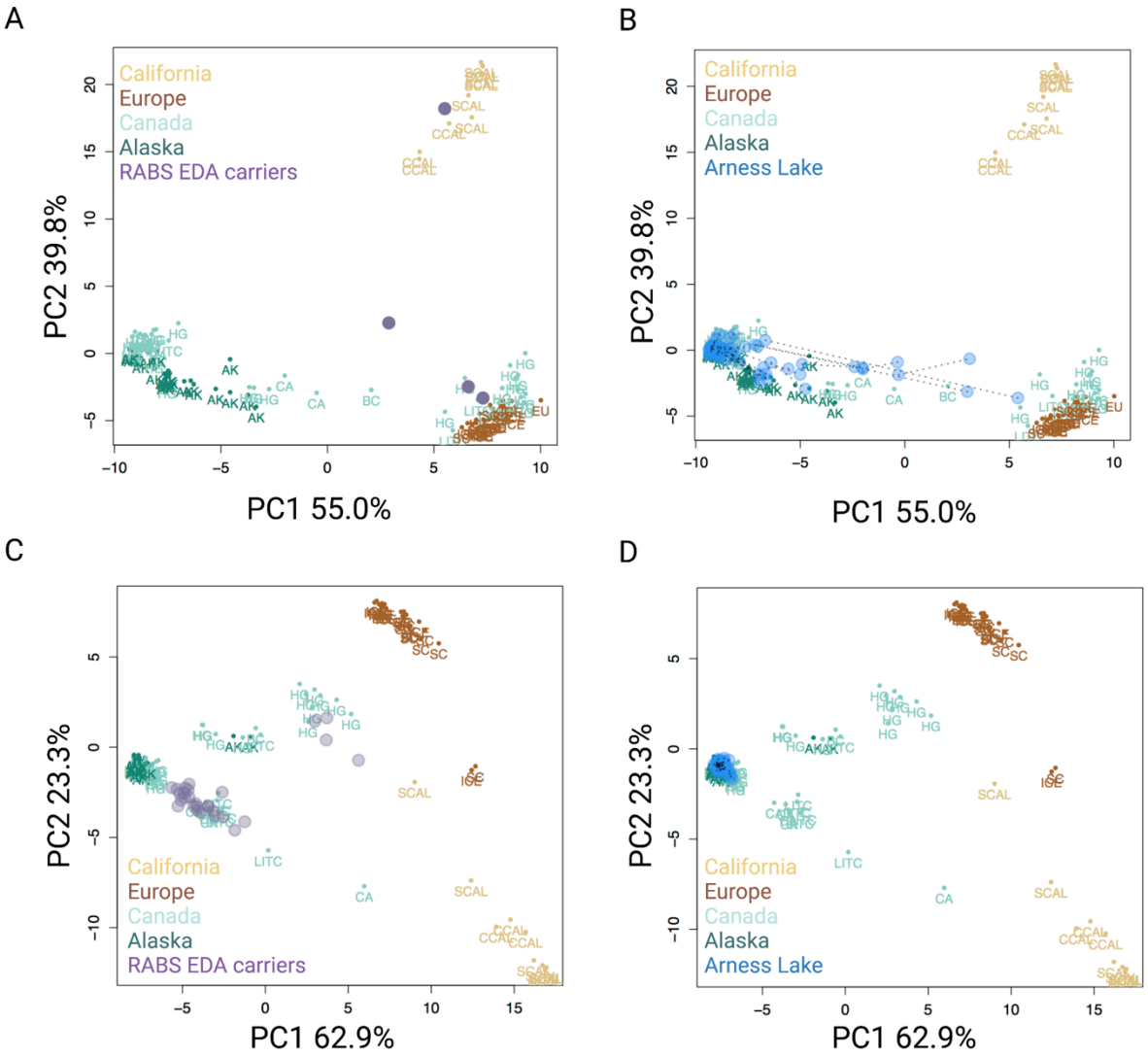


Figure 7: Freshwater-adaptive ancient haplotype blocks carried by marine individuals (Rabbit Slough, panels A,C) and carried at high frequency in a nearby freshwater population (Arness Lake, panels B,D) show similarity to haplotypes carried by geographically disparate freshwater populations around the world. (A,B) Principal component analysis (PCA) was performed for a central 6 kb block at the Eda region (chrIV: 12,818,000-12,824,000) for 103 global freshwater individuals from Alaska, Canada, Europe and California. (A) Projected haplotypes of the RABS EDA carriers cluster with samples from different geographic locations: one similar to the freshwater haplotype carried by Californian freshwater populations, while the other three more similar to the ancient haplotype carried by European and Canadian freshwater populations. (B) Projected haplotypes of freshwater individuals from an Alaskan freshwater population (Arness Lake) near to the RABS marine sampling location show that several different versions of the freshwater EDA haplotype are present at high frequencies in the population (similar to haplotypes found in other freshwater Alaskan and Canadian populations). (C,D) Principal component analysis (PCA) was performed for a 35 kb block within chrI inversion spanning the ATPase locus (chrI: 26,311,706 - 26,346,706) for 103 global freshwater individuals from Alaska, Canada, Europe and California. (C) Projected haplotypes of the RABS ATPase carriers cluster with two different Canadian samples suggesting Canadian populations are a likely source of the freshwater-adaptive standing genetic variation observed in RABS marine fish. (D) In contrast, the projected haplotypes of individuals carrying the freshwater adaptive haplotype in Alaskan freshwater lake (Arness Lake) all cluster together with one Alaskan haplotype cluster suggesting a hard sweep on Alaskan SGV in the freshwater population that is not observed in today's nearby RABS marine population.

To identify the mostly likely geographic origin of freshwater SGV carried by RABS marine fish, phased freshwater haplotypes carried by RABS individuals were subsequently projected on these PCAs. Three Eda freshwater haplotypes showed similarity to Canadian as well as European individuals, whereas one Eda freshwater haplotype clustered with Californian individuals (Figure 7A). 28 ATPase freshwater haplotypes located mostly within two separate Canadian clusters (Figure 7C). These results suggest that SGV is not introduced via one single hybrid zone in the marine population, but represents the variance globally present in freshwater populations suggesting that it originates from multiple, geographically distant source populations.

Soft and hard sweeps in freshwater population

In the following, phased haplotype carried by 39 freshwater individuals from Arness Lake, Alaska, were projected onto the PCAs previously performed with 103 global freshwater individuals. All pairs of haplotypes in the freshwater population at the Eda and ATPase locus were freshwater-adaptive, indicating that the freshwater-adaptive allele was fixed in the freshwater population at both loci. Most freshwater haplotypes at the Eda locus (chrIV:12,818,000 - 12,824,000) formed two separate clusters within the Alaskan/Canadian cluster implicating that two disparate, still mostly similar haplotypes are present in the freshwater population (Figure 7B). Several unique haplotypes were additionally present at the Eda locus sharing similarity with North-American individuals. This pattern supports that a soft sweep occurred at the Eda locus. The analysis for freshwater haplotypes at the ATPase locus (chrI: 26,311,706 – 26,346,706) revealed that all haplotypes of the freshwater population cluster tightly within one North-American cluster suggesting that a hard sweep happened (Figure 7D). The detected patterns of haplotype affiliation indicate that hard as well as soft sweeps occurred at adaptive loci in the freshwater population. Fixation of *de novo* mutations can cause hard sweeps in a population, whereas SGV can lead to soft as well as hard sweeps. Particularly the presence of soft sweeps seconds that SGV plays a crucial role in rapid adaptation of sticklebacks to freshwater environments.

Discussion

The results presented in this study offer new insights into evolutionary processes affecting the availability of SGV and thus, facilitating and/or constraining the rate at which future adaptation proceeds. Our study revealed that SGV is predominantly present at low frequencies in marine populations as suggested by previous studies (Colosimo *et al.*, 2005; Miller *et al.*, 2007; Kingman *et al.*, 2021). The large number of marine individuals enabled us to quantify the frequency of very rare alleles down to 1 out of 10 980 sampled chromosomes, whereby our data provided a significant advantage compared to previous studies which were performed

485 either with small sample sizes (impeding the detection of rare alleles) (Colosimo *et al.*, 2005;
486 Miller *et al.*, 2007; Hohenlohe *et al.*, 2010; Kingman *et al.*, 2021), with pooled data (precluding
487 individual-level information about linkage or haplotypes) or with SNP arrays (providing sparse
488 data for only a few hundred SNPs across the whole genome) (Kingman *et al.*, 2021).
489 In particular in RABS, freshwater-adaptive alleles are extremely rare with median allele
490 frequency estimated to be 0.005 for 232 loci showing the same magnitude as previous studies
491 (0.001 for 44 adaptive loci in Kingman *et al.*, 2021). In previous studies, RABS was used as
492 founder population for newly colonised freshwater lakes. Despite the low frequencies of
493 freshwater-adaptive alleles in RABS, contemporary time-series data depicted the successful
494 adaptation of these lake populations to the new environment (Kingman *et al.*, 2021) suggesting
495 that even the low availability of SGV provided sufficient evolutionary substrate for rapid
496 adaptation.
497 Different degrees of introgression and selective pressures among populations and loci
498 can lead to freshwater-adaptive alleles being prevalent or very rare in a marine population.
499 These factors can lead to very diverse availability of SGV between populations. Compared to
500 RABS, PGTS showed significantly higher frequencies of freshwater-adaptive alleles. Hence,
501 RABS experiences probably less introgression and/or higher selective pressures compared to
502 PGTS. Allele frequencies differing between marine populations have been reported in studies
503 before (Kingman *et al.*, 2021). The different availability of SGV indicates that some
504 environments are more favourable for the maintenance of SGV and might represent a more
505 promising “source” of SGV for conservation management.
506
507 Theoretically, adaptation to a new freshwater habitat would proceed most rapidly if one or more
508 founding individuals introduced all the necessary adaptive alleles in linkage (an “adaptive
509 ensemble”). Thus the frequency and linkage among freshwater adaptive alleles is an important
510 determinant of the rate at which a new population can adapt. Adaptation might occur most
511 efficiently through direct migration from a proximal, already adapted, freshwater population –
512 a process that is likely to be relevant when new tributaries or water bodies are formed in river
513 systems, e.g. the colonization of a newly formed oxbow lake or melting of glaciers up side
514 tributaries of a bigger river valley. While this mechanism of direct migration may be relevant to
515 rapid colonization of new habitats within the same river system, it cannot easily explain the
516 remarkable sharing of adaptive alleles between geographically disparate freshwater
517 populations across the Northern Hemisphere since freshwater resident individuals lack
518 migratory behaviour (Moyle, 1976; Wootton, 1984). Rather introgression is likely to play an
519 important role in the spread of adaptive alleles through a mismatched environment into a new
520 freshwater population (referred to as the “Transporter hypothesis” by Schluter & Conte (2009)).
521 With introgression via hybridisation, the movement of "freshwater adaptive ensembles of linked

alleles" in F1 hybrids or early generation admixed individuals offers the potential for rapid adaptation on colonisation of a new habitat. Although F1 hybrids of marine and freshwater sticklebacks are viable and interfertile in the laboratory (Hagen, 1967; McPhail, 1994), pre- and postzygotic barriers lead to a significant deficit of F1s in hybrid zones compared to the expectations under random mating and maintain reproductive isolation in the wild (Jones *et al.*, 2006). Few studies have been able to explore the presence of F1s in marine environments and the importance of linked freshwater adaptive allele ensembles. A previous study has shown that the freshwater adaptive Kit ligand allele is not carried in linkage with the freshwater adaptive allele at Eda in 107 Californian marine sticklebacks indicating that the alleles are not carried by F1 hybrids between marine and freshwater sticklebacks (Miller *et al.*, 2007). Consistently and in a much larger sample of more than 8 800 marine individuals our present study did not detect any F1 hybrids, indicating that selection against a completely linked freshwater adaptive allele ensemble is strong in the marine environment. Another sign of presumably strong selective pressures against some adaptive alleles is the excess of heterozygous carriers for some alleles in Figure 4A with no freshwater-homozygous carriers present. This doesn't seem to be an ascertainment error, but rather a sign of strong selection against freshwater-homozygous ancestry states (Slatkin, 2008). Our simulations show that marine individuals correspond to one or multigenerational backcrosses and suggest that freshwater adaptive SGV arriving in a new population by marine founders is for the most part present as dissembled alleles rather than as a pre-assembled "adaptive ensemble".

In order to explore the magnitude of association among freshwater-adaptive SGV carried in linkage by marine individuals (or lack thereof), we determined inter- as well as intra-chromosomal linkage disequilibrium (LD) among adaptive loci. In previous studies, evidence for considerable LD among adaptive loci was reported in a small pool of individuals from RABS and another Alaskan population (Resurrection Bay, Hohenlohe *et al.*, 2012). However, these results are difficult to interpret since they are confounded by pooling samples from both populations and population substructure within a sample can be a spurious cause of strong LD. Moreover, combined with the low frequency of freshwater adaptive alleles in marine populations, the extremely small samples sizes of 32 individuals severely limit the power and accuracy of LD estimates making the conclusions of this study questionable. In our earlier study (Kingman *et al.*, 2021), we reported modest LD in Alaskan populations including RABS and Resurrection Bay, but acknowledged that the analyses are still underpowered by relatively small sample sizes of 751 individuals in RABS and 655 in Resurrection Bay respectively. With large samples sizes of 3 329 PGTS and 5 490 RABS individuals in this study, we mostly overcame the limitations of small sample sizes and detected strong LD among some adaptive loci in physical proximity and especially for those within the same inversion. Further we

observed low but significantly more inter-chromosomal LD (among adaptive alleles located on different chromosomes) than expected by chance. These results indicate that parallel adaptation of marine individuals to freshwater environment is slightly facilitated by linkage accelerating the process of reassembling adaptive alleles in the new environment. However, intra-chromosomal LD, in general, was relatively low suggesting that freshwater adaptive allele ensembles within a chromosome are mostly broken down in smaller blocks. The process of reassembling these smaller blocks in the freshwater environment to a complete allele ensemble might take several generations of recombination and hence, longer than the process is reported to take in natural populations (Bell *et al.*, 2004; Lescak *et al.*, 2015; Terekhanova *et al.*, 2014; Bassham *et al.*, 2018). As argued in previous studies, rare carriers of long freshwater adaptive allele blocks might thus play an important role in rapid adaptation since they can significantly speed up the process of re-assembly (Kingman *et al.*, 2021; Hohenlohe *et al.*, 2012). In contrast, the somewhat uniform distribution of freshwater-adaptive alleles among marine individuals challenges the importance of single individuals that carry an excess of freshwater-adaptive alleles, so-called “jackpot carriers” (Kingman *et al.*, 2021; Hohenlohe *et al.*, 2012). It rather points to the crucial role of many founder individuals together combining their freshwater adaptive allele ensembles. Further studies are needed to disentangle whether the success of adaptation to the freshwater environment is dependent on the presence of a few relevant founder individuals, and/or on the adequate availability of SGV in the marine population in general and a respective number of founders from this population.

When a new freshwater habitat is colonised by marine sticklebacks, multiple freshwater haplotypes at the same locus might be available in the founder population. What features and mechanisms dictate whether one single or multiple haplotypes are used for adaptation to the new environment? If multiple freshwater haplotypes are equally beneficial, the adaptation process might result in soft sweeps with multiple haplotypes being present in the freshwater population (Hermisson & Pennings, 2005). In case one haplotype is significantly better than the others and provides higher evolutionary fitness for the carrier, only one haplotype might rise to high allele frequencies in the freshwater population (hard sweep) (Maynard-Smith & Haigh, 1974). Both scenarios can be affected by loss of beneficial alleles due to small effective population size and consequential demographic stochasticity (Kirch *et al.*, 2021). For instance, all freshwater haplotypes present in the source population can be lost leading to incomplete adaptation in the freshwater population (Kirch *et al.*, 2021). Aside from that, demographic stochasticity can result in hard sweeps of one haplotype despite the presence of multiple equally beneficial haplotypes in the founder population. Stochastic loss of alleles might also cause the fixation of a haplotype which is not the most beneficial one. In this study we show that both, hard and soft sweeps are present at adaptive loci in an Alaskan freshwater

population. Our results show the presence of several different North-American freshwater haplotypes at the Eda locus suggesting a soft sweep. Eda carriers in RABS representing a possible source population of the freshwater population depict North-American, as well as one Californian freshwater haplotype(s). The non-presence of Californian freshwater haplotypes in the Alaskan freshwater population could either point to the non-presence of this haplotype in the founder population of this lake, or it could indicate that the North-American haplotypes were more beneficial than the Californian one in this environment and therefore, outrivalled their competitor (Figure 7). Even though several North-American haplotypes are present in RABS at the ATPase locus, one Alaskan haplotype was fixed in the freshwater population displaying a hard sweep (Figure 7). Although different haplotypes were likely available in the founder population, both adaptive loci exhibit the (re-)use of a local haplotype supporting the theory that local haplotypes are more beneficial than haplotypes originating from more distant freshwater population and that effective population sizes are normally big enough so that local variants do not get stochastically lost.

Our data suggests that interconnectivity between marine populations as well as between marine and freshwater populations via hybrid zones leads to marine populations harbouring a conglomeration of ancient alleles from multiple sources. As founders, marine individuals serve a variety of different freshwater-adaptive haplotypes to choose from, facilitating soft and hard sweeps in the new environment and leading to a repeated adaptive success for threespine sticklebacks. While alleles that share haplotypes with largely geographic distant freshwater populations are present in RABS, there is so far no evidence for these alleles invading Alaskan freshwater populations. These results highlight that gene flow between local ecotypes is crucial for maintaining the availability of SGV and serve as basis of adaptability, whereas interconnectivity between global populations might be less fundamental.

In this study, targeted enrichment sequencing was performed to characterise SGV at previously identified divergent loci in many marine sticklebacks. Pull-down with probes from BAC libraries constructed from Canadian marine and freshwater individuals has the potential to favour the detection of Canadian marine and freshwater SGV and might lead to underestimates of the frequency of alleles from more evolutionary distant populations. Strong selection on derived alleles with hitchhiking of linked neutral variants could result in longer branch length among all possible “freshwater” haplotypes, and a relatively shorter branch length among marine haplotypes. Thus, freshwater haplotypes might depict more sequence differences than marine haplotypes compared to the Canadian pulldown probes. This bias could result in underestimating SGV frequencies in both marine populations. However, successful capture has been reported with target sequence differences up to 10% and the degree of sequence divergence (pi) within and between haplotypes in this study is estimated

633 to be considerably lower than that (Figure S13; Hedtke *et al.*, 2013). In conclusion, this effect,
634 if present, is likely to have a minimal effect on our data and interpretations.
635 Another limitation of the targeted enrichment data is the focus on 242 divergent regions
636 detected in global freshwater-marine stickleback pairs (Jones *et al.*, 2012), ignoring genomic
637 regions which are parallelly divergent in Eastern Pacific sticklebacks and span sevenfold more
638 of the genome (Kingman *et al.*, 2021; Fang *et al.*, 2020). These local adaptive alleles are most
639 likely introduced by local hybrid zones into the marine population and could be introgressed
640 more recently potentially leading to stronger linkage among them. Since we do not account for
641 these local divergent loci, we most probably underestimate the number of freshwater-adaptive
642 alleles carried by each marine individual in total.
643
644 Unfortunately, while natural selection leaves strong molecular signatures of selection around
645 adaptive loci in the genome, identifying the specific mutation(s) conferring a fitness advantage
646 in freshwater habitats requires functional genomic tests in a natural setting. Functional variants
647 are therefore still mostly unknown. Therefore, we inferred the carrier state for each divergent
648 region based on ecotype-informative SNPs (Bayesian approach) and the density of derived
649 alleles and heterozygous sites (HMM approach) – signatures that rely on adaptive mutations
650 and their linked hitchhiking variants. To avoid spurious inference driven by singleton sites of
651 variable pull-down sequencing coverage, carriers of freshwater adaptive alleles were identified
652 by two window-based approaches looking at 10 kb windows or 30 ecotype-divergent SNP
653 windows, respectively. This has the potential to obscure any finer-resolution signal (i.e. less
654 than 10kb or 30 ecotype-informative SNPs) that might be common in relatively young adaptive
655 loci. Thus our approaches have higher power to detect “older” ancient freshwater-adaptive
656 SGV and younger SGV may show entirely different patterns of frequencies, LD and geographic
657 origins in the marine population. The results of both methods show very high consistency when
658 applied to RABS individuals (93%), suggesting that any method-specific bias is neglectable.
659
660 In this study, we characterized the availability of SGV based on many marine
661 sticklebacks and identified soft as well as hard sweeps in a nearby freshwater population. Our
662 data suggests the dynamic reassembly of local freshwater-adaptive alleles in a freshwater
663 environment. The data obtained in this study opens the door to forward-in-time simulations for
664 the reassembly of freshwater-adaptive alleles based on empirical data of SGV. Whole genome
665 sequencing at individual resolution for a population in a newly colonised freshwater lake over
666 time should further make it possible to disentangle the conundrum of how SGV is reassembled
667 and how it facilitates rapid adaptation in threespine sticklebacks.

Acknowledgements

We thank Taylor Frierson and the Washington Department of Fish and Wildlife for providing stickleback samples from Puget Sound, Washington; Rahmiye Kürkcü, Jonas Schwickert, Dingwen Su, Moritz Peters, Volker Soltys, Cholpon Zhakshylykova and Julia Hagauer for assistance with processing several thousand stickleback samples in the laboratory. Christa Lanz, Heike Budde, Katrin Fritschi, and Ilja Bezrukov for support with Illumina sequencing and associated data processing; Sean Stankowski, Nick Barton, Frank Chan, Julia Hagauer, Cholpon Zhakshylykova, Moritz Peters, Dingwen Su, Volker Soltys, Detlef Weigel, Oliver Kohlbacher for discussions. We are grateful to the broader stickleback community many of whom contributed samples from geographically disparate populations for the genomic analyses reported in Jones *et al.* (2012) and Kingman *et al.* (2021). A number of these published datasets were downloaded from public read archives and used in the present study.

Funding: This study was supported by the Deutsche Forschungsgemeinschaft (DFG) SPP1819-JO1316. M.K. is supported by the International Max Planck Research School ‘‘From Molecules to Organisms’’ and the German Research Foundation (DFG). F.C.J. is supported by the Deutsche Forschungsgemeinschaft (DFG) SPP1819-JO1316 and the European Research Council FP7 CoG617279.

Competing interest: The authors declare that they have no competing interest.

References

Barrett, R. D. H., & Schluter, D. (2008). Adaptation from standing genetic variation. *Trends in Ecology & Evolution*, *23*(1), 38–44. https://doi.org/10.1016/j.tree.2007.09.008

Bassham, S., Catchen, J., Lescak, E., von Hippel, F. A., & Cresko, W. A. (2018). Repeated Selection of Alternatively Adapted Haplotypes Creates Sweeping Genomic Remodeling in Stickleback. *Genetics*, genetics.300610.2017. https://doi.org/10.1534/genetics.117.300610

Bell, M. A., Aguirre, W. E., & Buck, N. J. (2004). Twelve Years of Contemporary Armor Evolution in a Threespine Stickleback Population. *Evolution*, *58*(4), 814–824. https://doi.org/10.1111/j.0014-3820.2004.tb00414.x

Bell, M. A., & Foster, S. A. (1994). *The Evolutionary Biology of the Threespine Stickleback*. Oxford University Press.

Colosimo, P. F., Hosemann, K. E., Balabhadra, S., Villarreal, G., Dickson, M., Grimwood, J., Schmutz, J., Myers, R. M., Schluter, D., & Kingsley, D. M. (2005). Widespread Parallel Evolution in Sticklebacks by Repeated Fixation of Ectodysplasin Alleles. *Science*, *307*(5717), 1928–1933. https://doi.org/10.1126/science.1107239

Colosimo, P. F., Peichel, C. L., Nereng, K., Blackman, B. K., Shapiro, M. D., Schluter, D., & Kingsley, D. M. (2004). The Genetic Architecture of Parallel Armor Plate Reduction in Threespine Sticklebacks. *PLoS Biology*, *2*(5), e109. https://doi.org/10.1371/journal.pbio.0020109

Cresko, W. A., Amores, A., Wilson, C., Murphy, J., Currey, M., Phillips, P., Bell, M. A., Kimmel, C. B., & Postlethwait, J. H. (2004). Parallel genetic basis for repeated evolution of armor loss in Alaskan threespine stickleback populations. *Proceedings of the National Academy of Sciences*, *101*(16), 6050–6055. https://doi.org/10.1073/pnas.0308479101

Fang, B., Kemppainen, P., Momigliano, P., Feng, X., & Merilä, J. (2020). On the causes of geographically heterogeneous parallel evolution in sticklebacks. *Nature Ecology & Evolution*. https://doi.org/10.1038/s41559-020-1222-6

Hagen, D. W. (1967). Isolating Mechanisms in Threespine Sticklebacks (Gasterosteus). *Journal of the Fisheries Research Board of Canada*, *24*(8), 1637–1692. https://doi.org/10.1139/f67-138

Hedtke, S. M., Morgan, M. J., Cannatella, D. C., & Hillis, D. M. (2013). Targeted Enrichment: Maximizing Orthologous Gene Comparisons across Deep Evolutionary Time. *PLOS ONE*, *8*(7), e67908. https://doi.org/10.1371/journal.pone.0067908

Hermisson, J., & Pennings, P. S. (2005). Soft Sweeps: Molecular population genetics of adaptation from standing genetic variation. *Genetics*, *169*(4), 2335–2352. https://doi.org/10.1534/genetics.104.036947

Hohenlohe, P. A., Bassham, S., Currey, M., & Cresko, W. A. (2012). Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1587), 395–408. https://doi.org/10.1098/rstb.2011.0245

Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., & Cresko, W. A. (2010). Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags. *PLoS Genetics*, *6*(2), e1000862. https://doi.org/10.1371/journal.pgen.1000862

Jones, F. C., Brown, C., Pemberton, J. M., & Braithwaite, V. A. (2006). Reproductive isolation in a threespine stickleback hybrid zone. *Journal of Evolutionary Biology*, *19*(5), 1531–1544. https://doi.org/10.1111/j.1420-9101.2006.01122.x

Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., Swofford, R., Pirun, M., Zody, M. C., White, S., Birney, E., Searle, S., Schmutz, J., Grimwood, J., Dickson, M. C., Myers, R. M., Miller, C. T., Summers, B. R., Knecht, A. K., … Kingsley, D. M. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, *484*(7392), 55–61. https://doi.org/10.1038/nature10944

Kingman, G. A. R., Vyas, D. N., Jones, F. C., Brady, S. D., Chen, H. I., Reid, K., Milhaven, M., Bertino, T. S., Aguirre, W. E., Heins, D. C., Hippel, F. A. von, Park, P. J., Kirch, M., Absher, D. M., Myers, R. M., Palma, F. D., Bell, M. A., Kingsley, D. M., & Veeramah, K. R. (2021). Predicting future from past: The genomic basis of recurrent and rapid stickleback evolution. *Science Advances*, *7*(25), eabg5285. https://doi.org/10.1126/sciadv.abg5285

Kirch, M., Romundset, A., Gilbert, M. T. P., Jones, F. C., & Foote, A. D. (2021). Ancient and modern stickleback genomes reveal the demographic constraints on adaptation. *Current Biology*, *31*(9), 2027-2036.e8. https://doi.org/10.1016/j.cub.2021.02.027

Konečná, V., Bray, S., Vlček, J., Bohutínská, M., Požárová, D., Choudhury, R. R., Bollmann-Giolai, A., Flis, P., Salt, D. E., Parisod, C., Yant, L., & Kolář, F. (2021). Parallel adaptation in autopolyploid Arabidopsis arenosa is dominated by repeated recruitment of shared alleles. *Nature Communications*, *12*(1), 4979. https://doi.org/10.1038/s41467-021-25256-5

Lescak, E. A., Bassham, S. L., Catchen, J., Gelmond, O., Sherbick, M. L., von Hippel, F. A., & Cresko, W. A. (2015). Evolution of stickleback in 50 years on earthquake-uplifted islands. *Proceedings of the National Academy of Sciences*, *112*(52), E7204–E7212. https://doi.org/10.1073/pnas.1512020112

Maynard-Smith, J., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetics Research*, *23*(1), 23–35. https://doi.org/10.1017/S0016672300014634

McGee, M. D., Borstein, S. R., Meier, J. I., Marques, D. A., Mwaiko, S., Taabu, A., Kishe, M. A., O'Meara, B., Bruggmann, R., Excoffier, L., & Seehausen, O. (2020). The ecological and genomic basis of explosive adaptive radiation. *Nature*, *586*(7827), 75–79. https://doi.org/10.1038/s41586-020-2652-7

McPhail, J. D. (1994). Speciation and the evolution of reproductive isolation in the sticklebacks (Gasterosteus) of southwestern British Columbia. In *The evolutionary biology of the threespine stickleback. Edited by M.A. Bell and S.A. Foster.* Oxford University Press.

