Shared patterns of genome-wide differentiation are more strongly predicted by geography than by ecology

Diana J. Rennison[1,6,7]*, Kira E. Delmore[2]*, Kieran Samuk[3], Gregory L. Owens[4], Sara E. Miller[5]

1) Institute of Ecology and Evolution, University of Bern, Bern, Switzerland.
2) Department of Biology, Texas A&M University, College Station, Texas, United States of America.
3) Department of Biology, Duke University, Durham, North Carolina, United States of America.
4) Department of Integrative Biology, University of California Berkeley, Berkeley, California, United States of America.
5) Department of Neurobiology and Behavior, Cornell University, Ithaca, New York, United States of America.
6) Corresponding author email: dianajrenn@gmail.com
7) Present Address: Division of Biological Sciences, University of California, California, United States of America.

* Served as lead authors for this project.

Keywords: Parallel Evolution, Adaptive Evolution, Genomics, Natural Selection, Threespine stickleback, *Gasterosteus aculeatus*

ABSTRACT

Closely related populations often display similar patterns of genomic differentiation, yet it remains an open question which ecological and evolutionary forces generate these patterns. The leading hypothesis is that this similarity in divergence is driven by parallel natural selection. However, several recent studies have suggested that these patterns may instead be a product of the depletion of genetic variation that occurs as result of background selection (i.e. linked negative selection). To date, there have been few direct tests of these competing hypotheses. To determine the relative contributions of background selection and parallel selection to patterns of repeated differentiation, we examined 24 independently derived populations of freshwater stickleback occupying a variety of niches and estimated genomic patterns of differentiation in each relative to their common marine ancestor. Patterns of genetic differentiation were strongly correlated across pairs of freshwater populations adapting to the same ecological niche, supporting a role for parallel natural selection. In contrast to other recent work, our study comparing populations adapting to the same niche produced no evidence signifying that similar patterns of genomic differentiation are generated by background selection. We also found that overall patterns of genetic differentiation were considerably more similar for populations found in closer geographic proximity. In fact, the effect of geography on the repeatability of differentiation was greater than that of parallel selection. Our results suggest that shared selective landscapes and ancestral variation are the key drivers of repeated patterns of differentiation in systems that have recently colonized novel environments.

Introduction

The evolution of the same phenotypic traits in independent populations inhabiting similar environments, known as parallel or convergent evolution, is generally accepted as evidence for the action of parallel natural selection (Endler 1986; Schluter 2000; Losos 2011; Bonick et al., 2018) because chance processes are unlikely to yield repeated phenotypic evolution (Schluter and Nagel 1995). Using a similar logic, many recent studies have used genomic scans to search for evidence of parallel genetic differentiation among closely related species or populations adapting to similar environments (*e.g.* Fraser et al. 2015; Reid et al. 2016; Ravinet et al. 2016; Rougemont et al. 2017; Trucchi et al. 2017). In these studies, shared outliers and/or similar patterns of genetic differentiation (F_{ST} or D_{XY}) across the genome have been taken as evidence of parallel adaptation to local ecological conditions.

Similar differentiation landscapes across the genome have also been found to evolve in the absence of ecological or phenotypic parallelism (*e.g.* Burri et al. 2015; Martin et al. 2013; Renaut et al. 2014; Vijay et al. 2016). This has led to the argument that perhaps parallel natural selection alone does not drive repeatable genomic differentiation (*e.g.* Bank et al. 2014; Cruickshank and Hahn 2014; Haasl and Payseur 2016; Burri 2017). Rather, shared patterns of genomic differentiation could be generated by long-term linked selection in a heterogeneous recombination landscape that is shared among taxa due to synteny (Bank et al. 2014; Cruickshank and Hahn 2014; Haasl and Payseur 2016; Burri 2017). Linked selection occurs when a mutation at one locus affects the allele frequencies of loci in linkage disequilibrium (reviewed by Barton, 2000). Linked selection is referred to as genetic hitchhiking when selection on the focal locus is positive, and background selection when selection on the focal locus is negative (Charlesworth et al. 1993). It has been argued that background selection would more

often lead to similar patterns of differentiation than genetic hitchhiking because the opportunity for positive selection to affect the same genomic region multiple times may be limited (Burri 2017). In support of this argument, theoretical work suggests that genomic parallelism may only be seen when the selection landscape is highly parallel (Thompson et al., 2019). However, computer models suggest the effects of background selection on the divergence landscape may be modest (Zeng and Charlesworth 2011; Matthey-Doret and Whitlock 2018; Zeng and Corcoran 2018). Additionally, background selection has failed to explain some empirical patterns (*e.g.* Irwin et al. 2016; Irwin et al. 2018) and it may be the case that there hasn't been sufficient time for drift and/or negative selection to influence differentiation in recently diverged species (Burri 2017; Delmore et al. 2018). Given these differing theoretical and empirical results, we used a comparative approach to disentangle the contributions of background selection and parallel positive selection to the repeatability of genomic differentiation in a recently diverged species.

We further determined whether the source of genetic variation can influence the likelihood of observing shared genomic divergence (MacPherson and Nuismer 2017; Thompson et al., 2019). Shared standing genetic variation and/or introgression between populations experiencing parallel natural selection may more often facilitate the evolution of similar patterns of genomic differentiation compared to *de novo* mutation (MacPherson and Nuismer 2017), due to the fixation advantages conferred by higher initial allele frequencies (Innan and Kim 2004). To evaluate the role of shared standing genetic variation, we tested whether closer geographic proximity predicted an increased similarity of genomic differentiation, as nearby populations likely shared more initial standing genetic variation.

The recent diversification of threespine stickleback (*Gasterosteus aculeatus*) across the globe provides an excellent system to test the hypothesis that parallel selection is the dominant

force determining the degree of shared genomic differentiation. At the end of the last ice age (10 000 – 12 000 years ago) marine stickleback colonized newly formed freshwater habitats (Bell and Foster 1994). Phenotypically, these freshwater populations are more similar to one another than to their marine ancestors (Bell and Foster 1994) despite being independently derived. Among freshwater populations, there has been further parallel phenotypic differentiation, with benthic, limnetic, stream, and solitary lake ecotypes arising multiple times independently (Bell and Foster 1994; Taylor and McPhail 2000). These freshwater populations occur in both the Atlantic and Pacific Ocean basins, and span distances exceeding 20 000 km, with colonization occurring at similar times across the basins (Bell and Foster, 1994).

We leveraged existing genomic data from independent populations of stickleback to assess the contributions of parallel positive selection, background selection and shared standing genetic variation to patterns of genome-wide differentiation. First, we tested the prediction that population pairs with more similar ecology would exhibit a more similar pattern of genomic differentiation due to parallel positive selection. Second, we tested the prediction that in the absence of background selection, population pairs occupying the same niche, and thus diverging neutrally, would have largely dissimilar patterns of differentiation across the genome. Lastly, we tested the prediction that geographically proximate populations that have evolved from more genetically similar marine founders would exhibit more similar patterns of genomic differentiation, due to starting with more shared standing genetic variation.

Methods

Data Acquisition

We utilized a subset of the short-read dataset for threespine stickleback compiled by Samuk *et al.* (2017). The dataset consisted of individuals from 24 independent freshwater populations. These populations included solitary populations adapted to lakes (11) or streams (7), and sympatric benthic (3) and limnetic (3) species pairs (Figure 1a). As the Eastern Pacific marine population is generally considered to be panmictic, the marine reference population was an amalgamation of whole genome sequences from 9 Pacific Ocean locations collected along the West coast of North America. Additional population details can be found in Table S1.

Data preparation and variant calling

We focused on single nucleotide polymorphisms (SNPs) in our analysis, which we identified using a standard, reference-based bioinformatics pipeline comprised of custom R 3.2.2 (R Development Core Team, 2016) and Perl scripts (all scripts are available in the following GitHub repository https://github.com/ksamuk/gene_flow_linkage, see also Samuk et al., 2017). Briefly, we de-multiplexed the reads and used Trimmomatic 0.32 (Bolger et al., 2014) to filter out low quality sequences and adapter contamination. We then aligned reads to version the stickleback reference genome (Jones et al. 2012) using BWA 0.7.10 (Li et al., 2009), followed by realignment with STAMPY 1.0.23 (Lunter and Goodson 2011). The GATK 3.3.0 (McKenna et al., 2010) best practices workflow (DePristo et al. 2011) was followed except for the "MarkDuplicates" step which we skipped when reads were derived from reduced representation libraries (RAD and GBS). We realigned reads around indels ("RealignTargetCreator", "IndelRealigner"), identified SNPs in individuals using the "HaplotypeCaller", and jointly

genotyped the entire dataset using "GenotypeGVCFs". The results were then output as a VCF containing all genotyped sites (variant and invariant) and converted to tabular format for downstream analyses. For further details on the pipeline please see the scripts referenced above.

