Supplemental Information for:

Whole genome resequencing identifies local adaptation associated with environmental variation for redband trout

Kimberly R. Andrews, Travis Seaborn, Joshua P. Egan, Matthew W. Fagnan, Daniel D. New, Zhongqi Chen, Paul A. Hohenlohe, Lisette P. Waits, Christopher C. Caudill, Shawn R. Narum

Supplemental Text

Genome scans: GEA outliers

Environmental data for GEA analyses was obtained for a total of 26 climate and non-climate variables, including the 19 bioclimatic variables from WorldClim at a resolution of 30 arc seconds (Fick & Hijmans 2017), and seven variables from the NorWeST stream temperature data from 1993-2011 (Isaak *et al.* 2011): elevation, canopy cover, slope, precipitation, cumulative drainage, August flow, and mean August stream temperature. We then removed one of every pair of highly correlated variables (Spearman rank correlations >0.85, p < 0.05) using CARET v6.0-86 (Kuhn 2020) in R, resulting in a set of nine variables. One of the retained variables was mean annual air temperature from WorldClim, which was highly correlated with mean August stream temperature from NorWeST, which had been removed. Because stream temperature is more biologically relevant to trout, and because mean August stream temperature was not strongly correlated with any other retained variables, we replaced air temperature with mean August stream temperature to generate the final set of nine environmental variables (Table 2). Absolute values of final pairwise correlations of environmental variables ranged from 0.08 to 0.73 (Fig. S1).

For GEA analyses, temporal replicate samples were treated as a single population for each geographic site, with the exception of the two Mann Creek replicates, which were treated as separate populations. In contrast to the other sites, the Mann Creek replicates were collected from two separate tributaries. Therefore, treating these replicates as separate populations allowed the use of environmental data specific to each tributary. Thus, these analyses required separate allele frequency estimates for each Mann Creek tributary, rather than a single estimate for the two tributaries combined. To obtain these estimates, we ran the POOLPARTY "ppalign" and "ppanalyze" modules for all geographic sites as described in the main text, but treating each Mann Creek tributary as a separate site.

LFMM analyses were conducted accounting for population structure using K = 9 based on previous analyses of genetic structure for these populations (Kozfkay et al. 2011). These studies found Duncan Creek and Big Jacks Creek to cluster together; Mann Creek, Keithley Creek, and Little Weiser Creek to cluster together; and all other sites to cluster independently. K = 7 was used for the LFMM analyses conducted without the two Kootenai river watershed streams, which did not cluster together in previous studies.

MOLECULAR ECOLOGY

Literature Cited

Fick SE, Hijmans RJ (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* **37**, 4302-4315. https//doi.org/10.1002/joc.5086

Isaak DJ, Wenger SJ, Peterson EE*, et al.* (2011) NorWeST: an interagency stream temperature database and model for the Northwest United States. US Fish and Wildlife Service, Great Northern and North Pacific Landscape Conservation Cooperative grants. www.fs.fed.us/rm/boise/AWAE/projects/NorWeST.html

Kozfkay CC, Campbell MR, Meyer KA, Schill DJ (2011) Influences of habitat and hybridization on the genetic structure of redband trout in the upper Snake River Basin, Idaho. Transactions of the American Fisheries Society 140, 282-295. https//doi.org/10.1080/00028487.2011.567837

Kuhn M (2020) caret: Classification and Regression. Training. R package version 6.0-86. https://CRAN.R-project.org/package=caret.