Miller, C. T., Beleza, S., Pollen, A. A., Schluter, D., Kittles, R. A., Shriver, M. D., & Kingsley, D. M. (2007). Cis-Regulatory Changes in Kit Ligand Expression and Parallel Evolution of Pigmentation in Sticklebacks and Humans. *Cell*, *131*(6), 1179–1189. https://doi.org/10.1016/j.cell.2007.10.055

Moyle, P. B. (1976). Fish introductions in California: History and impact on native fishes. *Biological Conservation*, *9*(2), 101–118. https://doi.org/10.1016/0006-3207(76)90043-4

Reid, K., Bell, M. A., & Veeramah, K. R. (2021). Threespine Stickleback: A Model System For Evolutionary Genomics. *Annual Review of Genomics and Human Genetics*, *22*, 357–383. https://doi.org/10.1146/annurev-genom-111720-081402

Schluter, D., & Conte, G. L. (2009). Genetics and ecological speciation. *Proceedings of the National Academy of Sciences*, *106*(Supplement 1), 9955–9962. https://doi.org/10.1073/pnas.0901264106

Schluter, D., Marchinko, K. B., Arnegard, M. E., Zhang, H., Brady, S. D., Jones, F. C., Bell, M. A., & Kingsley, D. M. (2021). Fitness maps to a large-effect locus in introduced stickleback populations. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(3), e1914889118. https://doi.org/10.1073/pnas.1914889118

Slatkin, M. (2008). Linkage disequilibrium—Understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics*, *9*(6), Article 6. https://doi.org/10.1038/nrg2361

Terekhanova, N. V., Logacheva, M. D., Penin, A. A., Neretina, T. V., Barmintseva, A. E., Bazykin, G. A., Kondrashov, A. S., & Mugue, N. S. (2014). Fast Evolution from Precast Bricks: Genomics of Young Freshwater Populations of Threespine Stickleback Gasterosteus aculeatus. *PLOS Genetics*, *10*(10), e1004696. https://doi.org/10.1371/journal.pgen.1004696

Wooldridge, T. B., Kautt, A. F., Lassance, J.-M., McFadden, S., Domingues, V. S., Mallarino, R., & Hoekstra, H. E. (2022). An enhancer of *Agouti* contributes to parallel evolution of cryptically colored beach mice. *Proceedings of the National Academy of Sciences*, *119*(27), e2202862119. https://doi.org/10.1073/pnas.2202862119

795 Wootton, R. J. (1984). *A Functional Biology of Sticklebacks*. University of California Press.

Supplementary Materials for

Adaptive standing genetic variation from disparate sources provides a substrate for rapid adaptation via hard and soft sweeps

Kirch M, Avdievich E, Venkataramani K, Kucka M, Thome R, Philippos G, Gaspar L, Peichel C, Archambeault S, Bell M, Veeramah K, Kingsley DM, Jones FC

The PDF file includes:

Table of Contents

1. Material and Methods

1.1. Stickleback samples

More than 9 000 marine sticklebacks were collected for this study in order to quantify standing genetic variation in marine sticklebacks. Global Positioning System coordinates of the respective locations are listed in Table S1. 5 200 sticklebacks were caught with minnow traps at Rabbit Slough (RABS) under Glenn Highway and Arness (Upper Salamatof) Lake in Alaska in June and July 2018 under the permit SF 2018-080 from Alaska Department of Fish and Game. Caudal fin clips (approximately $2mm^2$) were taken from 5 000 RABS sticklebacks. 100 RABS samples as well as 100 additional freshwater sticklebacks caught at Arness Lake were euthanized using tricaine methanesulfonate (MS222, supplier PHARMA Ltd.) and preserved in ethanol for morphological and genomic DNA analyses. A further 606 marine sticklebacks were sampled from Rabbit Slough before 2013 by Professor Mike Bell. Another 3,593 marine sticklebacks were collected at Puget Sound, Washington in 2016 by the Washington Department of Fish and Wildlife. Most of the Puget Sound individuals were sampled nearshore in a beach seine in full saltwater in Clam Bay near Manchester (Archambeault *et al.*, 2020). PGTS samples were preserved frozen by the Washington Department of Fish and Wildlife and Sophie Archambeault, and shipped to Tübingen for DNA extraction and analyses. Jukka-Pekka Verta caught sticklebacks from Little Campbell River at 0.5 km and 23.32 km from the river mouth in 2016. *In vitro* fertilisation crosses between marine and freshwater sticklebacks were performed in the field and the offspring were raised in the fish facility at the Max Planck Campus in Tübingen (Competent authority: Regierungspräsidium Tübingen, Germany; Permit and notice numbers 35/9185.82-5, 35/9185.46; Verta & Jones, 2019). DNA of the parents as well as DNA extracted from fin clips of four offspring were used as controls in the described experiments.

1.2. Genomic DNA extraction

Fin clips of the collected sticklebacks were used for DNA extraction through a Solid Phase Reverse Immobilization (SPRI) bead-based protocol (DeAngelis *et al.*, 1995). Specifically, samples were lysed in 400 µl lysis buffer (50mM Tris-HCl pH=8, 0.1M NaCl, 10mM EDTA pH=8, 0.8% SDS, 15 µL of 10mg/mL proteinase K) for 5 hours at 58°C. Then 2 µL 10mg/mL RNAse A was added to each sample for RNA digestion at 37°C for 30 min. After adding 150 µl 5M potassium acetate, the samples were stored at 4°C overnight to precipitate proteins. The precipitate was discarded after centrifugation at 4000 g for 30 min on the next day and an equal volume of homemade SPRI beads were added to the supernatant to extract DNA. SPRI beads with DNA binding on them were collected using magnetic racks and were washed twice with 80% ethanol. Eventually, DNA was eluted from the beads with 10mM Tris pH 8.0 buffer.

The DNA extraction protocol as described above was carried out in 96 well plates, partly using TECAN® liquid handling robot for pipetting. The extracted DNA was quantified by TECAN® PicoGreen plate reader in order to verify whether the concentrations were sufficient for proceeding with the library preparation. If sufficient, subsequently, pull-down sequencing was performed with DNA from marine sticklebacks. In addition, 39 of 100 freshwater sticklebacks sampled at Arness Lake in Alaska as well as 57 identified marine carriers of freshwater-adaptive alleles on chrIV or within the chrI inversion were subjected to whole-genome sequencing using haplotype tagging to recover haplotype informative data (see chapter 1.7) following the protocol outlined in Meier *et al.* (2021).

1.3. Pull-down sequencing of marine stickleback samples

1.3.1. Library preparation

Sequencing libraries for high-throughput sequencing of almost 10 000 samples were prepared following the manufacturer's instruction with Nextera DNA Library Kit (Illumina, Inc, San Diego, USA). Tn5 transposase was expressed and purified in-house as described in Picelli *et al.* (2014). Tn5 transposome assembled with biotinyated Tn5-oligos (Table S4) were attached to Hydrophilic Streptavidin Magnetic Beads (NEB, S1421) to prepare Tn5-on-beads. Tn5-on-beads was then used for tagmenting 5-100 ng of input DNA to an average fragment size between 300 and 500 bp. Since Tn5 bound to magnetic beads breaks the DNA strands at regular intervals independent of the ratio of Tn5 to DNA, an additional normalisation step before library preparation was avoided.

During the tagmentation step, fragments were tagged with adaptors carrying complementary binding regions for the indexing primers used in the polymerase chain reaction (PCR). After tagmentation, beads were washed once with 0.4% SDS in washing buffer (10 mM Tris,pH=8, 30 mM NaCl, 0.1 % Triton X-100) and once with washing, the tagged fragments were amplified in a 9-cycle PCR using sample-specific index adapter sequences as primers (Table S4). (All oligonucleotides used in this study were synthesised by Integrated DNA Technologies, Coralville, Iowa, USA.) This way each sample was individually barcoded and the adaptor sequences served moreover as sequencing adaptors for the Illumina sequencing platform.

Following Ampure XP® (Beckman-Coulter) DNA size selection protocol fragments between 250-500bp were size-selected from the amplified DNA using in-house made SPRI beads. The size-selected PCR products were quantified by a TECAN® PicoGreen plate reader. Equal amounts of DNA from up to 384 amplified libraries from four 96 well plates were then pooled with each individual sample carrying a unique barcode combination (27 library pools total). Controls from Little Campbell River (marine mother, freshwater father and 4 F1s) were added into each library pool in order to quantify pull-down bias in the next step.

1.3.2. Preparation of biotinylated BAC probes

This project involves high throughput sequencing of targeted genomic regions in a large number (>9 000) of stickleback individuals. Since whole genome sequencing outnumbered the financial limits of this project, biotinylated probes were prepared from bacterial artificial chromosomes (BACs) and used for pulling down genomic regions of interest in the next step. Freshwater and marine stickleback BAC libraries with 72 000 clones and 107 087 clones were created in 2004 based on one freshwater benthic fish collected from Paxton Lake on Texada Island (CHORI-215) and based on approximately 60 individuals from an anadromous population from the Salmon River in British Columbia (CHORI-213), respectively (Kingsley *et al.*, 2004; Kingsley & Peichel, 2007). The ends of the BAC clone inserts were Sanger sequenced and paired BAC end sequences were mapped against the reference stickleback genome to infer the genomic locus carried by each BAC clone. BAC clones spanning 236 regions previously identified to be parallel divergent between global freshwater and marine sticklebacks (Jones *et al.*, 2012; FDR 0.05). 10 additional neutral regions were then selected and used as templates in the production of targeted pulldown probes. In fact, Jones *et al.* identified 242 divergent regions based on the reference genome gasAu1, which correspond to 240 regions in gasAcu1-4 coordinate space (Table S2). Since we wanted to identify heterozygous and homozygous carriers of rare freshwater-adaptive alleles in marine populations, it was crucial to maximise the likelihood to pull down freshwater and marine alleles in equal measure. To do so, a pair of marine and freshwater BAC clones was selected for each region to ensure probes capture sequence diversity present at adaptive loci and to minimise any allelic bias in targeted pulldown. A total of 278 BAC clones were picked from the BAC libraries stored in Stanford and transferred to Tübingen (Table S3).

After culturing each BAC clone overnight with chloramphenicol antibiotic, DNA was extracted from the culture with the Plasmid DNA Miniprep Kit following manufacturer's instruction (Thermo Fisher Scientific), whereby the purification of the BAC DNA was performed with in-house made SPRI beads. Subsequently, the ends of the stickleback DNA inserts were amplified with specifically designed primers to verify whether the BAC clone was carrying the corresponding insert of stickleback DNA.

In the following step, each selected BAC was used as a template to produce short biotinylated probes. Specifically, a dNTP mix with 20% of dTTP being replaced by Biotin-16-dUTP (Jena Biosciences, NU-803-BIO 16-S) as well as random octamer oligonucleotides were added to the 100ng of BAC DNA. After centrifugation for 5 min at 95°C, Klenow (exo-) fragment of DNA polymerase I (NEB) was added to the mixture and the tube was incubated at 37°C for 3 hours making biotinylated probes from the BAC DNA. After another incubation step (5 min at 95°C) probes of length 150-500 bp were selected by size selection with SPRI beads resulting in approximately 15-25 ng/µl of biotinylated probes in 40 µl for each BAC. The resulting probes

were measured on NanoDrop™ with the settings for ssDNA and equal amounts of all biotinylated BAC probes were pooled together into one BAC pool, which was used for the enrichment of 246 loci by biotinylation capture in the next step.

1.3.3. Enrichment of 248 loci by biotinylated BAC Hybridization

The biotinylated BAC probes were used for enrichment of previously identified divergent regions (Jones *et al.*, 2012; FDR 0.05) and flanking regions (mean flank size of 145 616 bp) for each marine individual and spanned in total 26.7 Mb corresponding to 5.8% of the stickleback genome (protocol adopted from Alvarado *et al.* (2014)). Sizes of divergent regions range from 175 bp to 129 261 bp with a mean size of 10 089 bp (Table S2). Enrichment was performed by a double pull-down for each of 27 nextera library pools, each containing up to 384 individual libraries and 6 controls from Little Campbell River (freshwater mother, marine father and 4 F1s that were used in quality control assessment of pulldown and enabled estimates of allelic bias in pulldown and variance across pulldown batches).

In the first step, the biotinylated BAC probe pool as well as the nextera library pool were prepared for the hybridization. 40 ng probes for each of 278 BAC clones, which corresponds to 11 120 µg of the biotinylated BAC probe pool, were used for one pull-down and pipetted into a PCR strip tube. 1 µg of a nextera library pool were combined with 15 µl of blocking 3-oligo pool containing 33 µM for each of 3 oligos blocking the i5 and i7 adapters attached to the nextera libraries in a PCR strip tube.

DNA in both pools was denatured at 95°C for 5 min. After incubation at 65°C for 15 min, the volume of both pools was doubled by adding pre-warmed 2x hybridization buffer (1.5 M NACl, 40mM sodium phosphate buffer pH 7.2, 10mM EDTA pH 8.0, 10x Denhardt's, 0.2% SDS) and the pools were pipetted together. The combined pool was incubated at 65°C for 70 hours to enable hybridization of the biotinylated BAC probes with the complementary parts of the nextera libraries.

After 70 hours of hybridization 300ul binding buffer (10mM Tris-HCl pH 7.5, 1mM EDTA pH 8, 1M NaCl) were added to the combined pool and everything was transferred to 100ul streptavidin C1 beads, which were previously washed with 200ul binding buffer. In order to bind the streptavidin beads to the biotinylated DNA, the mixture was incubated for 30 min at room temperature, while gently rotating. Beads were washed once with 1X SSC containing 0.1% SDS with rotation for 15 min at room temperature and another 3 times with 0.1X SSC containing 0.1% SDS, each time for 15 min at 65°C. To release the pulled-down library from the biotinylated probes, elution was performed with 50ul 0.15 M NaOH for 10 min at room temperature. The elution was transferred to a Lo-Bind tube and DNA was purified with in-lab-made SPRI beads. The pulled-down library was amplified by 9 cycles of PCR with Q5 polymerase and TrueSeq i5 and i7 primers resulting in approximately 1ug of DNA (Table S4).

Subsequently, another pull-down was performed to enrich from amplified DNA after 1st pull-down. This step helped minimise the amount of genomic background retained and maximise the ratio of on-target:off-target library for subsequent sequencing. For the second pull-down, the pulled-down library pool resulting from the first pull-down was used as input and all steps were the same as described above with the only difference of 6 cycles of PCR for amplification in the last step. The concentration of the double pulled-down library was measured on Invitrogen™ Qubit™ Fluorometer and adjusted for high-throughput sequencing by a HiSeq 3000 (Illumina) at the Genome Core Facility at the MPI Tübingen Campus.

1.4. Processing of pull-down sequencing data

1.4.1. Initial read processing

Pull-down libraries were sequenced in 27 HiSeq lanes generating a total of 40 784 million 150bp paired-end read-pairs. 86.4 % of reads could be confidently assigned to expected barcodes. Parallel computing at the Max Planck facilities in Tübingen enabled processing of this large dataset of 1.7 TB. The Illumina bcl2fastq Conversion Software converts BCL files into FASTQ files and was used to demultiplex the sequenced pull-down libraries, as well as to remove adapter sequences and Unique Molecular Identifier (UMI) bases for each read (representative command see 1.9.1). Subsequently, reads from each sample were aligned against the stickleback reference genome gasAcu1-4 (assembly generated from a female freshwater stickleback from Bear Paw Lake, Alaska; Jones *et al.*, 2012; Kingman *et al.*, 2021) using Burrows-Wheeler-Aligner (BWA) mem algorithm (Li & Durbin, 2009). Next, mapped reads were sorted and indexed using SAMtools (Li *et al.*, 2009; representative command see 1.9.2). Further read processing until variant calling was performed following the best practices recommended by Genome Analysis Tool Kit (GATK) version 4.1.8.1 (Auwera & O'Connor, 2020; DePristo *et al.*, 2011; McKenna *et al.*, 2010). Briefly, optical and technical read duplicates were identified using GATK MarkDuplicates (Picard) function with lenient validation stringency settings and including creating a BAM index (Auwera & O'Connor, 2020; McKenna *et al.*, 2010). Newly created bam files were sorted by coordinate using GATK SortSam function (Auwera & O'Connor, 2020; McKenna *et al.*, 2010) and adjacently indexed using SAMtools (Li *et al.*, 2009; representative command see 1.9.3). Base Quality Score Recalibration (BQSR) is a command to re-calibrate the estimates of accuracy of base calls in order to correct systematic errors made by the sequencing machine (DePristo *et al.*, 2011). The recalibration requires a database of known variants which is not yet available for sticklebacks and thus, this step was omitted in the pipeline. Representative commands for each step are available in chapter 1.9.

1.4.2. Variant calling and filtering

Following initial read processing, SNPs and indels were called for each sample individually by using the HaplotypeCaller function in GATK version 4.1.8.1. (Poplin *et al.*, 2017). For running HaplotypeCaller, the heterozygosity was set to 0.04 and the corresponding standard deviation to 0.1 (representative command see 1.9.4). Since the samples were sequenced at low coverage, all dangling branches were recovered and pruning was disabled in order to identify all variants. The resulting Genomic Variant Call Format (GVCF) file encompassed records for all sites. Subsequently, GVCF files for all individuals were imported into 112 Genome Databases by GATK version 4.1.8.1 function GenomicsDBimport (Auwera & O'Connor, 2020; McKenna *et al.*, 2010; representative command see 1.9.5). Each of the 112 Genome Databases corresponds to a proximity merged region spanning pull-down regions not further than 100 kb apart. In total, the proximity merged regions span 28.3 Mb, 6 % of the stickleback genome. Joint genotype files for each of these 112 proximity merged regions were exported from the respective Genome Database using the GenotypeGVCFs option in GATK version 4.1.8.1. (Auwera & O'Connor, 2020; McKenna *et al.*, 2010; representative command see 1.9.6). Heterozygosity for SNP and indel calling was set to 0.04 and standard deviation of heterozygosity to 0.1. Two alleles alternate to genotype were allowed at maximum. The resulting Variant Call Format (VCF) files comprised information for a total of 9 172 individuals including 3 545 PGTS and 5 627 RABS at 21 481 114 variant positions. Representative commands for each step are available in chapter 1.9.

1.5. Identifying freshwater-adaptive alleles

Two distinct approaches were adopted to identify carriers of freshwater-adaptive standing genetic variation: Bayesian probability of ancestry and Hidden Markov Models (Figure 2). The following analysis was partially performed in R version 4.0.3 (R Core Team, 2020).

1.5.1. Method 1: Bayesian probability of ancestry

Script names used in the following sections refer to the corresponding script available at https://github.com/MelanieKirch/standing-genetic-variation/tree/main/Bayesian%20Probability
For the Bayesian probability approach, SNPs with a minimal allele frequency difference of 0.5 between marine and freshwater populations were used to assign marine or freshwater ancestry to local genomic windows in each of the 9 172 fish. This approach works particularly well in genomic regions of highly divergent marine and freshwater haplotypes which contain many ancestry-informative SNPs.
A set of 581 194 ancestry informative SNPs (124 005 within proximity merged regions) was identified using criteria of allele frequency difference above 0.5 in 24 local Canadian marine and freshwater individuals (Table S5). The data for these samples were downloaded from the

Sequence Read Archive (www.ncbi.nlm.nih.gov/sra) at accession PRJNA247503 (Kingman *et al.*, 2021) and processed following the steps outlined in the markdown file PART1.md included in the GitHub repository. SNP data had to be available for at least 9 out of 12 individuals for each ecotype in order to be considered.

Following the GATK pipeline, variant call files for each of 112 proximity merged regions spanning all 240 divergent regions were created for all 9 172 fish (see section 1.4. above). Variant call positions from these files were intersected with the positions of ancestry informative SNPs. The Canadian ancestry informative SNP set only encompasses bi-allelic SNPs and in order to assign ancestry to each allele, variant call files for all 9 228 fish had to be adjusted. This means, in particular, that variant call files had to be downsampled to bi-allelic SNPs and further to SNPs with identical alleles compared to the Canadian ancestry informative SNP set. In order to minimise the loss of SNPs during downsampling, tri-allelic variants from the 9 172 fish were split into two bi-allelic SNPs respectively and indels with only one varying position were transformed into bi-allelic SNPs. All other indels were excluded. SNPs with data missing for more than 80% of the individuals were dismissed and 409 individuals with sparse data were excluded. This resulted in VCF files containing information for a total of 8 763 individuals including 3 273 PGTS and 5 490 RABS at 297 184 variant positions. VCF files for 7 proximity merged regions (spanning 8 divergent regions) comprised fewer than 35 ancestry informative SNPs and were excluded, since we would have little power to confidently assign local genomic ancestry. All commands used in this section are also outlined in the markdown file PART2.md included in the GitHub repository.

Calculations for the assignment of ancestry states

Bayesian probabilities were used to determine the likelihood of an individual being of homozygous marine, homozygous freshwater or heterozygous ancestry based on the given data. At first, the genotype likelihood, i.e. the probability of a genotype given the sequencing data P(G|D), for each fish and each position was used as input. This probability differs by a constant factor from the probability of getting certain sequencing data given a genotype (see Formula 1 based on the Bayes' theorem (Bayes & Price, 1763)).

$$P(D|G) = P(G|D) * \frac{P(D)}{P(G)} \qquad (1)$$

with D – data, G – genotype and P(D)/P(G) being constant.

Based on the genotype likelihood the probability of getting specific sequencing data and a specific genotype under a given ancestry P(D,G|A) was calculated following the Bayes' theorem (Bayes & Price, 1763) (Formula 2).

$$P(D,G|A) = P(D|G,A) * P(G|A) = \mathrm{P(D|G)} * \mathrm{P(G|A)}$$

$$P(D,G|A) = \mathrm{P(G|D)} * \frac{P(D)}{P(G)} * \mathrm{P(G|A)} \qquad (2)$$

with A – ancestry, D – data and G – genotype.

In the following *00, 01, 11* represent the different genotypes with *0* representing the marine allele and *1* the freshwater allele, respectively. *Mm, mf, ff* represent the different ancestries (marine homozygous, marine/freshwater heterozygous, freshwater homozygous). Furthermore, *p(m)* is the frequency of an allele in the marine population and *p(f)* is the frequency of the same allele in the freshwater environment.

Applying formula 2 to all combinations of genotypes and ancestries gives the following:

$P(D,00|mm) = p(m) * p(m) * P(00|D)$

$P(D,01|mm) = p(m) * \big(1 - p(m)\big) * 2 * P(01|D)$

$P(D,11|mm) = \big(1 - p(m)\big) * \big(1 - p(m)\big) * P(11|D)$

$P(D,00|mf) = p(m) * p(f) * P(00|D)$

$P(D,01|mf) = (p(m) * \big(1 - p(f)\big) * P(01|D)) + (p(f) * \big(1 - p(m)\big) * P(01|D))$

$P(D,11|mf) = \big(1 - p(m)\big) * \big(1 - p(f)\big) * P(11|D)$

$P(D,00|ff) = p(f) * p(f) * P(00|D)$

$P(D,01|ff) = p(f) * \big(1 - p(f)\big) * 2 * P(01|D)$

$P(D,11|ff) = \big(1 - p(f)\big) * \big(1 - p(f)\big) * P(11|D)$

The factor $\frac{P(D)}{P(G)}$ is cut later (see Formula 5) and is thus ignored in this calculation.

In the next step, the likelihood to observe the given data under a given ancestry *P(D|A)* was calculated by summing the respective probabilities for each genotype (see Formula 3).

$$P(D|A) = \sum_{G} P(D,G|A) \qquad (3)$$

with A – ancestry, D – data and G – genotype.

For instance, the probability of obtaining the data under a marine homozygous ancestry was calculated as presented in formula 4.

$$P(D|mm) = P(D,00|mm) + P(D,01|mm) + P(D,11|mm) \qquad (4)$$

Subsequently, the likelihood of each ancestry based on the given data *P(A|D)* was calculated based on the Bayes' theorem (Bayes & Price, 1763; Formula 5).

$$P(A|D) = \frac{P(D|A) * P(A)}{P(D)} \qquad (5)$$

with A – ancestry and D – data.

The probability $P(D|A)$ was already available from the previous calculations (see Formula 3). In order to avoid biasing the results towards one ancestry, the probability of ancestry $P(A)$ was set to ⅓ for each of the three ancestries. The likelihood of the data $P(D)$ was calculated by summing the probabilities of data given the ancestry over all ancestries (see Formula 6).

$$P(D) = \sum_{A} P(D|A) \qquad (6)$$

with A – ancestry and D – data.

$\frac{P(D)}{P(G)}$ is constant and is appearing as factor in the numerator and denominator in the fraction after inserting the respective formulas 1,2,3 and 6 in formula 5. It is thus irrelevant for the calculation.

Ultimately, the probabilities at each position are normalised in order to sum up to 1 over all three ancestries (script: 1_Prancestry_Likelihood_for-multiple-SNPs.awk). To level out the noise and smooth the data, probabilities of each ancestry (homozygous marine, heterozygous, freshwater homozygous) were multiplied and normalised within each window of 30 SNPs moving over a region with a step size of 1 SNP (script: 2_PrAncestry_30SNPwindows.R).

Correction factors to account for potential pull-down bias

One major challenge of enrichment data is the bias introduced by the pull-down. For instance, the marine allele can be favourably pulled down by the BAC clones and thus a fish is called to be marine homozygous in a given SNP window, although it is in fact freshwater/marine heterozygous. In order to account for this enrichment bias, a control family consisting of a marine mother, a freshwater father and four F1s was included into each pull-down providing information about how often a state was falsely assigned. The family allowed us to calculate the likelihood of the "true ancestry" tA given the assigned ancestry aA (see Formula 7).

$$P(tA|aA) = \frac{P(tA \wedge aA)}{P(tA \wedge aA) + P((\neg tA) \wedge aA)} \qquad (7)$$

At first, the most likely ancestry state was determined for each of the 26 replicates of each control. This resulted in a table indicating how often each state was assigned to each control. The further calculations then depended on the most frequent ancestry states present in the parents of the control family (Table S6).

	Percentage of affected SNP windows	Most frequent ancestry state assigned to marine mother	Most frequent ancestry state assigned to freshwater father
Case 1	91.4	marine homozygous	freshwater homozygous

Case 2	3.5	marine homozygous	heterozygous
Case 3	3.5	heterozygous	freshwater homozygous
Case 4	0.8	marine homozygous	marine homozygous

Table S6: Different cases for which ancestry states are most frequent in the parents of the control family in a given SNP window and their abundance i.e. the percentage of affected SNP windows for each case.

In case 1 (91.4% of the SNP windows) the marine mother is most frequently assigned to be marine homozygous and the father is most frequently assigned to be freshwater homozygous (Table S6). The F1s are thus expected to be marine/freshwater heterozygous. A hypothetical scenario for case 1 is shown in Table S7.

	Marine mother	Freshwater father	F1 Number 1	F1 Number 2	F1 Number 3	F1 Number 4
marine homozygous	24	0	2	1	1	2
heterozygous	2	0	24	25	25	24
freshwater homozygous	0	26	0	0	0	0

Table S7: Hypothetical scenario for case 1. Each column represents one of the six controls and each row represents one ancestry state. The numbers in the table are representing how many replicates of each control were assigned to each ancestry state.