Genomic differentiation and quantification of repeatability

We estimated average genetic differentiation (F_{ST}) between the marine reference population and each freshwater population for SNPs within 150 kb windows across the genome. A windowed approach was used to facilitate the comparison of F_{ST} among populations sequenced using different technologies. We also estimated genetic differentiation between pairs of lake populations using Weir & Cockerham's (1984) F_{ST}. We calculated these estimates by dividing the sum of the numerators of all SNP-wise F_{ST} estimates within the window by the sum of their denominators. In order to estimate average F_{ST} accurately, we dropped windows if they contained fewer than 3 SNPs. To estimate repeatability, we correlated F_{ST} values of all windows across the genome between population pairs using Pearson's correlation coefficients. Significance testing of individual correlation coefficients was done using the *Hmisc* package in R and correction for multiple testing was done using the BH method (Benjamini and Hochberg, 1995) with the *p.adjust* function. D_{xy} (Nei and Li, 1979) was also estimated for marine – freshwater comparisons in 150 kb windows, and these windows were used to estimated pairwise Pearson's correlation coefficients. The results based on the correlation coefficients of marine-freshwater comparisons using D_{xy} were qualitatively the same as those for F_{ST} and are reported in the supplementary materials (see Supplementary Figure 1). In recently diverged populations, segregating ancestral variation is an important determinant of D_{xy}. Correspondingly, shared ancestral variation will result in over-estimation of correlations in D_{xy} for population pairs in the

absence of divergent selection. Because we expected that the freshwater-freshwater comparisons used for the F_{ST} analysis are evolving neutrally, they were not a useful control for estimating the expected correlations in D_{xy}.

Correlations of genome-wide differentiation among marine -freshwater population pairs and neutrally evolving lake population pairs

The genome-wide pairwise Pearson's correlation coefficients of F_{ST} values were used to compare the effects of parallel selection (*i.e.* one Marine-Freshwater population pair compared to another Marine-Freshwater population pair) relative to the neutral expectation (*i.e.* one pair of freshwater lakes compared to another independent pair of freshwater lakes). See Figure 2 for a schematic of these two types of comparisons. To test for the effect of the type of selection (positive or background), we used linear models with average pairwise correlations of genome-wide differentiation as the response variable and divergent selection as the predictor (divergent or non-divergent). Significance testing was accomplished by resampling divergent selection categorizations 10,000 times and recalculating the mean correlation to form a null distribution to estimate the significance. For simplicity, this analysis was restricted to populations from the Pacific basin. We also exclude pseudo-replicated population comparisons, where the same lake population was included in both pairs (*e.g.* pair 1 = Boot lake and Roberts lake, pair 2 = Misty lake and Roberts lake would be excluded). However, we report a version of the analysis including these pairs in the supplementary materials to show that the pattern holds regardless of pruning.

Contribution of parallel natural selection and geographic proximity to genome-wide repeatability in differentiation

To test the effects of niche similarity and geography on genome-wide repeatability, we again used pairwise Pearson's correlation coefficients of windowed estimates of F_{ST}. We used multiple regression analyses implemented in the 'ecodist' package in R (999 permutations, MRM function) for this analysis, evaluating the contribution of distance matrices quantifying ecology and geography. First, we quantified ecology and geography using binomial variables, as the "same freshwater niche" or "different freshwater niche" for ecology and the "same ocean basin" (Pacific or Atlantic Ocean) or a "different ocean basin" for geography. Second, we quantified ecology and geography using continuous or ordinal variables. Previous work has shown that the ecology and diet of stream populations are more similar to benthic populations, while the solitary lake populations tend to be more similar to limnetic populations (Berner et al., 2008 & 2009). We gave populations occupying the same niche (*e.g.* benthic & benthic or stream & stream) a score of 3 (the maximum), populations occupying similar niches a score of 2 (*e.g.* benthic & stream or limnetic & lake), and populations occupying the most dissimilar niches a score of 1 (*e.g.* limnetic & stream or benthic & lake). We quantified geography as a continuous estimate of pairwise distance within the same ocean basin, determined by computing the Euclidean distance (square-root transformed for normality). Note, some population pairs included in these analyses come from the same watershed (*e.g.* stream and lake populations from Boot lake). Accordingly, we randomly sampled one population from each of these pairs to run our analyses. We repeated this down-sampling 512 times, which is all possible combinations of our pairs.

Results

Repeatability of genomic differentiation among marine-freshwater population pairs

There was considerable variation among marine-freshwater population pairs in the magnitude of genomic differentiation (see Figure S2 for a PCA of the populations); mean genome-wide F_{ST} ranged from 0.25 to 0.71 (mean F_{ST} = 0.47). There was also variation across the genome; genome-wide variance in F_{ST} ranged from 0.03 to 0.06 among population pairs (Figure 1b). Despite this variation, correlation coefficients (r) comparing windowed estimates of F_{ST} across population pairs ranged from 0.06 to 0.84 (mean r) = 0.38) and were significantly positive for all population pairs after correction for multiple testing ($P < 0.05$). Thus, the locations of genetic differentiation between marine and freshwater populations are often the same between independently derived population pairs.

Contribution of background selection to repeated genomic differentiation

There was also considerable variation in the magnitude of genomic differentiation between the independent freshwater lake populations used as our neutral reference populations. Values of F_{ST} spanned a range from 0.12 to 0.73 (mean F_{ST} = 0.48). However, F_{ST} was more similar between marine – freshwater population pairs (mean r) = 0.49) than between the lake – lake population pairs (mean r = 0.07). The difference in the average magnitude of correlation between the two types of comparisons was significant using a permutation test (difference in mean r = 0.42, $P < 0.0001$) (Figure 3A). Thus, background selection does not account for the full extent of genomic repeatability.

Contribution of parallel natural selection and geographic proximity to repeated genomic differentiation

Both ecology and geography explain a significant proportion of the variation in correlation coefficients comparing windowed estimates of F_{ST} across marine-freshwater pairs. For example, an MRM including both ecology and geography (quantified as "same" or "different" ecology and geography, respectively) explained 56% of the variation in F_{ST} correlation coefficients (R^2 = 0.56, P = 0.0001). There was a negative relationship between both matrices and F_{ST} correlation coefficients suggesting parallel natural selection is stronger when ecology is more similar (r = -0.07, P = 0.06) and populations are closer (r = -0.27, P = 0.0001 [averages over repeated down-sampling; ecology significant predictor in 57% of samples, geography in 100%]). The regression coefficient for geography was much higher than that for ecology, suggesting geography explains a larger portion of the variation. We obtained similar results when quantifying ecology (Figure S3) and geography (Figure S4) using continuous variables (MRM, geography quantified using Euclidean distance, R^2 = 0.54, P =0.0003, r and P ecology = -4.7x10-2 and 0.08, r and P geography = -3.7x10-5 and 0.0003 [ecology significant in 45% of down-samples and geography 100%]). Ecology and geography are not on the same scale in this analysis, preventing a direct comparison of correlation coefficients. However, when we reran this analysis excluding ecology, the model R^2 was 0.49 (vs. 0.54 when ecology was included), suggesting geography explains a larger portion of the variation in F_{ST} repeatability. When controlling for divergence time from the marine reference by limiting pairwise comparisons to only those within the same ocean basin, we still find that both more similar ecology and closer geographic proximity result in higher F_{ST} correlation coefficients (Figures S3 and S4).

Discussion

Repeatability of genome-wide differentiation is not due to background selection

In contrast to recent work in other taxa, we do not find strong evidence that similarity in genome-wide patterns of differentiation has been driven by background selection, as pairs of populations evolving in the absence of divergent selection show little similarity in the genomic locations of genetic differentiation. This finding was not an artefact of different magnitudes of divergence between population comparisons with and without parallel divergent selection, as average F_{ST} was essentially the same in marine – lake and lake – lake comparisons. This pattern also suggests that drift is unlikely to be a key player in generating strong correlations in genomic differentiation. Given our results, we argue that genome-wide correlations in differentiation are truly reflective of shared positive selection and likely parallel genetic evolution in some cases. Our results are in line with recent work suggesting the effects of background selection on differentiation are modest (Zeng and Charlesworth 2011; Matthey-Doret and Whitlock 2018; Zeng and Corcoran 2018) and cannot explain the extreme patterns of differentiation documented in empirical studies (Irwin et al. 2016, 2018).