Supplemental Figures
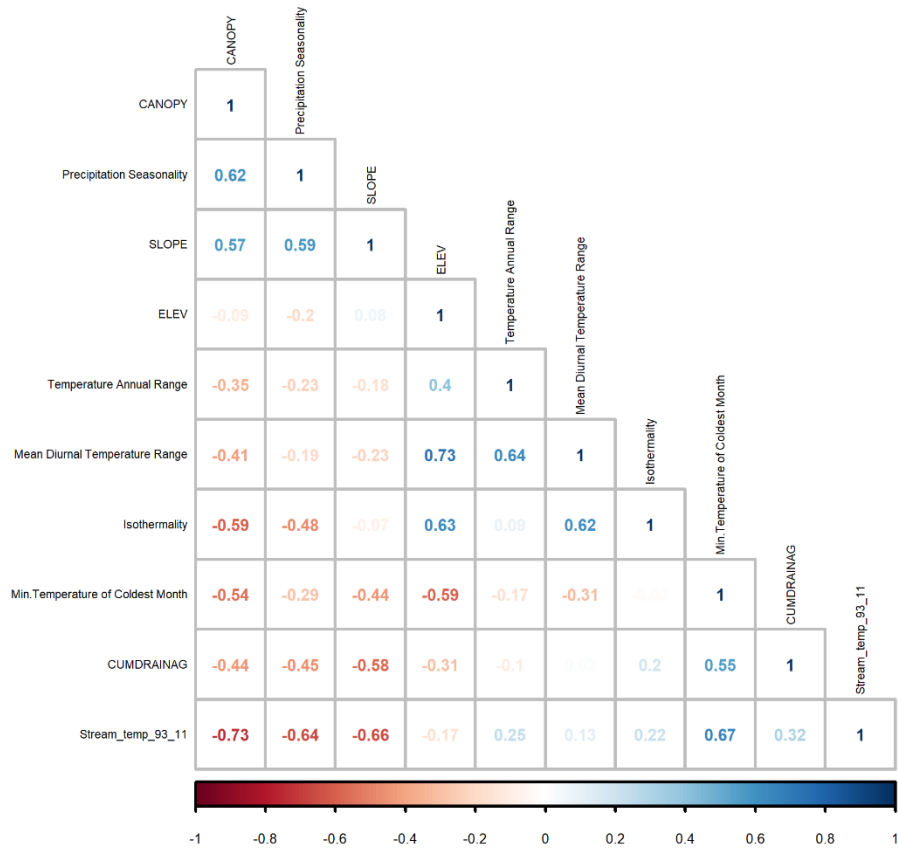
Figure S1. Spearman's rank correlation values for environmental variables included in the GEA's.
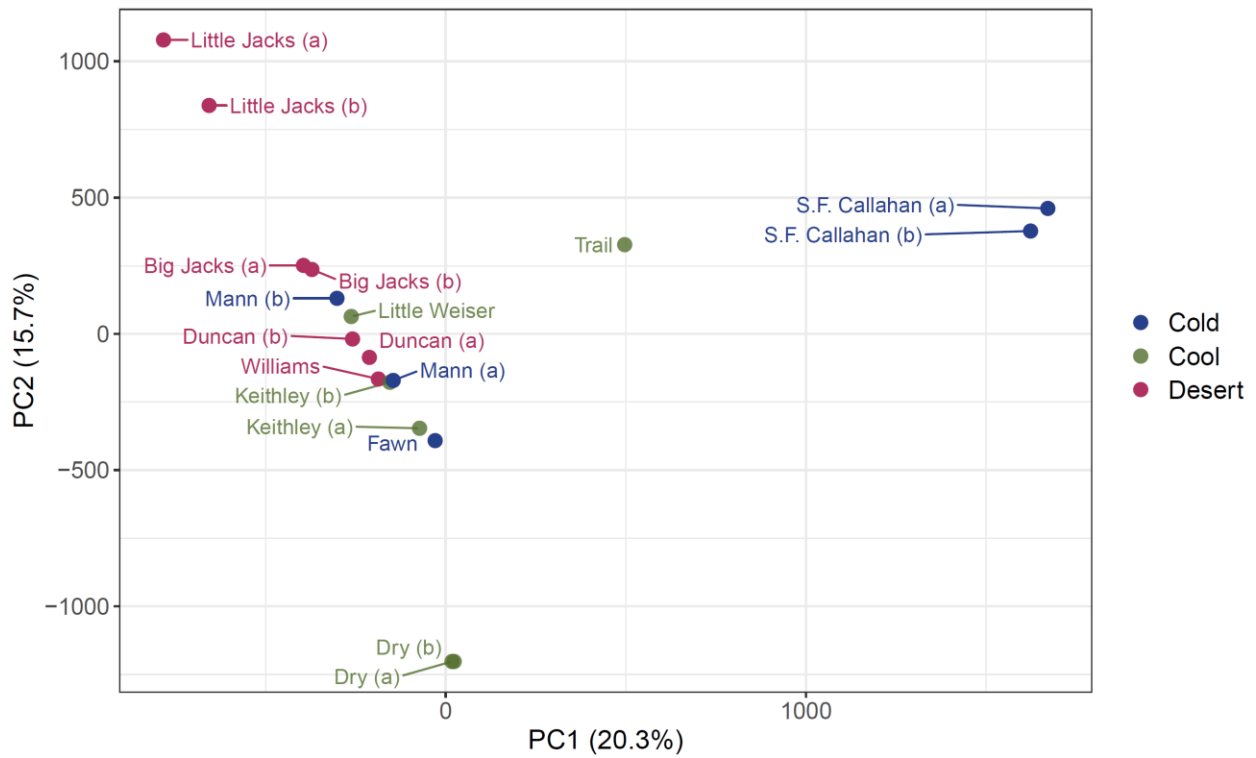
Figure S2. PCA with all genome-wide SNPs, treating each temporal replicate