AN ABSTRACT OF THE THESIS OF

William Hemstrom for the degree of Master of Science in Fisheries Science presented on May 26th, 2016.

Title: Thirty-four Kilometers and Fifteen Years: Rapid Adaptation at a Novel Chromosomal Inversion in Recently Introduced Deschutes River Three-spined Stickleback.

Abstract approved: _____

Michael A. Banks

Rapid adaptation and evolution based on standing genetic variation and novel mutations is likely to be one of the primary ways that species survive the widespread anthropogenic environmental changes expected of the next century. Three-spined stickleback (*Gasterosteus aculeatus*) are known to be capable of extremely quick, dramatic adaptation in response to similarly dramatic environmental changes, such as a transition from a salt to freshwater environment. In this study, we present evidence which shows selection has caused large increases in the frequency (1,784-6,482% relative increase) of a novel chromosomal inversion on LG IX in a recently introduced population of stickleback in the Deschutes River, Oregon, over approximately 35 years. This evidence suggests that subtle, watershed scale differences in environmental conditions can drive rapid evolution in stickleback, and possibly other fish species, over very short periods of time, which may aid those species both in surviving disturbances and in invading new ecosystems.

Thirty-four Kilometers and Fifteen Years: Rapid Adaptation at a Novel Chromosomal Inversion in Recently Introduced Deschutes River Three-spined Stickleback.

by
William Hemstrom

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Master of Science

Presented May 26, 2016
Commencement June 2016

Master of Science thesis of William Hemstrom presented on May 26, 2016

APPROVED:

Major Professor, representing Fisheries Science

Head of the Department of Fisheries and Wildlife

Dean of the Graduate School

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

William Hemstrom, Author

ACKNOWLEDGEMENTS

The author expresses sincere appreciation to Dr. William Cresko and Mark Currey of the University of Oregon, Dr. Barbara Shields of Bonneville Power Association, and Brett Hodgson of the Oregon Department of Fish and Wildlife or their contributions to this work, and to Jay Bowerman of the Sun River Nature Center and Eagle Crest Resort for facilitating sampling efforts on their properities.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Introduction:

It has recently been argued in *Nature* that the Earth is on the precipice of an anthropogenically driven mass extinction, driven by a "perfect storm" of broad spectrum ecological pressures and other anthropogenic environmental changes (Barnosky et al., 2011). To mitigate the effects of climate change and habitat destruction on global biodiversity, it is critical that we research how they will affect individual species and ecosystems. In response to climate and environmental changes, species generally react by migrating to more suitable habitat, adapting or evolving, or going locally extinct—each of which has happened repeatedly to many species in many habitats (Parmesan, 2006).

Adaptation and evolution in particular are both extremely important, since many species, especially those that are already endangered, cannot simply move. However, despite the large volume of research on the adaptive and evolutionary responses of populations to anthropogenic impacts, we do not yet have sufficient data to accurately predict how climate change and other global anthropogenic stresses will affect species at a genetic level (Franks & Hoffmann, 2012).

In their 2012 review, Franks and Hoffman argue that we are limited in our ability to predict genetic responses to environmental change by our incomplete understanding of how probable genetic evolution to rapid environmental change is and how fast that change will occur, among other things. While some theoretical work has been attempted to model evolution rates relative to climate change models (such as Kopp and Matuszewski (2014)), few direct comparisons have been made (Gienapp et al., 2013). Indeed, the research that has been conducted to study rapid, climate change driven

evolution has tended to focus the evolution of specific traits (as in (Karell et al., 2011)) or have been experimentally based (Franks & Hoffmann, 2012). Studies of these sorts may not properly elucidate the complex way in which genomic architecture influences the rates of evolution, or may neglect patterns of adaptation that are relevant to wild populations (Franks & Hoffmann, 2012). More studies on the rate of evolution of whole genomes in wild populations are therefore of interest, and could provide valuable information underlying mechanisms of rapid evolution and assist predictions on species responses to anthropogenic environmental changes.

Three-spined stickleback (*Gasterosteus aculeatus*) are an ideal species in which to study rapid adaptation to novel environments. A great deal of work has already been conducted on the species which has described their genome to an extremely high level of detail (Jones et al., 2012; Kingsley et al., 2004). Furthermore, stickleback are known to be extremely adaptable, even on very short time frames (Barrett et al., 2011; Gelmond et al., 2009; Hohenlohe et al., 2010). Indeed, a recent study by Bell and Aguirre (2013) has suggested that stickleback are capable of "measureable" evolution for many different traits over relatively short time frames, and that their genomic architecture is likely incredibly well suited to rapid adaptive radiation.

In the Deschutes River system of central Oregon, a natural "experiment" of sorts on stickleback evolution has been running for the last ~35-40 years. In the 1980s, stickleback were introduced to the system, and have since spread across nearly the entire region and into a multitude of radically different freshwater environments (Catchen,

Bassham, et al., 2013; Yake, 2003). The system, therefore, offers a unique opportunity to study stickleback evolution on a rapid time-scale in a natural environment.

We characterized the genomic divergence between six different Deschutes River sub-populations, and found strong evidence that a previously undescribed chromosomal inversion is being swept to a high frequency in two regions of the system at an extremely rapid rate (250+ percent change per generation), likely in response to solute rich groundwater inputs. Chromosomal inversions are well-known vehicles for adaptation in stickleback, particularly in the transition between freshwater and pelagic environments, lateral line development, and genomic sex determination, and have long been recognized as contributing to adaptation along environmental clines such as temperature in the various species of the *Drosophila* genus (Da Cunha et al., 1950; Dobzhansky & Sturtevant, 1938; Dubinin & Tiniakov, 1946; Jones et al., 2012; Krimbas, 1964; Rodríguez-Trelles et al., 1996; Roesti et al., 2015; Ross & Peichel, 2008; Spiess, 1950; Wark et al., 2012). Given that chromosomal inversions are likely candidates for differential selection due to their ability to maintain clusters of associated alleles and resist homogenization, it is not surprising that we observed rapid adaptation acting on the novel inversion that we observed (see Kirkpatrick and Barton (2006) and Yeaman (2013), for example). However, the rate of adaptation we observed was incredibly fast, particularly given that any putative sources of selection in this instance are relatively minor changes in stream conditions rather than dramatic environmental shifts such as a salt/freshwater transition.

Methods:

Sampling and Study Sites:

We collected a total of 465 three-spined stickleback from five locations in the Deschutes watershed in late April through early June 2015 and 55 stickleback from one additional site in November 2014 using unbaited minnow traps (Figure 1). Of these, sites, three were presumptive "cold water" sites and three were presumptive "warm water" sites.

We sampled 150 stickleback from two primary sites, both located on the Crooked River, which we selected because to the high level of thermal variance between them over a relatively small geographic range. The first of these, Opal Springs (OPL), is located in an approximately 800 foot canyon, about eight kilometers upstream from Lake Billy Chinook where the Crooked River drains into the Deschutes. Water temperatures are cold in this region in the summer (estimates vary from about 14°C in late July, to 12-14 degrees in August) because of an influx of cold water from multiple upstream springs (Isaak et al., 2011; Watershed Sciences, 2006). A small hydroelectric dam in this area creates a very small, shallow, and highly vegetated reservoir which hosts a sizable stickleback population. Stream temperatures rise steadily moving upstream between OPL and our second primary site, located approximately 34 river kilometers upstream at Smith Rocks State Park (SMR), where estimates place water temperatures between 24-26 °C in late July and 20-30 degrees in August. The portion of the river around SMR is more exposed, in a predominately agricultural area, and lacks significant cold water spring inputs, but is similarly highly vegetated. Surveys suggest that stickleback are likely only

recently (~10-15 years, surveys conducted in 2004 found only 3 stickleback, near OPL) introduced into the Crooked from the Deschutes proper, likely upstream of both SMR and OPL (Torgersen et al., 2003).

Our four secondary sites were selected to provide supporting data from additional warm and cold water locations. We sampled 55 individuals from the inlet of Aspen Lake (ASP) at the Sun River Nature Center (about 22-24 °C in July based on temperature logger data obtained from the Nature Center) and 55 from the Deschutes River at Cline Falls (CLF), which is estimated to be between 22 and 26 °C in late July and 20-30 °C in August (Isaak et al., 2011; Watershed Sciences, 2002). Interestingly, the Deschutes near CLF is geographically near to OPL (although distant by river kilometers), and receives cold water spring inputs from the same overarching geological feature, although these do not cool the river to the same degree. We also sampled 55 individuals from the upper region of the Deschutes adjacent the Sun River Nature Center (UPD), where cold water springs keep temperatures low in the summer (about 16-17 °C according to FLIR data taken in late July by the Oregon DEQ and 16-18 °C according to the NorWeST estimates). Lastly, we also sampled 55 individuals from Paulina Lake (PAL) directly adjacent to the Paulina Creek outlet. PAL is a high elevation (6, 350 ft) lake which fills a portion of the Newberry volcanic caldera southeast of Bend. It is fed primarily by rainfall, snowmelt, and by a hot spring in the northeast portion of the lake. While the hot springs serve to keep the lake relatively warm, the southwest portion likely harbors cold water refugia during the summer, as suggested by the cold waters which exit the lake into Paulina Creek (approx. 18 °C in late July, see also Koketsu (2004)). Previous work and

the lack of connectivity to downstream water systems suggests that the stickleback population in PAL represents a separate introduction into the Deschutes watershed, although both this introduction and the original likely share a common source population (Catchen, Bassham, et al., 2013).

Sequencing:

We extracted DNA from 451 stickleback (150 from SMR, 151 from OPL, and 50 from each secondary site) using phenol/chloroform (Maniatis et al., 1982). We then prepared standard (non-reduced representation) 2b-RAD libraries according to Wang *et al*. (2012) using the Alf1 restriction enzyme and 10 PCR cycles. We then combined samples into 12 pools for 50bp, single-end sequencing on an Illumina Hi Seq 3000.