The natural history of this species is consistent with our findings. Since all freshwater populations examined are postglacial, and therefore less than 12 000 years old, it is unlikely that there has been sufficient novel mutation in all populations to generate large-scale parallel divergence due to linked background selection alone. The idea that there has been insufficient novel mutations in this system is supported by the key role of standing genetic variation in adaptation to freshwater (*e.g.* Jones et al., 2012). Other species where there has been rapid adaptation aided by standing genetic variation (*e.g.* African cichlids), may also have only weakly

correlated patterns of genetic differentiation in the absence of parallel selection. Differences in the stage of speciation may explain the conflicting results observed across various taxa. Specifically, the populations of sticklebacks compared here are in the very early stages of speciation, and it was recently suggested that background selection will have a greater effect on genomic differentiation later in the speciation process. This suggestion derives from the fact that drift and negative selection will take time to influence differentiation (Burri 2017; Delmore et al. 2018). Regardless of the age of the study system, control comparisons, such as those we employ here with non-divergent lake-lake comparisons (and see Vijay et al., 2016), will provide an important reference point for researchers interested in measuring the contribution of parallel selection to the generation of genomic parallelism.

Parallel selection and geographic proximity both contribute to repeated genomic differentiation

Populations of stickleback with more similar ecological niches and in closer geographic proximity were found to have more similar patterns of genomic differentiation. This finding suggests that repeatable genome-wide patterns of genetic differentiation are indeed predicted to some degree by parallel natural selection. Interestingly, we find that geographic proximity is a much better predictor of repeated genome-wide differentiation (*e.g.*, correlation coefficient of -0.27 in quantitative MRM vs. -0.07 for ecology). This finding is consistent with mathematical modelling, which has suggested an important role for the geographic structure of populations in determining the probability of genetic convergence (Ralph and Coop 2015). The finding that ecological similarity explains less variation than geographic proximity is also consistent with previous empirical work suggesting the importance of factors other than ecology. For example, Renaut *et al*. (2014) examined genomic repeatability between three sister pairs of sunflowers and

found that while F_{ST} correlation coefficients were highest for pairs that diverged along same selective gradient (latitude), this factor explained a relatively small fraction of the variation (4%). Previous work in *Littorina* has also found limited support for ecological similarity explaining broad patterns of genomic repeatability (Ravinet et al. 2016). However, it is important to consider that our metric of ecological similarity is based only on niche type and therefore very coarse. Multidimensional variation in the selection landscape would perhaps explain more variation in the magnitude of repeatability exhibited among population pairs.

Similarity of abiotic agents of selection, gene flow and initial pools of genetic variation could drive the substantial contribution of geographic proximity to repeated genetic variation. Geographically proximate populations likely experience more similar abiotic selective pressures, for example temperature or mineral availability, which would conceivably lead to greater repeatability in genome-wide differentiation. In some watersheds there is also ongoing gene flow between marine and freshwater fish (*e.g.* Jones et al., 2012; Vines et al., 2016), which could lead to the sharing of adaptive alleles between geographically proximate freshwater populations. However, there is currently no evidence of direct gene flow between watersheds, as freshwater fish are often landlocked and are unlikely to survive migration through high salinity ocean waters.

More similar pools of initial variation among geographically proximate marine colonizers may also promote increased repeatability, given the importance of standing genetic variation in the marine ancestors for adaptation to freshwater in sticklebacks (Colosimo et al. 2005; Schluter and Conte 2009; Jones et al. 2012). Adaptation from standing genetic variation provides a reduced waiting time for fixation of beneficial alleles, relative to novel mutation, because beneficial alleles are immediately available and start at higher initial frequencies (Innan

and Kim 2004). Theoretical work has also shown that the probability of fixation is higher for alleles drawn from standing variation (Orr and Betancourt 2001). During rapid adaptation, the fixation advantages conferred by standing genetic variation may lead to biases in the frequency of gene use over the course of evolution, where loci with standing genetic variation contribute to adaptation more frequently than those which require variation to be generated through novel mutation. Thus, when comparing the location of differentiation among independently derived population pairs, there may be greater repeatability when there was a more similar initial pool of variation.

Theoretical work also predicts that when comparing across populations founded by ancestors with similar pools of standing genetic variation, loci with standing genetic variation will exhibit more similar patterns of evolution (MacPherson and Nuismer 2017). Laboratory experiments conducted under parallel selective regimes are also consistent with this prediction, as adaptation from standing genetic variation has been shown to lead to greater genetic parallelism than novel mutation (Teotónio et al., 2009). Thus, shared standing variation likely plays a key role in generating the considerable levels of repeatability seen in the locations of differentiation among freshwater stickleback populations. More broadly, our findings may suggest that parallel selection alone is unlikely to generate strong patterns of genomic parallelism and other genetic factors, such as the source of variation, that bias evolutionary trajectories may be important determinates of patterns of parallel evolution.

Acknowledgements

We are very grateful for the threespine stickleback research community, whose body of empirical work made this study possible. We also thank Catherine Peichel for useful comments on the manuscript. KMS, GLO, DJR, and KED were supported by Natural Sciences and Engineering Research Council graduate doctoral scholarships. All authors were supported by graduate fellowships from the University of British Columbia. DJR was subsequently supported by a Marie Skłodowska-Curie IF Fellowship.

Data Accessibility

Published genomic datasets: The original study references and accession numbers are listed in Table S1. Referenced Bioinformatic analysis code: https://github.com/ksamuk/gene_flow_linkage. Additional input files and R scripts are archived in the Dryad digital repository: 10.5061/dryad.34q91f0 (Rennison et al. 2020).

Literature Cited

Bank, C., G. B. Ewing, A. Ferrer-Admettla, M. Foll, and J. D. Jensen. 2014. Thinking too positive? Revisiting current methods of population genetic selection inference. Trends in Genetics 30:540–546.

Barton, N. H. 2000. Genetic hitchhiking. Philosophical Transactions of the Royal Society B 355:1553–1562.

Bell, M. A., and S. A. Foster. 1994. The evolutionary biology of the threespine stickleback. Oxford University Press, Oxford.

Benjamini Y, and Y. Hochberg. 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society 1:289–300.

Berner, D., D. C. Adams, A. C. Grandchamp, and A.P. Hendry, A.P. 2008. Natural selection drives patterns of lake–stream divergence in stickleback foraging morphology. Journal of Evolutionary Biology 21:1653–1665.

Berner, D., A. C. Grandchamp, and A. P. Hendry. 2009. Variable progress toward ecological speciation in parapatry: stickleback across eight lake–stream transitions. Evolution 63: 1740–1753.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.

Bolnick, D. I., R. D. H. Barrett, K. Oke, D. J. Rennison, and Y. E. Stuart. 2018. (Non)Parallel evolution. Annual Reviews of Ecology, Evolution & Systematics. 49: 303–330.

Burri, R. 2017. Interpreting differentiation landscapes in the light of long-term linked selection. Evolution Letters 1:118–131.

Burri, R., A. Nater, T. Kawakami, C. F. Mugal, P. I. Olason, L. Smeds, A. Suh, et al. 2015. Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. Genome Research 25:1656–1665.

Charlesworth, B., M. T. Morgan, and D. Charlesworth. 1993. The effect of deleterious mutations on neutral molecular variation. Genetics 134: 1289–1303.

Colosimo, P. F., K. E. Hosemann, S. Balabhadra, G. Villarreal, M. Dickson, J. Grimwood, J. Schmutz, et al. 2005. Widespread Parallel Evolution in Sticklebacks by Repeated Fixation of Ectodysplasin Alleles. Science 307:1928–1933.

Conte, G. L., M. E. Arnegard, C. L. Peichel, and D. Schluter. 2012. The probability of genetic parallelism and convergence in natural populations. Proceedings of the Royal Society B 279:5039–5047.

Cruickshank, T. E., and M. W. Hahn. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Molecular Ecology 23:3133–3157.

Delmore, K. E., J. S. L. Ramos, B. M. V. Doren, M. Lundberg, S. Bensch, D. E. Irwin, and M. Liedvogel. 2018. Comparative analysis examining patterns of genomic differentiation across multiple episodes of population divergence in birds. Evolution Letters 2:76–87.

DePristo, M. A., E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, G. Del Angel, M. A. Rivas, M. Hanna, and A. McKenna. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature Genetics 43:491–498.

Endler, J. A. 1986. Natural Selection in the Wild. Princeton University Press, Princeton NJ.

Fraser, B.A., A. Künstner, D. N. Reznick, C. Dreyer, and D. Weigel. 2015. Population genomics of natural and experimental populations of guppies (*Poecilia reticulata*). Molecular Ecology 24:389–408

Haasl, R. J., and B. A. Payseur. 2016. Fifteen years of genomewide scans for selection: trends, lessons and unaddressed genetic sources of complication. Molecular Ecology 25:5–23.

Innan, H., and Y. Kim. 2004. Pattern of polymorphism after strong artificial selection in a domestication event. Proceedings of the National Academy of Sciences of the USA 101:10667–10672.