independently. Temporal replicates are indicated by "(a)" and "(b)." Sample sites are color-coded by ecotype: "Cold" = cold montane forest, "Cool" = cool montane forest, "Desert" = high elevation desert.
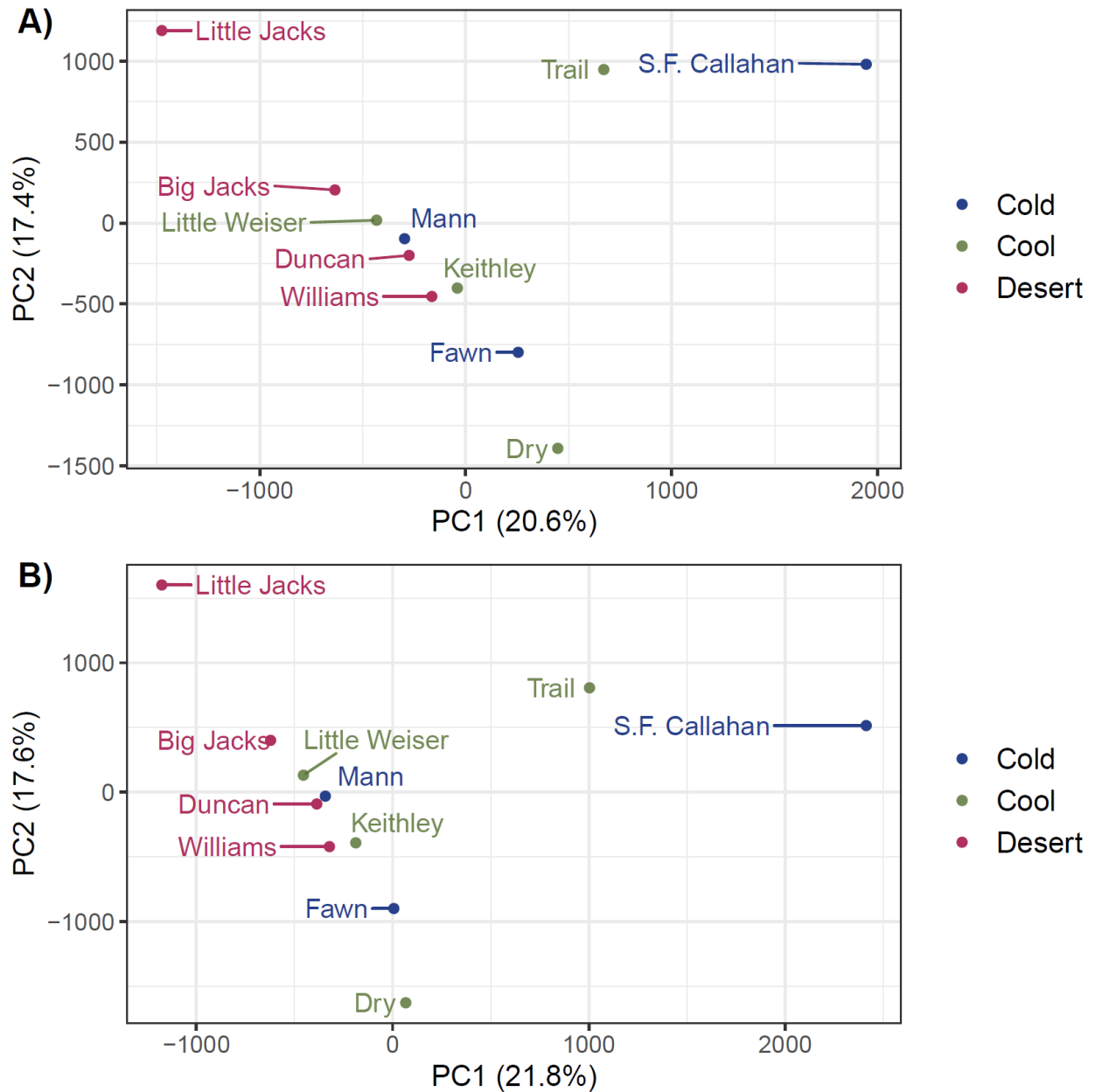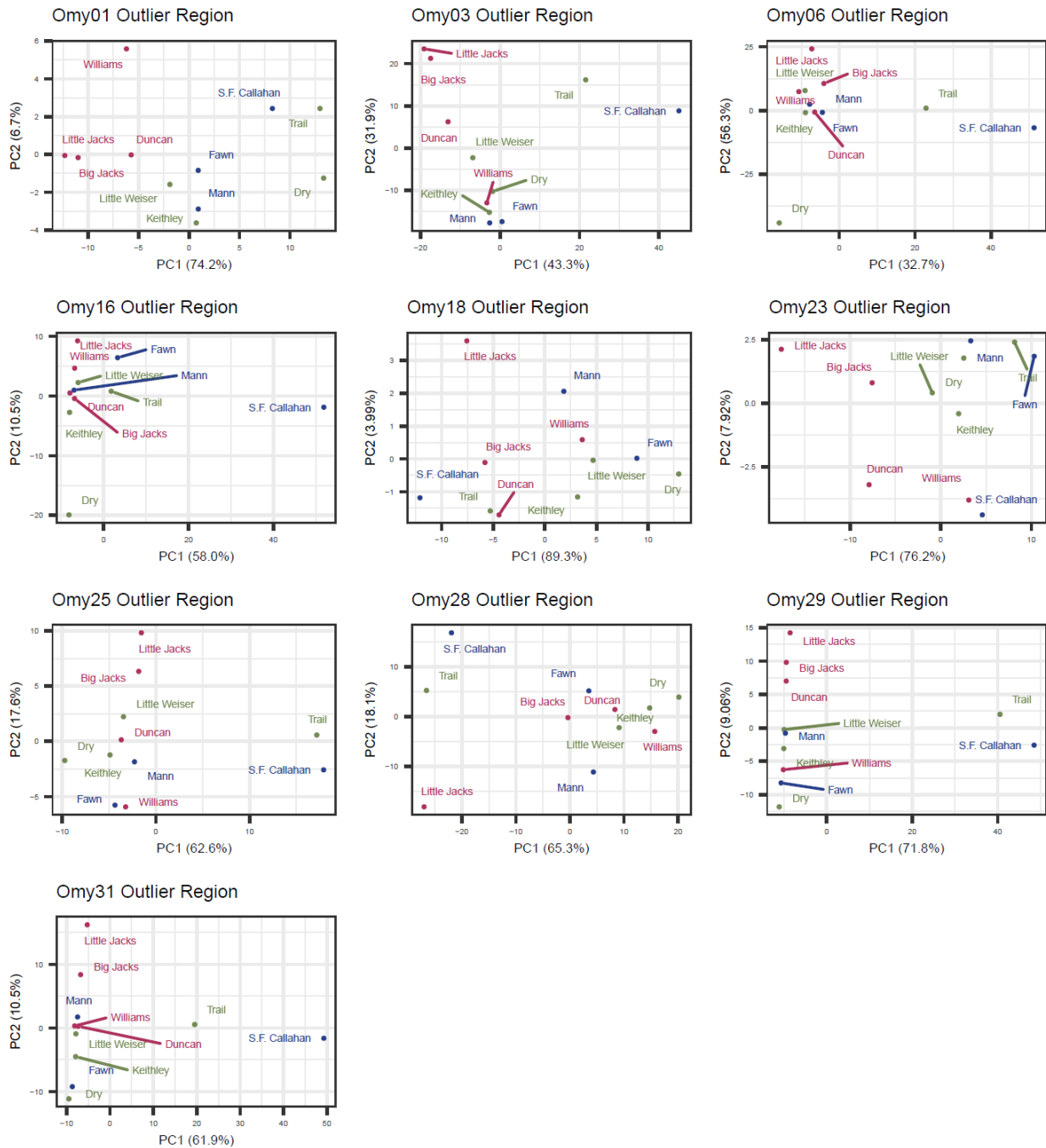
A)