In the example shown in Table S7 the marine mother is wrongly called to be heterozygous in 2 of the 26 replicates. Hence, the true marine homozygous ancestry was rightly assigned as marine homozygous ancestry 24 out of 26 times for the marine mother ($P(tA \wedge aA)$ in formula 7) and wrongly assigned as marine homozygous 6 out of 104 times for the F1s ($P((\neg tA) \wedge aA)$ in formula 7). The correction factor for the different ancestry states can thus be calculated as presented in formula 8-10.

$$P\big((tA = mm)\big|(aA = mm)\big) = \frac{\frac{24}{26}}{\frac{24}{26} + \frac{0}{26} + \frac{6}{104}} = 0.94 \quad (8)$$

$$P\big((tA = mf)\big|(aA = mf)\big) = \frac{\frac{98}{104}}{\frac{2}{26} + \frac{0}{26} + \frac{98}{104}} = 0.92 \quad (9)$$

$$P\big((tA = ff)\big|(aA = ff)\big) = \frac{\frac{26}{26}}{\frac{0}{26} + \frac{26}{26} + \frac{0}{104}} = 1 \quad (10)$$

In case 2 (3.5% of the SNP windows) the marine mother is most frequently assigned to be marine homozygous and the father is most frequently assigned to be heterozygous (Table S6). Since the underlying SNPs are identified to show a significant allele frequency difference between Canadian freshwater and marine individuals, they do not necessarily represent freshwater-marine divergence, but could also be identified due to local variance. The heterozygous state for the father is thus assumed to be true. The F1s are hence either expected to be marine homozygous or to be heterozygous. This case doesn't provide any information about whether the freshwater homozygous state is correctly assigned and the correction factor for this state is thus assumed to be 1.

In case 3 (3.5% of the SNP windows) the marine mother is most frequently assigned to be heterozygous and the father is most frequently assigned to be freshwater homozygous (Table S6). This case is treated analogously to case 2.

In case 4 (0.8% of the SNP windows) both parents are most frequently assigned to be homozygous marine (Table S6). In this case as in all other undescribed cases (0.8% of the SNP windows) all correction factors were set to 1.

Generation of final tables

The correction factors were calculated for each SNP window for each ancestry and multiplied with the probabilities for each ancestry state for each individual (script: 3_PrAncestry_CorrectionFactorControls.R). Subsequently, the ancestry genotype with the highest probability and with a difference >0.6 to the second highest probability was assigned to the respective SNP window for each individual (script: 3_PrAncestry_CorrectionFactorControls.R). NAs were assigned to the windows not fulfilling these criteria. The resulting tables for each of the 112 merged regions included data for 3 273 PGTS and 5 490 RABS at each SNP window and segmented the region for each individual into homozygous marine, heterozygous and freshwater homozygous ancestry states (Figure 2C). These tables were subsequently used for deciding if an individual is carrying a freshwater-adaptive allele at a given divergent region (see chapter 1.6.1). For determining the length of haplotypes NA windows were filled up (script: 5_PrAncestry_fillgaps.R). States for NA windows were chosen based on the state of the neighbouring windows. If the neighbouring windows displayed different ancestry genotypes, the most likely position with a switch between the two

different states was determined based on the underlying probabilities for each ancestry state. All commands used in this section are also outlined in the markdown file PART3.md. Scripts used in this work as well as markdown files outlining the steps are available at https://github.com/MelanieKirch/standing-genetic-variation/tree/main/Bayesian%20Probability.

1.5.2. Method 2: Hidden Markov Model

The existence of highly divergent marine and freshwater haplotypes among stickleback ecotypes provides detectable molecular features that we can use to identify fish carrying derived alleles in homozygous or heterozygous form. At loci underlying marine-freshwater adaptive divergence identified in Jones *et al.* (2012), the pairwise nucleotide diversity between marine and freshwater individuals is elevated (median $pi_{between}$ greater than 0.01; Figure S13A purple line) relative to pairwise nucleotide diversity within ecotypes (median pi_{within} 0.001; Figure S13A red and blue lines). In contrast, at loci that were not identified to be parallel adaptively divergent in Jones *et al.* (i.e. more "neutral loci"), there is only modest differences among marine and freshwater fish (pairwise nucleotide diversity between ecotypes is approximately equivalent to pairwise nucleotide diversity within, and considerably lower than that observed between ecotypes at adaptive loci; Figure S13B).

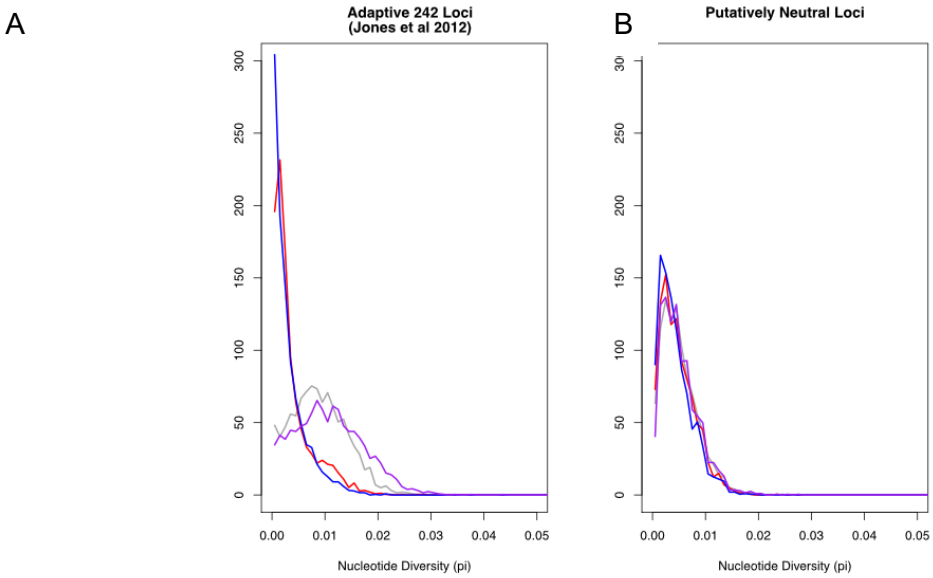


Figure S13: Elevated pairwise nucleotide diversity between marine and freshwater individuals at adaptive loci. The pairwise nucleotide diversity between marine and freshwater individuals from Rabbit Slough and Arness Lake 2018 samples respectively, shown for marine-freshwater parallel adaptive loci (Jones *et al.*, 2012; A) and putatively neutral loci (B, loci not detected as showing parallel adaptive divergence in Jones *et al.* (2012)). Pairwise nucleotide diversity was calculated within ecotypes (among Rabbit Slough individuals, red line; among Arness Lake individuals, blue line), between ecotypes (between Arness Lake and Rabbit Slough individuals, purple line), and among all individuals (grey line) using vcftools (Danecek *et al.*, 2011) windowed-pi command with 10kb windows and variant call files of whole genome short-read sequencing data.

This indicates deep evolutionary divergence among marine and freshwater individuals at adaptive loci, and has been reported at the haplotype level at focal loci underlying major differences in skeletal morphology (for example, EDA locus (Colosimo *et al.*, 2005)). In their study, Colosimo *et al.* (2005) used a molecular clock to date the divergence time between marine and freshwater haplotypes at the EDA locus to be 5 million years before present.

The deep evolutionary divergence provides two semi-related molecular signals in a fish carrying derived alleles: an excess of minor alleles (i.e. derived, non-marine alleles) and an excess of heterozygosity that when considered together can be used to segment the genome of individual fish into four different states: neutral/non-divergent, marine homozygous, marine/freshwater heterozygous and freshwater homozygous (see Table S8 and Figure S14 for examples from two genomic regions).

State	Derived allele density	Heterozygous site density
neutral/non-divergent	mid/low	mid/low
Marine homozygous	low	low
marine/freshwater heterozygous	mid	high
Freshwater homozygous	high	low

Table S8: A Hidden Markov Model with four different states was used to segment the genomes of each individual based on the molecular signatures of density of derived alleles and density of heterozygous sites. While the marine and freshwater homozygous states both have low density of heterozygous sites, freshwater homozygous regions are expected to have a comparatively high density of derived alleles. In contrast, heterozygous regions have comparatively high density of heterozygous sites and a moderate density of derived alleles. Neutral states are not enriched for derived alleles or heterozygous sites.

A multivariate Hidden Markov Model using both the density of derived alleles and the density of heterozygous sites per 10 kb window is trained to find these four states and segment the genomic data of individual marine fish. To enable direct comparisons between the three assignment categories used in the Bayesian ancestry assignment approach and the HMM method, a 3-state HMM was also performed in which the neutral state was excluded from the model. For both HMMs, variant call files were produced for genomic pulldown data of all fish using GATK pipeline (see section 1.4 above). For windows of 10 kb, stepping by 5 kb spanning the targeted pulldown regions, counts of the density of derived alleles and density of heterozygous sites were calculated using the bcftools +sampl-stats plugin (Danecek *et al.*, 2021). Specifically, for each variant, the ancestral allele was identified as the most frequent allele in all marine samples. The density of derived alleles per fish per window was then estimated by dividing the derived allele count by the total allele count, and excluding windows

with less than 2500 alleles called. Since targeted enrichment and sequencing of genomic regions by probe pull-down produces highly variable read coverage (mean depth per individual per site = 4.6+/- 1.9 SD), sites with diploid genotype calls but only 1 read coverage were treated as haploid. The density of heterozygous sites per fish per window was calculated by dividing the number of heterozygous sites by the total number sites, considering only genotype calls based on two or more reads and excluding windows with less than 1250 such sites. For both estimates, variants with excessively high or low heterozygosity across individuals within Rabbit Slough and Puget Sound were not considered (p-value for deviation from Hardy Weinberg Equilibrium <= 0.05 and/or the number of heterozygous genotype calls was greater than 1.15 times that expected under Hardy Weinberg Equilibrium). To reduce signal noise, the density of heterozygous sites and derived alleles was then transformed by smoothing and z-standardization, and values used in HMM described below.

	mean derived allele density	standard deviation derived allele density	mean heterozygous site density	standard deviation heterozygous site density
4-state HMM				
Homozygous marine	-0.07211	0.01649	-0.04447	0.01649
Heterozygous marine/freshwater	0.00899	0.01139	0.04215	0.01510
Homozygous Freshwater	0.06197	0.01044	-0.04611	0.02073
Neutral	-0.00250	0.00792	-0.00500	0.01375
3-state HMM				
Homozygous marine	-0.07231	0.01632	-0.04398	0.01698
Heterozygous marine/freshwater	0.00816	0.01157	0.04283	0.01568
Homozygous Freshwater	0.06127	0.01055	-0.04688	0.02149

Table S9. Emission matrices used in 4-state (upper rows) and 3-state (lower rows) Hidden Markov Models to segment individual marine genomes into homozygous marine, homozygous freshwater, heterozygous marine/freshwater and neutral states.

Hidden Markov Models were applied to the continuous data (derived allele density and heterozygous site density per 10 kb window) using R package depMixS4 (Visser & Speekenbrink, 2010). Specifically, the transformed density of derived alleles and density of heterozygous sites were used as dependent variables with gaussian distributions. Three- and four-state HMM model emission matrices were constructed using means and standard deviations of transformed derived allele density and heterozygous site density (described above) from a control F1 nuclear family comprising marine and freshwater parents and four F1 progeny (see Figure S14 for example genomic regions). The emission parameters (Table S9) for “marine homozygous”, “freshwater homozygous”, and “marine/freshwater heterozygous” were estimated from 10 kb windows overlapping adaptive loci in the marine parent, freshwater parent, and F1 progeny respectively while the emission parameters for the fourth neutral state were estimated from the marine and freshwater parents at neutral loci (Table S9).

The transition probability matrix was constructed to allow transitions between states by accounting for local genomic recombination rate of each region (probability of recombination switch event from one window to the next based on Mb-scale recombination map obtained from an marine x freshwater F2 intercross genetic mapping panel) and the expected genotype frequencies under a population at Hardy-Weinberg Equilibrium with starting allele frequencies of 0.0001 and 0.9999 for freshwater marine adaptive alleles, respectively. A viterbi algorithm was used to compute the maximum posterior state sequence for each genomic region in each individual fish.

Finally, we excluded genomic regions showing excessive variability in state assignment across the 26 technical genomic pull-down replicates performed for each individual of the nuclear family. Specifically, regions were retained if all 6 individuals in the family had 24 or more technical replicates called the same state. For the 3-state HMM model, the resulting tables for each of the 112 merged regions encompassed data for 3 273 PGTS and 5 490 RABS spanning 142 divergent regions and segmented the region for each individual into homozygous marine, heterozygous and freshwater homozygous ancestry states (Figure 2C). Scripts used in this work as well as a readme file outlining the steps are available at https://github.com/MelanieKirch/standing-genetic-variation/tree/main/Hidden%20Markov%20Model . Scripts include vcf_fill-tags.array.sh (annotating the VCF file with population specific allele frequencies), countingHetAndDerivedAlleles.v7.sh (counting heterozygous sites and derived alleles per fish per genomic window) and HMM.parts1to3.R (for HMM segmentation of data).

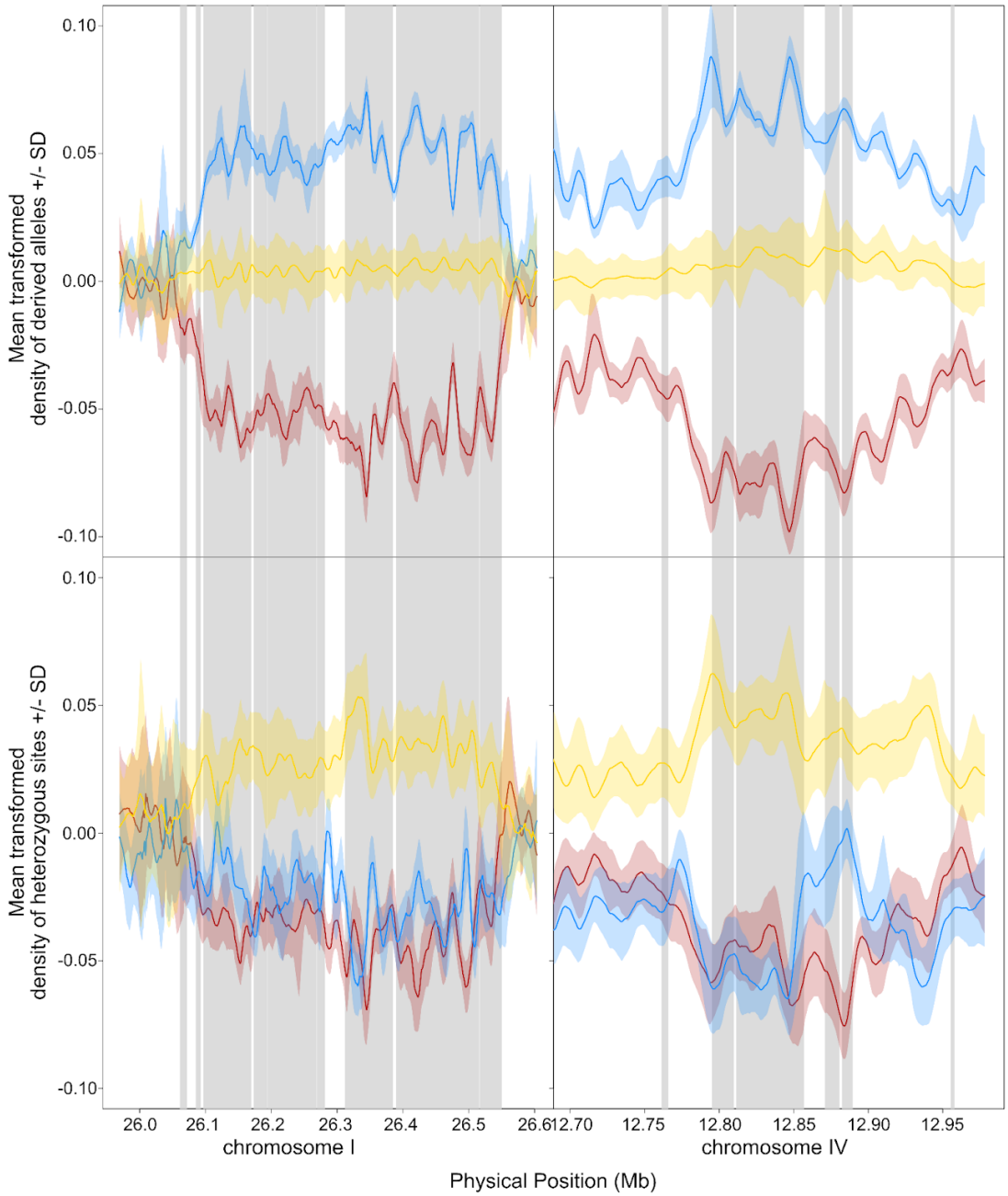


Figure S14: A Hidden Markov Model to segment individual genomes into homozygous marine, homozygous freshwater, heterozygous and neutral states was trained using control data from a marine x freshwater nuclear family. At adaptive loci (grey rectangles), the homozygous marine parent (red) shows a low density of both derived alleles and heterozygous sites, the homozygous freshwater parent (blue) a low density of heterozygous sites but high density of derived alleles, and the heterozygous F1 offspring (yellow) show a high density of heterozygous sites and medium density of derived alleles. Shading shows standard deviation around mean values obtained from 26 technical replicates of genomic pulldown for each individual. Neutral regions (non-adaptive) show medium values of both derived alleles and heterozygosity. Left panels show a region spanning an adaptive inversion on chromosome I. Right panels show the EDA region on chromosome IV.

1.6. Further data analysis

The following analyses were performed in R version 4.0.3 (R Core Team, 2020).

1.6.1. Carrier status classification

The Bayesian probability as well as the Hidden Markov Model approach subset each of 112 genomic regions into windows (30 SNP windows and 10 kb windows, respectively) and assigned marine homozygous, heterozygous and freshwater homozygous ancestry states to each window (see chapter 1.5). This segmentation allowed us to classify individuals into three groups (heterozygous, freshwater homozygous or non- carriers) at 232 divergent regions for the Bayesian probability approach and at 142 divergent regions for the Hidden Markov Model approach (Table S2; script: https://github.com/MelanieKirch/standing-genetic-variation/blob/main/Bayesian%20Probability/4_PrAncestry_carrier.R). To do so, all windows overlapping with a given divergent locus and their respective ancestry states were counted. An arbitrary threshold of 25% for the proportion of windows of freshwater ancestry in an individual was applied to the segmented data to assign this individual to a group. If the count of heterozygous windows added on the duplicated count of freshwater homozygous windows divided by the duplicated count of all windows was equal or bigger than 25%, an individual was classified as a carrier (Formula 11). In case the number of freshwater homozygous windows was higher than the number of heterozygous windows, the individual was classified as a freshwater homozygous carrier. Otherwise, it was a heterozygous carrier.

$$\frac{\#(heterozygous\ windows) + 2 * \#(homozygous\ windows)}{\#(all\ windows)} \geq 0.25 \quad (11)$$

Since the functional variants within the divergent regions are mostly unknown, it is unclear which part of a given locus is relevant for adaptation and represents the freshwater-ness in a stickleback genome. However, a classification into carriers and non-carriers enables us to focus on a subset of individuals which are carrying a significant freshwater ancestry fraction at a given locus. Furthermore, the threshold of 25% enables us to also include individuals that are carrying an incomplete haplotype.

1.6.2. Comparison of both methods

In order to compare the results of the Bayesian probability and Hidden Markov Model approach, the percentage of identical results in the carrier information obtained of both approaches was assessed for each divergent region. Since the HMM approach provided data for 142 divergent regions, these results were determined based on these 142 regions and were subsequently plotted in R (Figure 2D). For the following analysis, the results of the Bayesian

probability approach or the results of both approaches were respectively used for the following analysis.

1.6.3. Allele frequency of freshwater-adaptive alleles

For each divergent locus, heterozygous carriers as well as homozygous carriers within RABS and PGTS were counted, respectively. The count of heterozygous carriers added on the duplicated count of homozygous carriers divided by the duplicated number of individuals resulted in the allele frequency of the freshwater-adaptive allele for each locus (Formula 12).

$$allele\ frequency = \frac{\#(heterozygous\ carriers) + 2 * \#(homozygous\ carriers)}{2 * \#(individuals)} \quad (12)$$

1.6.4. Pearson correlation

The Pearson correlations between the allele frequencies and the freshwater fraction sums in the RABS and PGTS population were calculated with the cor() function from the stats package in R (R Core Team, 2020).

1.6.5. Haplotypes

In order to assess population differences in freshwater-adaptive haplotype block sizes, the sum of all freshwater-adaptive haplotype block sizes overlapping with the respective divergent region were determined for each individual at each of the 232 divergent regions (script: https://github.com/MelanieKirch/standing-genetic-variation/blob/main/Bayesian%20Probability/6_Haplotypes.R). In order to do so, final tables of the Bayesian Probability approach were used. For these final tables windows with missing data were already filled in based on the states of the surrounding windows (see chapter 1.5.1). Windows overlapping with each divergent region were thus segmented into the three ancestry states (marine homozygous, heterozygous, freshwater homozygous) for each individual. A freshwater-adaptive haplotype block was defined as the maximal consecutive sequence of freshwater-adaptive windows based on the ancestry states. Subsequently, the sum of all freshwater-adaptive haplotype block sizes was assessed for each carrier at a given divergent region. Only heterozygous states were considered for heterozygous carriers, whereas freshwater homozygous states were additionally considered for freshwater homozygous carriers. Since a sliding window approach with overlapping windows was used in the segmentation approach, the start and stop positions of a given haplotype block were assigned as the midpoint of the first or last window of the corresponding freshwater-adaptive haplotype block. The calculated values of both populations were compared by a one-tailed, paired t-test and plotted versus the allele frequency of each loci in the respective population with R. Correlation was calculated with cor() function from the stats package in R using the method Pearson (Figure 3).

1.6.6. Alleles carried by each individual

In order to assess whether alleles are predominantly carried by a few individuals or distributed on many individuals, the alleles carried by each individual were determined.

The alleles carried by each individual were summed up by looping through all 232 or 142 loci, in regard to the results of the Bayesian probability approach or the HMM approach, respectively. If an individual was a heterozygous carrier at a given locus, one allele was counted, if it was a homozygous carrier two alleles were counted. The results were plotted in histograms with R in the following (Figure 4B, Figure S4).

1.6.7. Alleles present in populations with different founder sizes

In order to assess how much freshwater adaptive genetic material would be present in a founder population depending on the size, different numbers of individuals (2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500) were randomly sampled from each population (RABS and PGTS). The divergent loci with freshwater adaptive alleles present within sampled individuals were counted based on the results of the Bayesian probability approach. 100 iterations were performed for this experiment and percentages of freshwater-adaptive alleles present were subsequently calculated by dividing the results by the total number of divergent regions present in each population (231 for PGTS and 232 for RABS). Finally, mean and standard deviations were plotted over the number of sampled individuals in R (Figure 4C).

1.6.8. Linkage disequilibrium among adaptive loci

The genomic basis of adaptation in sticklebacks is highly polygenic and involves loci on almost all chromosomes. In order to assess whether marine sticklebacks carry freshwater adaptive alleles within a chromosome and/or across multiple chromosomes in linkage, we estimated inter- and intra-chromosomal linkage disequilibrium (LD) in RABS and PGTS. To avoid interference with population structure, we reduced the analysis in RABS to 4 913 individuals sampled in 2018. Freshwater adaptive alleles are often present at very low frequencies and allele frequencies among loci can differ by an order of magnitude. Frequency dependent LD estimators such as r^2 can be confounded by this phenomena (Gaut & Long, 2003). To address this challenge, D' and chi-square p-values were calculated additionally to r^2. D' is standardised based on the allele frequencies of the analysed loci and is hence correcting for large differences in allele frequencies (Lewontin, 1964). Chi-square p-values indicates whether there is sufficient evidence that the observed results are not the same as the expected results. Calculations were performed in R based on the carrier information at 232 (231) divergent loci for RABS (PGTS) obtained with the Bayesian probability approach using the package genetics

and plotted in R as heatmap using the packages pheatmap and RColorBrewer (Figure 5, Figure S5-9; Kolde, 2019; Neuwirth, 2022).

1.6.9. Hybrid Plot

The breakdown of the freshwater-adaptive allele ensemble during multiple generations of introgression was simulated for each population separately in order to compare them to the empirical results for the respective population and thus, determine the degree of introgression. Freshwater-marine admixture in both marine populations was investigated based on the carrier information (freshwater homozygous carrier, heterozygous carrier or non-carrier) of each individual at 232 (231) divergent loci for RABS (PGTS). The starting points of the simulations were 26 technical replicates of the pulled-down Little Campbell River freshwater control as well as 26 marine individuals which were randomly selected from the respective population. Crosses were simulated based on the carrier data at 232 (231) divergent loci for these marine and freshwater individuals accounting for recombination with a constant recombination rate of 0.308 cM/Mb for all loci outside of chrI, chrXI and chrXXI inversions in which it was assumed to be zero. This way, 100 replicates of each of the following crosses were simulated: (a) F1; (b) F2; (c) F1 backcross with marine parent (F1xM); (d) backcross of F1xM with marine parent (Salces-Castellano *et al.*, 2021). Subsequently, freshwater ancestry and heterozygosity for each individual were calculated as the sum of freshwater-adaptive alleles and the sum of heterozygous carried alleles at all 232 (231) divergent loci, respectively. The simulated data for each population was plotted in R (Figure 6, S11). Freshwater ancestry and heterozygosity for the empirical data were calculated in the same way and plotted in R using the packages hexbin, RColorBrewer and ggplot2 (Figure 6, S11). Subsequently, the states for each divergent region were randomly re-assigned to the individuals keeping the allele frequencies for each divergent region. Similarly to the empirical data, the freshwater ancestry and heterozygosity of the shuffled empirical data was calculated and plotted in R using the packages hexbin, RColorBrewer and ggplot2 (Figure S10, S12; Dan Carr *et al.*, 2021; Neuwirth, 2022; Wickham, 2016). The script used in this work can be found in the GitHub repository: https://github.com/MelanieKirch/standing-genetic-variation/blob/main/HybridPlot.R.

1.7. Haplotype tagging of freshwater stickleback samples

39 freshwater samples from Arness Lake as well as 57 RABS individuals identified as carriers of freshwater-adaptive alleles on chrIV or within the chrI inversion were sequenced using haplotype tagging (Meier *et al.*, 2021) in order to obtain haplotype informative data. Briefly, per sample, 0.28 ng gDNA was tagmented with approximately 700 thousand sample-specific barcodes on the surface of approximately 700 thousand Haplotag beads. Haplotag beads are

M280-streptavidin modified beads (Thermo-Fischer) with immobilized and fully barcoded Tn5ME-A and Tn5ME-B-transposomes on their surface. Each Haplotag bead carries many copies of an unique segmented barcodes (Meier *et al.*, 2021). Tagmentation was peformed at 55°C for 10 minutes after mixing of gDNA, Haplotag beads, sterile water and 5x tagmentation buffer. After tagmentation, Tn5 protein was removed from the DNA with 0.2% SDS and the un-integrated barcoded Tn5-transposomes were removed with an Exonuclease-I (NEB, M0293) treatment at 48°C for 25 minutes. DNA library was then amplified with 10 cycles of PCR with NEBNext® High-Fidelity 2X PCR Master Mix (NEB, M0541) and TruSeq forward and reverse primers, size selected with AMPure beads and eluted in 10mM Tris, pH=8. To achieve the best linked-read results and low PCR duplication, the library pool was prepared from 27 ng of genomic DNA from the total of 96 samples, or 0.28 ng gDNA per sample.

1.8. Processing of haplotagged data

1.8.1. Initial read processing

The haplotagged library pool was sequenced on a single 2×150bp S4 lane of NovaSeq 6000 with 13bp index1 and 13bp index2 lengths resulting in 0.5 Gbp per genome and approximately 17x coverage per sample. Raw fastq data were demultiplexed using a custom C++ script to demultiplex all four 6bp-segments of the molecular and sample barcode stored in the 13bp index1 and 13bp index2 files. Demultiplexed molecular and sample barcode was then added in the header section of each R1 (read1) and R2 (read2) as the 'BX:Z: tag', e.g. BX:Z:A01C02B03D04, with C-barcode being the sample barcode, while B, D and A being the molecular barcode. After the data demultiplexing, adapter sequences were trimmed using cutadapt and the trimmed reads were placed against the stickleback reference genome gasAcu1-4 (Jones *et al.*, 2012; Kingman *et al.*, 2021; Martin, 2011) using Burrows-Wheeler-Aligner (BWA) mem algorithm (Li & Durbin, 2009)(representative command see 1.9.2). Mapped reads were subsequently sorted and indexed using SAMtools (Li *et al.*, 2009). Then, GATK MarkDuplicates (Picard) function with lenient validation stringency settings and including creating a BAM index was used to identify duplicates (representative command see 1.9.3). Finally, bam files were sorted by coordinate with GATK SortSam function (Auwera & O'Connor, 2020; McKenna *et al.*, 2010) and indexed using SAMtools (Li *et al.*, 2009).