Irwin, D. E., M. Alcaide, K. E. Delmore, J. H. Irwin, and G. L. Owens. 2016. Recurrent selection explains parallel evolution of genomic regions of high relative but low absolute differentiation in a ring species. Molecular Ecology 25:4488–4507.

Irwin, D. E., B. Milá, D. P. L. Toews, A. Brelsford, H. L. Kenyon, A. N. Porter, C. Grossen, et al. 2018. A comparison of genomic islands of differentiation across three young avian species pairs. Molecular Ecology 27:4839–4855.

Jones, F. C., M. G. Grabherr, Y. F. Chan, P. Russell, E. Mauceli, J. Johnson, R. Swofford, et al. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484:55–61.

Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25:1754–1760.

Losos, J. B. 2011. Convergence, adaptation, and constraint. Evolution 65:1827–1840.

Lunter, G., and M. Goodson. 2011. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Research 21:936–939.f

MacPherson, A., and S. L. Nuismer. 2017. The probability of parallel genetic evolution from standing genetic variation. Journal of Evolutionary Biology 30:326–337.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M. A. DePristo. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research 20:1297–1303.

Martin, S. H., K. K. Dasmahapatra, N. J. Nadeau, C. Salazar, J. R. Walters, F. Simpson, M. Blaxter, et al. 2013. Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. Genome Research 23:1817–1828.

Matthey-Doret, R., and M. C. Whitlock. 2018. Background selection and the statistics of population differentiation: Consequences for detecting local adaptation. bioRxiv 326256: https://doi.org/10.1101/326256.

Nei, M., and W-H Li. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proceedings of the National Academy of Sciences of the USA 79:5269–5273.

Orr, H. A., and A. J. Betancourt. 2001. Haldane's sieve and adaptation from the standing genetic variation. Genetics 157:875–884.

R Development Core Team. n.d. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.

Ralph, P. L., and G. Coop. 2015. The role of standing variation in geographic convergent adaptation. American Naturalist, 186: S5-S23.

Ravinet, M., A. Westram, K. Johannesson, R. Butlin, C. André, and M. Panova. 2016. Shared and nonshared genomic divergence in parallel ecotypes of *Littorina saxatilis* at a local scale. Molecular Ecology 25:287–305.

Reid, N. M., D. A. Proestou, B. W. Clark, W. C. Warren, J. K. Colbourne, J. R. Shaw, et al. 2016 The genomic landscape of rapid repeated evolutionary adaptation to toxic pollution in wild fish. Science 354:1305–1308.

Renaut, S., G. L. Owens, and L. H. Rieseberg. 2014. Shared selective pressure and local genomic landscape lead to repeatable patterns of genomic divergence in sunflowers. Molecular Ecology 23:311–324..

Rennison, D. J., K. E. Delmore, K. M. Samuk, G. L. Owens, and S. E. Miller. 2020. Data from: Shared patterns of genome-wide differentiation are more strongly predicted by geography than by ecology. American Naturalist, Dryad Digital Repository, https://doi.org/10.5061/dryad.34q91f0

Rougemont, Q., P. Gagnaire, C. Perrier, C. Genthon, A. Besnard, S. Launey, et al. 2017. Inferring the demographic history underlying parallel genomic divergence among pairs of parasitic and nonparasitic lamprey ecotypes. Molecular Ecology 26:142–62

Samuk, K., G. L. Owens, K. E. Delmore, S. Miller, D. J. Rennison, and D. Schluter. 2017. Gene flow and selection interact to promote adaptive divergence in regions of low recombination. Molecular Ecology 26:4378-4390.

Schluter, D. 2000. The Ecology of Adaptive Radiation. Oxford University Press, Oxford.

Schluter, D., and G. L. Conte. 2009. Genetics and ecological speciation. Proceedings of the National Academy of Sciences of the USA 106:9955–9962.

Schluter, D., and L. M. Nagel. 1995. Parallel Speciation by Natural Selection. American Naturalist 146:292–301.

Taylor, E. B., and J. D. McPhail. 2000. Historical contingency and ecological determinism interact to prime speciation in sticklebacks, *Gasterosteus*. Proceedings of the Royal Society B 267:2375–2384.

Teotónio, H., I. M. Chelo, M. Bradić, M. R. Rose, and A. D. Long. 2009. Experimental evolution reveals natural selection on standing genetic variation. Nature Genetics 41:251.

Thompson, K. A., M. M. Osmond, and D. Schluter. 2019. Parallel genetic evolution and speciation from standing variation. Evolution Letters 3:129–141

Trucchi, E., B. Frajman, T. H. A. Haverkamp, P. Schönswetter, and O. Paun. 2017. Genomic analyses suggest parallel ecological divergence in *Heliosperma pusillum* (caryophyllaceae). New Phytologist 216:267–78.

Vijay, N., C. M. Bossu, J. W. Poelstra, M. H. Weissensteiner, A. Suh, A. P. Kryukov, and J. B. Wolf. 2016. Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. Nature Communications 7:13195.

Vines, T. H., A. C. Dalziel, A. Y. Albert, T. Veen, P. M. Schulte, and D. Schluter. 2016. Cline coupling and uncoupling in a stickleback hybrid zone. Evolution 70:1023–1038.

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

Zeng, K., and B. Charlesworth. 2011. The Joint Effects of Background Selection and Genetic Recombination on Local Gene Genealogies. Genetics.111.130575.

Zeng, K., and P. Corcoran. 2018. The Effects of Background and Interference Selection on Patterns of Genetic Variation in Subdivided Populations. Genetics 210:1539–1554.

Figure Legends

Figure 1. *A,* Sampling locations for the 24 freshwater populations of stickleback. *B,* Marine – Freshwater F_{ST} profiles for linkage group 4 for two representative populations of each of the four freshwater ecotypes. Shades of grey correspond between panels *A* and *B.*

Figure 2. Schematic outlining the structure of pairwise population comparisons**.**

Figure 3. *A,* Comparison of pairwise correlation coefficients for F_{ST} between marine – freshwater population pairs (evolving with parallel selection) or lake – lake population pairs (diverging neutrally). *B,* Comparison of pairwise correlation coefficients for marine-freshwater F_{ST} between population pairs found within the same ocean basins and in different ocean basins. Populations pairs with the same ecology are indicated in light grey and those with different ecology are indicated in dark grey.