B)

Figure S3: Population structure of putatively neutral SNPs. PCA results after A) removing SNPs having the top 10% of F_{ST} values, and B) removing SNPs that were outliers for the Local Scores and/or GEA tests. Sample sites are color-coded by ecotype: "Cold" = cold montane forest, "Cool" = cool montane forest, "Desert" = high elevation desert.

Figure S4. PCAs for SNPs occurring in each of the ten Local Score outlier regions identified in the full dataset analyses (including Snake River and Kootenai River populations). Numbers of SNPs in each region is shown in Table 3. Sample sites are color-coded by ecotype: Blue = cold montane forest, Green = cool montane forest, Red = high elevation desert.
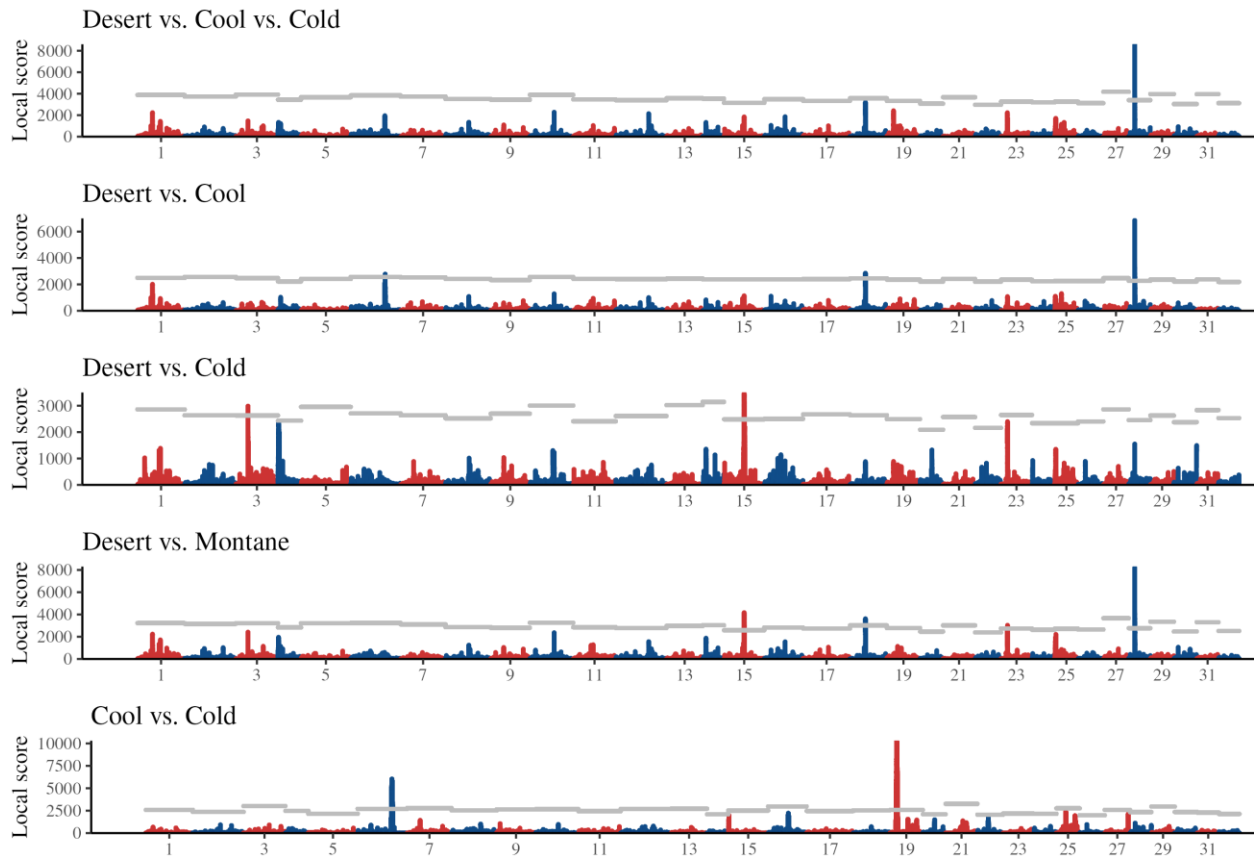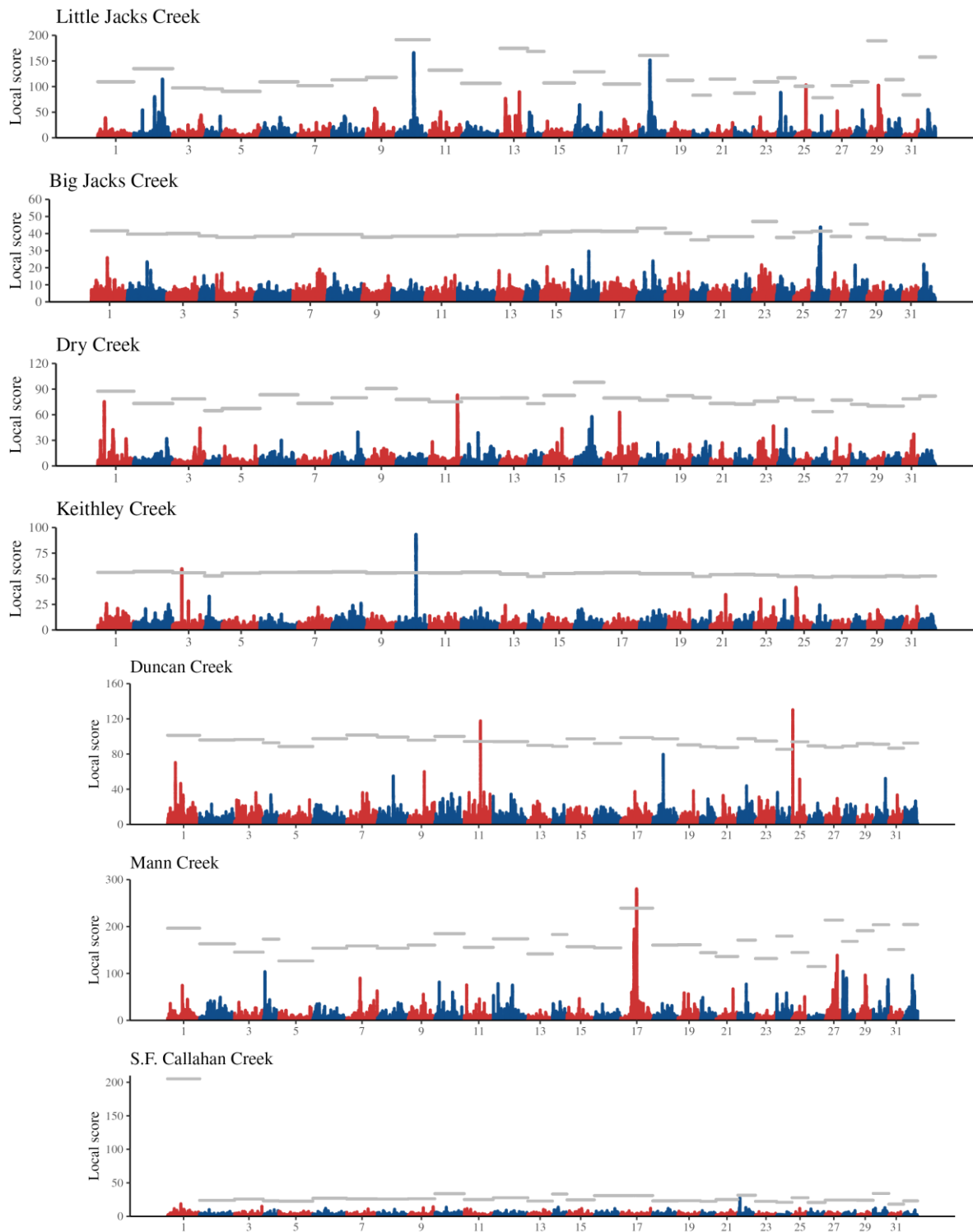
Figure S5. Manhattan plots of Local Scores across genome-wide SNPs for pairwise comparisons between habitat types in Snake River data subset analyses (excluding Kootenai River populations). Chromosome numbers are along the x-axis. Horizontal grey lines indicate average chromosome significance = 0.001 after correction for multiple tests.
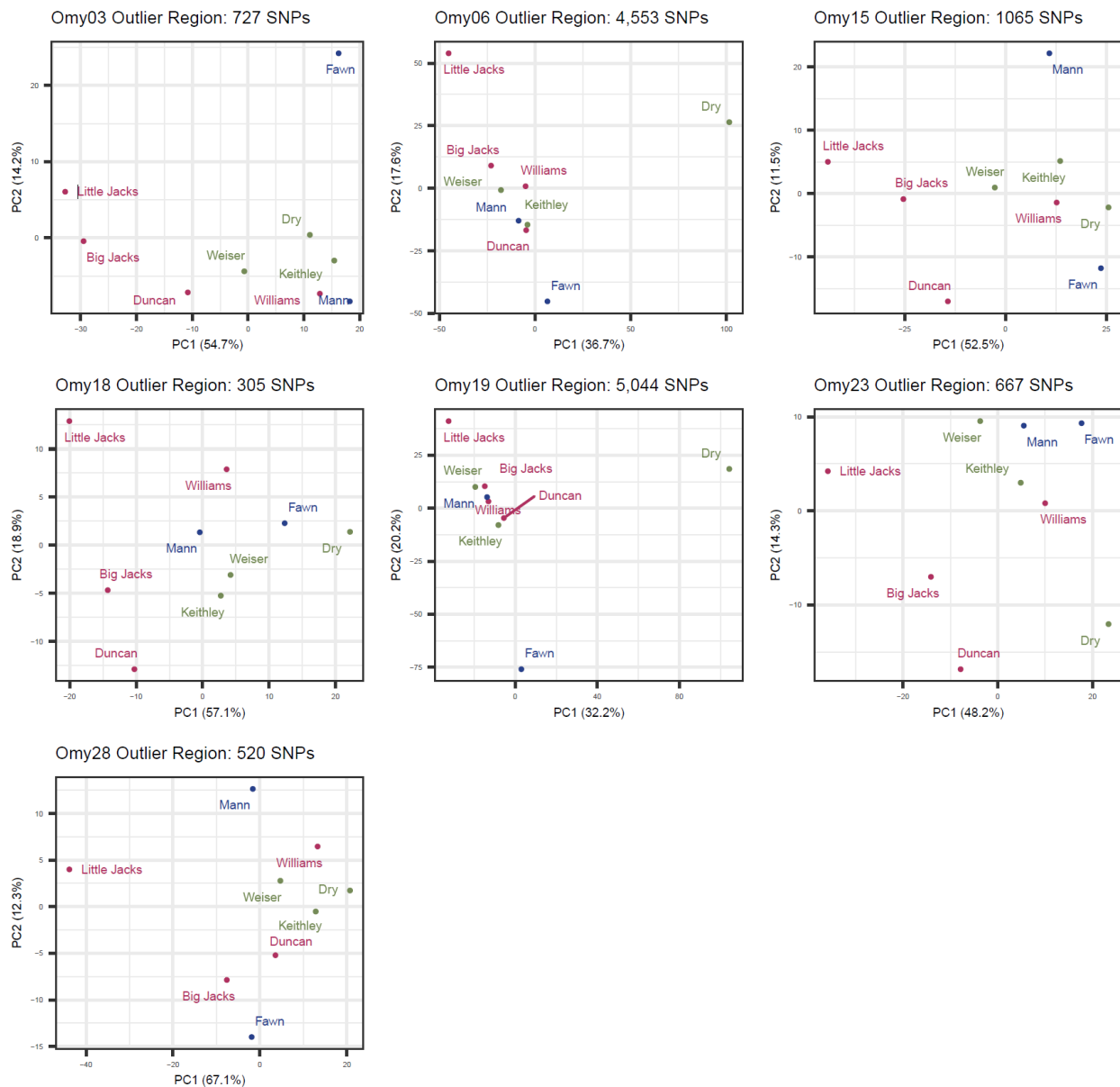
Figure S6. PCAs for SNPs occurring in each of the seven Local Score outlier regions identified by Snake River data subset analyses (excluding Kootenai River populations). Sample sites are color-coded by ecotype: Blue = cold montane forest, Green = cool montane forest, Red = high elevation desert. "Weiser" = "Little Weiser."