Sequence processing and genotyping:

We first filtered all reads for which 18 or more insert bases (out of 36) had sequencing quality scores lower than 30 (99.9% confidence), then removed all reads with an alignment score higher than 18 to adaptors. We then aligned the remaining reads to the Alf1 sites in the *G. aculeatus* genome (Jones et al., 2012), and filtered all alignments which spanned less than 32 bases and had fewer than 30 total matching bases. We then combined sequence data for sites where multiple tags overlapped a region within each individual sample, and genotyped all genomic sites with at least 15x coverage. Sites were determined to be homozygous if the frequency of any minor alleles were less than 1% and heterozygous if the frequency of each observed allele was between 25 and 75%. Sites where observed allele frequencies were found to be outside of this range were determined to be ambiguous, and removed. We used this genotyping method rather than one based on

a binomial/probabilistic framework, such as that employed by the STACKS pipeline (Catchen, Hohenlohe, et al., 2013), due to previous research which has suggested that these cut-offs agree closely with genotypes called by those methods where both call genotypes, and are more likely to reject genotype calls which probabilistic methods call incorrectly. (Wang et al., 2012). We then removed all individuals which were genotyped at less than half of the sequenced loci and all loci which were genotyped at less than half of the remaining individuals.

We then selected all sites where exactly two alleles were observed across all samples to form our list of observed SNPs. Since a high number of SNPs in a single RAD tag can be an indicator that the tag in question is repeated multiple times in the genome, we removed all SNPs from tags on which we observed three or more SNPs to remove those which were potentially incorrectly mapped to the genome. Lastly, we removed one SNP from each tag with two SNPs, since these loci are likely to be in very close linkage and thus provide redundant information, and removed all SNPs sequenced in less than 50% of samples and all individuals sequenced in a small number of SNPs from our dataset.

Data Analysis:

To determine the overall level of population sub-structuring between populations, we selected the 1000 SNPs which were sequenced in the most individuals across all populations. We then used this data to conduct ten 100,000 iteration runs using the program STRUCTURE for each value of *k* (or the putative number of clusters) between one and seven (Pritchard et al., 2000).

We calculated pairwise F_{ST} estimates at each SNP between each pair of populations using Wier and Cockerham's (1984) weighted analysis of variance via the GENEPOP software package (Raymond & Rousset, 1995). Since our sampling groups differed in size by a substantial degree (~50 individuals for the secondary sites vs. ~125 individuals in the primary sites), we also used the weighted estimator proposed by Nielsen *et al* (2009) as adapted by Hohenlohe (2010) and employed in the STACKS package using custom R scripts, which is supposedly less biased by differences in sample size (Catchen, Hohenlohe, et al., 2013). As an alternative to these methods, we also calculated log10 posterior odds and alpha values for every SNP between each pair of populations using the program BayeScan with prior odds for the neutral model set to 100 and other settings at their default (Foll & Gaggiotti, 2008), which uses a Bayesian Reversible Jump-MCMC framework to estimate the likelihood that each loci is under selection (log10(PO)) and the direction of said selection (positive alpha values indicate diversifying selection, negative values indicate purifying selection). We then used the BayeScan package to identify outlier loci based on a 0.05 false discovery rate.

We also calculated genetic variation (π) for each SNP in each population as employed in the STACKS package using custom R scripts. We calculated the observed heterozygosity (H_o) at each SNP in each population as a simple proportion of heterozygotes across all individuals within that population. We also determined the presence of private alleles (P_A) by simply checking for alleles which were only present in samples from one location. If a private allele was identified, we assigned that SNP for that group of samples a 1. If not, we assigned the SNP a 0. To determine how tightly π

and H_o were correlated across the genome between populations, we also calculated the Pearson correlation coefficients for π and H_o between each population at the same SNP aside from those on the sex determining linkage group (LG) XIX, where genome wide stats were heavily biased by sampling sex ratios. In order to estimate the local levels of linkage disequilibrium across each LG within each population, we calculated the average of both D' and r^2 for each pair of SNPs within a 6σ sliding window centered on each SNP (Hill & Robertson, 1966; Lewontin, 1964), where σ was equal to 200kb. We also calculated the same estimators of pairwise linkage disequilibrium (LD) between every pair of SNPs in each LG within each population using custom R scripts to determine how full patterns of inter-LG LD varied across the genome and between populations.

We then used the Gaussian smoothing method described by Hohenlohe *et al* (2010) and employed by the STACKS package to calculate the local average values of π, pairwise F_{ST}, q-value, alpha, local r^2, local D', number of private alleles, and H_o in 6σ sliding windows centered around each SNP, weighted by the proximity of each SNP within that window to the central SNP. As before, we used σ = 200kb, which was small enough to preserve relatively narrow bands of divergence while not being overly biased by regions of the genome with few SNPs. Where windows overlapped the edges of LGs, we truncated them. To reduce bias due to differences in sequencing success at different SNPs, we additionally weighted every SNP's contribution to local average values using the formula $(nk - 1)$, where *nk* is the total count of observed alleles at that SNP in all relevant samples (for example, for pairwise F_{ST} between SMR and OPL, *nk* is the sum count of both the major and minor alleles in all SMR and OPL samples), again as

described in Hohenlohe 2010. To determine if differences in *nk* per SNP or SNP density per sliding window had any effect on pairwise F_{ST}, we fit multiple linear regression models ($F_{ST} \sim nk + count$) to each pairwise F_{ST}.

To calculate *p* values for each smoothed value of F_{ST}, π, r^2, H_o, and P_A, we used a bootstrap resampling approach to create an expected null distribution of these values from the observed genome-wide values of each statistic. To do so, we first calculated the number of SNPs in each sliding window observed above. To calculate each bootstrapped value, we then randomly sampled with replacement one of these counts of SNPs within a window, then randomly sampled with replacement that number of values of the statistic of interest from across the genome and randomly designated one as the value about which to center the bootstrapped window. Since a small portion of the observed sliding windows overlapped the ends of LGs (for SNPs within 3σ of the edge of a LG), we first selected a random position on one of the observed LGs and assigned this position to the central value. Every other value was randomly given a position relative to that central SNP within a 6σ window. If one of these randomly assigned values was outside the range of the LG (a negative position or a position greater than the size of the LG), we selected a new random relative position for it. We then calculated a smoothed value for the window based on each sampled value and its randomly derived relative position according to the smoothing equation given by Hohenlohe *et al* (2010). To better reflect the variability observed in the data, we also randomly selected an observed *nk* value. For each of these statistics, we calculated one million bootstrapped smoothed values, which was sufficient

to create an acceptable degree of uniformity in the tails of the bootstrapped distributions for each statistic.

Since the bootstrapped distributions tended to be extremely long-tailed rather than normal, we calculated *p*-values for each observed smoothed statistic by direct comparison to the bootstrapped distribution. Observed statistics that were more extreme than any observed in the bootstrapped distribution for that statistic were assigned a conservative *p*-value equal to one divided by the number of bootstraps (the lowest possible *p*-value calculable from the distribution).

For one region of interest on LG IX, we also divided SNPs into a "divergent" ($F_{ST} > 0.1$) or a "not divergent" ($F_{ST} < 0.1$) group based pairwise F_{ST} scores for the OPL/SMR comparison, and mapped π and H_o across the region for both groups. OPL/SMR was the most conservative choice for this, given that smoothed pairwise F_{ST} scores for this comparison were the lowest for the region that were still significant. We compared results using this baseline to those produced using a different comparison as a baseline or different cutoff values of F_{ST}, and found no significant difference. We then used one-way ANOVA and two-sample *t* tests to determine if π and/or H_o differed between sampling sites for both divergent and non-divergent SNPs. We also fit a multiple regression model to determine if pairwise-r^2 measures of LD were different on average for comparisons between SNPs in the region of interest and/or between diverged SNPs than between other comparisons. To estimate the frequency of the probable chromosomal inversion (f_I) in this region, we calculated the mean frequency of the rare allele (q) for each divergent SNP in the region after removing SNPs for which the frequency of q was greater than

0.05 in ASP (those which were likely not fixed differences between the inversion and the typical version of the genome).

We used f_I in ASP plus UPD as a proxy for baseline f_I in the ancestral population to calculate the change in f_I between the baseline and CLF, OPL, and SMR by simple subtraction, relative changes in f_I by division, and relative rates of increase of f_I over the years since establishment of OPL and SMR by a simple exponential growth function $f = f_0(1 + r)^t$, where *f* is the frequency of the inversion in the population, f_0 is the frequency of the inversion in ASP/UPD, *r* is the average percent increase per generation, and *t* is the best estimate of the number of generations since establishment. We estimated the average percent increase per generation of f_I between OPL and SMR in the same way, with the frequency in SMR as the baseline. We used *t* = 35 years/generations to calculate rates of increase in f_I between the ancestral population and our samples, and both *t* = 10 and *t* = 15 to calculate the rate of increase between SMR and OPL, based on our knowledge of the timing of stickleback introductions into the Deschutes and Crooked Rivers, as described above. We estimated the selection coefficient (*S*) assuming complete dominance and a constant rate of selection using the equation $s = \frac{1-\frac{p_0}{p_1}}{q^2}$, adapted from Gillespie (2010) equation 3.2.

Lastly, we used the GSR option in the program ermineJ to determine if any functional gene groups we associated with higher pairwise F_{ST} values for each pair of populations (Lee et al., 2005). We recovered gene annotations using the Ensemble BioMart tool and estimated F_{ST} for each gene on each LG via interpolation of smoothed F_{ST} values using custom R scripts (Smedley et al., 2015). We discarded genes which were

outside the range of observed SNPs and those which were on scaffolds rather than established LGs, since we recovered few SNPs per scaffold on average. We used F_{ST} between our primary sites of OPL and SMR as our gene score, although we also ran several other pairwise F_{ST} values as scores to validate these results.