Figure 1

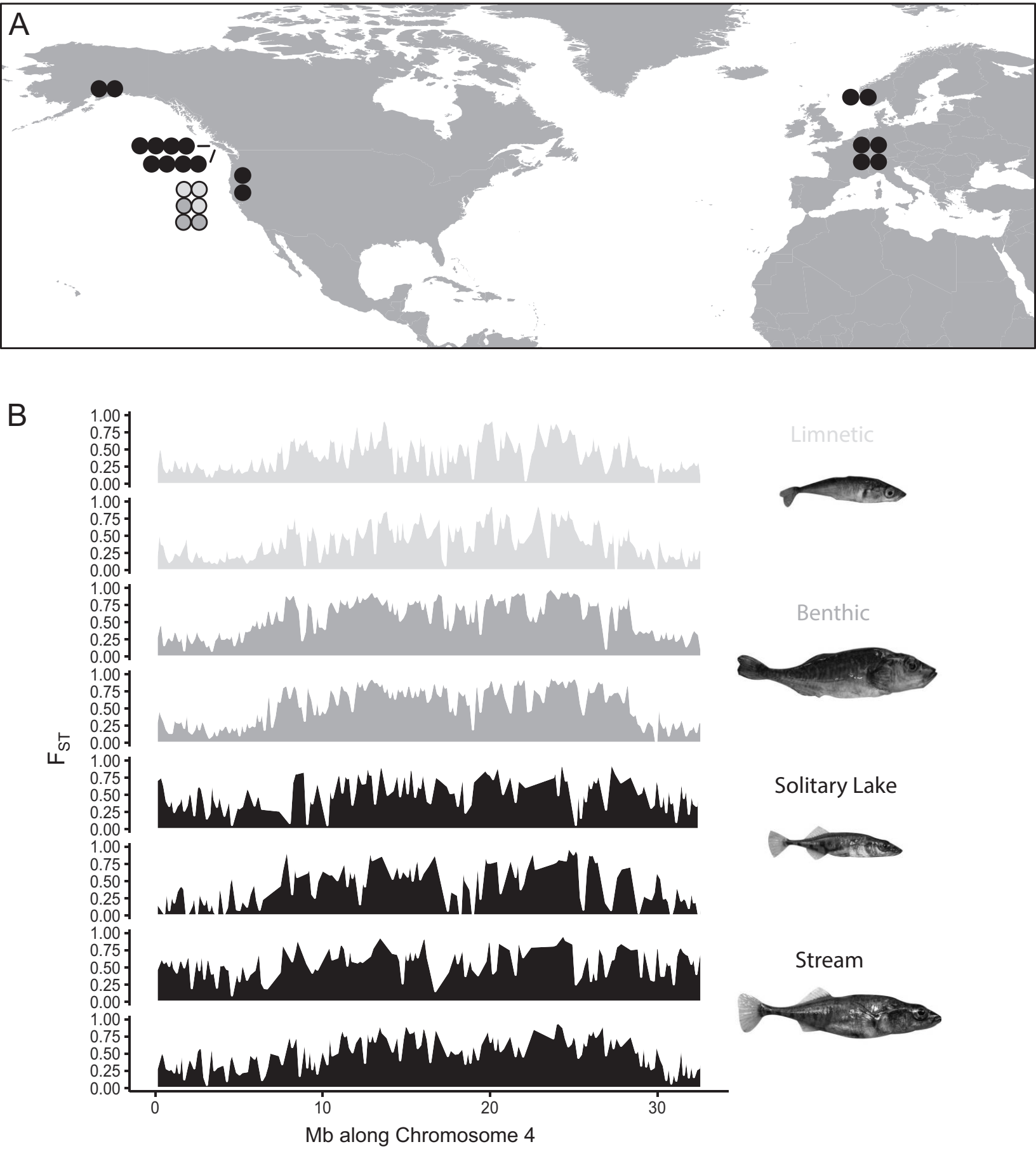


Figure 1 (online version)