Figure S7. Manhattan plots of Local Scores across genome-wide SNPs for temporal replicate comparisons within populations. Chromosome numbers are along the x-axis. Horizontal grey lines indicate average chromosome significance = 0.001 after correction for multiple tests.

Omy01 Outlier Region: 30 SNPs
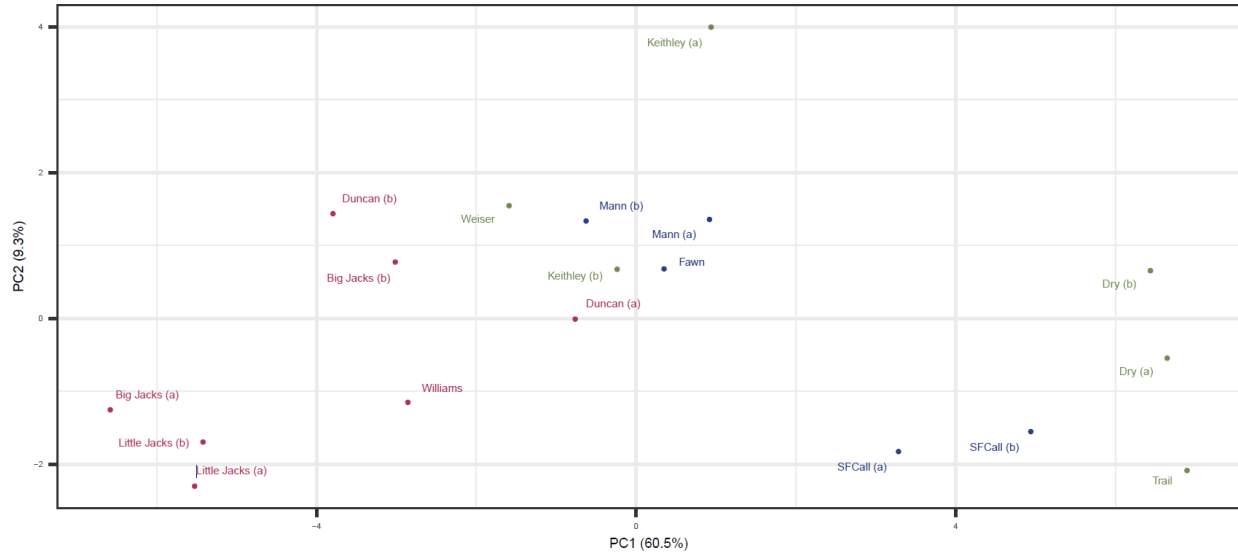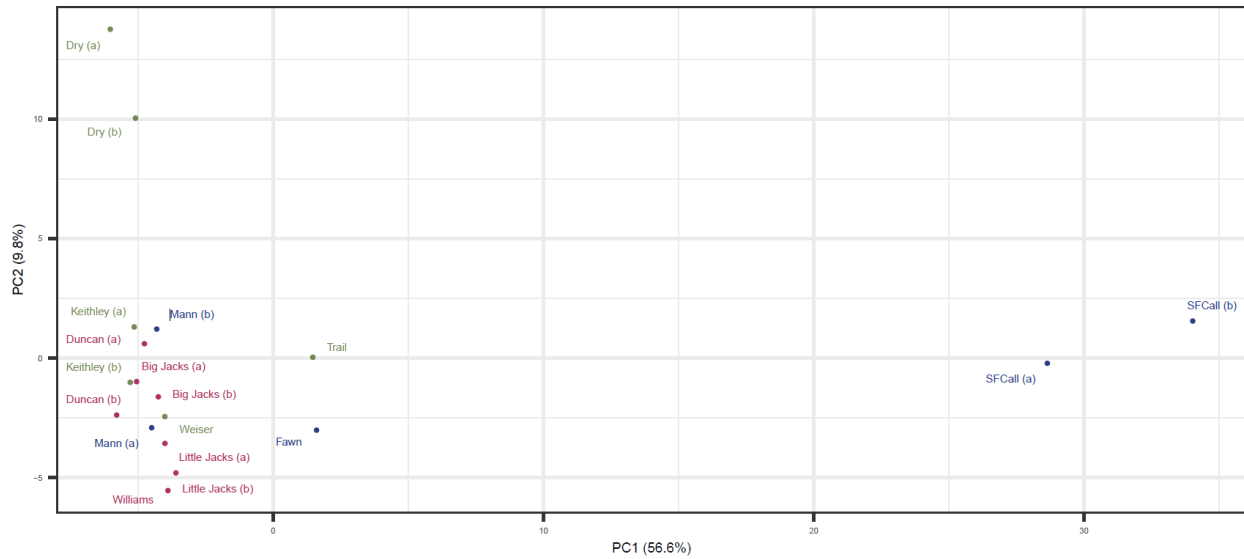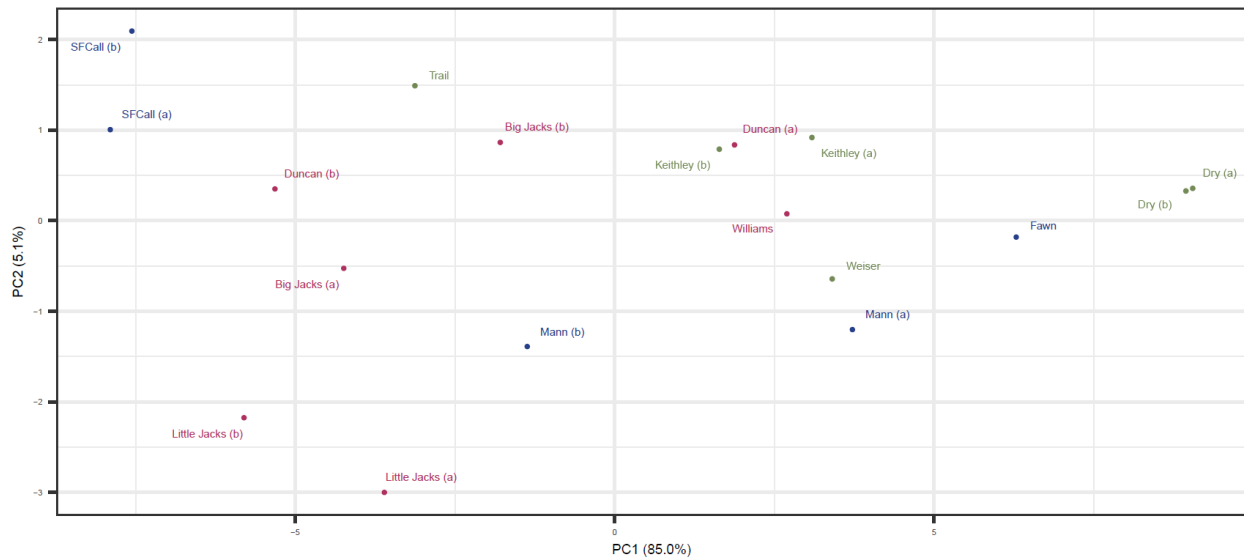