1.8.2. Download and processing of 207 global samples

The data for 206 global sticklebacks was downloaded from the Sequence Read Archive (www.ncbi.nlm.nih.gov/sra) at accession PRJNA247503 using prefetch from the SRA toolkit (Kingman *et al.*, 2021; SRA Toolkit Development Team, 2022; representative command see 1.9.7). Using fastq-dump from the SRA toolkit downloaded sra files were transformed into fastq files (representative command see 1.9.8). Fastq files were subsequently processed as

described in 1.4.1., whereby bam files corresponding to one sample were merged using samtools merge after read mapping (representative command see 1.9.9). Representative commands for each step are available in chapter 1.9.

1.8.3. Variant calling and filtering

Variants were called at the Eda locus (chrIV: 12,818,000 - 12,824,000) and the ATPase locus (chrI: 26,311,706 - 26,346,706) based on bamfiles for 206 global sticklebacks, 39 freshwater sticklebacks from Arness Lake as well as 57 RABS carrier individuals using graphtyper with the subcommand genotype (Eggertsson *et al.*, 2017). Representative command line is available in section 1.9.10.

1.8.4. Haplotype phasing

In the next step, haplotypes were phased for individuals with haplotype tagging data (39 freshwater sticklebacks from Arness Lake, 57 RABS carriers) and the annotated data was included in the previously created VCF files using HapCUT2 (Edge *et al.*, 2017). The script used for haplotype phasing was loosely based on https://github.com/evolgenomics/HeliconiusHaplotagging/blob/main/HAPCUT2/hapcutVcf.helera1_demo.sh (Meier *et al.*, 2021). The script used in this work can be found here: https://github.com/MelanieKirch/standing-genetic-variation/blob/main/Principal%20component%20analysis%20for%20geographic%20origin/1_hapcut2.sh and a representative command for calling the script can be found in section 1.9.11.

1.8.5. Principal component analysis for assessing geographic origin

Firstly, the VCF files for each individual created in section 1.8.4 were transformed into a table with genotype information (see section 1.9.12). The following steps were performed in R version 4.0.3 (R Core Team, 2020). Subsequently, a distance matrix between all pairs of fish was calculated based on the genotype information at each position within the Eda and ATPase locus, respectively. Genotype information was obtained of the VCF file created for 207 global samples as well as haplotype phased data of freshwater samples from Arness Lake and RABS carriers. Then a principal component analysis (PCA) was performed with 103 global freshwater samples that were known to have a low-plated phenotype based on the distance matrix (Table S10). Data of freshwater samples and RABS carriers were projected onto this PCA, respectively, and plotted in R. The scripts used in this work include:

- https://github.com/MelanieKirch/standing-genetic-variation/blob/main/Principal%20component%20analysis%20for%20geographic%20origin/2_DistanceMatrix.R

- https://github.com/MelanieKirch/standing-genetic-variation/blob/main/Principal%20component%20analysis%20for%20geographic%20origin/3_PCAplot.R

Representative commands for calling these scripts can be found in section 1.9.13.

1.9. Representative commands

R version 4.0.3 was used for the following analyses performed in R (R Core Team, 2020). Pipelines and scripts used in this work are available at https://github.com/MelanieKirch/standing-genetic-variation/tree/main.

1.9.1. Demultiplexing and trimming

Command used for demultiplexing and removing adapter sequences and Unique Molecular Identifier (UMI) bases

```
$BCL2FASTQ --input-dir Data/Intensities/BaseCalls/ --output-dir $outputDir --sample-sheet $sampleSheet_file --barcode-mismatches $mismatchDistance --use-bases-mask $baseMask --tiles $tiles
```

$outputDir is an output directory. $sampleSheet_file is a file indicating the adapters for each sample. $mismatchDistance was either 0 or 1 depending on the used adapters. $baseMasek depends on the read and index read length. The usual setting here was Y150,I8N5,I8N4,Y150 meaning use the first 150 cycles, then use 8 cycles for index and ignore the next 5, use 8 cycles for the 2nd index, and ignore the next. Finally, use the next 150 cycles. $tiles indicates the lane.

1.9.2. Read mapping

Command for aligning the raw reads against the stickleback reference genome with subsequent sorting and indexing.

```
bwa mem -t 30 gasAcu1-4.fa $file ${file/R1_001/R2_001} -R "@RG\tID:$name\tSM:$name\tLB:ManyMarineFish\tPL:Illumina.HiSeq3000.2x150" | samtools view -bh - > $name.gasAcu1-4.bam

samtools sort -l 9 -T $name.tmpsort -o $name.sorted.gasAcu1-4.bam $name.gasAcu1-4.bam

samtools index $name.sorted.gasAcu1-4.bam
```

$file represents one of the fastq files created in 1.9.1.. $name is the name of the analysed sample.

1.9.3. Mark duplicates

Command for identifying sequence duplicates in the aligned reads with subsequent sorting and indexing.

```
GenomeAnalysisTK-4.1/gatk --java-options "-Xmx2g" MarkDuplicates -I $name.sorted.gasAcu1-4.bam -O $name.mkdup.bam -M $name.mkdup.metrics -VALIDATION_STRINGENCY LENIENT -CREATE_INDEX TRUE

GenomeAnalysisTK-4.1/gatk --java-options "-Xmx2g" SortSam -I $name.mkdup.bam -O $name.mkdup.sorted.bam -SORT_ORDER coordinate

samtools index $name.mkdup.sorted.bam
```

$name is the name of the analysed sample.

1.9.4. Haplotypecaller

Command for calling potential variant sites per sample.

```
GenomeAnalysisTK/gatk --java-options "-Xmx32g" HaplotypeCaller --emit-ref-confidence GVCF --output $name.g.vcf --intervals pulldown_genome.bed --input $name.mkdup.sorted.bam --reference gasAcu1-4.fa --recover-all-dangling-branches true --min-pruning 0 --heterozygosity 0.04 --heterozygosity-stdev 0.1
```

$name is the name of the analysed sample.

1.9.5. Import into Genomics Database

Command for importing all previously created GVCF files into a database for each proximity merged region.

```
GenomeAnalysisTK/gatk --java-options "-Xmx40g" GenomicsDBImport --genomicsdb-workspace-path $folder --batch-size 50 --intervals chrI:7587000-7775009 --sample-name-map cohort.sample_map.txt --reference gasAcu1-4.fa --tmp-dir GenomsDB_tmp --reader-threads 5
```

$folder is the folder for the Genome Database of the respective proximity merged region. cohort.sample_map.txt is a list of all GVCF files which are imported into the Genome Database.

1.9.6. Export joint genotype files

Command for exporting joint VCF files from each database.

```
GenomeAnalysisTK/gatk GenotypeGVCFs -R gasAcu1-4.fa -V
gendb:///$folder -O 10k_pull-down_Region1.vcf --heterozygosity 0.04 -
-heterozygosity-stdev 0.1 --max-alternate-alleles 2 --verbosity ERROR
```

$folder is the folder for the Genome Database of the respective proximity merged region.

1.9.7. Download 207 global samples from Sequence Read Archive

Command used to download 1758 files for all 207 genomes from Sequence Read Archive (www.ncbi.nlm.nih.gov/sra).

```
sratoolkit/bin/prefetch -O 206genomes SRR1462568
```

1.9.8. Transform SRA files into fastq files

Command to turn downloaded sra files into fastq files.

```
sratoolkit/bin/fastq-dump SRR1462568.sra --split-3
```

1.9.9. Merging bam files

Command to merge all files corresponding to one sample into one file.

```
samtools merge BARW_X_2012_04.merged.gasAcu1-4.bam
SRR1437857.mkdup.gasAcu1-4.bam SRR1437985.mkdup.gasAcu1-4.bam
```

1.9.10. Variant calling with graphtyper

Command used to call variants based on bam files at a specific region.

```
graphtyper genotype gasAcu1-4.fa --sams=AllFish.bamlist
--region=chrI:26311706-26346706 --threads=20 --output outputfolder
```

1.9.11. Haplotype phasing

Command used to haplotype phase the data for one individual.

```
./hapcut2.sh $bam $allfishvcf $region $chrom $fishID $tmp $outdir
$local_lib
```

$bam is the bam file of the respective individual, $allfishvcf is the VCF file with data for all fish, $region is the loci of interest, $chrom is the respective chromosome of the loci of interest. $fishID is the ID of the analysed individual in the VCF file, $tmp is a folder for temporary files, $outdir is a folder for output files, $local_lib is the local library for commands.

1.9.12. Transforming VCF files to genotype tables

Command used to transform VCF files into genotype tables for PCA analysis.

```
for i in $path/*.PL.AD.HAPCUT2.vcf.gz;
do awk '(/^chr/){split($10,a,":");
if($10~/\|/)
{if($9~/^GT/)                {split(a[1],b,"|");                print
$1"\t"$2"\t"$3"\t"$4"\t"$5"\t"b[1]"/"b[1]"\t"b[2]"/"b[2]}
else                     {split(a[6],b,"|");                    print
$1"\t"$2"\t"$3"\t"$4"\t"$5"\t"b[1]"/"b[1]"\t"b[2]"/"b[2]}}
else if ($10~/\//) {print $1"\t"$2"\t"$3"\t"$4"\t"$5"\tNA\tNA"}}'
<(zcat $i) >${i/.PL.AD.HAPCUT2.vcf.gz/.geno.table.out}; done;
```

1.9.13. Principal component analysis

Command used to create a distance matrix for a PCA and plot the data subsequently.

```
Rscript 2_DistanceMatrix.R $count
Rscript 3_PCAplot.R $count
```

with $count being the respective number of the loci of interest (normally multiple regions are processed in parallel).

2. Supplementary Figures

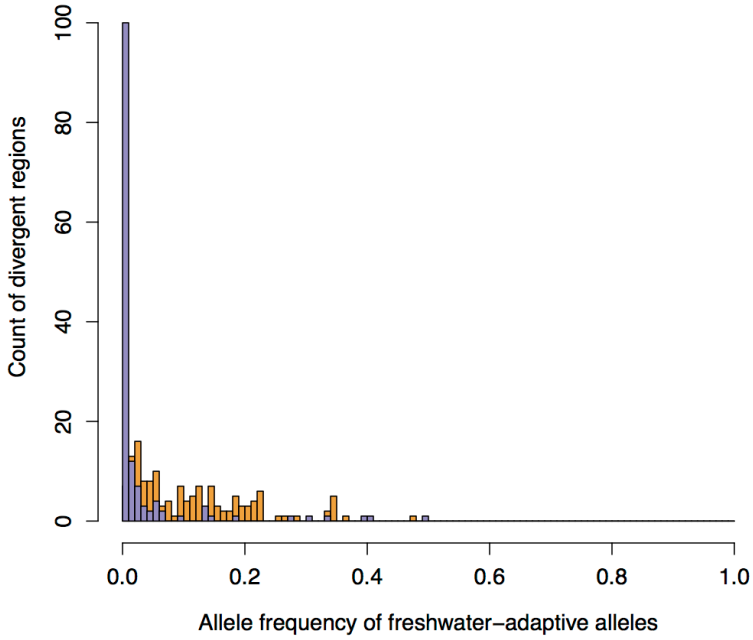


Figure S1: Availability of SGV. Results presented here were quantified based on a Hidden Markov Model approach (see 1.5.2). Histogram of allele frequencies of freshwater-adaptive alleles for each of 142 divergent loci in two marine populations (PGTS and RABS).

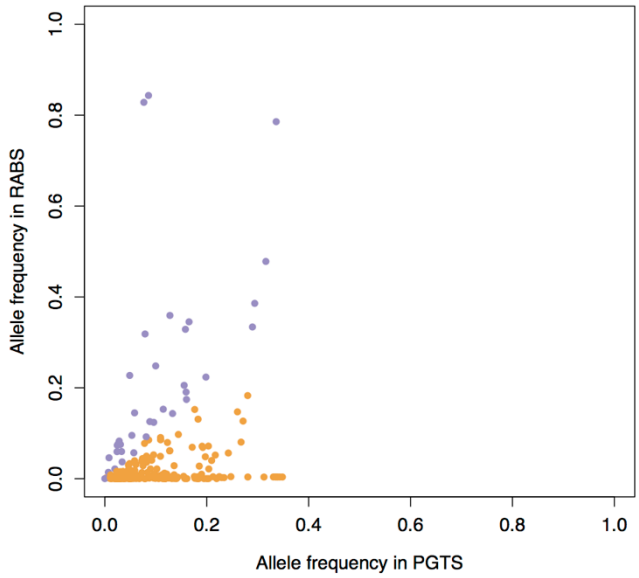
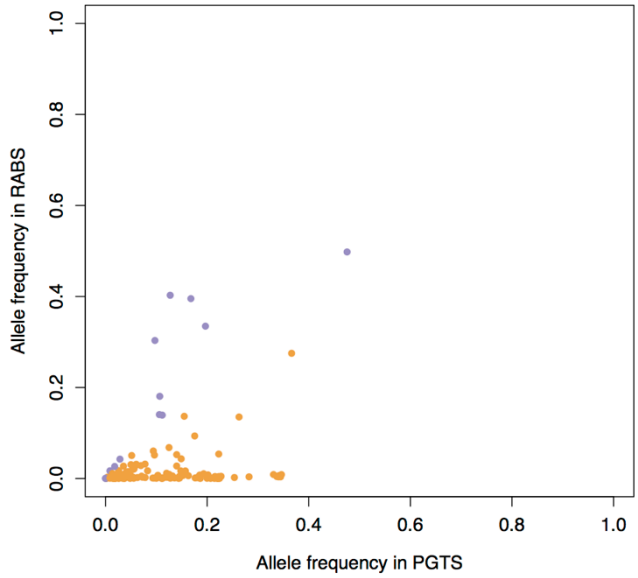


Figure S2: Joint frequency plot. Allele frequencies of freshwater-adaptive alleles in RABS and PGTS estimated by Bayesian Probability approach are plotted against each other (see 1.5.1). Each dot represents one of 232 loci previously identified to be divergent between marine and freshwater sticklebacks (Jones *et al.*, 2012; Table S2). Allele frequencies of freshwater-adaptive alleles which are higher in the PGTS are coloured in orange, the ones higher in RABS are coloured in purple.

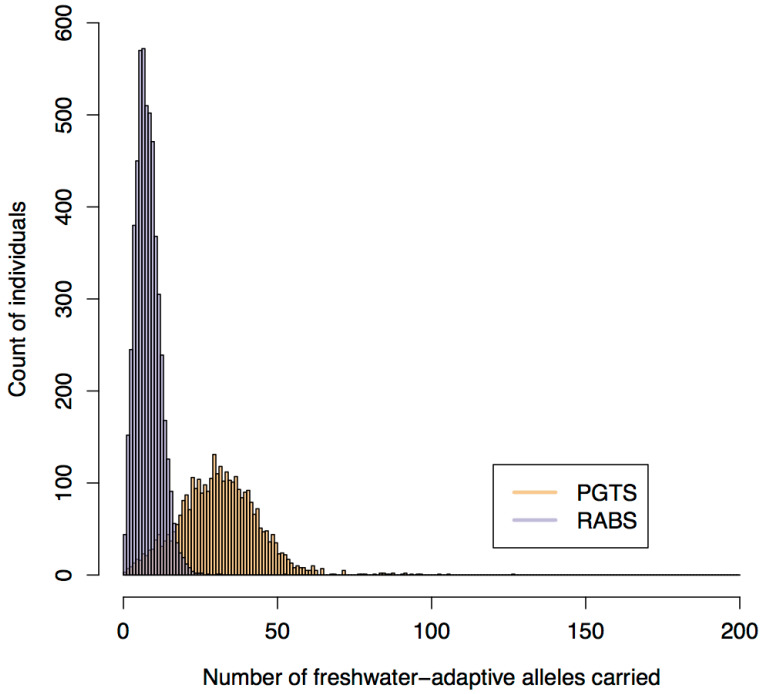


798

799 Figure S3: Joint frequency plot. Allele frequencies of freshwater-adaptive alleles in RABS and PGTS
800 estimated by HMM approach are plotted against each other (see 1.5.2). Each dot represents one of 142
801 loci previously identified to be divergent between marine and freshwater sticklebacks (Jones *et al.*, 2012,
802 Table S2). Allele frequencies of freshwater-adaptive alleles which are higher in the PGTS are coloured
803 in orange, the ones higher in RABS are coloured in purple.

804

805



806

807 Figure S4: Individual marine fish carry multiple freshwater-adaptive alleles across the genome. A
808 histogram shows the number of freshwater-adaptive alleles carried by the RABS and PGTS individuals
809 genome-wide (142 divergent regions in total). Results were quantified based on a Hidden Markov Model
810 approach (see 1.5.2).

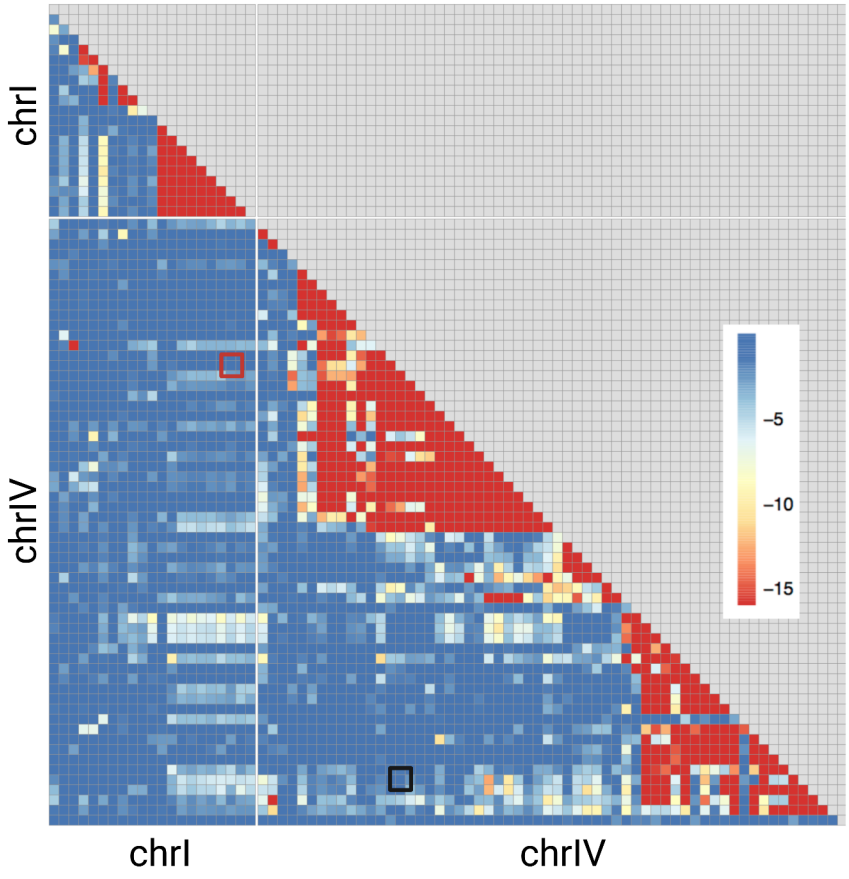


Figure S5: Estimates of linkage disequilibrium among adaptive alleles on chrI and chrIV for PGTS individuals based on carrier information (see 1.6.8). The lower triangle shows chi-square p-values. The red box marks the inter-chromosomal LD between Eda (chrIV:12,811,394-12,856,894) and ATPase (chrIV:26,703,207-26,767,873) and the black box the intra-chromosomal LD between Eda (chrIV:12,811,394-12,856,894) and Wnt7b (chrI:26,312,062-26,384,706).

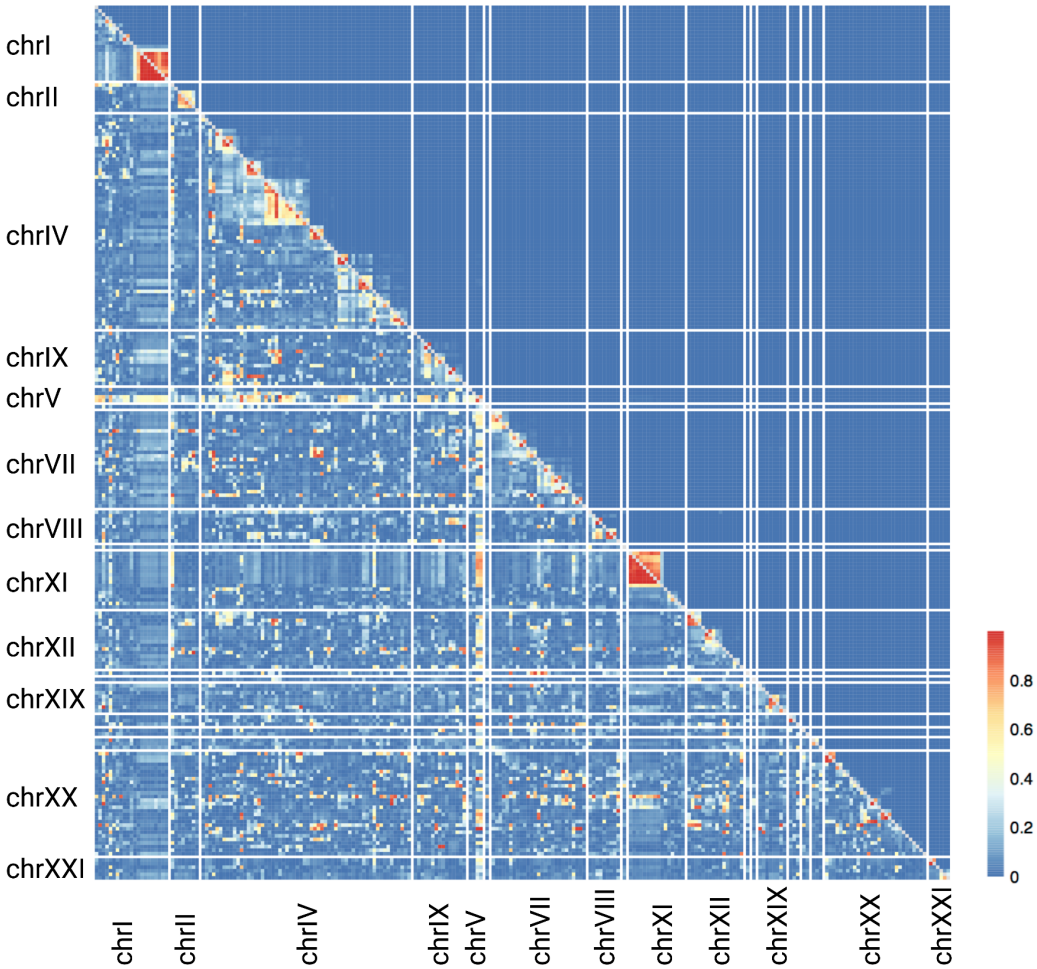


Figure S6: Estimates of linkage disequilibrium among 231 adaptive alleles across the genome for PGTS individuals based on carrier information (see 1.6.8). The upper triangle shows r^2 and the lower triangle shows D'.

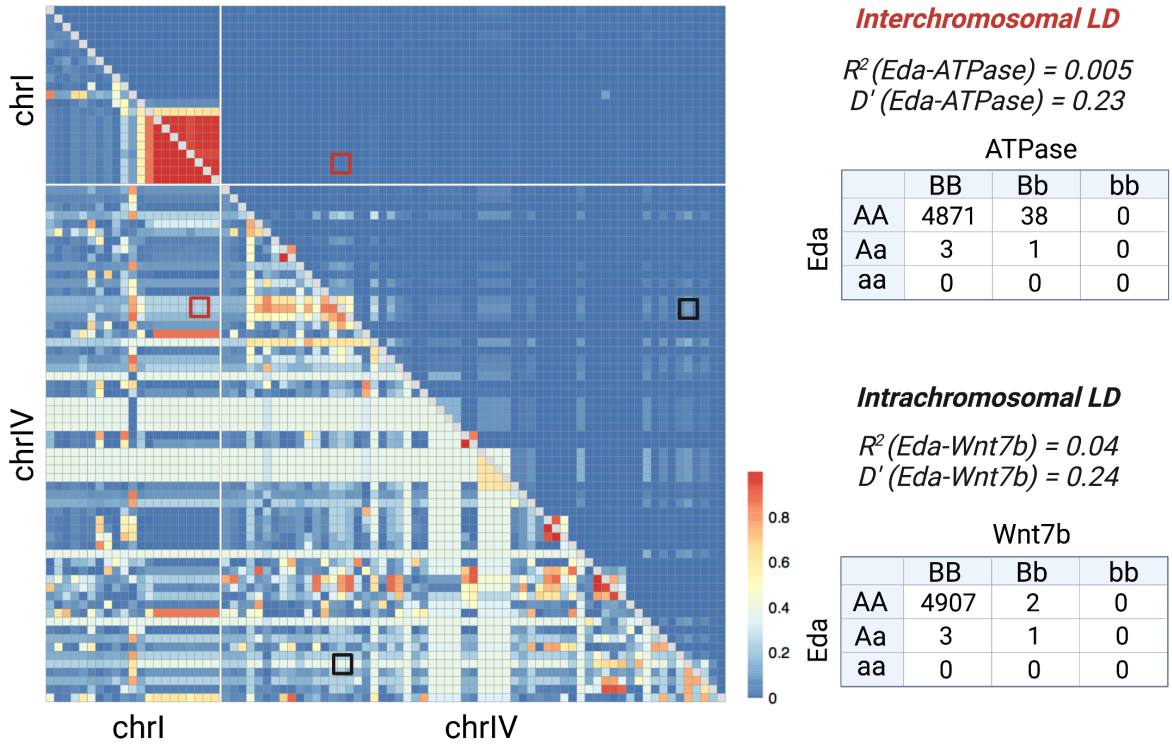


Figure S7: Estimates of linkage disequilibrium among adaptive alleles in chrI and chrIV for RABS individuals based on carrier information (see 1.6.8). The upper triangle shows r^2 and the lower triangle shows D'. Red boxes mark the inter-chromosomal LD between Eda (chrIV:12,811,394-12,856,894) and ATPase (chrIV:26,703,207-26,767,873) and black boxes the intra-chromosomal LD between Eda (chrIV:12,811,394-12,856,894) and Wnt7b (chrI:26,312,062-26,384,706).

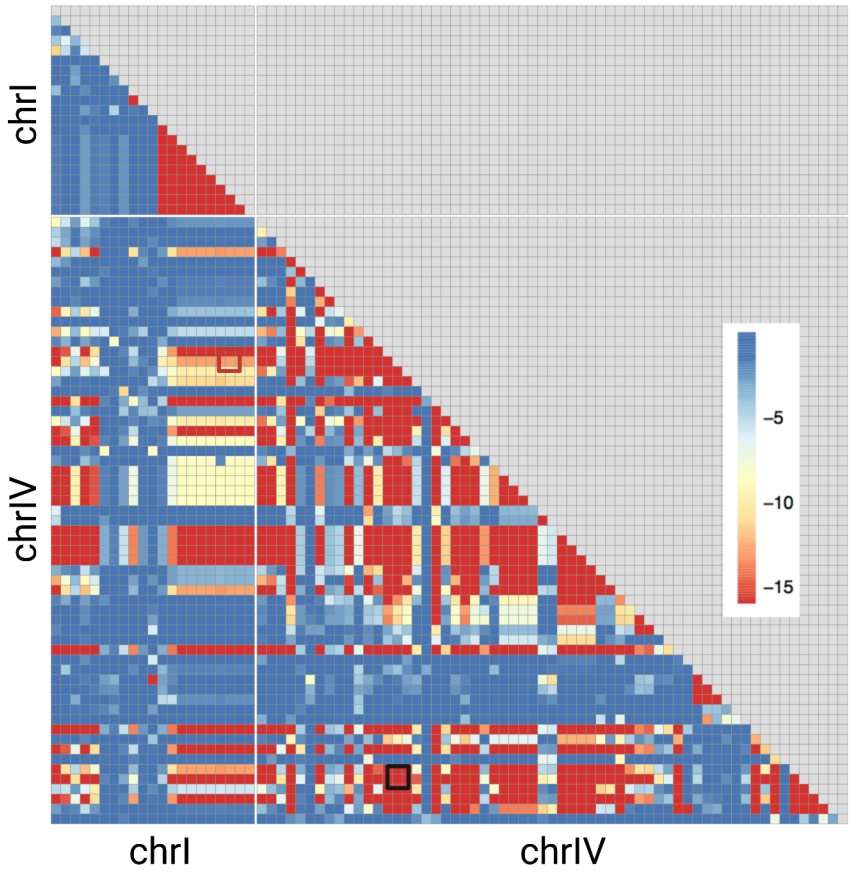


Figure S8: Estimates of linkage disequilibrium among adaptive alleles on chrI and chrIV for RABS individuals based on carrier information (see 1.6.8). The lower triangle shows chi-square p-values. The red box marks the inter-chromosomal LD between Eda (chrIV:12,811,394-12,856,894) and ATPase (chrIV:26,703,207-26,767,873) and the black box the intra-chromosomal LD between Eda (chrIV:12,811,394-12,856,894) and Wnt7b (chrI:26,312,062-26,384,706).