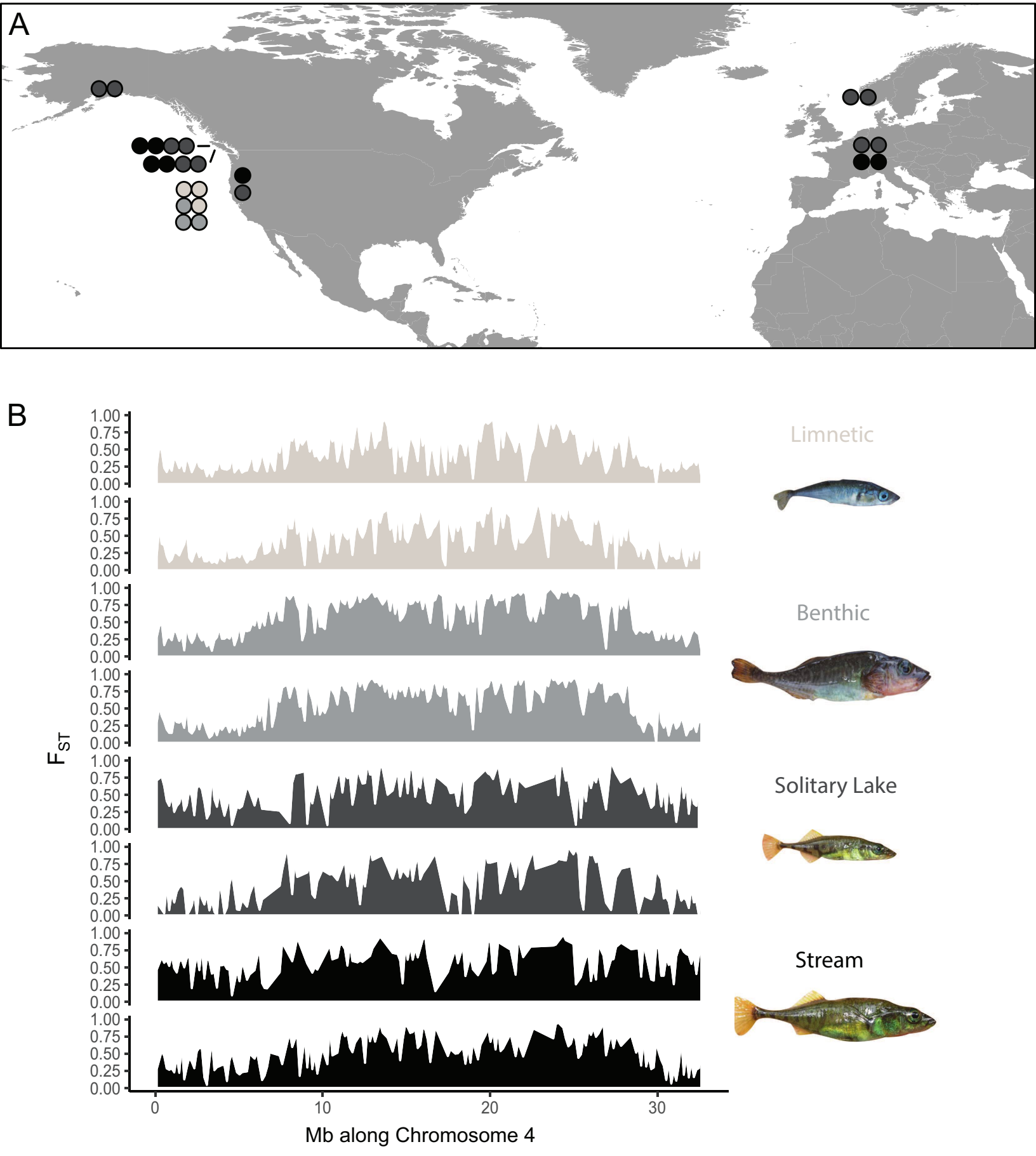


Figure 2

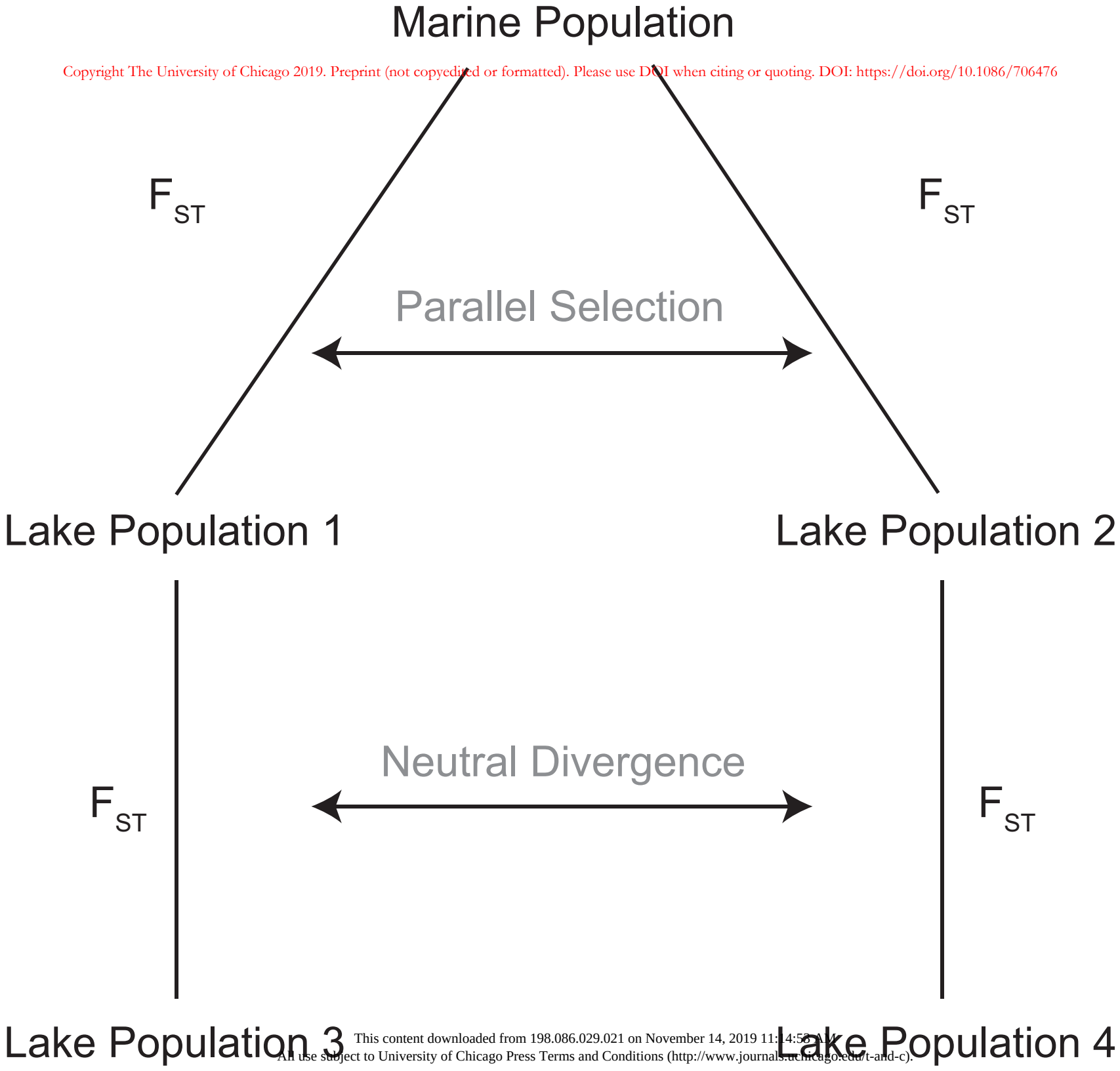
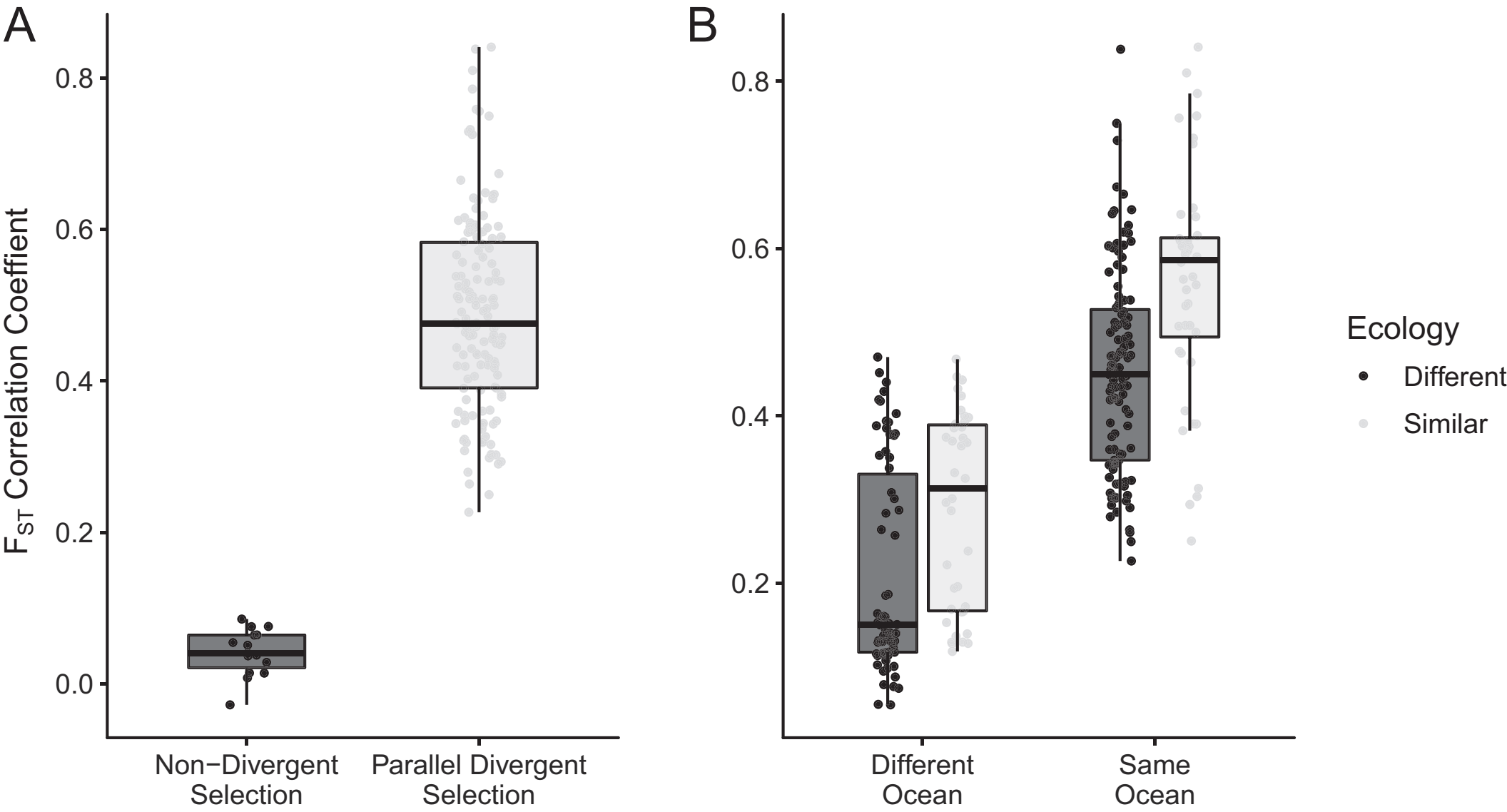


Figure 3



Shared patterns of genome-wide differentiation are more strongly predicted by geography than by ecology.

Diana J. Rennison[1,6]*, Kira E. Delmore[2]*, Kieran Samuk[3], Gregory L. Owens[4], Sara. E. Miller[5].

1) Institute of Ecology and Evolution, University of Bern, Bern, Switzerland.
2) Department of Biology, Texas A&M University, College Station, Texas, United States of America.
3) Department of Biology, Duke University, Durham, North Carolina, United States of America.
4) Department of Integrative Biology, University of California Berkeley, Berkeley, California, United States of America.
5) Department of Neurobiology and Behavior, Cornell University, Ithaca, New York, United States of America.

Corresponding author email: dianajrenn@gmail.com

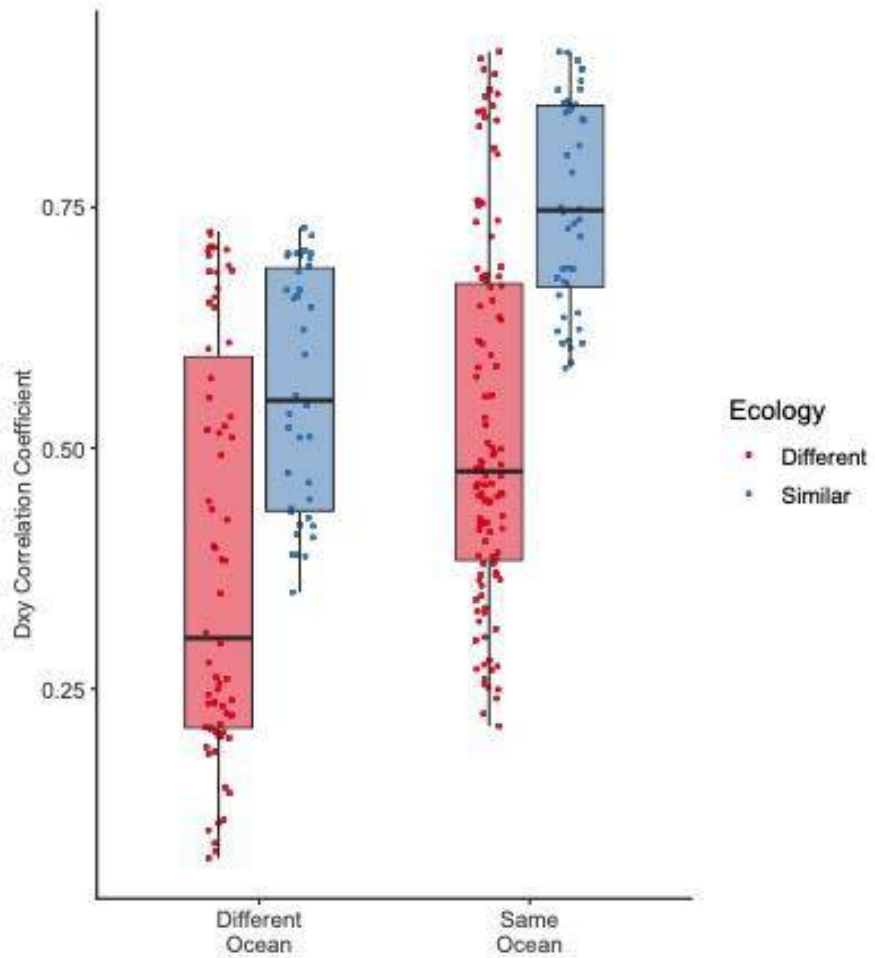
Supplementary Materials

Absolute estimates of differentiation of marine and freshwater population pairs.