Results:

Sequence processing and genotyping:

We recovered 11,962,114,596 raw reads across all 451 samples. Of these, 11,941,593,624 remained after initial filtering steps (99.8%). After mapping these reads to the *G. aculeatus* genome and removing data for low-coverage (<15X) or poorly aligned loci, we retained data for 2,506,606 bases at an average coverage of 55.9x after removing 15 poorly sequenced individuals and 506,682 poorly sequenced loci (less than five individuals were removed from each population). We genotyped a total of 19,854 SNPs across all individuals prior to filtering, of which we retained 10,952 after filtering and removing poorly sequenced individuals and poorly sequenced loci. These were sequenced at an average depth of 53.2x. 10,196 of these SNPs were on large established LGs rather than scaffolds.

Overall diversity and differentiation:

Overall pairwise F_{ST} was greatest between PAL and every other sampling location (mean F_{ST} = 0.104) and smallest between both SMR/OPL (F_{ST} = 0.0110) and ASP/UPD (F_{ST} = 0.000926), following expected patterns given the history of introductions in the region and local geography (Table 1). F_{ST} was also lower when comparing either OPL or SMR to either ASP or UPD than comparing OPL or SMR to CLF, contrary to their geographic proximities but not their hypothesized introduction pattern (mean F_{ST} = 0.0291 vs mean F_{ST} = 0.0240, respectively). STRUCTURE produced similar results, separating PAL from the other sites at a *k* of 2 and grouped SMR and OPL apart at a *k* of 3 (data not shown). Overall SNP π was lowest in PAL (π = 0.213) but similar overall

(Table 2). Overall SNP H_o was also lowest in PAL (H_o = 0.222), highest in UPD (H_o = 0.330), and intermediate but similar in the other sites (between 0.278 in CLF to 0.300 in ASP). There were generally very few private alleles, although there were substantially more observed in SMR (Table 2). The average degree of local LD measured by r^2 was small overall, and higher in PAL than in other sites (r^2 = 0.0428 in PAL, 0.0142-0.0217 elsewhere). Local LD as measured by D' differed little between populations (D = 0.198-0.223, Table 2). All of the above statistics are given after the removal of the sex-determining LG XIX.

Genome-wide variation in F_{ST} and diversity:

On a genome wide level, pairwise F_{ST}, genetic diversity (π and H_o), and local LD varied greatly within each population and between each pair of populations. Wier and Cockerham's (1984) method of estimating F_{ST} seemed to perform best overall, producing fairly small F_{ST} between populations where biogeography and introduction history would suggest little overall differentiation, regardless of population size. Surprisingly, the method adapted by Hohenlohe *et al*. (2010) and used in the STACKS package performed notably worse, producing much higher overall F_{ST} scores and many more large peaks in divergence, particularly when comparing between sampling sites with different sample sizes (data not shown). Genome-wide, we observed one large block of repeated elevated F_{ST} on LG IX, between ~ 13.97-17.75mb, covering roughly 18.5% of the LG (Figure 2). Pairwise F_{ST} values in this block were elevated across most of the region for each of the upper basin sites (ASP, UPD, and the outgroup PAL) compared to both OPL and CLF, as well as between the closely related OPL and SMR. These F_{ST} values were significant at *p*

< 0.00005 for these elevated comparisons, save for those involving UPD, where it was nearly significant ($p \sim 0.00075$). Within this region, for example, the average pairwise F_{ST} between OPL and SMR was 0.0871, nearly nine times higher than the average pairwise F_{ST} between SMR and OPL at all other LG SNPs excluding those on LG XIX (0.0097). This difference in average F_{ST} was even more pronounced in each of the other comparisons listed above. Interestingly, we did not find elevated pairwise F_{ST} values between SMR and CLF in this region. We found only four other small peaks of significant smoothed F_{ST} across all comparisons, none of which were repeated (data not shown). BayeScan log10(PO) values generally mirrored but were less extreme than pairwise F_{ST} values given by Wier and Cockerham's method, and outliers were scarce and only found in LG XIX and in the same region of interest in LG IX, aside from a small handful on scaffolds for which little other data is available (data not shown).

We found that genome-wide π was strongly correlated between sampling locations save PAL, with ρ ranging between 0.807 for OPL/CLF and 0.944 for UPD/ASP after removing LG XIX SNPs and those from the high F_{ST} region of LG IX (all p – values < 0.0001). π in the PAL samples, while generally lower than in the other samples, was strongly divergent from the others throughout the genome, showing a less strongly, but still significant positive correlation to each other group of smaples (ρ = 0.535 for CLF to 0.600 for ASP, all p – values < 0.0001). Substantial divergence in π was observed at the same location as the above mentioned peak of F_{ST} divergence on LG IX, where π was lowest in PAL (mean π = 0.0756), followed by ASP (0.133), then UPD (0.142), SMR (0.217), and CLF (0.288), and was highest in OPL (0.350, see Figure 3). Overall, π was

variable enough that few genomic regions had significantly elevated or reduced π relative to the genome-wide bootstrapped null distribution.

We also observed a high level of genome-wide correlation in H_o between SNPs across populations, with ρ between 0.73 for CLF/UPD to 0.86 for ASP/UPD (all p – values < 0.0001). Again, the degree of correlation was much smaller but still statistically significant when comparing the PAL samples to those from the other locations (ρ between 0.49 and 0.56, all p-values < 0.0001). H_o was also substantially different between sampling sites for the region between ~ 13.97 and 17.75mb on LG IX, following the same pattern as with π in the region (H_o = 0.0809, 0.137, 0.173, 0.227, 0.302, and 0.362 for PAL, ASP, UPD, SMR, CLF, and OPL, respectively, see Figure 4). Other than in this region, H_o varied across the genome in each sampling group, but no major divergences between sampling sites or regions of drastically reduced or increased H_o were observed.

P_A was highly variable due to the rarity of private alleles outside of SMR, and no strong trends were observable (data not shown).

We found substantially different trends in π and H_o among both divergent and non-divergent SNPs in the differentiated region of LG IX. For both non-divergent and divergent SNPs, we found strong evidence to indicate that mean π was not the same for all sampling locations (p = 0.0005 and $p << 0.00001$, respectively). However, for the non-divergent SNPs, we found only evidence that π in PAL differed from that in the other sampling locations (p = 0.000156 – 0.00551 from two-sample t-tests, all significant after Bonferroni correction). For locations other than PAL, there was no evidence to indicate

that average π differed between locations (*p*-value = 0.9978, Figure 5a). For the divergent SNPs, however, there was convincing evidence to indicate that π differed between multiple sets of sampling locations, with π differing significantly on average between all sampling locations save internal comparisons between ASP, PAL, and UPD (all *p* values << 0.0001). On average, π in this region for the divergent SNPs was much higher in OPL and CLF (π = 0.439 and 0.326) than in UPD, ASP, and PAL (π = 0.0349, 0.0206, and 0.0113), and intermediate in SMR (π = 0.179) (Figure 5b). H_o mirrored these trends, and was extremely low (just above 0) on average in ASP, PAL, and UPD (Figure 5c and Figure 5d). The average frequency of *q* for these divergent SNPs was substantially different across sampling sites after removing SNPs which likely represent non-fixed differences between the inversion and the typical version of the genome: 0 in PAL, 0 in ASP, 0.0098 in UPD, 0.092 in SMR, 0.20 in CLF, and 0.32 in OPL (standard deviation = 0, 0, 0.041, 0.017, 0.025, and 0.038, respectively, see Figure 5e).

In general, LD was very low across the genome, with only two major peaks in local r^2. D' was extremely variable across the genome, likely due to the high number of private alleles we observed, which will produce a D' of 1 for a SNP when present. As such, we used r^2 as our primary metric of LD. We observed two peaks of LD outside of LG XIX, one at the region of high divergence on LG IX, and one at ~12mb in LG VII (see Figure 6 for LG IX, LG VII not shown). The peak on LG IX was present in all sub-populations save PAL and ASP (the populations where no putative inversions were observed), and was considerably higher than the background levels of LD (mean r^2 among relevant sub-populations = 0.080 vs 0.017). Smoothed local average r^2 values in

this region were significant at $\alpha = 0.00005$ in UPD, SMR, OPL, and CLF. LG-specific full pairwise LD revealed clean edges to this peak of r^2, consistent with the position of the edges of the region of elevated pairwise F_{ST} and divergence in π/H_o between sub-populations (Figure 7). The differences in r^2 between divergent SNPs, SNPs in the region of interest, and background SNPs was pronounced in SMR, where pairwise-r^2 between divergent SNPs was 0.19 higher than between other SNPs in the region, and 0.21 higher than the background r^2, after accounting for the effect of SNP proximity. The differences in pairwise-r^2 were also substantial in CLF, OPL, and UPD, although to a smaller degree (0.15, 0.18, 0.13, 0.17, and X, respectively, Figure 8a). The differences in pairwise-r^2 between these groups was much smaller, although still statistically significant, in samples from ASP, and PAL (between +0.041 and -0.0046, Figure 8b). The peak of local LD on LG VII was substantial but much narrower. The region of LD here was tightly contained and degraded gradually farther from its center (data not shown).

We found very different estimated inversion frequencies between samples. f_I was zero in both PAL and ASP, and very low in UPD (0.0098), but much higher in SMR, CLF, and OPL (0.092, 0.20, and 0.32, respectively). Joint f_I in ASP/UPD was 0.0049, equating to a single incidence of the inversion across all samples. We found an absolute increase in f_I (Δf_I) of 0.0874 between SMR and ASP/UPD, 0.195 between CLF and ASP/UPD, and 0.318 between OPL and ASP/UPD. Estimated Δf_I between SMR and OPL was 0.0230. We observed a relative percent increase in f_I between SMR and ASP/UPD of 1,784%, between CLF and ASP/UPD of 3,989%, between OPL and ASP/UPD of 6,482%, and between SMR and OPL of 249%. For $t = 35$ (35 years/generations), we

estimated an average per generation relative increase in f_I of 8.75% between SMR and the ancestral proxy, 11.2% between CLF and the ancestral proxy, and 12.7% between OPL and the ancestral proxy. We estimated an average per generation relative increase in f_I of 13.3% between SMR and OPL assuming $t = 10$ and 8.7% assuming $t = 15$. We estimated an S of 0.114, 0.081, and 0.101 for OPL, SMR, and CLF vs the ancestral proxy, respectively, and an S of 0.119 for SMR vs OPL.