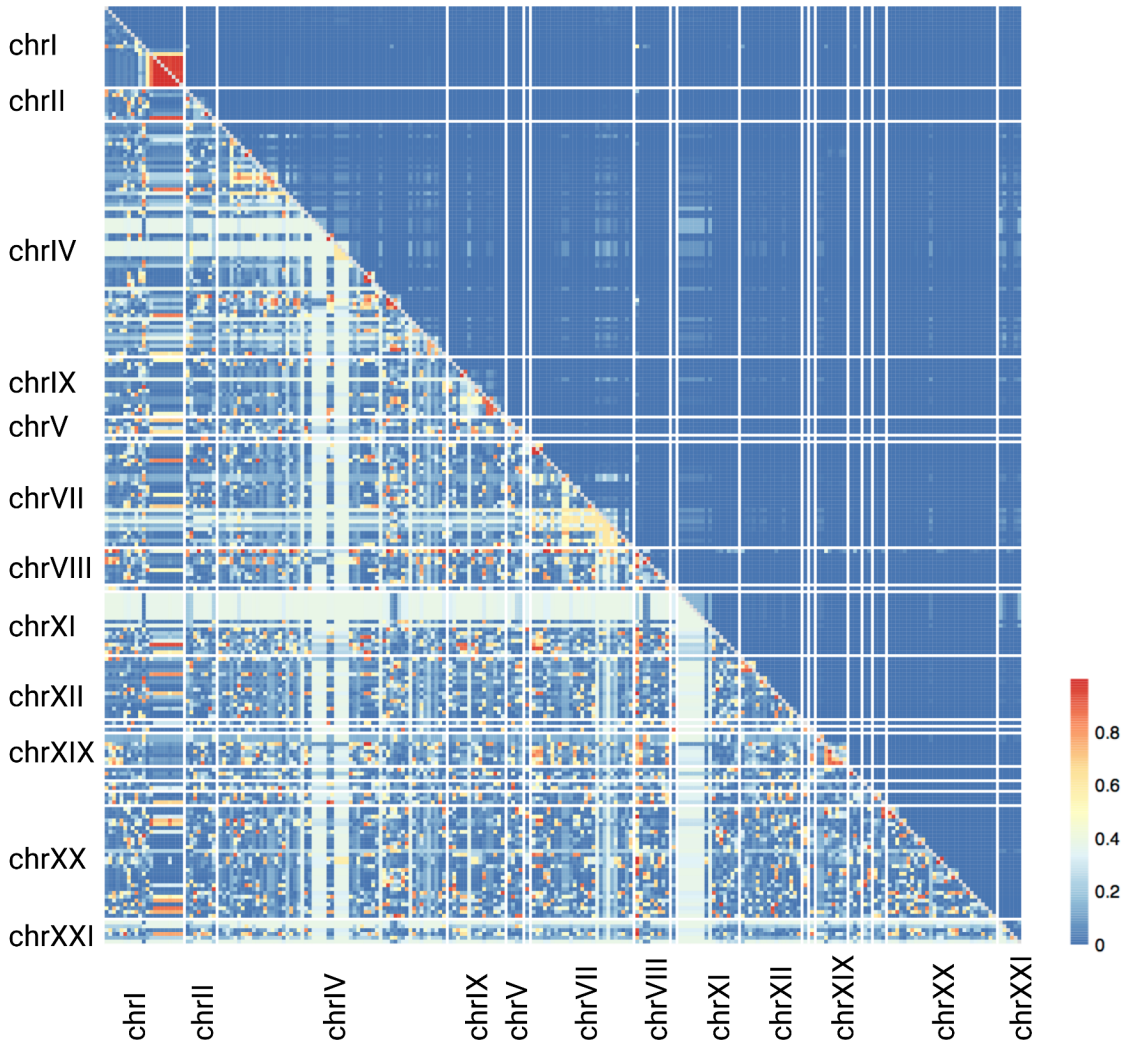
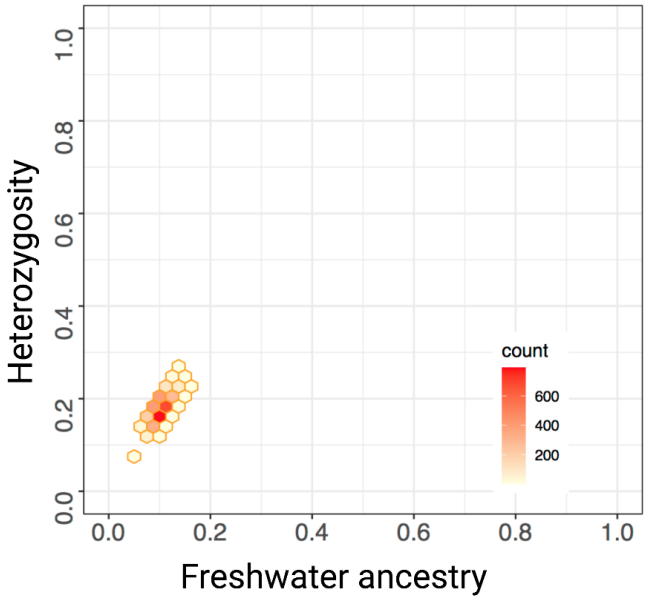
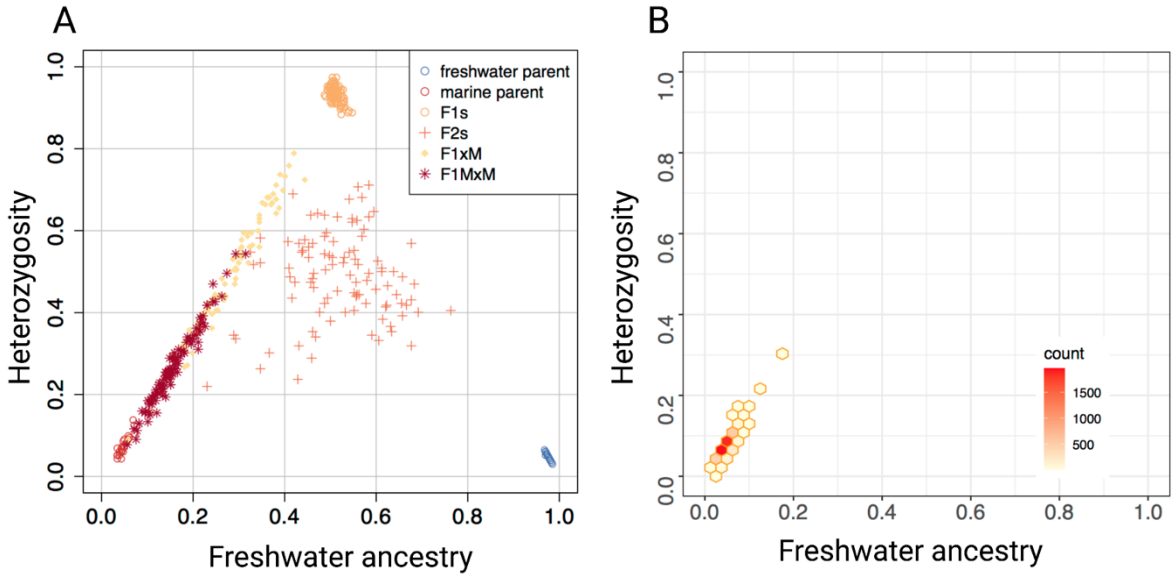


Figure S9: Estimates of linkage disequilibrium among 232 adaptive alleles across the genome for RABS individuals based on carrier information (see 1.6.8). The upper triangle shows r^2 and the lower triangle shows D'.



838

839 Figure S10: Empirical data of PGTS population with freshwater-adaptive alleles randomly re-distributed
840 based on allele frequencies at each divergent loci (see 1.6.9). Results presented here were quantified
841 based on a Bayesian Probability approach (see 1.5.1).



842

843 Figure S11: Marine fish carrying freshwater-adaptive alleles are products of multiple generations of
844 introgression into an ancestral marine genomic background. (A) Simulations for the breakdown of the
845 freshwater-adaptive allele ensemble during multiple generations of introgression. Simulations were
846 performed based on carrier information at 232 divergent loci. (B) Comparison of empirical data of RABS
847 population to simulated data in panel A suggest RABS marine individuals represent predominantly F1-
848 Marine backcrosses with marine ancestors (F1MxM). Results presented here were quantified based on
849 a Bayesian Probability approach (see 1.5.1).

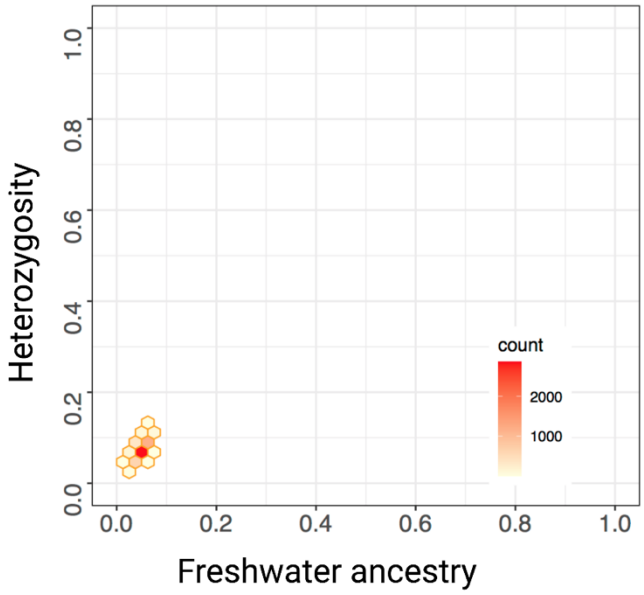


Figure S12: Empirical data of RABS population with freshwater-adaptive alleles randomly re-distributed based on allele frequencies at each divergent loci (see 1.6.9). Results presented here were quantified based on a Bayesian Probability approach (see 1.5.1).

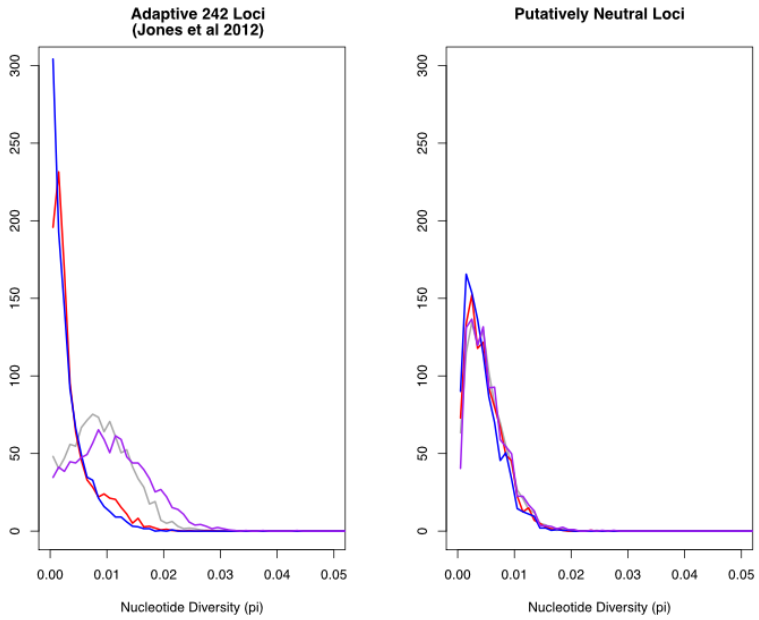


Figure S13: Elevated pairwise nucleotide diversity between marine and freshwater individuals at adaptive loci. The pairwise nucleotide diversity between marine and freshwater individuals from Rabbit Slough and Arness Lake 2018 samples respectively, shown for marine-freshwater parallel adaptive loci (Jones *et al.*, 2012; A) and putatively neutral loci (B, loci not detected as showing parallel adaptive divergence in Jones *et al.* (2012)). Pairwise nucleotide diversity was calculated within ecotypes (among Rabbit Slough individuals, red line; among Arness Lake individuals, blue line), between ecotypes (between Arness Lake and Rabbit Slough individuals, purple line), and among all individuals (grey line) using vcftools (Danecek *et al.*, 2011) windowed-pi command with 10kb windows and variant call files of whole genome short-read sequencing data.

3. Supplementary Tables

SITE	LOCATION	POPULATION	LATITUDE/ LONGITUDE	SAMPLING YEAR	TARGETED PULLDOWN OR WHOLE GENOME SEQUENCED SAMPLES	EXCLUDED SAMPLES	SAMPLES USED FOR ANALYSES
PUGET SOUND	Washington	Clam Bay	47.568164 N, 122.547364 E	2016	3 485	216	3 273
		Hale Passage	47.27126 N, 122.62953 E	2016	22	0	
		Thompson Lagoon	48.097210 N, 122.941190 E	2016	32	0	
		Port Angeles Bay	48.140432 N ,123.445805 E	2016	6	0	
RABBIT SLOUGH	Alaska		61.53583 N, 149.25292 E	2018	5 021	108	5 490
			61.53583 N, 149.25292 E	2004	606	29	
TOTAL MARINE SAMPLES					9 172	409	8 763
LITTLE CAMPBELL RIVER, MARINE	British Columbia		49.0161194 N, 122.7795472 E	2016	1	0	26 replicates
LITTLE CAMPBELL RIVER, FRESHWATER	British Columbia		49.012125 N, 122.6251222 E	2016	1	0	26 replicates
ARNESS LAKE	Alaska		60.691 N, 151.294 E	2018	39	0	39

Table S1: Information on sample locations including their latitude and longitude, sampling time, number of samples sequences, excluded and used for analysis.

Number	chromosome	start position	end position	included in pull-down	included in HMM	included in BP	mean depth	standard deviation
1	chrI	7747712	7750212	1	0	1	3,98	+/-1,9 SD
2	chrI	8516712	8520212	1	1	1	2,55	+/-1,18 SD
3	chrI	9096712	9099212	1	1	1	3,6	+/-1,7 SD
4	chrI	12076212	12078712	1	1	1	6,49	+/-2,94 SD
5	chrI	12419212	12421712	1	1	1	9,1	+/-4,14 SD
6	chrI	12627212	12630712	1	0	1	6,46	+/-2,63 SD
7	chrI	12716212	12720712	1	0	1	7,32	+/-3,01 SD
8	chrI	12828712	12831212	1	1	1	4,64	+/-2 SD
9	chrI	14303796	14308440	1	1	1	6,14	+/-2,52 SD
10	chrI	14310712	14313212	1	1	1	7,04	+/-3,04 SD
11	chrI	17483086	17483982	1	0	1	24,74	+/-15,98 SD
12	chrI	26061393	26071817	1	0	1	1,06	+/-0,68 SD
13	chrI	26085364	26092819	1	0	1	1,55	+/-0,75 SD
14	chrI	26096706	26169519	1	1	1	4,68	+/-2,03 SD
15	chrI	26173123	26176392	1	1	1	3,96	+/-1,83 SD
16	chrI	26176404	26194329	1	1	1	5,01	+/-2,13 SD
17	chrI	26194610	26268938	1	1	1	4,85	+/-2 SD
18	chrI	26269206	26281621	1	1	1	4,19	+/-1,77 SD
19	chrI	26312062	26384706	1	1	1	6,94	+/-2,89 SD
20	chrI	26389706	26515706	1	1	1	6,62	+/-2,78 SD
21	chrI	26516512	26550206	1	1	1	3,74	+/-1,66 SD
22	chrII	957684	973691	1	0	1	5,98	+/-2,5 SD
23	chrII	2872077	2874577	1	0	1	7,26	+/-3,14 SD
24	chrII	6788577	6791077	1	0	0	5,61	+/-2,55 SD
25	chrII	15226600	15234243	1	1	1	4,69	+/-2 SD
26	chrII	15240501	15247685	1	0	1	4,9	+/-2,07 SD
27	chrII	15254522	15260728	1	0	1	7,18	+/-2,99 SD
28	chrII	15310987	15326849	1	1	1	5,08	+/-2,09 SD
29	chrII	15339236	15346066	1	1	1	6,27	+/-2,62 SD
30	chrII	16829759	16832259	1	0	1	0,64	+/-0,78 SD
31	chrIV	8332093	8344936	1	1	1	8,21	+/-3,5 SD
32	chrIV	8570693	8585028	1	1	1	6,98	+/-2,97 SD
33	chrIV	9338602	9346000	1	1	1	2,47	+/-1,2 SD
34	chrIV	11220289	11221697	1	0	1	7,82	+/-3,51 SD
35	chrIV	12003059	12015299	1	1	1	4,89	+/-2,14 SD
36	chrIV	12015504	12020013	1	1	1	5,26	+/-2,36 SD
37	chrIV	12245212	12248167	1	0	1	3,71	+/-1,64 SD
38	chrIV	12250449	12251281	1	0	1	3,55	+/-1,88 SD
39	chrIV	12251599	12253564	1	0	1	3,8	+/-1,79 SD
40	chrIV	12340067	12343218	1	1	1	3,26	+/-1,92 SD
41	chrIV	12454094	12456397	1	0	1	6,78	+/-2,91 SD

42	chrIV	12601697	12607290	1	1	1	6,37	+/-2,62 SD
43	chrIV	12761394	12765894	1	1	1	4,76	+/-2,05 SD
44	chrIV	12795280	12809894	1	1	1	8,84	+/-3,58 SD
45	chrIV	12811394	12856894	1	1	1	6,24	+/-2,53 SD
46	chrIV	12870894	12880606	1	1	1	8,53	+/-3,54 SD
47	chrIV	12882394	12889548	1	1	1	5,94	+/-2,5 SD
48	chrIV	12955394	12957894	1	1	1	7,15	+/-3,09 SD
49	chrIV	13371594	13373485	1	1	1	1,72	+/-0,97 SD
50	chrIV	13395394	13397894	1	1	1	3,87	+/-1,81 SD
51	chrIV	13581894	13584394	1	1	1	5,05	+/-2,29 SD
52	chrIV	13667894	13670394	1	1	1	4,65	+/-2,18 SD
53	chrIV	13833979	13835713	1	1	1	5,12	+/-2,3 SD
54	chrIV	13861443	13868894	1	1	1	5,1	+/-2,15 SD
55	chrIV	13914800	13918841	1	1	1	6,03	+/-2,52 SD
56	chrIV	13925394	13936394	1	1	1	5,2	+/-2,17 SD
57	chrIV	13936533	13939899	1	1	1	4,51	+/-2,01 SD
58	chrIV	13950886	13957772	1	0	1	4,99	+/-2,11 SD
59	chrIV	13957940	13963394	1	0	1	4,5	+/-1,99 SD
60	chrIV	13977550	13998894	1	0	1	2,64	+/-1,17 SD
61	chrIV	14000894	14003394	1	0	1	4,3	+/-1,98 SD
62	chrIV	15042726	15048156	1	1	1	4,36	+/-2,12 SD
63	chrIV	15048963	15057312	1	1	1	4,99	+/-2,18 SD
64	chrIV	15066293	15071914	1	1	1	3,84	+/-1,73 SD
65	chrIV	15076812	15082812	1	1	1	3,6	+/-1,61 SD
66	chrIV	15358312	15361812	1	1	1	4,55	+/-2,01 SD
67	chrIV	15442312	15446312	1	0	1	3,08	+/-1,41 SD
68	chrIV	15493404	15498850	1	1	1	2,92	+/-1,33 SD
69	chrIV	17849542	17849817	0	0	0	0,68	+/-0,97 SD
70	chrIV	19636055	19638555	1	0	1	3,69	+/-1,76 SD
71	chrIV	20009159	20070647	1	1	1	5,19	+/-2,13 SD
72	chrIV	20073713	20158690	1	1	1	5,09	+/-2,13 SD
73	chrIV	20159176	20185905	1	1	1	5,02	+/-2,06 SD
74	chrIV	20311308	20312293	1	1	1	5,73	+/-2,83 SD
75	chrIV	20972014	20973807	1	1	1	6,35	+/-2,87 SD
76	chrIV	21276555	21279055	1	1	1	3,09	+/-1,46 SD
77	chrIV	22204058	22237555	1	1	1	5,64	+/-2,31 SD
78	chrIV	22238555	22253942	1	1	1	4,26	+/-1,79 SD
79	chrIV	22256844	22262284	1	1	1	4,26	+/-1,88 SD
80	chrIV	22262616	22272327	1	1	1	5,19	+/-2,19 SD
81	chrIV	22824055	22826555	1	0	1	5,68	+/-2,46 SD
82	chrIV	24312373	24314873	1	0	1	4,37	+/-2,11 SD
83	chrIV	24478100	24491692	1	1	1	4,46	+/-1,87 SD
84	chrIV	24813314	24819953	1	1	1	4,63	+/-2,03 SD

85	chrIV	25004693	25006569	1	1	1	7,42	+/-3,39 SD
86	chrIV	25008373	25014873	1	1	1	8,3	+/-3,76 SD
87	chrIV	26354220	26358990	1	1	1	7,03	+/-2,94 SD
88	chrIV	26703207	26767873	1	1	1	6,44	+/-2,65 SD
89	chrIV	26770873	26775373	1	1	1	3,84	+/-1,69 SD
90	chrIV	26780373	26809873	1	1	1	5,62	+/-2,32 SD
91	chrIV	26846873	26854361	1	1	1	5,82	+/-2,49 SD
92	chrIV	28885031	28889573	1	0	1	5,76	+/-2,49 SD
93	chrIX	4914304	4925074	1	0	1	8,15	+/-3,55 SD
94	chrIX	9748541	9749572	1	0	1	5,48	+/-2,65 SD
95	chrIX	11898250	11905480	1	1	1	6,78	+/-2,8 SD
96	chrIX	13027283	13030582	1	1	1	2,15	+/-1,14 SD
97	chrIX	13030650	13032620	1	1	1	3,11	+/-1,54 SD
98	chrIX	13049287	13053244	1	0	1	3,85	+/-1,73 SD
99	chrIX	13336484	13340193	1	1	1	1,56	+/-0,98 SD
100	chrIX	13388160	13393132	1	1	1	2,52	+/-1,31 SD
101	chrIX	13393999	13394542	1	1	1	0,56	+/-0,65 SD
102	chrIX	13533062	13535943	1	0	1	8,66	+/-3,64 SD
103	chrIX	13644183	13650772	1	1	1	3,7	+/-1,62 SD
104	chrIX	13652065	13662399	1	1	1	6,11	+/-2,63 SD
105	chrIX	13701104	13716307	1	1	1	6,19	+/-2,52 SD
106	chrIX	13720479	13725873	1	1	1	3,71	+/-1,63 SD
107	chrIX	13834303	13834716	1	0	1	5,63	+/-2,85 SD
108	chrIX	19819095	19823595	0	0	0	0,7	+/-0,55 SD
109	chrUn	2576721	2579221	0	0	0	0,21	+/-0,19 SD
110	chrV	4027794	4034252	1	1	1	8,51	+/-3,54 SD
111	chrV	5943571	5945130	1	0	1	5,89	+/-3,8 SD
112	chrV	8350629	8357129	1	0	1	2,92	+/-1,53 SD
113	chrV	8359655	8360427	1	0	1	1,62	+/-1,04 SD
114	chrVI	6471586	6474586	1	0	0	6,63	+/-2,95 SD
115	chrVI	10405624	10406900	1	0	1	1,13	+/-0,7 SD
116	chrVII	8560397	8565477	1	1	1	7,14	+/-3,01 SD
117	chrVII	8569313	8571796	1	1	1	4,99	+/-2,2 SD
118	chrVII	8578414	8594013	1	1	1	5,3	+/-2,43 SD
119	chrVII	8802165	8805165	1	0	1	7,04	+/-2,99 SD
120	chrVII	8813263	8815408	1	0	1	2,21	+/-1,08 SD
121	chrVII	10706165	10709665	1	0	1	12,69	+/-5,45 SD
122	chrVII	14809829	14811082	1	0	1	5,42	+/-2,58 SD
123	chrVII	14863132	14866806	1	1	1	3,02	+/-1,42 SD
124	chrVII	16027268	16034901	1	1	1	6,34	+/-2,62 SD
125	chrVII	16048512	16050868	1	1	1	7,83	+/-3,42 SD
126	chrVII	17432570	17436070	1	1	1	13,58	+/-7,19 SD
127	chrVII	17645722	17647665	1	1	1	3,53	+/-1,65 SD

128	chrVII	17648319	17655257	1	1	1	8,53	+/-3,86 SD
129	chrVII	18782070	18784570	1	1	1	5,45	+/-2,44 SD
130	chrVII	18863070	18865570	1	0	1	3,47	+/-1,69 SD
131	chrVII	19691070	19693570	1	1	1	3,18	+/-1,53 SD
132	chrVII	19694019	19701109	1	1	1	3,87	+/-1,69 SD
133	chrVII	19751070	19753570	1	1	1	5,5	+/-2,46 SD
134	chrVII	19816716	19821636	1	1	1	5,57	+/-2,36 SD
135	chrVII	19823099	19831570	1	1	1	6,06	+/-2,57 SD
136	chrVII	19999215	20027490	1	1	1	4,05	+/-1,66 SD
137	chrVII	20035494	20044743	1	1	1	7,46	+/-3,09 SD
138	chrVII	20059178	20061421	1	1	1	5,72	+/-2,45 SD
139	chrVII	20320070	20323070	1	1	1	3,86	+/-1,8 SD
140	chrVII	20632533	20637570	1	0	1	5,01	+/-2,37 SD
141	chrVII	20638231	20656570	1	0	1	4,33	+/-1,79 SD
142	chrVII	21991130	21992627	1	1	1	5,06	+/-2,36 SD
143	chrVIII	2695921	2699921	1	1	1	12,92	+/-6,89 SD
144	chrVIII	9448414	9470181	1	1	1	5,74	+/-2,32 SD
145	chrVIII	10072427	10076242	1	0	1	3,53	+/-1,59 SD
146	chrVIII	10078257	10079747	1	0	1	3,83	+/-1,86 SD
147	chrVIII	10602414	10604914	1	1	1	4,88	+/-2,17 SD
148	chrVIII	10948414	10955059	1	0	1	5,22	+/-2,23 SD
149	chrVIII	10958914	10964414	1	0	1	4,49	+/-1,94 SD
150	chrVIII	10967939	10977350	1	1	1	5,2	+/-2,14 SD
151	chrVIII	13542741	13545241	1	0	1	4,91	+/-2,23 SD
152	chrX	3190147	3192647	1	0	0	4,61	+/-3,2 SD
153	chrX	11039087	11040519	1	1	1	7,04	+/-3,19 SD
154	chrXI	5439626	5440496	1	0	0	7,09	+/-3,89 SD
155	chrXI	6184800	6282017	1	1	1	6,87	+/-2,88 SD
156	chrXI	6284313	6292517	1	1	1	4,05	+/-1,76 SD
157	chrXI	6293517	6297017	1	1	1	8,32	+/-3,53 SD
158	chrXI	6298517	6331444	1	1	1	6,01	+/-2,51 SD
159	chrXI	6333177	6336517	1	1	1	7,64	+/-3,31 SD
160	chrXI	6336785	6412585	1	1	1	6,93	+/-2,89 SD
161	chrXI	6413562	6542823	1	1	1	4,35	+/-1,85 SD
162	chrXI	6543542	6549121	1	1	1	4,22	+/-1,88 SD
163	chrXI	6549368	6617396	1	1	1	2,96	+/-1,27 SD
164	chrXI	6658892	6664532	1	1	1	6,02	+/-2,57 SD
165	chrXI	6890017	6893017	1	0	1	4,85	+/-2,27 SD
166	chrXI	8007017	8009517	1	0	1	4,68	+/-2,1 SD
167	chrXI	8385295	8390809	1	1	1	6,73	+/-2,9 SD
168	chrXI	8494017	8497017	1	0	1	3,95	+/-1,82 SD
169	chrXI	8769517	8773517	1	0	1	4,52	+/-2,07 SD
170	chrXI	11100431	11102931	1	0	1	10,88	+/-5,17 SD