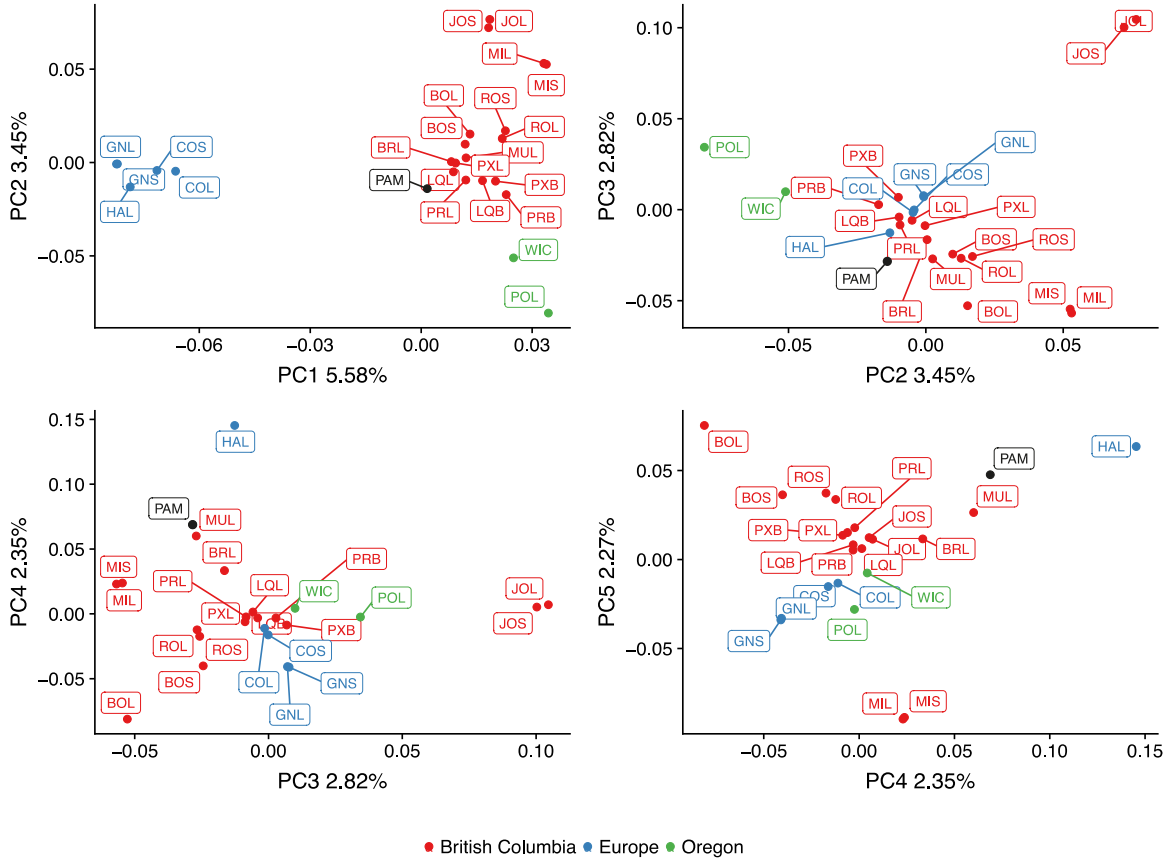
Correlation coefficients comparing windowed estimates of D_{xy} across marine-freshwater population pairs ranged from 0.07 to 0.91 (mean correlation coefficient, mean r = 0.53), further suggesting that the locations of genetic differentiation between marine and freshwater populations are often the same between independently derived population pairs. The average magnitude of the correlation of D_{xy} was significantly higher ($P < 0.001$) for population pairs found within the same ocean basin (average correlation coefficients (r) = 0.59) than for population pairs found in different ocean basins ($r = 0.44$) ($P < 0.001$). Pairs of populations occupying the same ecological niche had significantly more similar patterns ($P < 0.001$) for D_{xy} ($r = 0.64$) than population pairs occupying different niches ($r = 0.49$).

Repeated genomic differentiation with and without shared divergent selection when including all possible population pairs.

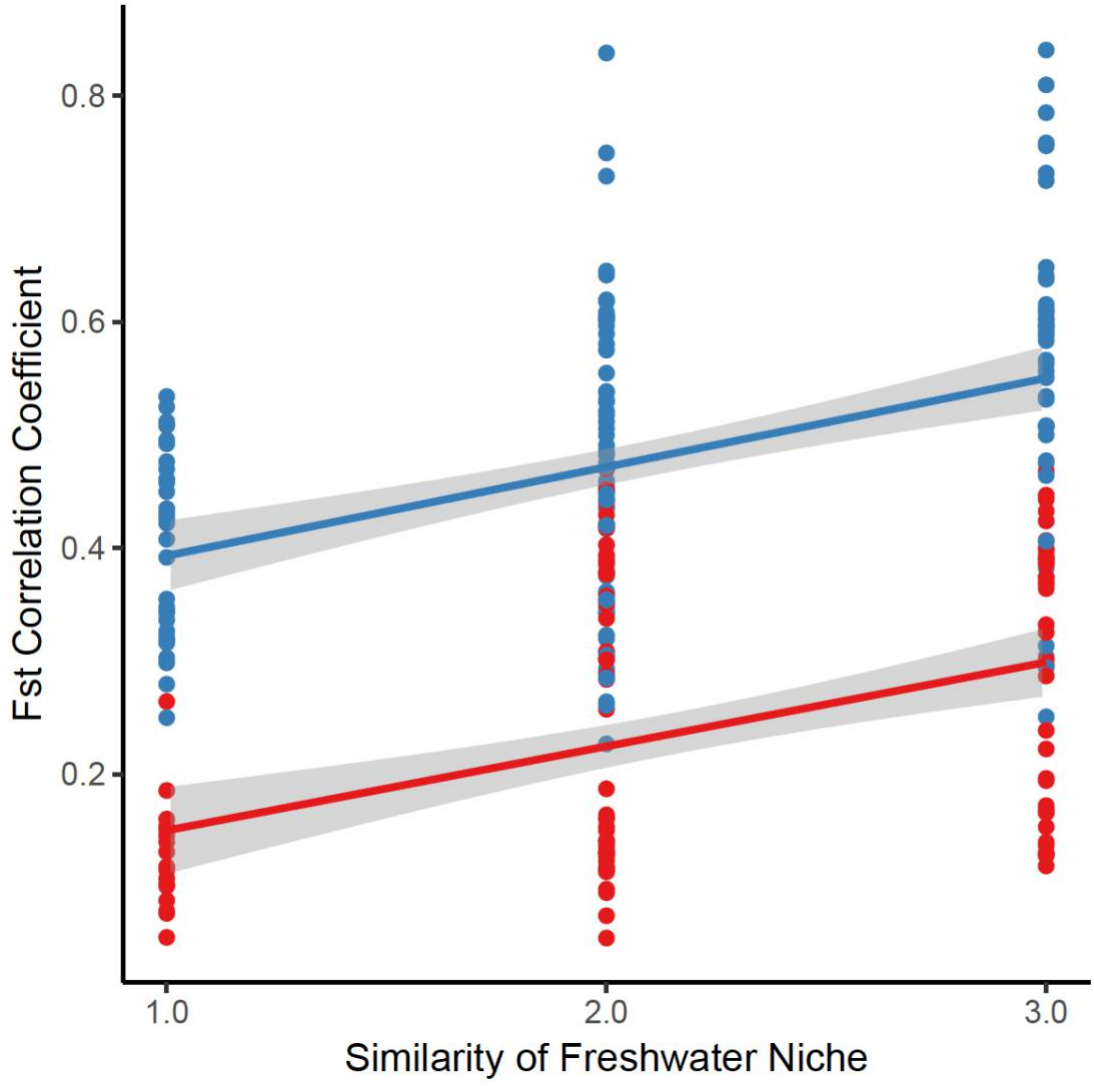
F_{ST} was more similar between marine – freshwater population pairs (average correlation coefficients (r) = 0.49) than between reference lake – lake population pairs (*i.e.* the neutral expectation) (mean r = 0.27). The difference in the average magnitude of correlation between the two types of comparisons was significant in permutation tests (difference in mean r = 0.22, $P < 0.0001$)