Only the related "membrane coat" and "coated membrane" protein families (GO:0030117 and GO:0048475, respectively) were significantly associated with high SMR/OPL pairwise F_{ST} scores via GSR in ermine (corrected p-value 0.0239). The high scores of this family were primarily driven by the presence of five (out of 42) members of these families in the high F_{ST} region of LG IX. This result did not vary strongly between GSR analyses using other pairwise F_{ST} values as gene scores if those values were from a comparison with a significant differentiation in the IX region.

Discussion:

Introduction History of Deschutes Stickleback and Broad-Scale Structuring:

At a basic level, our results suggest a complex history of repeated stickleback introductions in the Deschutes River. Our finding that the average pairwise-F_{ST} values between PAL and the other sampling sites was considerably higher than that between all other locations (~ 0.1 vs. ~ 0.027, respectively) is consistent with the hypothesis favored by Catchen *et al*. (2013) that the Paulina Lake population of stickleback is the result of a separate introduction. Given the relatively low genetic diversity that we also observed in PAL compared to the rest of our sampling locations (average $\pi = 0.204$ in PAL vs. 0.273 average π for the other locations), which is to be expected in the case of a recent population bottleneck, it seems very likely that this introduction is more recent in comparison to the ~1980s introduction to the rest of the system.

Our evidence also supports a separate introduction of stickleback into the Crooked River, from a source population relatively high upstream in the Deschutes. We found that stickleback in the Crooked River (SMR and OPL) had relatively higher pairwise-F_{ST} values when compared to fish from the geographically proximate CLF sampling site than when compared to fish from ASP and UPD, which are located hundreds of river kilometers upstream. This confirms management speculations that stickleback in the Crooked River were artificially introduced rather than natural migrants from the Deschutes proper (ODFW, personal communication), and that the fish that founded the population were taken from high up the Deschutes rather than from the nearby portions of the larger river. This is not entirely surprising, given that the

hydroelectric dam at OPL likely blocks upstream passage of stickleback into the Crooked River from the Deschutes proper. Our findings are consistent with a natural origin for the stickleback at CLF, however, since the relatively high pairwise-F_{ST} values we found between CLF and ASP/UPD would be expected given the greater geographic distance between the two locations that should limit and slow geneflow according to Wright's theory of Isolation by Distance (1943).

The overall picture of stickleback establishment in the Deschutes is therefore likely one of human-mediated transportation followed by natural, downstream establishment. Stickleback were likely introduced into the system initially in the 1980s as an illegal food supplementation for game fish (Yake, 2003), then spread downstream, establishing at ASP/UPD and eventually at CLF over the next ~30-40 years. During that period, they were subsequently introduced into PAL, likely from the same source population (likely the Upper Willamette River) as into the Crane Prairie Reservoir given their close relative genetic relatedness compared to outgroups (Catchen, Bassham, et al., 2013), and to the Crooked River upstream of SMR from a source somewhere high upstream in the Deschutes. They then naturally dispersed downstream to OPL. Reverse migration of Crooked River stickleback to CLF via the confluence of the Crooked River with the Deschutes is likely prohibited by Cline Falls itself. Despite this mosaic history of introductions and dispersal, the population is still very closely related on average, as indicated by the low-overall pairwise-F_{ST} scores between locations, the relative lack of differentiation by STRUCTURE, and the tight correlation between SNP diversity across the genome between all locations.

Extremely Rapid Chromosomal Inversion Mediated Repeated Adaptation:

We found convincing evidence to indicate that a large region of LG (chromosome) IX, likely a chromosomal inversion, has been selected up to dramatically increased frequencies on at least two separate occasions in Deschutes River stickleback. No previous work has found evidence for an inversion in the latter part of LG IX.

There are several lines of evidence that are highly suggestive that a large, novel chromosomal inversion (~3.78mb, approximately 18% of LG IX) is present at a moderately high frequency in stickleback at several of our sites. Most conclusively, we found that pairwise LD (as measured by r^2) was substantially elevated (+0.17-0.21) between SNPs in this region in SMR, OPL, and CLF, especially between SNPs which had substantial OPL/PAL pairwise-F_{ST} values. Pairwise F_{ST} dropped sharply outside this region to background levels. Regions of cleanly defined LD have previously been taken as an indicator of an inversion in a population, given that inversions act to maintain sets of alleles by resisting recombination and thus cause co-inheritance and increased LD, particularly towards their edges (Roesti et al., 2015; Sturtevant & Beadle, 1936).

This region of elevated LD coincided with a drastic increase in F_{ST} between CLF/OPL and ASP/UPD as well as between OPL and SMR, and a slight increase in F_{ST} between SMR and ASP/UPD. F_{ST} was also high in this region between OPL/SMR/CLF and PAL (Figure 2). At the boundaries of the region, F_{ST} in these groups drops off very quickly rather than gradually decaying, as we would expect if the region was instead undergoing a traditional selectional sweep (Pokalyuk, 2012). This is consistent, however,

with changes in the frequency of an inversion, which would cause consistent levels of F_{ST} increase across the length of the inversion.

Plotting the frequencies of the major (f_p) and minor (f_q) alleles, diversity (π) and heterozygosity (H_o) across the region also provides clear evidence of an inversion. Both π and H_o are distinctly segregated between sampling locations for divergent LG IX SNPs (OPL > CLF > SMR > UPD > ASP/PAL), and are nearly horizontally linear with respect to genomic position, as would be expected for co-inherited SNPs on an inversion found at different frequencies in different sub-populations (Figures 5a and 45). Non-divergent SNPs in the same genomic region, conversely, are completely randomly distributed for π and H_o with respect to sampling location and genomic position, as would be expected of background, neutral SNPs with little population structuring. f_q is also nearly constant across the length of the region for divergent SNPs and similarly segregated by sampling location.

The linearity of f_q across the inversion also allows for the frequency of the inversion itself (f_I) to be estimated in each population. The frequency of the inversion in a population should be equal to the f_q in instances where SNP alleles are different and fixed between the inversion and the typical version of the genome (for example, A in the inversion, C in the typical genome). We observed that f_q was part very low (zero or near to zero) for divergent but not for non-divergent SNPs in ASP, UPD, and PAL, with a handful of exceptions. If we assume that f_I is extremely low or zero in the samples we collected from PAL and ASP, where π and H_o are lowest (at or near zero) for divergent SNPs and pairwise-r^2 amongst SNPs in the region of the inversion are not elevated, it

would follow that divergent SNPs for which f_q is near zero in the region of the inversion are likely fixed differences between the inversion and the typical version of the genome. The other divergent SNPs are likely instances where the major allele in the inversion is present as a rare alleles on the inversion or vice versa. Mean f_q for these SNPs should therefore be a reasonable estimator of f_I.

We observed striking differences in estimated f_I between even closely related and nearby sampling locations. f_I was zero in ASP and PAL, indicating that none of the stickleback we collected and sequenced from those locations carried the inversion. We sampled a single individual from UPD which appeared to be a heterozygote for the inversion (nearly all divergent SNPs had a single observation of *q*), out of 49 sequenced individuals after filtering. In UPD and ASP, which are nearly genetically identical, the LG IX inversion is therefore clearly present at an extremely low frequency. It seems reasonable to assume, therefore, that the inversion itself was a very rare allele in the founding Deschutes/PAL population, or occurred early after establishment. We observed a dramatically higher f_I at SMR, CLF, and OPL (0.092, 0.20, and 0.32, respectively, as given above).

These changes in f_I across the Deschutes/Crooked river systems are evidence of extremely rapid, parallel selection. We observed very high relative Δf_I between SMR, CLF, and OPL and ASP/UPD, our proxy for the baseline f_I (1,784-6,482% increase). Our estimates of the relative rate of increase in f_I are therefore also large (up to 12.7%). These numbers are likely underestimates, given that stickleback must have taken some time to establish in these areas. We also found a large relative increase in f_I between SMR and

OPL (250%), for a Δf_I per generation of between 13.3% and 8.7%, despite their proximity (less than 34 river kilometers apart) and likely extensive gene-flow. These rates of adaptation are slower than those reported for the marine-freshwater transition (Barrett *et al* (2008) estimated a selection coefficient of 0.52 for a SNP associated with the transition, while Terekhanova *et al* (2014) estimated a coefficient of 0.27). The *S* values that we estimated for selection on the inversion were considerably smaller (0.018 to 0.119) were considerably smaller, but still substantial given that any environmental drivers of this selection are certainly less drastic than the radical environmental shift between the freshwater and pelagic ecosystems. The *S* values we calculated for OPL and CLF, in particular, are still quite large when compared to the typical *S* values found by meta-analyses on both animal and plant QTEs in experimental but not artificially selected populations (between 0.01 and 0.065 for animals, 0.11 on average in plants (Morjan & Rieseberg, 2004; Rieseberg & Burke, 2001)). Individual studies have noted a wider range of selection coefficients in animals, such as 0.078–0.103 for a selective sweep and 0.0001 to 0.01 for adaptation following range expansion in *Drosophila melanogaster* (Harr et al., 2002; Sáez et al., 2003). The latter study in particular is similar in context to ours, but reported much lower *S* values. Comparatively the *S* values we observe constitute evolution on an extremely rapid time-scale, and is a testament to the ability of stickleback to adapt via standing genetic variation and may help explain their success as an invasive species.