171	chrXII	4380418	4386368	1	1	1	2,59	+/-1,31 SD
172	chrXII	4393617	4394847	1	0	1	2,94	+/-1,71 SD
173	chrXII	4412905	4413663	1	0	1	2,14	+/-1,73 SD
174	chrXII	4415573	4420839	1	0	1	0,79	+/-0,49 SD
175	chrXII	6135530	6138030	1	0	1	3,31	+/-1,58 SD
176	chrXII	6609939	6618530	1	1	1	2,55	+/-1,15 SD
177	chrXII	6618652	6622839	1	1	1	1,94	+/-0,95 SD
178	chrXII	6628530	6631030	1	1	1	2,26	+/-1,16 SD
179	chrXII	6735707	6740042	1	0	1	2,58	+/-1,2 SD
180	chrXII	6924030	6926530	1	0	1	2,24	+/-1,11 SD
181	chrXII	7806030	7825864	1	1	1	4,03	+/-1,68 SD
182	chrXII	8184917	8186817	1	1	1	4,39	+/-2,11 SD
183	chrXII	8223678	8225428	1	1	1	3,03	+/-1,59 SD
184	chrXII	8945972	8947832	1	0	1	5,6	+/-2,54 SD
185	chrXII	8970195	8974096	1	1	1	4,35	+/-1,94 SD
186	chrXII	10886530	10889030	1	0	1	4,91	+/-2,22 SD
187	chrXIII	10169683	10172183	1	0	1	3,29	+/-1,54 SD
188	chrXIV	11774577	11785908	1	0	1	4,86	+/-2,06 SD
189	chrXIX	2613275	2672911	1	1	1	3,73	+/-1,64 SD
190	chrXIX	2673911	2741294	1	1	1	3,04	+/-1,37 SD
191	chrXIX	7301028	7302353	1	1	1	3,39	+/-1,73 SD
192	chrXIX	9616100	9616868	1	0	1	4,63	+/-2,35 SD
193	chrXIX	9617425	9633425	1	0	1	6,49	+/-2,71 SD
194	chrXIX	9639425	9647425	1	1	1	6,29	+/-2,78 SD
195	chrXIX	10282425	10284925	1	1	1	7,85	+/-3,58 SD
196	chrXIX	10579425	10581925	1	0	1	4,4	+/-2,02 SD
197	chrXV	7513929	7514427	1	0	1	1,93	+/-1,32 SD
198	chrXV	7514457	7516989	1	0	1	3,33	+/-1,56 SD
199	chrXV	8273701	8274622	1	0	1	2,92	+/-1,66 SD
200	chrXVI	13915496	13915671	1	0	1	4,3	+/-3,1 SD
201	chrXVI	17841073	17844573	1	0	1	1,63	+/-1,01 SD
202	chrXVIII	869710	881820	1	0	1	6,05	+/-2,49 SD
203	chrXVIII	887500	898000	1	0	1	6,68	+/-2,75 SD
204	chrXVIII	4290500	4293000	1	0	1	2,82	+/-1,4 SD
205	chrXX	618036	639311	1	1	1	5,78	+/-2,46 SD
206	chrXX	654064	660088	1	1	1	5,88	+/-2,56 SD
207	chrXX	674015	686169	1	0	1	5,3	+/-2,46 SD
208	chrXX	5544497	5546904	1	0	1	2,26	+/-1,3 SD
209	chrXX	5547995	5549707	1	0	1	3,68	+/-2,02 SD
210	chrXX	5890884	5896221	1	0	1	5,47	+/-2,42 SD
211	chrXX	6041173	6045173	1	0	1	3,64	+/-1,63 SD
212	chrXX	7909794	7916201	1	0	1	4,91	+/-2,11 SD
213	chrXX	7935151	7948173	1	0	1	3,02	+/-1,29 SD

214	chrXX	7950173	7952673	1	0	1	3,15	+/-1,47 SD
215	chrXX	10219173	10221673	1	0	1	2,56	+/-1,23 SD
216	chrXX	10575173	10577673	1	1	1	5,04	+/-2,18 SD
217	chrXX	10635673	10639173	1	1	1	3,54	+/-1,63 SD
218	chrXX	10649002	10657888	1	0	1	7,69	+/-3,52 SD
219	chrXX	10659173	10661673	1	0	1	2,78	+/-1,53 SD
220	chrXX	10734673	10737173	1	0	1	8,05	+/-3,59 SD
221	chrXX	11054673	11057173	1	1	1	3,4	+/-1,56 SD
222	chrXX	11064673	11067173	1	1	1	7,53	+/-3,23 SD
223	chrXX	11068673	11071173	1	1	1	6	+/-2,63 SD
224	chrXX	11171173	11173673	1	1	1	4,41	+/-1,99 SD
225	chrXX	11677673	11680173	1	0	1	6,61	+/-2,88 SD
226	chrXX	11726173	11728673	1	1	1	7,53	+/-3,27 SD
227	chrXX	12458658	12459092	1	0	1	0,62	+/-0,57 SD
228	chrXX	13220173	13222673	1	0	1	3,03	+/-1,62 SD
229	chrXX	13596772	13599772	1	1	1	6,35	+/-2,67 SD
230	chrXX	14316606	14325920	1	1	1	8,07	+/-3,38 SD
231	chrXX	14343772	14346272	1	1	1	5,29	+/-2,37 SD
232	chrXX	14477643	14478364	1	0	1	7,73	+/-3,82 SD
233	chrXX	15916272	15928046	1	1	1	4,61	+/-1,94 SD
234	chrXX	18703126	18703364	0	0	0	10,4	+/-9,07 SD
235	chrXXI	4904027	4922176	1	1	1	3,05	+/-1,35 SD
236	chrXXI	4923634	4932158	1	1	1	1,87	+/-0,89 SD
237	chrXXI	7694615	7697115	1	0	1	0,85	+/-0,58 SD
238	chrXXI	9874605	9881330	1	0	1	7,19	+/-3,05 SD
239	chrXXI	9934979	9936866	1	1	1	15,06	+/-6,57 SD
240	chrXXI	11410944	11413444	1	1	1	4,38	+/-2,01 SD
neutral	chrII	8956684	9136929	1	0	0	6,54	+/-2,65 SD
neutral	chrII	14636813	14814756	1	0	0	6,71	+/-2,73 SD
neutral	chrIX	4124152	4292881	1	0	0	6,21	+/-2,54 SD
neutral	chrV	8849423	9036047	1	0	0	4,67	+/-1,92 SD
neutral	chrVIII	14441450	14628866	1	0	0	1,19	+/-0,6 SD
neutral	chrX	13543958	13702754	1	0	0	7,84	+/-3,18 SD
neutral	chrXII	17691673	17859299	1	0	0	7,19	+/-2,94 SD
neutral	chrXIII	17307239	17474739	1	0	0	5,11	+/-2,08 SD
neutral	chrXIX	7647713	7819503	1	0	0	5,93	+/-2,54 SD
neutral	chrXVI	311403	509357	1	0	0	5,88	+/-2,39 SD

Table S2: List of divergent and non-divergent (neutral) regions. The table includes the number of each divergent region, the genomic coordinates and the mean depth as well as the standard deviation (SD). Information which regions were included in the pull-down, in the processing with the HMM approach and the Bayesian Probability approach (BP) are also displayed in the table.

BACname	*BAClibrary*	*chromosome*	*start position*	*end position*	*divergent regions within BAC region*
STB59-F15_CH215	freshwater	chrI	7589810	7768138	1
STA206-F07_CH213	marine	chrI	7596322	7775009	1
STB239-J14_CH215	freshwater	chrI	8348185	8564529	2
STA222-J19_CH213	marine	chrI	8404587	8572773	2
STB26-B04_CH215	freshwater	chrI	9005181	9213199	3
STA225-H08_CH213	marine	chrI	9008872	9225194	3
STB36-A09_CH215	freshwater	chrI	11995268	12173769	4
STA235-F21_CH213	marine	chrI	11997738	12164623	4
STB209-F07_CH215	freshwater	chrI	12377893	12599674	5
STA62-M22_CH213	marine	chrI	12384587	12589132	5
STA198-G18_CH213	marine	chrI	12559956	12728044	6,7
STB55-A13_CH215	freshwater	chrI	12562068	12727736	6,7
STB39-N14_CH215	freshwater	chrI	12721016	12890817	8
STA05-F10_CH213	marine	chrI	12721041	12909741	8
STA21-A15_CH213	marine	chrI	14256386	14437038	9,1
STB08-B09_CH215	freshwater	chrI	14262689	14442935	9,1
STA38-E03_CH213	marine	chrI	17333546	17527530	11
STB222-H02_CH215	freshwater	chrI	17439745	17552042	11
STA216-J14_CH213_b	marine	chrI	25974514	26105482	12,13,14
STB22-P10_CH215	freshwater	chrI	26089646	26218526	13,14,15,16,17
STA04-F01_CH213_b	marine	chrI	26105482	26233774	14,15,16,17
STA12-L03_CH213	marine	chrI	26136012	26346208	14,15,16,17,18,19
STB09-L01_CH215	freshwater	chrI	26208117	26389256	17,18,19
STA60-L13_CH213	marine	chrI	26315853	26516333	19,2
STB39-G08_CH215	freshwater	chrI	26378914	26556288	19,20,21
STA216-J14_CH213_a	marine	chrI	26501739	26544039	20,21
STA04-F01_CH213_a	marine	chrI	26544039	26611075	21
STA06-N21_CH213	marine	chrII	851737	1029815	22
STB49-P23_CH215	freshwater	chrII	879372	1029842	22
STA57-O06_CH213	marine	chrII	2767605	2946237	23
STB33-C05_CH215	freshwater	chrII	2768277	2930799	23
STB37-I08_CH215	freshwater	chrII	6706857	6870191	24
STA10-H20_CH213	marine	chrII	6709150	6892589	24
STA24-E22_CH213	marine	chrII	8956658	9136929	non-divergent/neutral
STB40-A24_CH215	freshwater	chrII	8956684	9136934	non-divergent/neutral
STA199-L10_CH213	marine	chrII	14636792	14814804	non-divergent/neutral
STB214-K22_CH215	freshwater	chrII	14636813	14814756	non-divergent/neutral
STB236-A08_CH215	freshwater	chrII	15160340	15358480	25,26,27,28,29
STA43-C22_CH213	marine	chrII	15162374	15355788	25,26,27,28,29

STB230-J15_CH215	freshwater	chrII	16701468	16906864	30
STA203-N20_CH213	marine	chrII	16729668	16910246	30
STB14-F09_CH215	freshwater	chrIV	8284488	8466473	31
STA234-B14_CH213	marine	chrIV	8287000	8458606	31
STA243-K02_CH213	marine	chrIV	8453066	8636809	32
STB60-L02_CH215	freshwater	chrIV	8453694	8622294	32
STA205-A17_CH213	marine	chrIV	9196828	9355394	33
STB12-O14_CH215	freshwater	chrIV	9208191	9371190	33
STA14-J02_CH213	marine	chrIV	11148599	11349897	34
STB237-M15_CH215	freshwater	chrIV	11150140	11324948	34
STA44-F20_CH213	marine	chrIV	11916909	12130371	35,36
STB207-B07_CH215	freshwater	chrIV	11922120	12121413	35,36
STA56-B11_CH213	marine	chrIV	12087086	12284111	37,38,39
STB242-H11_CH215	freshwater	chrIV	12120468	12313653	37,38,39
STA241-K05_CH213	marine	chrIV	12256009	12430797	40
STB28-N12_CH215	freshwater	chrIV	12299965	12475210	40,41
STA02-F16_CH213	marine	chrIV	12428290	12637812	41,42
STB09-J21_CH215	freshwater	chrIV	12460006	12658063	42
STA20-L12_CH213	marine	chrIV	12606958	12810158	42,43,44
STB35-G06_CH215	freshwater	chrIV	12612110	12789509	43
STB37-K19_CH215	freshwater	chrIV	12789533	12959093	44,45,46,47,48
STA81-B05_CH213	marine	chrIV	12799621	12985769	44,45,46,47,48
STA230-K18_CH213	marine	chrIV	13297944	13473500	49,5
STB52-B19_CH215	freshwater	chrIV	13304774	13467122	49,5
STB229-P07_CH215	freshwater	chrIV	13529269	13730805	51,52
STA217-C01_CH213	marine	chrIV	13539886	13714207	51,52
STB23-I24_CH215	freshwater	chrIV	13739612	13918722	53,54,55
STA205-B21_CH213	marine	chrIV	13780645	13930313	53,54,55,56
STA52-I06_CH213	marine	chrIV	13876523	14045806	55,56,57,58,59,60,61
STB21-E22_CH215	freshwater	chrIV	13885304	14013680	55,56,57,58,59,60,61
STB34-E05_CH215	freshwater	chrIV	14983338	15156390	62,63,64,65
STA59-O13_CH213	marine	chrIV	14983354	15186944	62,63,64,65
STB198-I01_CH215	freshwater	chrIV	15320678	15523961	66,67,68
STA229-F23_CH213	marine	chrIV	15349379	15509042	66,67,68
STA214-G10_CH213	marine	chrIV	19560011	19721400	70
STB40-N13_CH215	freshwater	chrIV	19560743	19710608	70
STB252-P01_CH215	freshwater	chrIV	19959822	20164602	71,72,73
STA50-J21_CH213	marine	chrIV	19984775	20203382	71,72,73
STB39-M13_CH215	freshwater	chrIV	20033543	20215390	71,72,73
STA57-I15_CH213	marine	chrIV	20186620	20384312	74
STB22-L02_CH215	freshwater	chrIV	20203421	20368602	74
STB10-L21_CH215	freshwater	chrIV	20903442	21066786	75
STA194-P07_CH213	marine	chrIV	20904148	21077279	75

STB221-F11_CH215	freshwater	chrIV	21166278	21390289	76
STA13-B04_CH213	marine	chrIV	21181204	21387953	76
STA126-N06_CH213	marine	chrIV	22038636	22261646	77,78,79
STB28-H07_CH215	freshwater	chrIV	22137114	22338156	77,78,79,80
STA58-P14_CH213	marine	chrIV	22266715	22469009	80
STA09-D23_CH213	marine	chrIV	22700977	22879074	81
STB39-K22_CH215	freshwater	chrIV	22717462	22879121	81
STB27-E11_CH215	freshwater	chrIV	24180625	24343166	82
STA56-I01_CH213	marine	chrIV	24273009	24501851	82,83
STB223-H17_CH215	freshwater	chrIV	24333135	24536104	83
STA56-M05_CH213	marine	chrIV	24768437	24958640	84
STB14-C22_CH215	freshwater	chrIV	24774547	24936856	84
STB60-I23_CH215	freshwater	chrIV	24926843	25097948	85,86
STA209-P20_CH213	marine	chrIV	24944378	25124178	85,86
STB209-N14_CH215	freshwater	chrIV	26285717	26443230	87
STA56-B01_CH213	marine	chrIV	26286992	26440066	87
STA251-D20_CH213	marine	chrIV	26696325	26874679	88,89,90,91
STB38-L13_CH215	freshwater	chrIV	26699639	26873742	88,89,90,91
STB56-G23_CH215	freshwater	chrIV	28808328	28975384	92
STA64-N11_CH213	marine	chrIV	28808383	29023007	92
STB40-H15_CH215	freshwater	chrIX	4124106	4292881	non-divergent/neutral
STA29-H17_CH213	marine	chrIX	4124152	4292886	non-divergent/neutral
STA200-F21_CH213	marine	chrIX	4838390	5006912	93
STB231-E01_CH215	freshwater	chrIX	4849686	5001941	93
STA251-K10_CH213	marine	chrIX	9630327	9825012	94
STB59-C24_CH215	freshwater	chrIX	9655172	9826703	94
STB43-B15_CH215	freshwater	chrIX	11796905	11975924	95
STA196-G18_CH213	marine	chrIX	11820791	11981454	95
STA05-D19_CH213	marine	chrIX	12933038	13139125	96,97,98
STB07-N04_CH215	freshwater	chrIX	12944906	13104233	96,97,98
STA52-O19_CH213	marine	chrIX	13289061	13453235	99,100,101
STB27-G08_CH215	freshwater	chrIX	13306928	13456749	99,100,101
STA39-O21_CH213	marine	chrIX	13500771	13710539	102,103,104,105
STB36-B02_CH215	freshwater	chrIX	13515199	13690180	102,103,104
STA155-D12_CH213	marine	chrIX	13669375	13849507	105,106,107
STB227-B21_CH215	freshwater	chrIX	13691418	13841847	105,106,107
STB28-F15_CH215	freshwater	chrV	3956022	4116628	110
STA239-P18_CH213	marine	chrV	3966590	4119950	110
STA243-E14_CH213	marine	chrV	5844858	6029114	111
STB27-I15_CH215	freshwater	chrV	5856439	6039196	111
STA230-C24_CH213	marine	chrV	8261096	8425999	112,113
STB24-N11_CH215	freshwater	chrV	8282399	8432777	112,113
STB60-B06_CH215	freshwater	chrV	8849384	9036077	non-divergent/neutral

STA46-O09_CH213	marine	chrV	8849423	9036047	non-divergent/neutral
STB50-I21_CH215	freshwater	chrVI	6381409	6553171	114
STA196-I10_CH213	marine	chrVI	6384167	6555705	114
STB38-O22_CH215	freshwater	chrVI	10332141	10487614	115
STA206-M19_CH213	marine	chrVI	10335624	10484966	115
STB38-C07_CH215	freshwater	chrVII	8507133	8679481	116,117,118
STA59-O16_CH213	marine	chrVII	8507135	8685154	116,117,118
STA01-I07_CH213	marine	chrVII	8674093	8870372	119,12
STB252-O24_CH215	freshwater	chrVII	8677476	8876247	119,12
STB227-C12_CH215	freshwater	chrVII	10660339	10887117	121
STA200-I23_CH213	marine	chrVII	10667912	10893322	121
STA09-F20_CH213	marine	chrVII	14689616	14897833	122,123
STB209-B13_CH215	freshwater	chrVII	14718950	14924340	122,123
STA210-M24_CH213	marine	chrVII	15962329	16116926	124,125
STB52-C10_CH215	freshwater	chrVII	15963491	16127041	124,125
STA54-F12_CH213	marine	chrVII	17361748	17553241	126
STB19-H14_CH215	freshwater	chrVII	17387466	17553243	126
STA25-K07_CH213	marine	chrVII	17538879	17732112	127,128
STB31-K03_CH215	freshwater	chrVII	17540273	17698759	127,128
STA247-J23_CH213	marine	chrVII	18738445	18911821	129,13
STB27-D16_CH215	freshwater	chrVII	18742834	18898963	129,13
STA213-J03_CH213	marine	chrVII	19621140	19796576	131,132,133
STB07-M17_CH215	freshwater	chrVII	19637010	19789620	131,132,133
STB22-D18_CH215	freshwater	chrVII	19769111	19949339	134,135
STA246-F18_CH213	marine	chrVII	19792703	19958543	134,135
STA55-A01_CH213	marine	chrVII	19925184	20119689	136,137,138
STB33-A03_CH215	freshwater	chrVII	19937975	20108538	136,137,138
STA231-B06_CH213	marine	chrVII	20209155	20375645	139
STB223-D19_CH215	freshwater	chrVII	20232044	20372228	139
STB42-P10_CH215	freshwater	chrVII	20556942	20726637	140,141
STA233-B04_CH213	marine	chrVII	20563594	20720067	140,141
STA47-E23_CH213	marine	chrVII	21869084	22067214	142
STB35-J20_CH215	freshwater	chrVII	21894698	22069254	142
STA19-N15_CH213	marine	chrVIII	2617666	2828336	143
STB247-A09_CH215	freshwater	chrVIII	2618398	2813543	143
STB59-C20_CH215	freshwater	chrVIII	9359128	9532189	144
STA28-N14_CH213	marine	chrVIII	9364954	9555932	144
STA217-H20_CH213	marine	chrVIII	9986314	10157015	145,146
STB45-N21_CH215	freshwater	chrVIII	9986385	10157525	145,146
STB41-C05_CH215	freshwater	chrVIII	10507261	10665322	147
STA51-O02_CH213	marine	chrVIII	10511978	10697211	147
STA33-C22_CH213	marine	chrVIII	10803712	11001944	148,149,150
STB23-D04_CH215	freshwater	chrVIII	10821863	11001977	148,149,150

STB194-I06_CH215	freshwater	chrVIII	13445787	13659358	151
STA243-B15_CH213	marine	chrVIII	13448411	13627002	151
STB35-O20_CH215	freshwater	chrVIII	14441410	14628909	non-divergent/neutral
STA09-A13_CH213	marine	chrVIII	14441450	14628866	non-divergent/neutral
STA205-D03_CH213	marine	chrX	3035864	3234922	152
STB222-L07_CH215	freshwater	chrX	3056436	3255664	152
STA194-H02_CH213	marine	chrX	10961870	11131054	153
STB27-M04_CH215	freshwater	chrX	10996591	11153809	153
STA18-L05_CH213	marine	chrX	13543957	13702754	non-divergent/neutral
STB33-G04_CH215	freshwater	chrX	13543958	13702767	non-divergent/neutral
STA02-E23_CH213	marine	chrXI	5380481	5561396	154
STA244-H12_CH213_a	marine	chrXI	6011602	6192966	155
STB226-F04_CH215	freshwater	chrXI	6131805	6346793	155,156,157,158,159,160
STA39-D21_CH213_a	marine	chrXI	6192966	6212960	non-divergent/neutral
STA49-A16_CH213	marine	chrXI	6198057	6409078	155,156,157,158,159,160
STB207-B23_CH215	freshwater	chrXI	6326161	6514954	158,159,160,161
STA53-G14_CH213	marine	chrXI	6368766	6567990	160,161,162,163
STB216-K04_CH215	freshwater	chrXI	6502300	6704347	161,162,163,164
STA244-H12_CH213_b	marine	chrXI	6604178	6604235	non-divergent/neutral
STA39-D21_CH213_b	marine	chrXI	6604235	6783245	163,164
STB200-A02_CH215	freshwater	chrXI	6679084	6875249	non-divergent/neutral
STA26-F11_CH213	marine	chrXI	6770678	6986746	165
STB225-F16_CH215	freshwater	chrXI	6848317	7044561	165
STA50-H23_CH213	marine	chrXI	7896251	8104766	166
STB27-E23_CH215	freshwater	chrXI	7913620	8088723	166
STA246-D14_CH213	marine	chrXI	8362362	8531157	167,168
STB198-C20_CH215	freshwater	chrXI	8363765	8521763	167,168
STB195-A06_CH215	freshwater	chrXI	8680918	8901951	169
STA167-F17_CH213	marine	chrXI	8684009	8873965	169
STA16-M11_CH213	marine	chrXI	10973542	11173939	170
STB228-H15_CH215	freshwater	chrXI	10998802	11174847	170
STB43-O17_CH215,76	freshwater	chrXII	4422513	18760880	175,176,177,178,179,180,181,182,183,184,185,186
STA24-N18_CH213	marine	chrXII	4436686	18760880	175,176,177,178,179,180,181,182,183,184,185,186
STB26-F18_CH215	freshwater	chrXII	6052099	6217791	175
STA216-E19_CH213	marine	chrXII	6055689	6216565	175
STA164-J08_CH213	marine	chrXII	6576454	6760814	176,177,178,179
STB15-J23_CH215	freshwater	chrXII	6601834	6751492	176,177,178,179
STA05-B03_CH213	marine	chrXII	6837002	7045169	180
STB235-F08_CH215	freshwater	chrXII	6842897	7057539	180
STB28-E02_CH215	freshwater	chrXII	7689208	7875090	181

STA17-E08_CH213	marine	chrXII	7744434	7940080	181
STB235-B08_CH215	freshwater	chrXII	8120320	8313131	182,183
STA50-P19_CH213	marine	chrXII	8137609	8331157	182,183
STA223-I20_CH213	marine	chrXII	8879578	9061093	184,185
STB34-N06_CH215	freshwater	chrXII	8894446	9047851	184,185
STB41-D17_CH215	freshwater	chrXII	10789039	10953824	186
STA244-J19_CH213	marine	chrXII	10793409	10969434	186
STB43-E17_CH215	freshwater	chrXII	17691667	17859299	non-divergent/neutral
STA156-K15_CH213	marine	chrXII	17691673	17859343	non-divergent/neutral
STB218-A22_CH215	freshwater	chrXIII	10081534	10272390	187
STA229-H09_CH213	marine	chrXIII	10083725	10268162	187
STB26-I16_CH215	freshwater	chrXIII	17307202	17474761	non-divergent/neutral
STA29-N13_CH213	marine	chrXIII	17307239	17474739	non-divergent/neutral
STB01-O05_CH215	freshwater	chrXIV	11693075	11868676	188
STA224-L05_CH213	marine	chrXIV	11703028	11870607	188
STB208-E12_CH215	freshwater	chrXIX	2605912	2805363	189,19
STA251-E01_CH213	marine	chrXIX	2610468	2806870	189,19
STB29-A15_CH215	freshwater	chrXIX	7208826	7356092	191
STA47-A21_CH213	marine	chrXIX	7213469	7397399	191
STB23-A02_CH215	freshwater	chrXIX	7647681	7819512	non-divergent/neutral
STA11-J13_CH213	marine	chrXIX	7647713	7819503	non-divergent/neutral
STB28-B11_CH215	freshwater	chrXIX	9533838	9721439	192,193,194
STA34-C13_CH213	marine	chrXIX	9533844	9721428	192,193,194
STB212-D9_CH215	freshwater	chrXIX	10198230	10373229	195
STA250-B05_CH213	marine	chrXIX	10209740	10376628	195
STB215-F02_CH215	freshwater	chrXIX	10499483	10653699	196
STA252-O08_CH213	marine	chrXIX	10504548	10688314	196
STA42-H22_CH213	marine	chrXV	7420542	7627590	197,198
STB231-D05_CH215	freshwater	chrXV	7449670	7663198	197,198
STA30-G09_CH213	marine	chrXV	8171372	8358182	199
STB56-O17_CH215	freshwater	chrXV	8184906	8363698	199
STB29-F09_CH215	freshwater	chrXVI	311395	509382	non-divergent/neutral
STA124-I22_CH213	marine	chrXVI	311403	509357	non-divergent/neutral
STB194-A15_CH215	freshwater	chrXVI	13752331	13951638	200
STA205-D10_CH213	marine	chrXVI	13768343	13930222	200
STB07-L06_CH215	freshwater	chrXVI	17754415	17948538	201
STA55-J08_CH213	marine	chrXVIII	774893	968032	202,203
STB222-E07_CH215	freshwater	chrXVIII	779761	959628	202,203
STB17-K15_CH215	freshwater	chrXVIII	4188929	4372368	204
STA175-K16_CH213	marine	chrXVIII	4203972	4367486	204
STA33-J18_CH213	marine	chrXX	479173	703079	205,206,207
STB41-K13_CH215	freshwater	chrXX	560377	746612	205,206,207
STB229-A16_CH215	freshwater	chrXX	5428413	5632154	208,209

STA217-E21_CH213	marine	chrXX	5478354	5646072	208,209
STA03-I03_CH213	marine	chrXX	5850311	6060818	210,211
STB230-K23_CH215	freshwater	chrXX	5879618	6055889	210,211
STA197-D03_CH213	marine	chrXX	7859836	8028008	212,213,214
STB42-K18_CH215	freshwater	chrXX	7860074	8030531	212,213,214
STB206-I05_CH215	freshwater	chrXX	10127409	10314302	215
STA14-F03_CH213	marine	chrXX	10138777	10315947	215
STA151-E14_CH213	marine	chrXX	10558500	10762754	216,217,218,219,220
STB209-P17_CH215	freshwater	chrXX	10574677	10753927	216,217,218,219,220
STA216-J06_CH213	marine	chrXX	11008694	11186455	221,222,223,224
STB28-I15_CH215	freshwater	chrXX	11015670	11181019	221,222,223,224
STA249-J14_CH213	marine	chrXX	11612934	11797212	225,226
STB13-P12_CH215	freshwater	chrXX	11617336	11780682	225,226
STB59-K16_CH215	freshwater	chrXX	12377884	12536949	227
STA201-C02_CH213	marine	chrXX	12380622	12529807	227
STB32-C12_CH215	freshwater	chrXX	13156263	13324689	228
STA242-L02_CH213	marine	chrXX	13161951	13322403	228
STA206-A01_CH213	marine	chrXX	13510718	13678630	229
STB237-O12_CH215	freshwater	chrXX	13532386	13686181	229
STA31-E14_CH213	marine	chrXX	14276056	14486812	230,231,232
STB221-F04_CH215	freshwater	chrXX	14292465	14485031	230,231,232
STA215-G06_CH213	marine	chrXX	15854426	16025558	233
STB46-C17_CH215	freshwater	chrXX	15854608	16005022	233
STB207-P12_CH215	freshwater	chrXXI	4769817	4968265	235,236
STA226-O16_CH213	marine	chrXXI	4814255	4996635	235,236
STA246-N08_CH213	marine	chrXXI	7526316	7947002	237
STB250-E07_CH215	freshwater	chrXXI	7553900	7852823	237
STA93-C04_CH213_a	marine	chrXXI	9788036	9948192	238,239
STB211-L11_CH215	freshwater	chrXXI	9826876	10023048	238,239
STA228-E24_CH213	marine	chrXXI	11323296	11495790	240
STB59-K01_CH215	freshwater	chrXXI	11324891	11500715	240
STA93-C04_CH213_b	marine	chrXXI	11615969	11641129	non-divergent/neutral

Table S3: List of BAC clones used for targeted enrichment. The table includes the name of each BAC, the respective library (freshwater or marine), the genomic position of the BAC clone and the number of the divergent regions the BAC is spanning. BAC clones spanning the end of chrI, chrXI and chrXXI inversion are each split into two parts (marked with the suffix _a and _b).