Omy03 Outlier Region: 553 SNPs
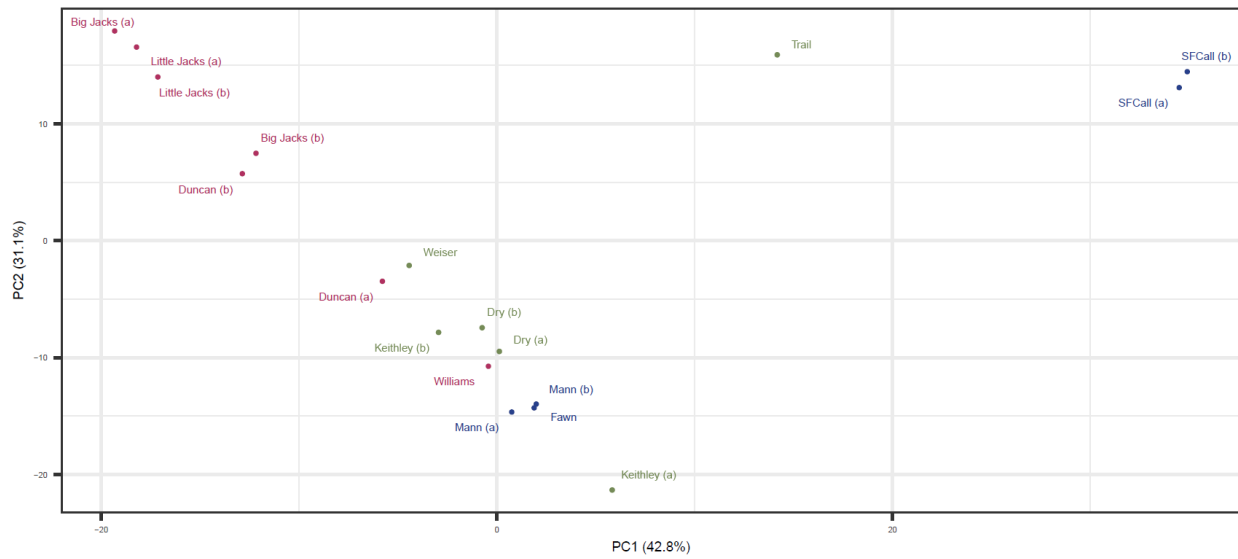
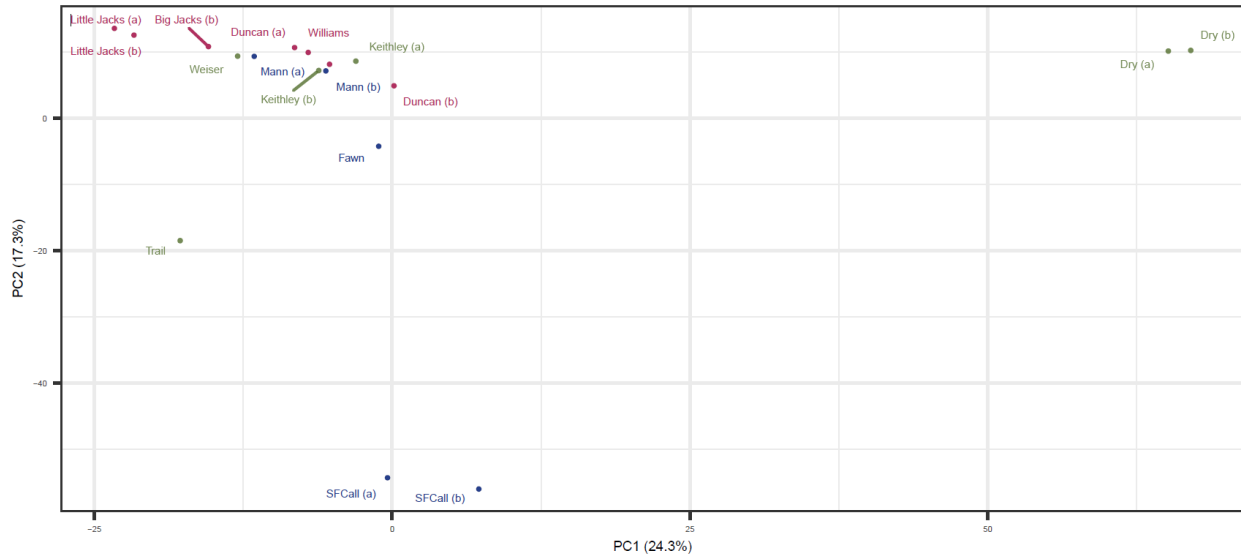Omy06 Outlier Region: 625 SNPs
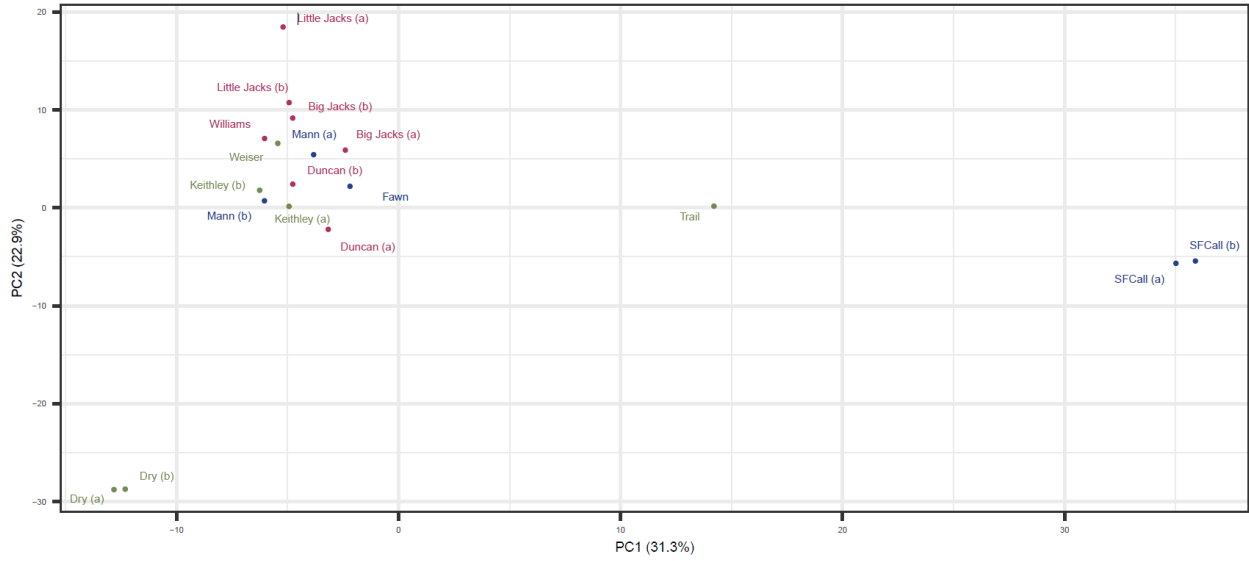
Omy15 Outlier Region: 532 SNPs
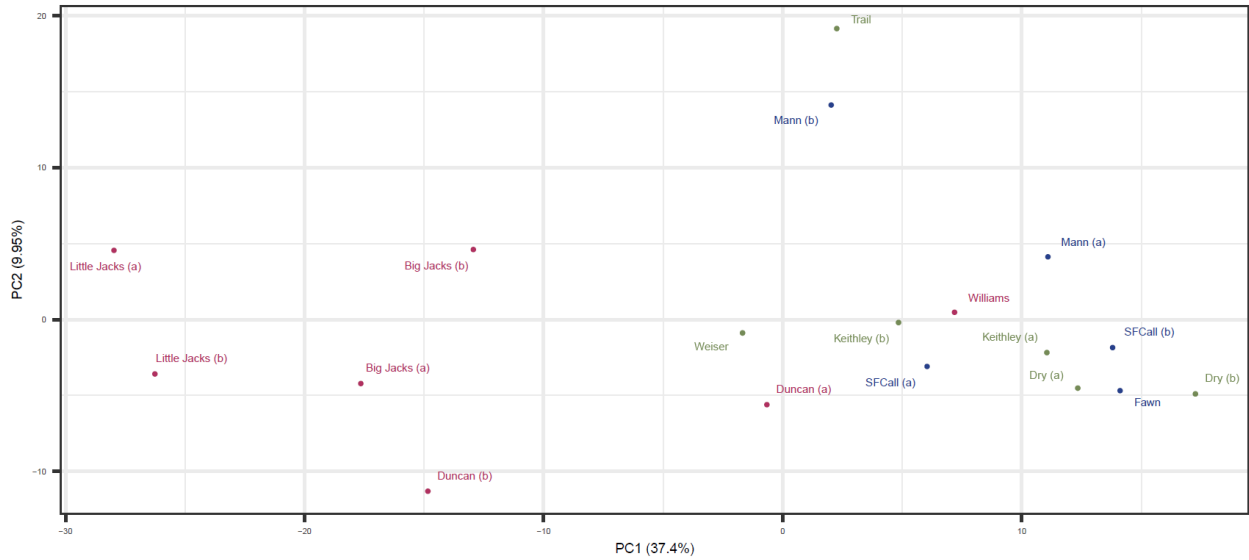
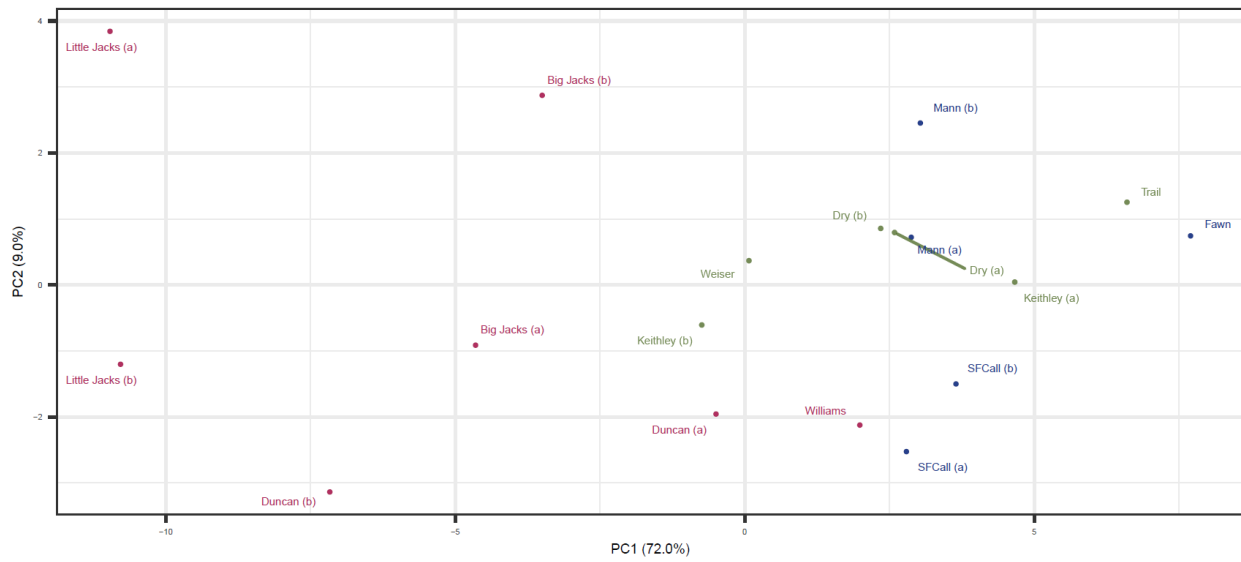Omy16 Outlier Region: 239 SNPs