Causes of Selection:

The selection on the LG IX inversion that has occurred is likely not in response to temperature, as we had predicted. OPL and CLF have the coldest and warmest summer water temperatures, yet have the two highest f_I. Temperature at other times of the year is also unlikely to explain selection on the inversion, since OPL stays at a roughly constant temperature due to spring inputs, whereas similar inputs to CLF are not significant enough in comparison to the larger Deschutes River flow to maintain constant temperatures. Indeed, although it is possible that temperature at some other period during the year may be driving selection, OPL and CLF are, in many ways, probably the least similar of our sampling sites.

However, the geology at both of these sites is dominated by the recently deposited Deschutes volcanic formation, which is extremely permeable (Lite Jr & Gannett, 2002). SMR, although nearby to OPL, is just off of the Deschutes formation. Since the overall direction of groundwater flow is northward in the region, groundwater in the region around OPL, CLF, and, to a lesser extent SMR, is as a result much higher in total dissolved solids and salts than is groundwater located near ASP and UPD (Caldwell, 1998). This undoubtedly causes drastic change in water chemistry at OPL and CLF, since the majority of river flow from below Bend to Lake Billy Chinook on the Deschutes and in the lower reaches of the Crooked is dominated by groundwater inputs from April to December (Caldwell, 1998). CLF and SMR, which are farther upstream, is likely less impacted by groundwater/spring inputs than OPL.

Dissolved-solid rich and saline groundwater inputs are a likely driver of the drastic change f_I between our sampling sites. Previous studies have found total dissolved

solid (TDS) concentrations to be a strong predictor of fish yield and decreases in fish community health in streams and lakes (Hanson & Leggett, 1982; Meador & Goldstein, 2003), so physiological impacts on the stickleback in the regions of the Deschutes with large inputs of TDS and mineral rich water is likely. We did observe some direct evidence of this, since the only gene ontology (GO) groups strongly associated with pairwise-F_{ST} between SMR and OPL were the redundant coated membrane and membrane coat protein groups, one of which (membrane coat, GO:0030117) has been previously identified as responding to changes in water salinity and quality in the diatom *Phaeodactylum tricornutum* (De Martino et al., 2011). Five genes in the GO group are located on the observed LG IX inversion. This is circumstantial evidence and speculation at best, and more work is needed to verify the proximate cause of the changes in f_I across locations in Deschutes stickleback are needed.

Conclusion:

We found evidence that a novel chromosomal inversion on LG IX has been selected to a high frequency in stickleback in two portions of the Deschutes River system since their introduction between 30-40 years ago. The increase in frequency is quite dramatic in the region of the Crooked River near Opal Springs (+6,482%) and in the Deschutes River near Cline Falls (+3,989%). We also observed a large increase in frequency between Opal Springs and stickleback from Smith Rocks State Park, which is located approximately 34 river kilometers upstream (+248%). The selection coefficients that we estimated from these increases were as high as 0.119, which, while lower than that reported for freshwater-saltwater transition genes (such as armor plate genes), is still quite high, and may help explain the ability of stickleback to adapt to novel environments by leveraging alleles at a low-frequency in the population. We suggest that the proximate driver of selection at the inversion may be changes in water chemistry due to the influx of solute-heavy spring water lower in the system upstream of Opal Springs and Cline Falls, but more work is needed to confirm this or develop alternate explanations.

If similar standing variation exists in stickleback populations in the Deschutes and elsewhere, it is likely that stickleback are readily capable of invading other systems, and likely have a high capacity to adapt to anthropogenic environmental changes. Given the high *S* values observed in other stickleback populations undergoing a transition from fresh to saltwater or vice-versa, it seems likely that this is indeed the case. However, it is entirely possible that further studies on other fish species will reveal a similar capacity for adaptation, which may serve to help those species survive environmental changes. Given

the increasing degree of anthropogenic habitat modification occurring worldwide, more studies of this sort are urgently needed.

Tables:

Population	SMR	CLF	ASP	UPD	PAL
OPL	0.011	0.044	0.034	0.044	0.104
SMR		0.032	0.026	0.023	0.089
CLF			0.044	0.035	0.125
ASP				0.004	0.094
UPD					0.096

Population	SMR	CLF	ASP	UPD	PAL
OPL	0.011	0.044	0.034	0.044	0.104
SMR		0.032	0.026	0.023	0.089
CLF			0.044	0.035	0.125
ASP				0.004	0.094
UPD					0.096

Table 2: Mean π, H_o, and P_A/n, as well as local LD as measured by both r^2 and D' across all SNPs in each population.

Population	π	H_o	P_A/n	r^2	D'
OPL	0.262	0.296	0.699	0.015	0.184
SMR	0.260	0.281	4.462	0.014	0.204
CLF	0.259	0.278	0.578	0.022	0.218
ASP	0.273	0.300	1.620	0.015	0.198
UPD	0.273	0.330	1.571	0.016	0.202
PAL	0.204	0.222	0.804	0.043	0.223

Figures:

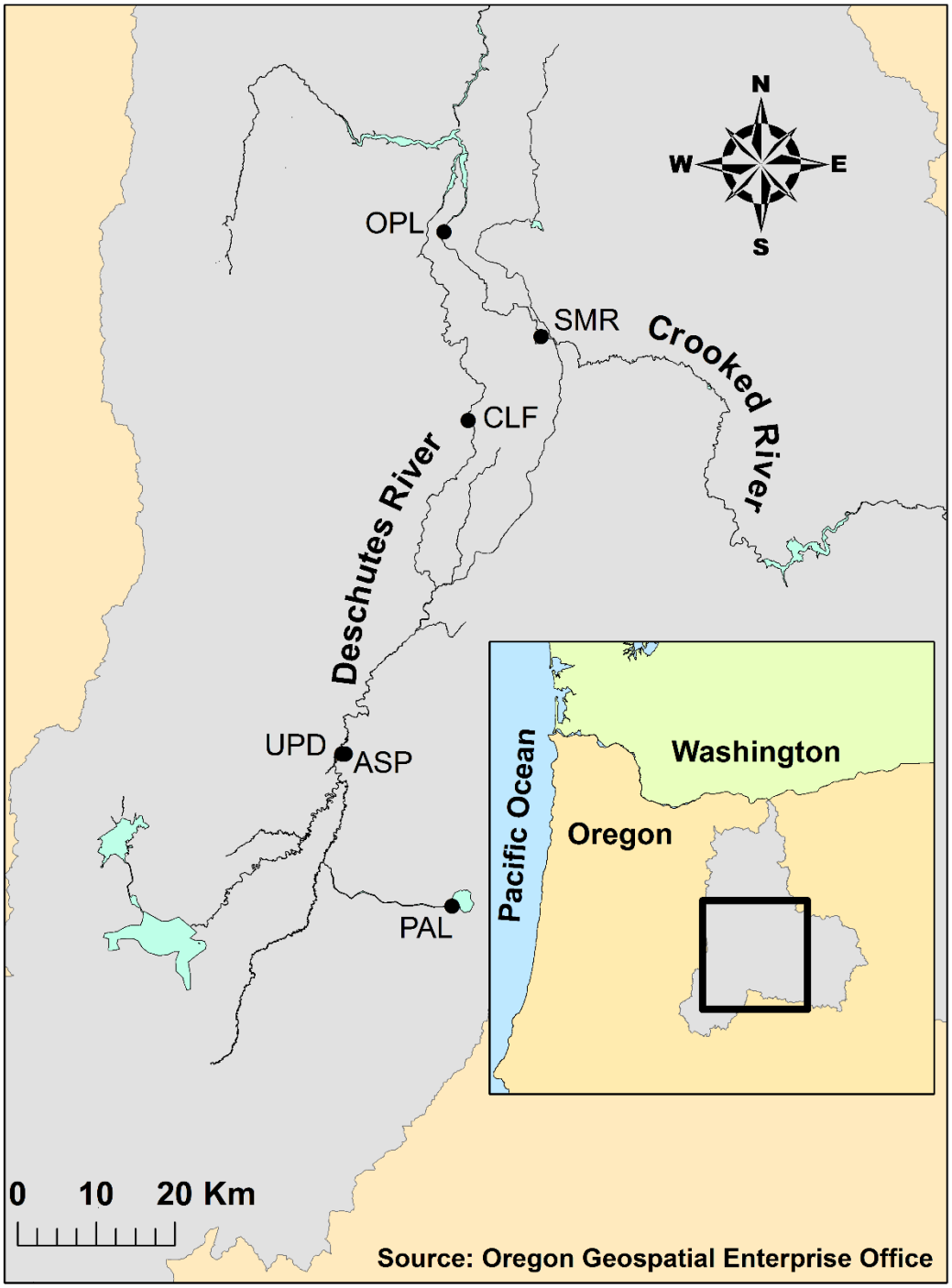

Figure 1: Map of the Deschutes and Crooked Rivers, Oregon, with sampling sites labeled.

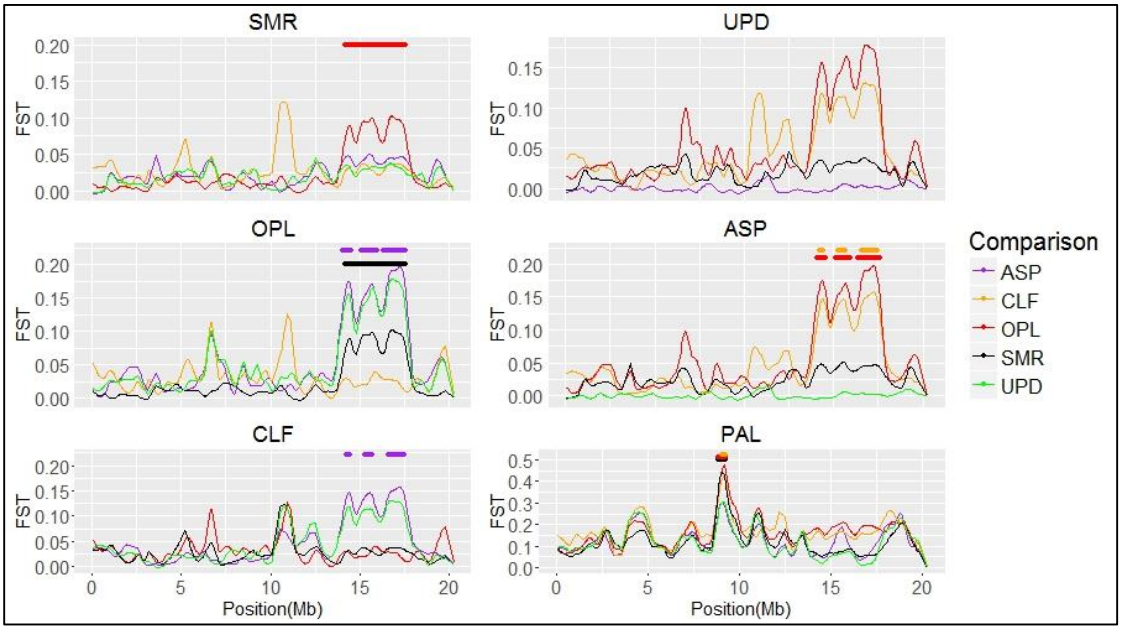

Figure 2: Pairwise F_{ST} vs. every other population across LG IX for each population. Dots above lines denote smoothed values that are significantly elevated vs. the genome wide average at $p < 0.00005$.