880

name of adapter	i5 or i7 barcodes within adapter	adapter
Nextera_N501	TAGATCGC	AATGATACGGCGACCACCGAGATCTACACTAGATCGCTCGTCGGCAGCGTC
Nextera_N502	CTCTCTAT	AATGATACGGCGACCACCGAGATCTACACCTCTCTATTCGTCGGCAGCGTC
Nextera_N503	TATCCTCT	AATGATACGGCGACCACCGAGATCTACACTATCCTCTTCGTCGGCAGCGTC
Nextera_N504	AGAGTAGA	AATGATACGGCGACCACCGAGATCTACACAGAGTAGATCGTCGGCAGCGTC
Nextera_N505	GTAAGGAG	AATGATACGGCGACCACCGAGATCTACACGTAAGGAGTCGTCGGCAGCGTC
Nextera_N506	ACTGCATA	AATGATACGGCGACCACCGAGATCTACACACTGCATATCGTCGGCAGCGTC
Nextera_N507	AAGGAGTA	AATGATACGGCGACCACCGAGATCTACACAAGGAGTATCGTCGGCAGCGTC
Nextera_N508	CTAAGCCT	AATGATACGGCGACCACCGAGATCTACACCTAAGCCTTCGTCGGCAGCGTC
Nextera_N701	TCGCCTTA	CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTCTCGTGGGCTCGG
Nextera_N702	CTAGTACG	CAAGCAGAAGACGGCATACGAGATCTAGTACGGTCTCGTGGGCTCGG
Nextera_N703	TTCTGCCT	CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTCTCGTGGGCTCGG
Nextera_N704	GCTCAGGA	CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTCTCGTGGGCTCGG
Nextera_N705	AGGAGTCC	CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTCTCGTGGGCTCGG
Nextera_N706	CATGCCTA	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTCTCGTGGGCTCGG
Nextera_N707	GTAGAGAG	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTCTCGTGGGCTCGG
Nextera_N708	CCTCTCTG	CAAGCAGAAGACGGCATACGAGATCCTCTCTGGTCTCGTGGGCTCGG
Nextera_N709	AGCGTAGC	CAAGCAGAAGACGGCATACGAGATAGCGTAGCGTCTCGTGGGCTCGG
Nextera_N710	CAGCCTCG	CAAGCAGAAGACGGCATACGAGATCAGCCTCGGTCTCGTGGGCTCGG
Nextera_N711	TGCCTCTT	CAAGCAGAAGACGGCATACGAGATTGCCTCTTGTCTCGTGGGCTCGG
Nextera_N712	TCCTCTAC	CAAGCAGAAGACGGCATACGAGATTCCTCTACGTCTCGTGGGCTCGG
Nextera_N714	TCATGAGC	CAAGCAGAAGACGGCATACGAGATTCATGAGCGTCTCGTGGGCTCGG
Nextera_N715	CCTGAGAT	CAAGCAGAAGACGGCATACGAGATCCTGAGATGTCTCGTGGGCTCGG

Nextera_N716	TAGCGAGT	CAAGCAGAAGACGGCATACGAGATTAGCGAGTGTCTCGTGGGCTCGG
Nextera_N718	GTAGCTCC	CAAGCAGAAGACGGCATACGAGATGTAGCTCCGTCTCGTGGGCTCGG
Nextera_N719	TACTACGC	CAAGCAGAAGACGGCATACGAGATTACTACGCGTCTCGTGGGCTCGG
Nextera_N720	AGGCTCCG	CAAGCAGAAGACGGCATACGAGATAGGCTCCGGTCTCGTGGGCTCGG
Nextera_N721	GCAGCGTA	CAAGCAGAAGACGGCATACGAGATGCAGCGTAGTCTCGTGGGCTCGG
Nextera_N722	CTGCGCAT	CAAGCAGAAGACGGCATACGAGATCTGCGCATGTCTCGTGGGCTCGG
Nextera_N723	GAGCGCTA	CAAGCAGAAGACGGCATACGAGATGAGCGCTAGTCTCGTGGGCTCGG
Nextera_N724	CGCTCAGT	CAAGCAGAAGACGGCATACGAGATCGCTCAGTGTCTCGTGGGCTCGG
Nextera_N726	GTCTTAGG	CAAGCAGAAGACGGCATACGAGATGTCTTAGGGTCTCGTGGGCTCGG
Nextera_N727	ACTGATCG	CAAGCAGAAGACGGCATACGAGATACTGATCGGTCTCGTGGGCTCGG
Nextera_N728	TAGCTGCA	CAAGCAGAAGACGGCATACGAGATTAGCTGCAGTCTCGTGGGCTCGG
Nextera_N729	GACGTCGA	CAAGCAGAAGACGGCATACGAGATGACGTCGAGTCTCGTGGGCTCGG
Nextera_N730	TGTGGTTG	CAAGCAGAAGACGGCATACGAGATTGTGGTTGGTCTCGTGGGCTCGG
Nextera_N731	TAGTCTTG	CAAGCAGAAGACGGCATACGAGATTAGTCTTGGTCTCGTGGGCTCGG
Nextera_N732	TTCCATTG	CAAGCAGAAGACGGCATACGAGATTTCCATTGGTCTCGTGGGCTCGG
Nextera_N733	TCGAAGTG	CAAGCAGAAGACGGCATACGAGATTCGAAGTGGTCTCGTGGGCTCGG
Nextera_N734	TAACGCTG	CAAGCAGAAGACGGCATACGAGATTAACGCTGGTCTCGTGGGCTCGG
Nextera_N735	TTGGTATG	CAAGCAGAAGACGGCATACGAGATTTGGTATGGTCTCGTGGGCTCGG
Nextera_N736	TGAACTGG	CAAGCAGAAGACGGCATACGAGATTGAACTGGGTCTCGTGGGCTCGG
Nextera_N737	TACTTCGG	CAAGCAGAAGACGGCATACGAGATTACTTCGGGTCTCGTGGGCTCGG
Nextera_N738	TCTCACGG	CAAGCAGAAGACGGCATACGAGATTCTCACGGGTCTCGTGGGCTCGG
Nextera_N739	TCAGGAGG	CAAGCAGAAGACGGCATACGAGATTCAGGAGGGTCTCGTGGGCTCGG
Nextera_N740	TAAGTTCG	CAAGCAGAAGACGGCATACGAGATTAAGTTCGGTCTCGTGGGCTCGG
Nextera_N741	TCCAGTCG	CAAGCAGAAGACGGCATACGAGATTCCAGTCGGTCTCGTGGGCTCGG

Nextera_N742	TGTATGCG	CAAGCAGAAGACGGCATACGAGATTGTATGCGGTCTCGTGGGCTCGG
Nextera_N743	TCATTGAG	CAAGCAGAAGACGGCATACGAGATTCATTGAGGTCTCGTGGGCTCGG
Nextera_N744	TGGCTCAG	CAAGCAGAAGACGGCATACGAGATTGGCTCAGGTCTCGTGGGCTCGG
Nextera_N745	TATGCCAG	CAAGCAGAAGACGGCATACGAGATTATGCCAGGTCTCGTGGGCTCGG
Nextera_N746	TCAGATTC	CAAGCAGAAGACGGCATACGAGATTCAGATTCGTCTCGTGGGCTCGG
Nextera_N747	TACTAGTC	CAAGCAGAAGACGGCATACGAGATTACTAGTCGTCTCGTGGGCTCGG
Nextera_N748	TTCAGCTC	CAAGCAGAAGACGGCATACGAGATTTCAGCTCGTCTCGTGGGCTCGG
Nextera_N749	TGTCTATC	CAAGCAGAAGACGGCATACGAGATTGTCTATCGTCTCGTGGGCTCGG
Nextera_N750	TATGTGGC	CAAGCAGAAGACGGCATACGAGATTATGTGGCGTCTCGTGGGCTCGG
Nextera_N751	TTACTCGC	CAAGCAGAAGACGGCATACGAGATTTACTCGCGTCTCGTGGGCTCGG
Nextera_N752	TCGTTAGC	CAAGCAGAAGACGGCATACGAGATTCGTTAGCGTCTCGTGGGCTCGG
Nextera_N753	TACCGAGC	CAAGCAGAAGACGGCATACGAGATTACCGAGCGTCTCGTGGGCTCGG
Nextera_N754	TGTTCTCC	CAAGCAGAAGACGGCATACGAGATTGTTCTCCGTCTCGTGGGCTCGG
Nextera_N755	TTCGCACC	CAAGCAGAAGACGGCATACGAGATTTCGCACCGTCTCGTGGGCTCGG
Nextera_N756	TTGCGTAC	CAAGCAGAAGACGGCATACGAGATTTGCGTACGTCTCGTGGGCTCGG
Nextera_N757	TCTACGAC	CAAGCAGAAGACGGCATACGAGATTCTACGACGTCTCGTGGGCTCGG
Nextera_N758	TGACAGAC	CAAGCAGAAGACGGCATACGAGATTGACAGACGTCTCGTGGGCTCGG
Nextera_N759	TAGAACAC	CAAGCAGAAGACGGCATACGAGATTAGAACACGTCTCGTGGGCTCGG
Nextera_N760	TCATCCTA	CAAGCAGAAGACGGCATACGAGATTCATCCTAGTCTCGTGGGCTCGG
Nextera_N761	TGCTGATA	CAAGCAGAAGACGGCATACGAGATTGCTGATAGTCTCGTGGGCTCGG
Nextera_N762	TAGACGGA	CAAGCAGAAGACGGCATACGAGATTAGACGGAGTCTCGTGGGCTCGG
Nextera_N763	TGTGAAGA	CAAGCAGAAGACGGCATACGAGATTGTGAAGAGTCTCGTGGGCTCGG
Nextera_N764	TCTCTTCA	CAAGCAGAAGACGGCATACGAGATTCTCTTCAGTCTCGTGGGCTCGG
Nextera_N765	TTGTTCCA	CAAGCAGAAGACGGCATACGAGATTTGTTCCAGTCTCGTGGGCTCGG

Nextera_N766	TGAAGCCA	CAAGCAGAAGACGGCATACGAGATTGAAGCCAGTCTCGTGGGCTCGG
Nextera_N767	TACCACCA	CAAGCAGAAGACGGCATACGAGATTACCACCAGTCTCGTGGGCTCGG
Nextera_N768	TGCGTGAA	CAAGCAGAAGACGGCATACGAGATTGCGTGAAGTCTCGTGGGCTCGG
Nextera_N769	GGTGAGTT	CAAGCAGAAGACGGCATACGAGATGGTGAGTTGTCTCGTGGGCTCGG
Nextera_N770	GATCTCTT	CAAGCAGAAGACGGCATACGAGATGATCTCTTGTCTCGTGGGCTCGG
Nextera_N771	GTGTCCTT	CAAGCAGAAGACGGCATACGAGATGTGTCCTTGTCTCGTGGGCTCGG
Nextera_N772	GACGGATT	CAAGCAGAAGACGGCATACGAGATGACGGATTGTCTCGTGGGCTCGG
Nextera_N773	GCAACATT	CAAGCAGAAGACGGCATACGAGATGCAACATTGTCTCGTGGGCTCGG
Nextera_N774	GGTCGTGT	CAAGCAGAAGACGGCATACGAGATGGTCGTGTGTCTCGTGGGCTCGG
Nextera_N775	GAATCTGT	CAAGCAGAAGACGGCATACGAGATGAATCTGTGTCTCGTGGGCTCGG
Nextera_N776	GTACATCT	CAAGCAGAAGACGGCATACGAGATGTACATCTGTCTCGTGGGCTCGG
Nextera_N777	GAGGTGCT	CAAGCAGAAGACGGCATACGAGATGAGGTGCTGTCTCGTGGGCTCGG
Nextera_N778	GCATGGCT	CAAGCAGAAGACGGCATACGAGATGCATGGCTGTCTCGTGGGCTCGG
Nextera_N779	GTTAGCCT	CAAGCAGAAGACGGCATACGAGATGTTAGCCTGTCTCGTGGGCTCGG
Nextera_N780	GTCGCTAT	CAAGCAGAAGACGGCATACGAGATGTCGCTATGTCTCGTGGGCTCGG
Nextera_N781	GGAATGAT	CAAGCAGAAGACGGCATACGAGATGGAATGATGTCTCGTGGGCTCGG
Nextera_N782	GAGCCAAT	CAAGCAGAAGACGGCATACGAGATGAGCCAATGTCTCGTGGGCTCGG
Nextera_N783	GCTCCTTG	CAAGCAGAAGACGGCATACGAGATGCTCCTTGGTCTCGTGGGCTCGG
Nextera_N784	GTAAGGTG	CAAGCAGAAGACGGCATACGAGATGTAAGGTGGTCTCGTGGGCTCGG
Nextera_N785	GAGGATGG	CAAGCAGAAGACGGCATACGAGATGAGGATGGGTCTCGTGGGCTCGG
Nextera_N786	GTTGTCGG	CAAGCAGAAGACGGCATACGAGATGTTGTCGGGTCTCGTGGGCTCGG
Nextera_N787	GGATTAGG	CAAGCAGAAGACGGCATACGAGATGGATTAGGGTCTCGTGGGCTCGG
Nextera_N788	GATAGAGG	CAAGCAGAAGACGGCATACGAGATGATAGAGGGTCTCGTGGGCTCGG
Nextera_N789	GTGTGTCG	CAAGCAGAAGACGGCATACGAGATGTGTGTCGGTCTCGTGGGCTCGG

Nextera_N790	GCAATCCG	CAAGCAGAAGACGGCATACGAGATGCAATCCGGTCTCGTGGGCTCGG
Nextera_N791	GACCTTAG	CAAGCAGAAGACGGCATACGAGATGACCTTAGGTCTCGTGGGCTCGG
Nextera_N792	GCCTGTTC	CAAGCAGAAGACGGCATACGAGATGCCTGTTCGTCTCGTGGGCTCGG
Nextera_N793	GCACTGTC	CAAGCAGAAGACGGCATACGAGATGCACTGTCGTCTCGTGGGCTCGG
Nextera_N794	GCTAACTC	CAAGCAGAAGACGGCATACGAGATGCTAACTCGTCTCGTGGGCTCGG
Nextera_N795	GATTCATC	CAAGCAGAAGACGGCATACGAGATGATTCATCGTCTCGTGGGCTCGG
Nextera_N796	GTCTTGGC	CAAGCAGAAGACGGCATACGAGATGTCTTGGCGTCTCGTGGGCTCGG
Nextera_N797	TGCGATCT	CAAGCAGAAGACGGCATACGAGATTGCGATCTGTCTCGTGGGCTCGG
Nextera_N798	TTCCTGCT	CAAGCAGAAGACGGCATACGAGATTTCCTGCTGTCTCGTGGGCTCGG
Nextera_N799	TAGTGACT	CAAGCAGAAGACGGCATACGAGATTAGTGACTGTCTCGTGGGCTCGG
Nextera_N800	TACAGGAT	CAAGCAGAAGACGGCATACGAGATTACAGGATGTCTCGTGGGCTCGG
Nextera_N801	TCCTCAAT	CAAGCAGAAGACGGCATACGAGATTCCTCAATGTCTCGTGGGCTCGG
TrueSeq_i5		GATCGGAAGAGCACACGTCTGAACTCCAGTCACCAACCACAATCTCGTATGCCGTCTTCTGCTTG
TrueSeq_i7		TCTAGCCTTCTCGCAGCACATCCCTTTCTCACATCTAGAGCCACCAGCGGCATAGTAA

881

882 Table S4: Primer used in this study for nextera library preparation including respective barcodes.

883

Sample name	*Drainage*	*Region*	*Country*	*Salinity*	*Accession*
DRIZ_L_2009#02	Drizzle Lake	Haiida Gwai	Canada	freshwater	SAMN02781075
BNST|X|2006#01	Bonsall Creek	Vancouver Island	Canada	freshwater	SAMN02864868
BNST|X|2006#06	Bonsall Creek	Vancouver Island	Canada	freshwater	SAMN02869621
BNST|X2006#08	Bonsall Creek	Vancouver Island	Canada	freshwater	SAMN02781106
BNST|X|2006#09	Bonsall Creek	Vancouver Island	Canada	freshwater	SAMN02864898
BNST|X|2006#10	Bonsall Creek	Vancouver Island	Canada	freshwater	SAMN02864899
COND|X|2002#14	Connor Creek Site D	Washington State	USA	freshwater	SAMN02864876
LITC_23_32_2008#306	Little Campbell River	White Rock	Canada	freshwater	SAMN02781109
LITC_23_32_2008#324	Little Campbell River	White Rock	Canada	freshwater	SAMN02781085
LITC_23_32_2008#347	Little Campbell River	White Rock	Canada	freshwater	SAMN02781068
LITC_23_32_2008#356	Little Campbell River	White Rock	Canada	freshwater	SAMN02781678
LITC_23_32_2008#744	Little Campbell River	White Rock	Canada	freshwater	SAMN02781104
DAWS_X_2009#02	Dawson Marine Pond	Haiida Gwai	Canada	marine	SAMN02781101
BNMA|X|2006#01	Bonsall Creek	Vancouver Island	Canada	marine	SAMN02864903
BNMA|X|2006#02	Bonsall Creek	Vancouver Island	Canada	marine	SAMN02869625
BNMA|X|2006#03	Bonsall Creek	Vancouver Island	Canada	marine	SAMN02864851
BNMA|X|2006#05	Bonsall Creek	Vancouver Island	Canada	marine	SAMN02866134
BNMA|X|2006#07	Bonsall Creek	Vancouver Island	Canada	marine	SAMN02864845
MANC|X|X#05	Manchester Clam Bay	Washington State	USA	marine	SAMN02864902
LITC_0_05_2008#841	Little Campbell River	White Rock	Canada	marine	SAMN02781694
LITC_0_05_2008#842	Little Campbell River	White Rock	Canada	marine	SAMN02781680
LITC_0_05_2008#FL	Little Campbell River	White Rock	Canada	marine	SAMN02781063
LITC_0_05_2008#FO	Little Campbell River	White Rock	Canada	marine	SAMN02781107
LITC_0_45_2008#789	Little Campbell River	White Rock	Canada	marine	SAMN02781691

884

885 Table S5: Canadian individuals used for identifying ecotype-informative SNPs for Bayesian Probability
886 approach. The data was downloaded from the Sequence Read Archive (www.ncbi.nlm.nih.gov/sra).
887 Accessions for each sample are included in the table together with sample names, drainage, region,
888 country as well as respective ecotype.

Sample name	*Drainage*	*Region*	*Country*	*Salinity*	*armor plates*	*used for analysis*	*short-cut*
AKMA|X|2001#102	AlaskaMarine	Alaska	USA	mar	high	no	
AKST|X|2001#03	AlaskaStream	Alaska	USA	fw	low	yes	AK
ANSR_X_2009#01	Anser Lake	Haiida Gwai	Canada	fw	low	yes	HG
BARW|X|2012#04	Barrow	Alaska	USA	unknown	high	no	
BHAR|X|2011#02	Loch a'Bharpa	Uist	Scotland	unknown	unknown	no	
BIGR_1_32_2007#03	Big River	Northern California	USA	mar	high	no	
BIGR_52_54_2008#02	Big River	Northern California	USA	fw	high	no	
BIGR|1_32|2007#01	Big River	Northern California	USA	mar	high	no	
BIGR|1_32|2007#02	Big River	Northern California	USA	mar	high	no	
BIGR|3_63|2007#08	Big River	Northern California	USA	mar	high	no	
BIGR|3_63|2007#14	Big River	Northern California	USA	mar	high	no	
BIGR|52_54|2007#04	Big River	Northern California	USA	fw	high	no	
BIGR|52_54|2007#05	Big River	Northern California	USA	fw	high	no	
BIGR|52_54|2007#12	Big River	Northern California	USA	fw	high	no	
BIGR|52_54|2007#17	Big River	Northern California	USA	fw	high	no	
BITJ|X|X#17	Bitrufjördur	Northern	Iceland	mar	high	no	
BK70_X_2010#02	Peninsula Banks	Haiida Gwai	Canada	fw	low	yes	HG
BKW2_X_2010#01	Peninsula Banks	Haiida Gwai	Canada	fw	low	no	
BLAU|X|2002#08	Blautaver	Central	Iceland	fw	low	yes	ICE
BNMA|X|2006#01	Bonsall Creek	Vancouver Island	Canada	mar	high	no	
BNMA|X|2006#02	Bonsall Creek	Vancouver Island	Canada	mar	high	no	
BNMA|X|2006#03	Bonsall Creek	Vancouver Island	Canada	mar	high	no	
BNMA|X|2006#05	Bonsall Creek	Vancouver Island	Canada	mar	high	no	
BNMA|X|2006#07	Bonsall Creek	Vancouver Island	Canada	mar	high	no	
BNST|X|2006#01	Bonsall Creek	Vancouver Island	Canada	fw	low	yes	CA
BNST|X|2006#06	Bonsall Creek	Vancouver Island	Canada	fw	low	yes	CA
BNST|X|2006#09	Bonsall Creek	Vancouver Island	Canada	fw	low	yes	CA
BNST|X|2006#10	Bonsall Creek	Vancouver Island	Canada	fw	low	yes	CA
BNST|X2006#08	Bonsall Creek	Vancouver Island	Canada	fw	low	yes	CA
BOOT|X|2011#05	Boot Lake	Alaska	USA	fw	low	yes	AK
BOUL_X_2010#01	Boulton Lake	Haiida Gwai	Canada	fw	low	yes	HG
BOUL_X_2010#02	Boulton Lake	Haiida Gwai	Canada	fw	low	yes	HG
BRLY|X|2012#01	Barley Lake	Alaska	USA	unknown	low	yes	AK
BRNT_X_2009#05	Branta Lake	Haiida Gwai	Canada	fw	low	yes	HG
BSEA|55|2011#01	Bering Sea	Alaska	USA	mar	high	no	
CERC|X|X#04	Cerrito Creek	BayArea California	USA	fw	unknown	no	
CHRU|X|2011#04	Loch a Chadha Ruaidh	Uist	Scotland	unknown	unknown	no	

CMCB|X|2011#02	Community Club	Alaska	USA	fw	low	yes	AK
COAT|X|2009#90234	Coates Lake	Haiida Gwai	Canada	fw	low	yes	HG
COND|X|2002#14	Connor Creek Site D	Washington State	USA	fw	low	yes	USA
CORC|X|2011#01	Corcoran	Alaska	USA	fw	low	yes	AK
DAIM|X|2011#03	Loch an Daimh	Uist	Scotland	unknown	unknown	no	
DANS|X|X#05	Daniels Lake	Alaska	USA	fw	low	yes	AK
DARW_X_2009#03	Darwin Lake	Haiida Gwai	Canada	fw	low	no	
DAWS_X_2009#02	Dawson Marine Pond	Haiida Gwai	Canada	mar	low	no	
DNSE|X|2011#07	Denise Lake	Alaska	USA	fw	low	yes	AK
DRIZ_I_2010#01	Drizzle Inlet	Haiida Gwai	Canada	fw	low	yes	HG
DRIZ_L_2009#02	Drizzle Lake	Haiida Gwai	Canada	fw	low	yes	HG
DRIZ_L_2009#15	Drizzle Lake	Haiida Gwai	Canada	fw	low	yes	HG
DRIZ_O_2009#03	Drizzle Outlet	Haiida Gwai	Canada	fw	low	yes	HG
ECHO|X|2011#03	Echo Lake	Alaska	USA	fw	low	yes	AK
EDEN_X_2010#01	Eden	Haiida Gwai	Canada	fw	low	yes	HG
ERBC|X|X#8770	El Rosa Boca	Southern California	Mexico	fw	low	yes	SCAL
ESCP|X|1993#75_8	Escarpment	British Columbia	USA	unknown	unknown	no	
FADA_FJ_2001#17	Loch Fada	Uist	Scotland	fw	low	yes	SC
FRIC_X_2003#C10	Friant River Complete	Central California	USA	fw	high	no	
FRIL|X|X#05	Friant River Low Plated	Central California	USA	fw	low	yes	CCAL
GARC|X|X#711	Garrity Creek	BayArea California	USA	fw	unknown	no	
GIFU|X|2000#03	NAKA	Gifu	Japan	fw	unknown	no	
GOLD_X_2009#02	Gold Creek	Haiida Gwai	Canada	fw	low	yes	HG
GOLD|X|X#12	Gold Creek	Haiida Gwai	Canada	fw	low	yes	HG
HRUN|X|X#02	Hraun	Southwest	Iceland	fw	low	yes	ICE
HSTA|X|2011#03	Loch Hosta	Uist	Scotland	fw	low	yes	SC
JADE|X|2011#05	Jake Lake	Alaska	USA	fw	low	yes	AK
KALD|X|2010#03	Kaldbak	Faroe	Denmark	mar	partially	no	
KFSY|X|2011#05	Kalifonsky Lake	Alaska	USA	fw	low	yes	AK
KODK|X|2004#04	Kodiak Island	Aleutian Islands	USA	mar	low	yes	USA
LAMH|X|2001#09	Loch a' Mhuillin	West Highlands	Scotland	fw	low	yes	SC
LANO|X|2001#05	Loch an Nostarie	West Highlands	Scotland	fw	low	yes	SC
LAUR|X|1993#9_5	Laurel Pond	British Columbia	USA	unknown	unknown	no	
LITC_0_05_2008#841	Little Campbell River	White Rock	Canada	mar	high	no	