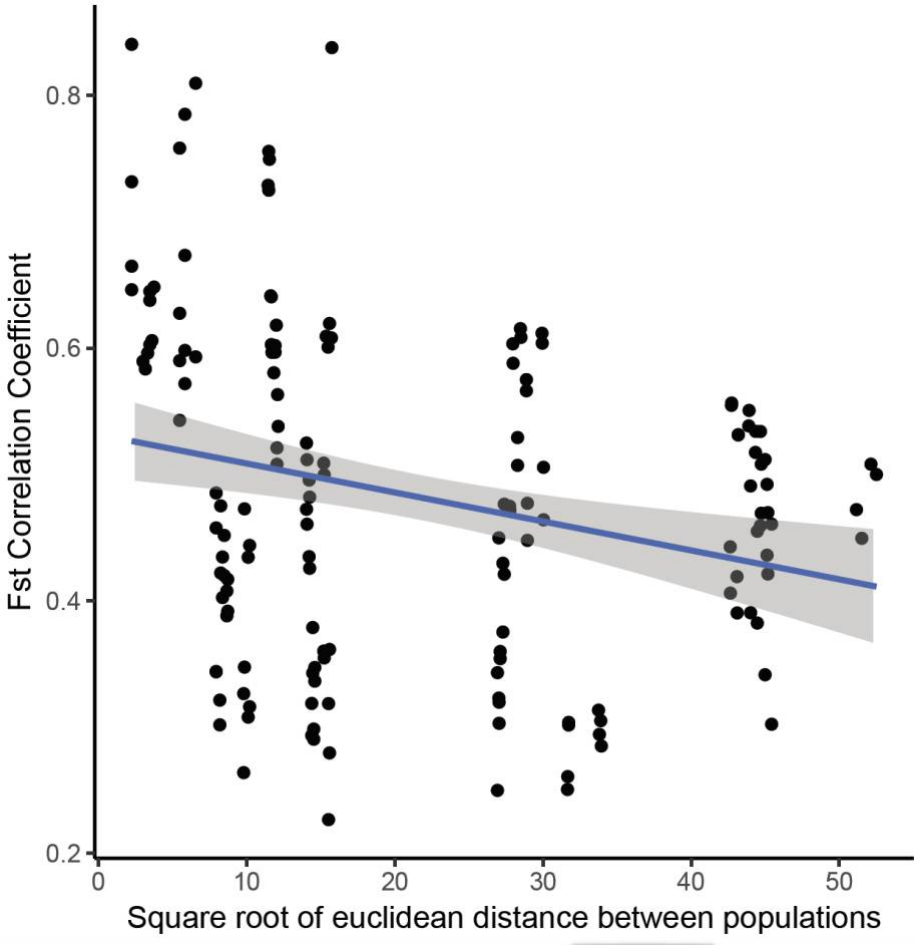
Supplementary Figure 1. Comparison of pairwise correlation coefficients for marine-freshwater D_{xy} between population pairs found within the same ocean basins and in different ocean basins. Populations pairs with the same ecology are indicated in blue and those with different ecology are indicated in red.



Supplementary Figure 2. Principal component analysis of genotypic data for the marine reference and derived freshwater populations. Each point represents the centroid of the population. The marine reference population is colored black. Population abbreviations are given in Supplementary Table 1.



Supplementary Figure 3. Relationship between pairwise correlation coefficients for marine-freshwater F_{ST} genome-wide and similarity of freshwater niche in the same (Blue) or different (Red) Ocean basins.



Supplementary Figure 4. Relationship between pairwise correlation coefficients for marine-freshwater F_{ST} genome-wide and pairwise Euclidean distance between populations within an ocean basin.

Table S1 Collection locations, names and metadata for all samples included in the study. Citations for each study are noted for the first occurrence of the study only. Data source abbreviates: Short Read Archive (SRA) and the European Nucleotide Archive (ENA). Sequencing technology abbreviations: Whole Genome Sequencing (WGS), Restriction Amplified Digest (RAD), Genotyping-by-Sequencing (GBS). BC refers to British Columbia, Canada, AK to Alaska, WS to Washington and 0R to Oregon, SW to Switzerland & DK to Denmark.

Study	*Latitude*	*Longitude*	*Region*	*Population*	*Abbreviation*	*Ecotype*	*Data Source*	*Acces. No.*	*Technology*	*No. Individuals*
Jones	*49.013*	*-122.778*	*BC*	*Little Campbell River*	*PAM*	*Marine*	*SRA*	*PRJNA247503*	*WGS*	*5*
Miller	*48.824*	*-125.136*	*BC*	*Bamfield*	*PAM*	*Marine*	*SRA*	*SRS2235303*	*WGS*	*1*
Miller	*49.612*	*-124.032*	*BC*	*Oyster Bay*	*PAM*	*Marine*	*SRA*	*SRS2235302*	*WGS*	*1*
Jones	*56.965*	*-151.35*	*AK*	*Kodiak*	*PAM*	*Marine*	*SRA*	*PRJNA247503*	*WGS*	*1*
Miller	*49.013*	*122.778*	*BC*	*Little Campbell River*	*PAM*	*Marine*	*SRA*	*SRS2235301*	*WGS*	*1*
Jones	*47.567*	*-122.547*	*WS*	*Manchestor*	*PAM*	*Marine*	*SRA*	*PRJNA247503*	*WGS*	*1*
Miller	*49.175*	*-122.778*	*BC*	*Salmon River*	*PAM*	*Marine*	*SRA*	*SRS2228867*	*WGS*	*1*
Miller	*50.373*	*-125.929*	*BC*	*Seyward*	*PAM*	*Marine*	*SRA*	*SRS2228870*	*WGS*	*1*
Miller	*49.191*	*-122.655*	*BC*	*West Creek*	*PAM*	*Marine*	*SRA*	*SRS2228871*	*WGS*	*1*
Catchen	*43.145*	*-124.190*	*OR*	*Winchester Creek*	*WIC*	*Stream*	*SRA*	*SRA070979*	*RAD*	*22*
Catchen	*43.424*	*-121.153*	*OR*	*Pony Creek Reservoir*	*POL*	*Lake*	*SRA*	*SRA070979*	*RAD*	*68*
Samuk	*49.663*	*-124.109*	*BC*	*Little Quarry Lake*	*LQB*	*Benthic*	*SRA*	*SRP107890*	*GBS*	*20*
Samuk	*49.663*	*-124.109*	*BC*	*Little Quarry Lake*	*LWL*	*Limnetic*	*SRA*	*SRP107890*	*GBS*	*10*
Samuk	*49.709*	*-124.525*	*BC*	*Paxton Lake*	*PXL*	*Limnetic*	*SRA*	*SRP107890*	*GBS*	*20*
Samuk	*49.709*	*-124.525*	*BC*	*Paxton Lake*	*PXB*	*Benthic*	*SRA*	*SRP107890*	*GBS*	*20*
Samuk	*49.745*	*-124.566*	*BC*	*Priest Lake*	*PRL*	*Limnetic*	*SRA*	*SRP107890*	*GBS*	*20*
Samuk	*49.745*	*-124.566*	*BC*	*Priest Lake*	*PRB*	*Benthic*	*SRA*	*SRP107890*	*GBS*	*20*

Roesti	*46.205*	*6.544*	*SW*	*Lake Geneva*	*GNS*	*Stream*	*SRA*	*SRP007695*	*RAD*	*27*
Roesti	*46.313*	*6.344*	*SW*	*Lake Geneva*	*GNL*	*Lake*	*SRA*	*SRP007695*	*RAD*	*27*
Roesti	*47.332*	*9.225*	*SW*	*Lake Constance*	*COL*	*Lake*	*SRA*	*SRP007695*	*RAD*	*27*
Roesti	*47.333*	*9.164*	*SW*	*Lake Constance*	*COS*	*Stream*	*SRA*	*SRP007695*	*RAD*	*27*
Roesti	*50.022*	*-125.336*	*BC*	*Boot Lake*	*BOS*	*Stream*	*SRA*	*SRP007695*	*RAD*	*27*
Roesti	*50.030*	*-125.323*	*BC*	*Boot Lake*	*BOL*	*Lake*	*SRA*	*SRP007695*	*RAD*	*26*
Roesti	*50.134*	*-125.331*	*BC*	*Roberts Lake*	*ROL*	*Lake*	*SRA*	*SRP007695*	*RAD*	*27*
Roesti	*50.143*	*-125.352*	*BC*	*Roberts Lake*	*ROS*	*Stream*	*SRA*	*SRP007695*	*RAD*	*27*
Roesti	*50.363*	*-127.156*	*BC*	*Misty Lake*	*MIL*	*Lake*	*SRA*	*SRP007695*	*RAD*	*27*
Roesti	*50.365*	*-127.322*	*BC*	*Joes Lake*	*JOS*	*Stream*	*SRA*	*SRP007695*	*RAD*	*26*
Roesti	*50.366*	*-127.170*	*BC*	*Misty Lake*	*MIS*	*Stream*	*SRA*	*SRP007695*	*RAD*	*27*
Roesti	*50.373*	*-127.291*	*BC*	*Joes Lake*	*JOL*	*Lake*	*SRA*	*SRP007695*	*RAD*	*27*
Ferchaud	*56.330*	*10.048*	*DK*	*Hadsten Lake*	*HAD*	*Lake*	*SRA*	*SRX437379*	*RAD*	*20*
Ferchaud	*56.383*	*9.354*	*DK*	*Hald Lake*	*HAL*	*Lake*	*SRA*	*SRX437379*	*RAD*	*20*
Hohenlohe	*61.563*	*-148.949*	*AK*	*Mud Lake*	*MUL*	*Lake*	*SRA*	*SRP001747*	*RAD*	*19*
Hohenlohe	*61.614*	*-149.756*	*AK*	*Bear Paw Lake*	*BRL*	*Lake*	*SRA*	*SRP001747*	*RAD*	*28*