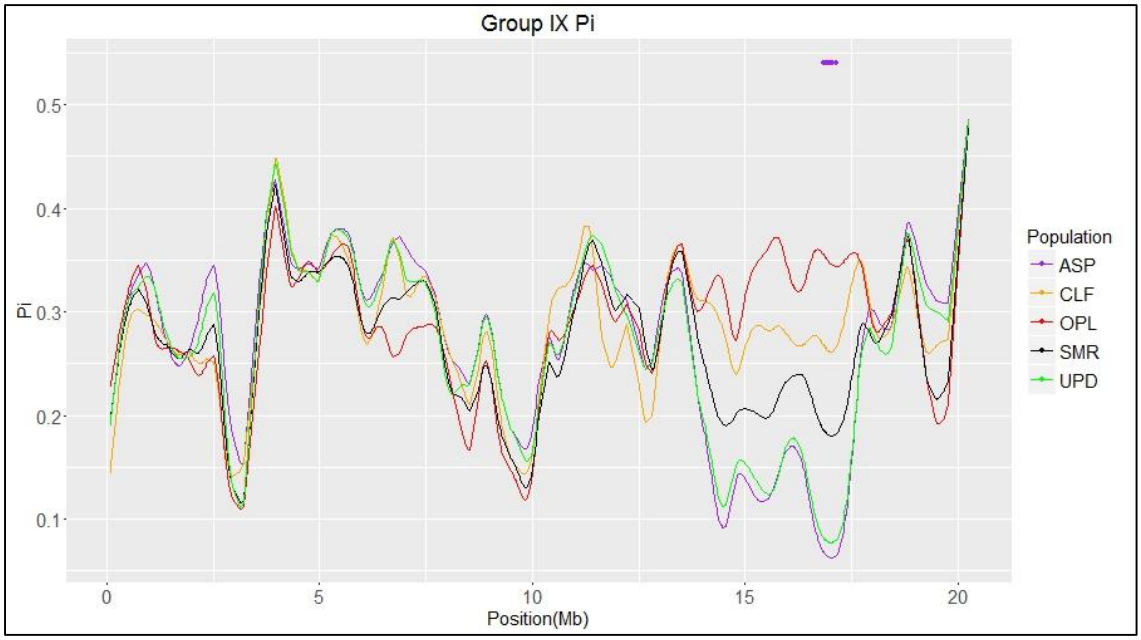

Figure 3: π across LG IX in every population. Dots above line indicate regions where smoothed π values are significantly depressed vs. the genome wide average at $p < 0.00005$.

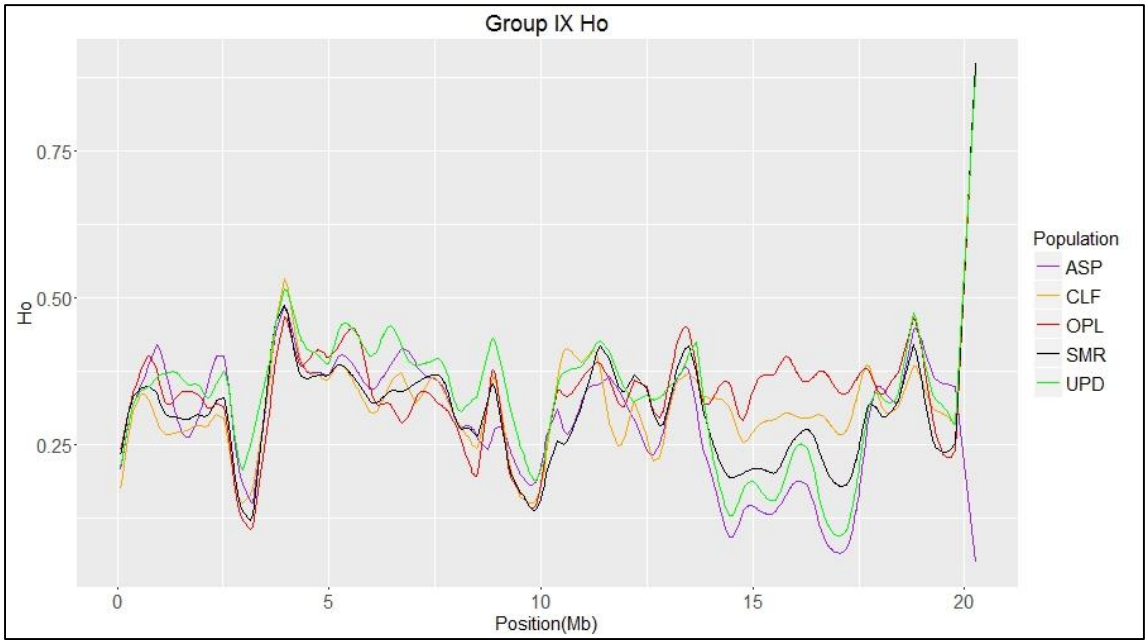

Figure 4: Ho across LG IX in every population.

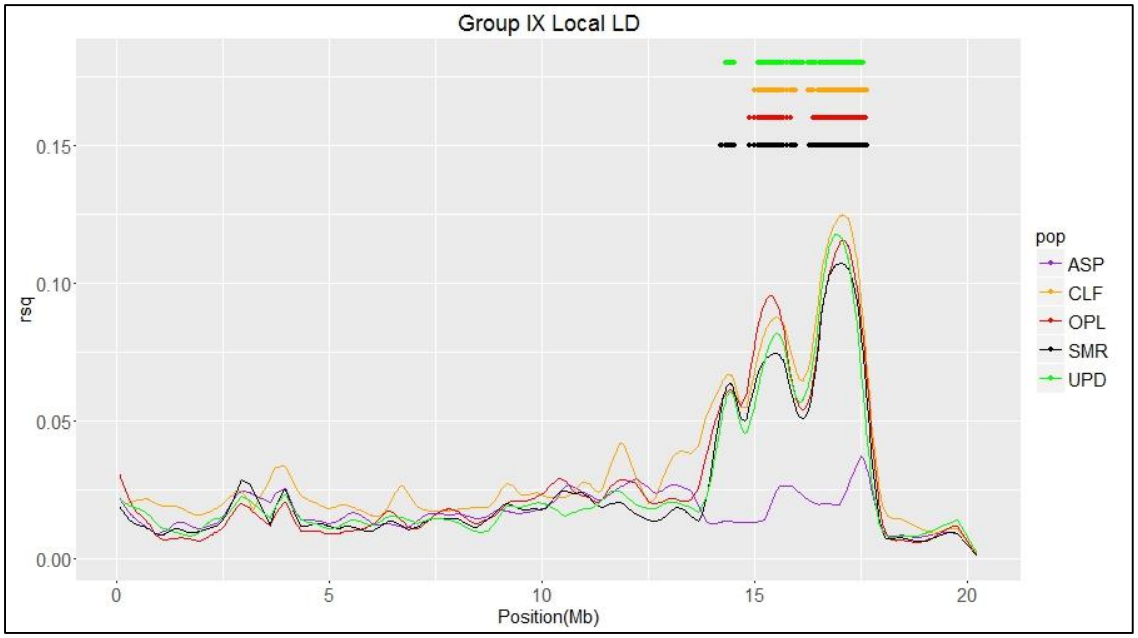

Figure 4: Local LD as measured by r^2 on LG IX for each population. Dots above lines indicate regions where smoothed local r^2 values are elevated vs. the genome wide average at $p < 0.00005$.