LITC_0_05_2008#842	Little Campbell River	White Rock	Canada	mar	high	no	
LITC_0_05_2008#FL	Little Campbell River	White Rock	Canada	mar	high	no	
LITC_0_05_2008#FO	Little Campbell River	White Rock	Canada	mar	high	no	
LITC_0_45_2008#789	Little Campbell River	White Rock	Canada	mar	high	no	
LITC_23_32_2008#306	Little Campbell River	White Rock	Canada	fw	low	yes	CA
LITC_23_32_2008#324	Little Campbell River	White Rock	Canada	fw	low	yes	CA
LITC_23_32_2008#347	Little Campbell River	White Rock	Canada	fw	low	yes	CA
LITC_23_32_2008#356	Little Campbell River	White Rock	Canada	fw	low	yes	CA
LITC_23_32_2008#744	Little Campbell River	White Rock	Canada	fw	low	yes	CA
LLYD|X|2003#03	Lloyd State Park	East Coast	USA	mar	high	no	
LMCK|X|X#03	Little Meadow Creek	Alaska	USA	fw	low	yes	AK
LMIA|X|X#02	Limia	NA	Spain	fw	low	yes	EU
LMOR|X|2001#12	Loch Morar	West Highlands	Scotland	fw	low	yes	SC
LNEA|M|2001#03	Loch nan Eala	West Highlands	Scotland	fw	low	yes	SC
LNGH|X|2011#01	Lough Neagh	County Wicklow	Ireland	fw	unknown	no	
LOBG|X|1999#09	Loberg Lake	Alaska	USA	fw	partially	no	
LOBG|X|1999#11	Loberg Lake	Alaska	USA	fw	low	no	
LOBG|X|1999#12	Loberg Lake	Alaska	USA	fw	low	no	
LOBG|X|1999#13	Loberg Lake	Alaska	USA	fw	low	no	
LOBG|X|1999#32	Loberg Lake	Alaska	USA	fw	low	no	
LOBG|X|1999#36	Loberg Lake	Alaska	USA	fw	high	no	
LOBG|X|1999#38	Loberg Lake	Alaska	USA	fw	high	no	
LOBG|X|1999#39	Loberg Lake	Alaska	USA	fw	partially	no	
LOBG|X|1999#43	Loberg Lake	Alaska	USA	fw	low	no	
LOBG|X|1999#46	Loberg Lake	Alaska	USA	fw	partially	no	
LOBG|X|2012#02	Loberg Lake	Alaska	USA	fw	partially	no	
LOBG|X|2012#03	Loberg Lake	Alaska	USA	fw	low	no	
LOBG|X|2012#04	Loberg Lake	Alaska	USA	fw	low	no	
LOBG|X|2012#05	Loberg Lake	Alaska	USA	fw	low	no	

LOBG|X|2012#06	Loberg Lake	Alaska	USA	fw	low	no	
LOBG|X|2012#08	Loberg Lake	Alaska	USA	fw	low	no	
LOBG|X|2012#11	Loberg Lake	Alaska	USA	fw	low	no	
LOBG|X|2012#17	Loberg Lake	Alaska	USA	fw	partially	no	
LOBG|X|2012#26	Loberg Lake	Alaska	USA	fw	low	no	
LOBG|X|2012#28	Loberg Lake	Alaska	USA	fw	low	no	
LONG|X|2011#01	Long Lake	Alaska	USA	fw	low	yes	AK
LSHP|X|2011#01	L Shape Lake	Alaska	USA	fw	low	yes	AK
LSOL|X|2012#04	Lake Solano	Northern California	USA	unknow n	low	no	
LSWL|X|2011#05	Lough Swilly	County Donegal	Ireland	mar	unknown	no	
LTLT|X|2011#06	Lough Talt	County Sligo	Ireland	fw	unknown	no	
LUTE|X|2003#76_4	Lutea Lake	Haiida Gwai	Canada	fw	low	yes	HG
MANC|X|X#05	Manchester Clam Bay	Washington State	USA	mar	high	no	
MAYR_X_2009#05	Mayer Lake	Haiida Gwai	Canada	fw	low	yes	HG
MAYR_X_2009#11	Mayer Lake	Haiida Gwai	Canada	fw	low	yes	HG
MAYR|X|2004#03	Mayer Lake	Haiida Gwai	Canada	fw	low	yes	HG
MAYR|X|2004#04	Mayer Lake	Haiida Gwai	Canada	fw	low	yes	HG
MAYR|X|2004#09	Mayer Lake	Haiida Gwai	Canada	fw	low	yes	HG
MAYR|X|2004#13	Mayer Lake	Haiida Gwai	Canada	fw	low	no	
MAYR|X|2004#14	Mayer Lake	Haiida Gwai	Canada	fw	low	no	
MAYR|X|2004#17	Mayer Lake	Haiida Gwai	Canada	fw	low	no	
MAYR|X|2004#21	Mayer Lake	Haiida Gwai	Canada	fw	low	no	
MAYR|X|2004#24	Mayer Lake	Haiida Gwai	Canada	fw	low	no	
MAYR|X|2004#26	Mayer Lake	Haiida Gwai	Canada	fw	low	no	
MAYR|X|2004#30	Mayer Lake	Haiida Gwai	Canada	fw	low	no	
MDPC_X_1993#01	MidPacific	Pacific Ocean	USA	mar	low	no	
MIDF|BDVW|2011#01	Midfjardara River	Iceland	Iceland	mar	high	no	
MIDF|BDVW|2011#02	Midfjardara River	Iceland	Iceland	mar	high	no	
MIDF|BLUP|2011#01	Midfjardara River	Iceland	Iceland	mar	high	no	
MIDF|REND|2011#01	Midfjardara River	Iceland	Iceland	fw	low	yes	ICE
MIDF|REND|2011#04	Midfjardara River	Iceland	Iceland	fw	low	yes	ICE
MIDF|REND|2011#05	Midfjardara River	Iceland	Iceland	fw	low	yes	ICE
MIDF|REND|2011#06	Midfjardara River	Iceland	Iceland	fw	low	yes	ICE
MIDF|REND|2011#10	Midfjardara River	Iceland	Iceland	fw	low	yes	ICE
MIDF|S101|2011#05	Midfjardara River	Iceland	Iceland	mar	high	no	
MIDF|S101|2011#06	River	Iceland	Iceland	mar	high	no	

MNKA|X|X#02	Matanuska	Alaska	USA	fw	low	yes	AK
MNYN|X|X#29997	Menyanthes	Haiida Gwai	Canada	fw	low	yes	HG
NHBR|X|2003#05	New Harbor River	East Coast	USA	mar	high	no	
NTRW|X|X#02	Norway Sea Trawler	North Sea	Norway	mar	high	no	
OBSE|X|2011#05	Ob nan Stearnain	Uist	Scotland	mar	high	no	
OLNY_X_2007#03	Olney Creek	Northern California	USA	fw	high	no	
ORPH|X|2011#05	Orphea Lake	Alaska	USA	fw	low	yes	AK
PINC|X|X#03	Pinhole Creek	BayArea California	USA	fw	unknown	no	
POQU|X|2009#29995	Poque Lake	Haiida Gwai	Canada	fw	low	yes	HG
QUIN_X_2003#02	Quinalt Lake	Washington State	USA	fw	partially	no	
RCAL|X|2001#24	River Callop	West Highlands	Scotland	fw	low	yes	SC
RDSP_X_2010#02	Road Side Pond	Haiida Gwai	Canada	fw	low	no	
RDSP|X|2012#70307	Road Side Pond	Haiida Gwai	Canada	fw	unknown	no	
RDSP|X|2012#70311	Road Side Pond	Haiida Gwai	Canada	fw	unknown	no	
RDSP|X|2012#70314	Road Side Pond	Haiida Gwai	Canada	fw	low	no	
RDSP|X|2012#70332	Road Side Pond	Haiida Gwai	Canada	fw	low	no	
RDSP|X|2012#70337	Road Side Pond	Haiida Gwai	Canada	fw	low	no	
RDSP|X|2012#70346	Road Side Pond	Haiida Gwai	Canada	fw	low	no	
RDSP|X|2012#70353	Road Side Pond	Haiida Gwai	Canada	fw	low	no	
RDSP|X|2012#70364	Road Side Pond	Haiida Gwai	Canada	fw	low	no	
RDSP|X|2012#70368	Road Side Pond	Haiida Gwai	Canada	fw	low	no	
RDSP|X|2012#70385	Pond	Haiida Gwai	Canada	fw	low	no	
REIV|X|2011#04	Loch na Reivil Rivery	Uist	Scotland	unknow n	unknown	no	
RGLN|X|2011#04	Glenavy	County Antrim	Ireland	fw	unknown	no	
ROUG|X|2000#f03	Rouge Lake	Haiida Gwai	Canada	fw	low	yes	HG
SACK|X|2010#0898	San Antonio Creek	Santa Barbara Co.	USA	fw	low	yes	
SALS|X|X#01	Salinas River	Central California	USA	fw	low	yes	CCAL
SAPC|X|X#01	San Pablo Creek	Central California	USA	fw	unknown	no	
SCRM|X|2010#873	Santa Clara River Mouth	Southern California	USA	fw	low	yes	SCAL
SCRS|X|2009#8253	Santa Clara River	Southern California	USA	fw	low	yes	SCAL

SDPY_X_2006#24	Serendipity Pond	Haiida Gwai	Canada	fw	low	yes	HG
SHEL_L_2008#01	Loch Shiel	West Highlands	Scotland	fw	low	yes	SC
SJCR|X|2009#8212	San Jacinto River	San Bernardino Co.	USA	fw	low	yes	
SKID_X_2009#01	Skidegate Lake	Haiida Gwai	Canada	fw	low	yes	HG
SKON|X|X#12	Skonun Lake	Haiida Gwai	Canada	fw	low	yes	HG
SLMW|X|2001#0918	Sugarloaf Meadow	Southern California	USA	fw	low	yes	SCAL
SLTC|X|X#90458	Solstice	Haiida Gwai	Canada	fw	low	yes	HG
SLVR|X|2009#29996	Silver Lake	Haiida Gwai	Canada	fw	low	yes	HG
SPNC_L_2009#01	Spence Lake	Haiida Gwai	Canada	fw	low	yes	HG
SPNC_L_2009#02	Spence Lake	Haiida Gwai	Canada	fw	low	yes	HG
SPNC_O_2010#08	Spence Lake Outlet	Haiida Gwai	Canada	fw	low	yes	HG
SPNC_O_2010#14	Spence Lake Outlet Creek	Haiida Gwai	Canada	fw	low	yes	HG
SRLY|X|2011#01	South Rolly	Alaska	USA	fw	low	yes	AK
SRST|S1|2000#05	Salmon River Steam	British Columbia	Canada	fw	low	yes	CA
SSMC|X|2010#01	San Simeon Creek	Central California	USA	fw	low	yes	CCAL
STIU_X_2009#03	Stiu Lake	Haiida Gwai	Canada	fw	low	no	
STMY|X|2011#03	Stormy Lake	Alaska	USA	fw	low	yes	AK
SUNC|X|X#04	Suisan Creek	BayArea California	USA	fw	unknown	no	
TERN|X|2011#02	Tern Lake	Alaska	USA	fw	low	yes	AK
TORM|X|2011#03	Loch Tormasad	Uist	Scotland	unknow n	low	yes	SC
TYNE_1_2001#07	River Tyne	East Lothian	Scotland	mar	high	no	
TYNE_1_2001#08	River Tyne	East Lothian	Scotland	mar	high	no	
TYNE_1_2001#09	River Tyne	East Lothian	Scotland	mar	high	no	
TYNE_1_2001#10	River Tyne	East Lothian	Scotland	mar	high	no	
TYNE_1_2001#14	River Tyne	East Lothian	Scotland	mar	high	no	
TYNE_8_2003#902	River Tyne	East Lothian	Scotland	fw	low	yes	SC
TYNE_8_2003#906	River Tyne	East Lothian	Scotland	fw	low	yes	SC
TYNE_8_2003#908	River Tyne	East Lothian	Scotland	fw	low	yes	SC
TYNE_8_2003#919	River Tyne	East Lothian	Scotland	fw	low	yes	SC
TYNE_8_2003#920	River Tyne	East Lothian	Scotland	fw	low	yes	SC
URRI|X|X#07	Urridavatn	Reykjavik	Iceland	fw	low	yes	ICE
VIFC|X|X#07	Vifissta_avatn Pelvic Complete	Reykjavik	Iceland	fw	low	yes	ICE
VIFR|X|X#08	Vifissta_avatn Pelvic Reduced	Reykjavik	Iceland	fw	low	yes	ICE

WALC|X|2005#05	Wallace Lake Pelvic Complete	Alaska	USA	fw	low	yes	AK
WALR|X|2005#32	Wallace Lake Pelvic Reduced	Alaska	USA	fw	low	yes	AK
WATT_X_2009#05	Watt Lake	Haiida Gwai	Canada	fw	low	yes	HG
WDPL_X_2009#05	Wood Pile Creek	Haiida Gwai	Canada	fw	low	yes	HG
WMSO_X_2002#bigf	Williamsonni	Southern California	USA	fw	low	yes	SCAL
WOLF|X|2011#04	Wolf	Alaska	USA	fw	low	yes	AK
ZERO|X|2011#05	Zero	Alaska	USA	fw	low	yes	AK

Table S10: Information for 207 genomes downloaded from the Sequence Read Archive (www.ncbi.nlm.nih.gov/sra) at accession PRJNA247503. The table includes sample names, drainage, region, country, salinity (fw means freshwater, mar means marine) and armour plate phenotype. Low-plated, freshwater individuals used for the PCA anslyis are indicated in the respective column and their short-cut used in the PCA plot (Figure 7) is included.

4. References

Alvarado, D. M., Yang, P., Druley, T. E., Lovett, M., & Gurnett, C. A. (2014). Multiplexed direct genomic selection (MDiGS): A pooled BAC capture approach for highly accurate CNV and SNP/INDEL detection. *Nucleic Acids Research*, *42*(10), e82–e82. https://doi.org/10.1093/nar/gku218

Archambeault, S. L., Bärtschi, L. R., Merminod, A. D., & Peichel, C. L. (2020). Adaptation via pleiotropy and linkage: Association mapping reveals a complex genetic architecture within the stickleback *Eda* locus. *Evolution Letters*, *4*(4), 282–301. https://doi.org/10.1002/evl3.175

Auwera, G. A. V. der, & O'Connor, B. D. (2020). *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra* (1st ed.). O'Reilly Media.

Barrett, R. D. H., & Schluter, D. (2008). Adaptation from standing genetic variation. *Trends in Ecology & Evolution*, *23*(1), 38–44. https://doi.org/10.1016/j.tree.2007.09.008

Bassham, S., Catchen, J., Lescak, E., von Hippel, F. A., & Cresko, W. A. (2018). Repeated Selection of Alternatively Adapted Haplotypes Creates Sweeping Genomic Remodeling in Stickleback. *Genetics*, genetics.300610.2017. https://doi.org/10.1534/genetics.117.300610

Bayes, T., & Price, null. (1763). LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S. *Philosophical Transactions of the Royal Society of London*, *53*, 370–418. https://doi.org/10.1098/rstl.1763.0053

Bell, M. A., Aguirre, W. E., & Buck, N. J. (2004). Twelve Years of Contemporary Armor Evolution in a Threespine Stickleback Population. *Evolution*, *58*(4), 814–824. https://doi.org/10.1111/j.0014-3820.2004.tb00414.x

Bell, M. A., & Foster, S. A. (1994). *The Evolutionary Biology of the Threespine Stickleback*. Oxford University Press.

Colosimo, P. F., Hosemann, K. E., Balabhadra, S., Villarreal, G., Dickson, M., Grimwood, J., Schmutz, J., Myers, R. M., Schluter, D., & Kingsley, D. M. (2005). Widespread Parallel Evolution in Sticklebacks by Repeated Fixation of Ectodysplasin Alleles. *Science*, *307*(5717), 1928–1933. https://doi.org/10.1126/science.1107239

Colosimo, P. F., Peichel, C. L., Nereng, K., Blackman, B. K., Shapiro, M. D., Schluter, D., & Kingsley, D. M. (2004). The Genetic Architecture of Parallel Armor Plate Reduction in Threespine Sticklebacks. *PLoS Biology*, *2*(5), e109. https://doi.org/10.1371/journal.pbio.0020109

Cresko, W. A., Amores, A., Wilson, C., Murphy, J., Currey, M., Phillips, P., Bell, M. A., Kimmel, C. B., & Postlethwait, J. H. (2004). Parallel genetic basis for repeated evolution of armor loss in Alaskan threespine stickleback populations. *Proceedings of the National Academy of Sciences*, *101*(16), 6050–6055. https://doi.org/10.1073/pnas.0308479101

Dan Carr, Nicholas Lewin-Koh, Martin Maechler, & Deepayan Sarkar. (2021). *Hexbin : Hexagonal Binning Routines. R package version 1.28.2.* https://CRAN.R-project.org/package=hexbin

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2), giab008. https://doi.org/10.1093/gigascience/giab008

DeAngelis, M. M., Wang, D. G., & Hawkins, T. L. (1995). Solid-phase reversible immobilization for the isolation of PCR products. *Nucleic Acids Research*, *23*(22), 4742–4743. https://doi.org/10.1093/nar/23.22.4742

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., & Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, *43*(5), 491–498. https://doi.org/10.1038/ng.806

Edge, P., Bafna, V., & Bansal, V. (2017). HapCUT2: Robust and accurate haplotype assembly for diverse sequencing technologies. *Genome Research*, *27*(5), 801–812. https://doi.org/10.1101/gr.213462.116

Eggertsson, H. P., Jonsson, H., Kristmundsdottir, S., Hjartarson, E., Kehr, B., Masson, G., Zink, F., Hjorleifsson, K. E., Jonasdottir, A., Jonasdottir, A., Jonsdottir, I., Gudbjartsson, D. F., Melsted, P., Stefansson, K., & Halldorsson, B. V. (2017). Graphtyper enables population-scale genotyping using pangenome graphs. *Nature Genetics*, *49*(11), Article 11. https://doi.org/10.1038/ng.3964

Fang, B., Kemppainen, P., Momigliano, P., Feng, X., & Merilä, J. (2020). On the causes of geographically heterogeneous parallel evolution in sticklebacks. *Nature Ecology & Evolution*. https://doi.org/10.1038/s41559-020-1222-6

Gaut, B. S., & Long, A. D. (2003). The Lowdown on Linkage Disequilibrium. *The Plant Cell*, *15*(7), 1502–1506. https://doi.org/10.1105/tpc.150730

967 Hagen, D. W. (1967). Isolating Mechanisms in Threespine Sticklebacks (Gasterosteus).
968 *Journal of the Fisheries Research Board of Canada*, *24*(8), 1637–1692.
969 https://doi.org/10.1139/f67-138
970 Hedtke, S. M., Morgan, M. J., Cannatella, D. C., & Hillis, D. M. (2013). Targeted Enrichment:
971 Maximizing Orthologous Gene Comparisons across Deep Evolutionary Time. *PLOS*
972 *ONE*, *8*(7), e67908. https://doi.org/10.1371/journal.pone.0067908
973 Hermisson, J., & Pennings, P. S. (2005). Soft Sweeps: Molecular population genetics of
974 adaptation from standing genetic variation. *Genetics*, *169*(4), 2335–2352.
975 https://doi.org/10.1534/genetics.104.036947
976 Hohenlohe, P. A., Bassham, S., Currey, M., & Cresko, W. A. (2012). Extensive linkage
977 disequilibrium and parallel adaptive divergence across threespine stickleback
978 genomes. *Philosophical Transactions of the Royal Society B: Biological Sciences*,
979 *367*(1587), 395–408. https://doi.org/10.1098/rstb.2011.0245
980 Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., & Cresko, W. A. (2010).
981 Population Genomics of Parallel Adaptation in Threespine Stickleback using
982 Sequenced RAD Tags. *PLoS Genetics*, *6*(2), e1000862.
983 https://doi.org/10.1371/journal.pgen.1000862
984 Jones, F. C., Brown, C., Pemberton, J. M., & Braithwaite, V. A. (2006). Reproductive isolation
985 in a threespine stickleback hybrid zone. *Journal of Evolutionary Biology*, *19*(5), 1531–
986 1544. https://doi.org/10.1111/j.1420-9101.2006.01122.x
987 Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., Swofford, R.,
988 Pirun, M., Zody, M. C., White, S., Birney, E., Searle, S., Schmutz, J., Grimwood, J.,
989 Dickson, M. C., Myers, R. M., Miller, C. T., Summers, B. R., Knecht, A. K., … Kingsley,
990 D. M. (2012). The genomic basis of adaptive evolution in threespine sticklebacks.
991 *Nature*, *484*(7392), 55–61. https://doi.org/10.1038/nature10944
992 Kingman, G. A. R., Vyas, D. N., Jones, F. C., Brady, S. D., Chen, H. I., Reid, K., Milhaven, M.,
993 Bertino, T. S., Aguirre, W. E., Heins, D. C., Hippel, F. A. von, Park, P. J., Kirch, M.,
994 Absher, D. M., Myers, R. M., Palma, F. D., Bell, M. A., Kingsley, D. M., & Veeramah,
995 K. R. (2021). Predicting future from past: The genomic basis of recurrent and rapid
996 stickleback evolution. *Science Advances*, *7*(25), eabg5285.
997 https://doi.org/10.1126/sciadv.abg5285
998 Kingsley, D. M., & Peichel, C. L. (2007). The molecular genetics of evolutionary change in
999 sticklebacks. In *Östlund-NilssonS, MayerI, HuntingfordF, editors. Biology of the three-*
1000 *spined sticklebacks* (pp. 44–81). Boca Raton (FL): CRC Press.
1001

Kingsley, D. M., Zhu, B., Osoegawa, K., De Jong, P. J., Schein, J., Marra, M., Peichel, C., Amemiya, C., Schluter, D., Balabhadra, S., Friedlander, B., Cha, Y. M., Dickson, M., Grimwood, J., Schmutz, J., Talbot, W. S., & Myers, R. (2004). New Genomic Tools for Molecular Studies of Evolutionary Change in Threespine Sticklebacks. *Behaviour*, *141*(11/12), 1331–1344.

Kirch, M., Romundset, A., Gilbert, M. T. P., Jones, F. C., & Foote, A. D. (2021). Ancient and modern stickleback genomes reveal the demographic constraints on adaptation. *Current Biology*, *31*(9), 2027-2036.e8. https://doi.org/10.1016/j.cub.2021.02.027

Kolde, R. (2019). *Pheatmap: Pretty Heatmaps. R package version 1.0.12.* https://CRAN.R-project.org/package=pheatmap

Konečná, V., Bray, S., Vlček, J., Bohutínská, M., Požárová, D., Choudhury, R. R., Bollmann-Giolai, A., Flis, P., Salt, D. E., Parisod, C., Yant, L., & Kolář, F. (2021). Parallel adaptation in autopolyploid Arabidopsis arenosa is dominated by repeated recruitment of shared alleles. *Nature Communications*, *12*(1), 4979. https://doi.org/10.1038/s41467-021-25256-5

Lescak, E. A., Bassham, S. L., Catchen, J., Gelmond, O., Sherbick, M. L., von Hippel, F. A., & Cresko, W. A. (2015). Evolution of stickleback in 50 years on earthquake-uplifted islands. *Proceedings of the National Academy of Sciences*, *112*(52), E7204–E7212. https://doi.org/10.1073/pnas.1512020112

Lewontin, R. C. (1964). The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models. *Genetics*, *49*(1), 49–67. https://doi.org/10.1093/genetics/49.1.49

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), Article 1. https://doi.org/10.14806/ej.17.1.200

Maynard-Smith, J., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetics Research*, *23*(1), 23–35. https://doi.org/10.1017/S0016672300014634

McGee, M. D., Borstein, S. R., Meier, J. I., Marques, D. A., Mwaiko, S., Taabu, A., Kishe, M. A., O'Meara, B., Bruggmann, R., Excoffier, L., & Seehausen, O. (2020). The ecological and genomic basis of explosive adaptive radiation. *Nature*, *586*(7827), 75–79. https://doi.org/10.1038/s41586-020-2652-7

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella,

K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

McPhail, J. D. (1994). Speciation and the evolution of reproductive isolation in the sticklebacks (Gasterosteus) of southwestern British Columbia. In *The evolutionary biology of the threespine stickleback. Edited by M.A. Bell and S.A. Foster.* (pp. 399–437). Oxford University Press.

Meier, J. I., Salazar, P. A., Kučka, M., Davies, R. W., Dréau, A., Aldás, I., Box Power, O., Nadeau, N. J., Bridle, J. R., Rolian, C., Barton, N. H., McMillan, W. O., Jiggins, C. D., & Chan, Y. F. (2021). Haplotype tagging reveals parallel formation of hybrid races in two butterfly species. *Proceedings of the National Academy of Sciences*, *118*(25), e2015005118. https://doi.org/10.1073/pnas.2015005118

Miller, C. T., Beleza, S., Pollen, A. A., Schluter, D., Kittles, R. A., Shriver, M. D., & Kingsley, D. M. (2007). Cis-Regulatory Changes in Kit Ligand Expression and Parallel Evolution of Pigmentation in Sticklebacks and Humans. *Cell*, *131*(6), 1179–1189. https://doi.org/10.1016/j.cell.2007.10.055

Moyle, P. B. (1976). Fish introductions in California: History and impact on native fishes. *Biological Conservation*, *9*(2), 101–118. https://doi.org/10.1016/0006-3207(76)90043-4

Neuwirth, E. (2022). *RColorBrewer: ColorBrewer Palettes. R package version 1.1-3.* https://CRAN.R-project.org/package=RColorBrewer

Picelli, S., Björklund, Å. K., Reinius, B., Sagasser, S., Winberg, G., & Sandberg, R. (2014). Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Research*, *24*(12), 2033–2040. https://doi.org/10.1101/gr.177881.114

Poplin, R., Ruano-Rubio, V., DePristo, M. A., Fennell, T. J., Carneiro, M. O., Van der Auwera, G. A., Kling, D. E., Gauthier, L. D., Levy-Moonshine, A., Roazen, D., Shakir, K., Thibault, J., Chandran, S., Whelan, C., Lek, M., Gabriel, S., Daly, M. J., Neale, B., MacArthur, D. G., & Banks, E. (2017). *Scaling accurate genetic variant discovery to tens of thousands of samples* [Preprint]. Genomics. https://doi.org/10.1101/201178

R Core Team. (2020). *A language and environment for statistical computing. R Foundation for Statistical Computing*. https://www.R-project.org/

Reid, K., Bell, M. A., & Veeramah, K. R. (2021). Threespine Stickleback: A Model System For Evolutionary Genomics. *Annual Review of Genomics and Human Genetics*, *22*, 357383. https://doi.org/10.1146/annurev-genom-111720-081402

Salces-Castellano, A., Stankowski, S., Arribas, P., Patiño, J., Karger, D. N., Butlin, R., & Emerson, B. C. (2021). Long-term cloud forest response to climate warming revealed by insect speciation history*. *Evolution*, *75*(2), 231–244. https://doi.org/10.1111/evo.14111

Schluter, D., & Conte, G. L. (2009). Genetics and ecological speciation. *Proceedings of the National Academy of Sciences*, *106*(Supplement 1), 9955–9962. https://doi.org/10.1073/pnas.0901264106

Schluter, D., Marchinko, K. B., Arnegard, M. E., Zhang, H., Brady, S. D., Jones, F. C., Bell, M. A., & Kingsley, D. M. (2021). Fitness maps to a large-effect locus in introduced stickleback populations. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(3), e1914889118. https://doi.org/10.1073/pnas.1914889118

Slatkin, M. (2008). Linkage disequilibrium—Understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics*, *9*(6), Article 6. https://doi.org/10.1038/nrg2361

SRA Toolkit Development Team. (2022). *SRA Toolkit*. https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=software

Terekhanova, N. V., Logacheva, M. D., Penin, A. A., Neretina, T. V., Barmintseva, A. E., Bazykin, G. A., Kondrashov, A. S., & Mugue, N. S. (2014). Fast Evolution from Precast Bricks: Genomics of Young Freshwater Populations of Threespine Stickleback Gasterosteus aculeatus. *PLOS Genetics*, *10*(10), e1004696. https://doi.org/10.1371/journal.pgen.1004696

Verta, J.-P., & Jones, F. C. (2019). Predominance of cis-regulatory changes in parallel expression divergence of sticklebacks. *ELife*, *8*, e43785. https://doi.org/10.7554/eLife.43785

Visser, I., & Speekenbrink, M. (2010). depmixS4: An R Package for Hidden Markov Models. *Journal of Statistical Software*, *36*, 1–21. https://doi.org/10.18637/jss.v036.i07

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York.

Wooldridge, T. B., Kautt, A. F., Lassance, J.-M., McFadden, S., Domingues, V. S., Mallarino, R., & Hoekstra, H. E. (2022). An enhancer of *Agouti* contributes to parallel evolution of cryptically colored beach mice. *Proceedings of the National Academy of Sciences*, *119*(27), e2202862119. https://doi.org/10.1073/pnas.2202862119

Wootton, R. J. (1984). *A Functional Biology of Sticklebacks*. University of California Press.