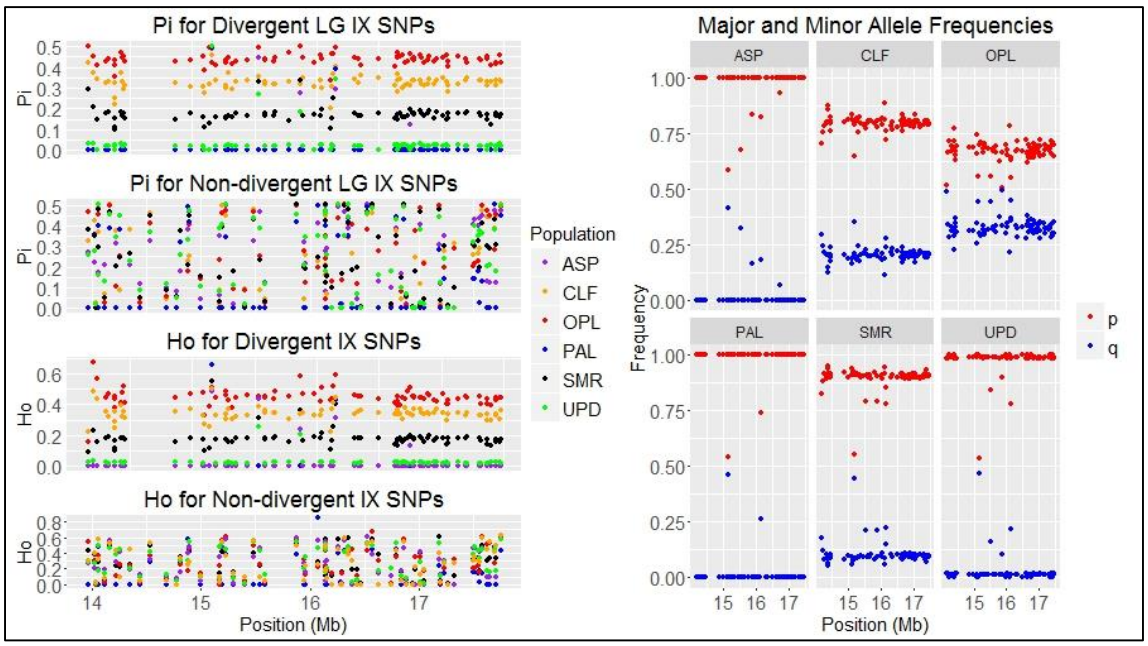

Figure 6: π, H_o, and major and minor allele frequencies in the region of LG IX containing the inversion. Left side, from top to bottom: (a) π in divergent SNPs, (b) π in non- divergent SNPs, (c) H_o in divergent SNPs, (d) H_o in non- divergent. Right side: (e) major and minor allele frequencies of divergent SNPs for each population.

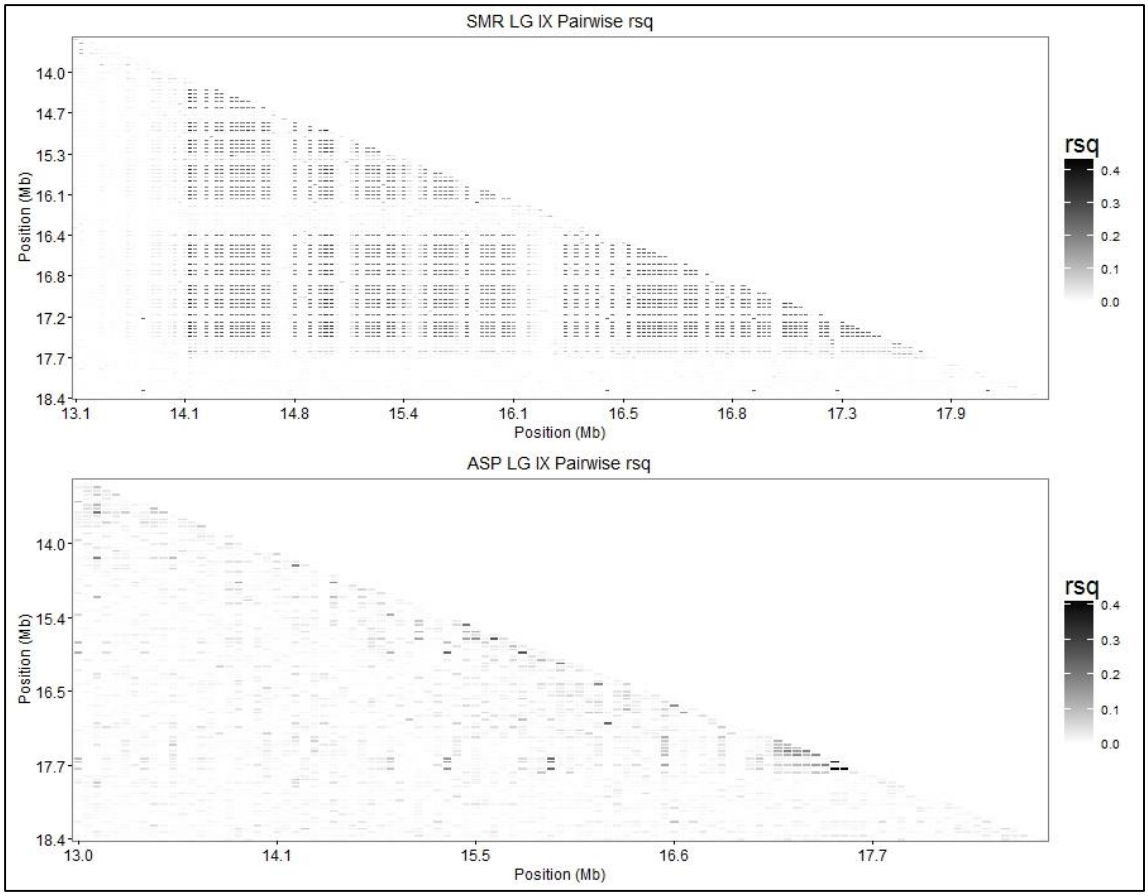

Figure 7: Heatmaps of the r^2 measure of LDfor every pairwise comparison between 13 and 18.5mb on LG IX. (a, top) SMR and (b, bottom) ASP.

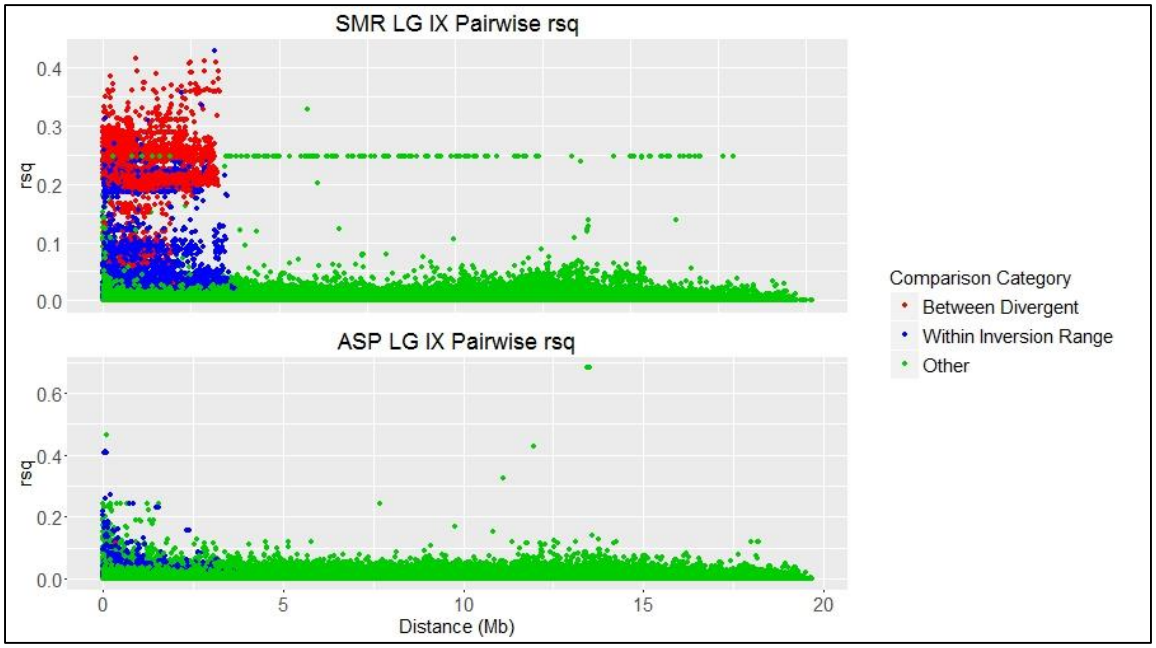

Figure 8: LD as measured by r^2 for each pairwise comparison between SNPs on LG IX vs. proximity in mb between the compared SNPs. SMR (top, a), and ASP (bottom, b). Points in red are comparisons between two divergent SNPs and points in blue are comparisons between non-divergent SNPs in the range of the inversion. All other comparisons are in green.

Bibliography:

Barnosky, A. D., Matzke, N., Tomiya, S., Wogan, G. O., Swartz, B., Quental, T. B., Marshall, C., McGuire, J. L., Lindsey, E. L., & Maguire, K. C. (2011). Has the Earth/'s sixth mass extinction already arrived? *Nature*, *471*(7336), 51-57.

Barrett, R. D., Paccard, A., Healy, T. M., Bergek, S., Schulte, P. M., Schluter, D., & Rogers, S. M. (2011). Rapid evolution of cold tolerance in stickleback. *Proceedings of the Royal Society of London B: Biological Sciences*, *278*(1703), 233-238.

Barrett, R. D., Rogers, S. M., & Schluter, D. (2008). Natural selection on a major armor gene in threespine stickleback. *Science*, *322*(5899), 255-257.

Bell, M. A., & Aguirre, W. E. (2013). Contemporary evolution, allelic recycling, and adaptive radiation of the threespine stickleback. *Evolutionary Ecology Research*, *15*(4), 377-411.

Caldwell, R. R. (1998). *Chemical study of regional ground-water flow and ground-water/surface-water interaction in the upper Deschutes Basin, Oregon*. Retrieved from

Catchen, J., Bassham, S., Wilson, T., Currey, M., O'Brien, C., Yeates, Q., & Cresko, W. A. (2013). The population structure and recent colonization history of Oregon threespine stickleback determined using restriction-site associated DNA-sequencing. *Molecular ecology*, *22*(11), 2864-2883.

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular ecology*, *22*(11), 3124-3140.

Da Cunha, A. B., Burla, H., & Dobzhansky, T. (1950). Adaptive chromosomal polymorphism in Drosophila willistoni. *evolution*, 212-235.

De Martino, A., Bartual, A., Willis, A., Meichenin, A., Villazán, B., Maheswari, U., & Bowler, C. (2011). Physiological and molecular evidence that environmental changes elicit morphological interconversion in the model diatom Phaeodactylum tricornutum. *Protist*, *162*(3), 462-481.

Dobzhansky, T., & Sturtevant, A. H. (1938). Inversions in the chromosomes of Drosophila pseudoobscura. *Genetics*, *23*(1), 28.

Dubinin, N. P., & Tiniakov, G. (1946). Natural selection and chromosomal variability in populations of Drosophila funebris. *Journal of heredity*, *37*(2), 39-44.

Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, *180*(2), 977-993.

Franks, S. J., & Hoffmann, A. A. (2012). Genetics of climate change adaptation. *Genetics*, *46*.

Gelmond, O., Von Hippel, F., & Christy, M. (2009). Rapid ecological speciation in three-spined stickleback Gasterosteus aculeatus from Middleton Island, Alaska: the roles of selection and geographic isolation. *Journal of fish biology*, *75*(8), 2037-2051.

Gienapp, P., Lof, M., Reed, T. E., McNamara, J., Verhulst, S., & Visser, M. E. (2013). Predicting demographically sustainable rates of adaptation: can great tit breeding time keep pace with climate change? *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 368*(1610), 20120289.

Gillespie, J. H. (2010). *Population genetics: a concise guide*: JHU Press.

Hanson, J. M., & Leggett, W. C. (1982). Empirical prediction of fish biomass and yield. *Canadian Journal of Fisheries and Aquatic Sciences, 39*(2), 257-263.

Harr, B., Kauer, M., & Schlötterer, C. (2002). Hitchhiking mapping: a population-based fine-mapping strategy for adaptive mutations in Drosophila melanogaster. *Proceedings of the National Academy of Sciences, 99*(20), 12949-12954.

Hill, W. G., & Robertson, A. (1966). The effect of linkage on limits to artificial selection. *Genetical research, 8*(03), 269-294.

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 Genet, 6*(2), e1000862.

Isaak, D., Wenger, S., Peterson, E., Ver Hoef, J., Hostetler, S., Luce, C., Dunham, J., Kershner, J., Roper, B., & Nagel, D. (2011). NorWeST: an interagency stream temperature database and model for the Northwest United States. *US Fish and Wildlife Service, Great Northern Landscape Conservation Cooperative Grant. Project website: www. fs. fed. us/rm/boise/AWAE/projects/NorWeST. html*.

Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., Swofford, R., Pirun, M., Zody, M. C., & White, S. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature, 484*(7392), 55-61.

Karell, P., Ahola, K., Karstinen, T., Valkama, J., & Brommer, J. E. (2011). Climate change drives microevolution in a wild bird. *Nature Communications, 2*, 208.

Kingsley, D. M., Zhu, B., Osoegawa, K., De Jong, P. J., Schein, J., Marra, M., Peichel, C., Amemiya, C., Schluter, D., & Balabhadra, S. (2004). New genomic tools for molecular studies of evolutionary change in threespine sticklebacks. *Behaviour, 141*(11), 1331-1344.

Kirkpatrick, M., & Barton, N. (2006). Chromosome inversions, local adaptation and speciation. *Genetics, 173*(1), 419-434.

Koketsu, W. (2004). Environmental correlates of parasitism in introduced threespine stickleback (Gasterosteus aculeatus) in the Upper Deschutes River Basin, Oregon.

Kopp, M., & Matuszewski, S. (2014). Rapid evolution of quantitative traits: theoretical perspectives. *Evolutionary Applications, 7*(1), 169-191.

Krimbas, C. B. (1964). The genetics of Drosophila subobscura populations. I. Inversion polymorphism in populations of southern Greece. *evolution*, 541-552.

Lee, H. K., Braynen, W., Keshav, K., & Pavlidis, P. (2005). ErmineJ: tool for functional analysis of gene expression data sets. *BMC bioinformatics, 6*(1), 1.

Lewontin, R. (1964). The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics, 49*(1), 49-67.

Lite Jr, K. E., & Gannett, M. W. (2002). *Geologic framework of the regional ground-water flow system in the upper Deschutes Basin, Oregon*. Retrieved from

Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982). *Molecular cloning: a laboratory manual* (Vol. 545): Cold Spring Harbor Laboratory Cold Spring Harbor, NY.

Meador, M. R., & Goldstein, R. M. (2003). Assessing water quality at large geographic scales: relations among land use, water physicochemistry, riparian condition, and fish community structure. *Environmental Management, 31*(4), 0504-0517.

Morjan, C. L., & Rieseberg, L. H. (2004). How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. *Molecular ecology, 13*(6), 1341-1356.

Nielsen, R., Hubisz, M. J., Hellmann, I., Torgerson, D., Andrés, A. M., Albrechtsen, A., Gutenkunst, R., Adams, M. D., Cargill, M., & Boyko, A. (2009). Darwinian and demographic forces affecting human protein coding genes. *Genome research, 19*(5), 838-849.

Parmesan, C. (2006). Ecological and evolutionary responses to recent climate change. *Annual Review of Ecology, Evolution, and Systematics*, 637-669.

Pokalyuk, C. (2012). The effect of recurrent mutation on the linkage disequilibrium under a selective sweep. *Journal of mathematical biology, 64*(1-2), 291-317.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics, 155*(2), 945-959.

Raymond, M., & Rousset, F. (1995). GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of heredity, 86*(3), 248-249.

Rieseberg, L. H., & Burke, J. M. (2001). The biological reality of species: gene flow, selection, and collective evolution. *Taxon*, 47-67.

Rodríguez-Trelles, F., Alvarez, G., & Zapata, C. (1996). Time-series analysis of seasonal changes of the O inversion polymorphism of Drosophila subobscura. *Genetics, 142*(1), 179-187.

Roesti, M., Kueng, B., Moser, D., & Berner, D. (2015). The genomics of ecological vicariance in threespine stickleback fish. *Nature Communications, 6*.

Ross, J. A., & Peichel, C. L. (2008). Molecular cytogenetic evidence of rearrangements on the Y chromosome of the threespine stickleback fish. *Genetics, 179*(4), 2173-2182.

Sáez, A. G., Tatarenkov, A., Barrio, E., Becerra, N. H., & Ayala, F. J. (2003). Patterns of DNA sequence polymorphism at Sod vicinities in Drosophila melanogaster: unraveling the footprint of a recent selective sweep. *Proceedings of the National Academy of Sciences, 100*(4), 1793-1798.

Smedley, D., Haider, S., Durinck, S., Pandini, L., Provero, P., Allen, J., Arnaiz, O., Awedh, M. H., Baldock, R., Barbiera, G., Bardou, P., Beck, T., Blake, A., Bonierbale, M., Brookes, A. J., Bucci, G., Buetti, I., Burge, S., Cabau, C., Carlson, J. W., Chelala, C., Chrysostomou, C., Cittaro, D., Collin, O., Cordova, R., Cutts, R. J., Dassi, E., Genova, A. D., Djari, A., Esposito, A., Estrella, H., Eyras, E., Fernandez-Banet, J., Forbes, S., Free, R. C., Fujisawa, T., Gadaleta, E., Garcia-Manteiga, J. M., Goodstein, D., Gray, K., Guerra-Assunção, J. A., Haggarty, B., Han, D.-J., Han, B. W., Harris, T., Harshbarger, J., Hastings, R. K., Hayes, R. D., Hoede, C., Hu, S., Hu, Z.-L., Hutchins, L., Kan, Z., Kawaji, H., Keliet, A., Kerhornou, A., Kim, S., Kinsella, R., Klopp, C., Kong, L., Lawson, D.,

Lazarevic, D., Lee, J.-H., Letellier, T., Li, C.-Y., Lio, P., Liu, C.-J., Luo, J., Maass, A., Mariette, J., Maurel, T., Merella, S., Mohamed, A. M., Moreews, F., Nabihoudine, I., Ndegwa, N., Noirot, C., Perez-Llamas, C., Primig, M., Quattrone, A., Quesneville, H., Rambaldi, D., Reecy, J., Riba, M., Rosanoff, S., Saddiq, A. A., Salas, E., Sallou, O., Shepherd, R., Simon, R., Sperling, L., Spooner, W., Staines, D. M., Steinbach, D., Stone, K., Stupka, E., Teague, J. W., Dayem Ullah, A. Z., Wang, J., Ware, D., Wong-Erasmus, M., Youens-Clark, K., Zadissa, A., Zhang, S.-J., & Kasprzyk, A. (2015). The BioMart community portal: an innovative alternative to large, centralized data repositories. *Nucleic Acids Research, 43*(W1), W589-W598. doi:10.1093/nar/gkv350

Spiess, E. B. (1950). Experimental populations of Drosophila persimilis from an altitudinal transect of the Sierra Nevada. *evolution*, 14-33.

Sturtevant, A., & Beadle, G. (1936). The relations of inversions in the X chromosome of Drosophila melanogaster to crossing over and disjunction. *Genetics, 21*(5), 554.

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 Genet, 10*(10), e1004696.

Torgersen, C., Hockman-Wert, D., Batemen, D., & Gresswell, R. (2003). *Longitudinal Patterns of Fish Assemblages, Aquatic Habitat, and Water Temperature in the Lower Crooked River, Oregon*.

Wang, S., Meyer, E., McKay, J. K., & Matz, M. V. (2012). 2b-RAD: a simple and flexible method for genome-wide genotyping. *Nature methods, 9*(8), 808-810.

Wark, A. R., Mills, M. G., Dang, L.-H., Chan, Y. F., Jones, F. C., Brady, S. D., Absher, D. M., Grimwood, J., Schmutz, J., & Myers, R. M. (2012). Genetic architecture of variation in the lateral line sensory system of threespine sticklebacks. *G3: Genes| Genomes| Genetics, 2*(9), 1047-1056.

Watershed Sciences, I. (2002). *Aerial Surveys in the Deschutes River Basin: Thermal Infrared and Color Videography*

Watershed Sciences, I. (2006). *Airborne Thermal Infrared Remote Sensing Crooked River, OR*

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *evolution*, 1358-1370.

Wright, S. (1943). Isolation by distance. *Genetics, 28*(2), 114.

Yake, K. (2003). *Upper Deschutes Subbasin Assessment* Retrieved from

Yeaman, S. (2013). Genomic rearrangements and the evolution of clusters of locally adaptive loci. *Proceedings of the National Academy of Sciences, 110*(19), E1743-E1751.