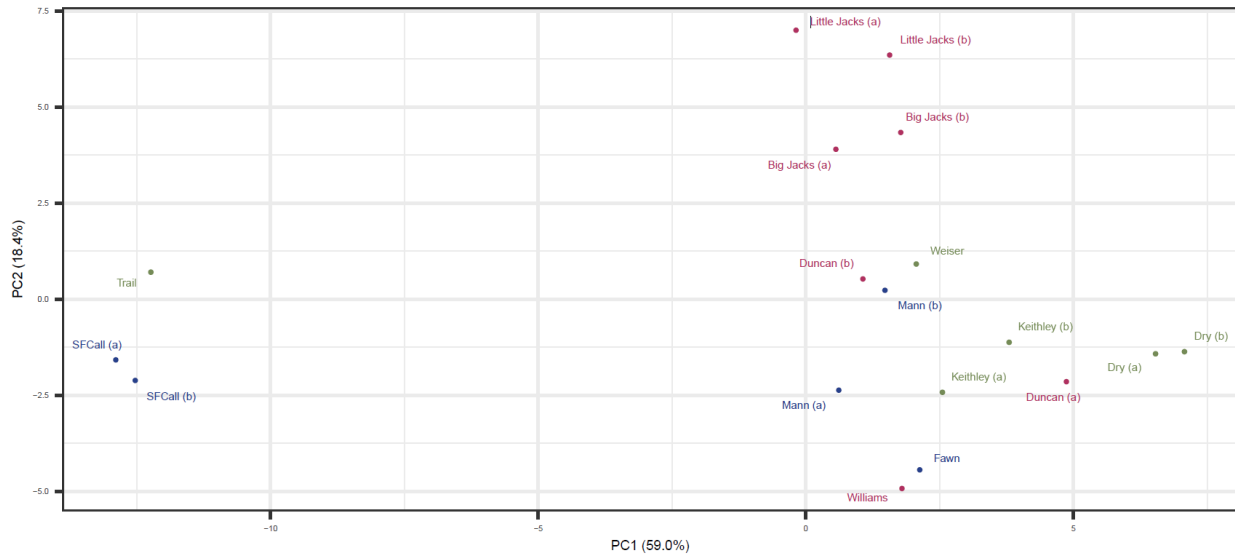Omy18 Outlier Region: 33 SNPs

Omy19 Outlier Region: 2635 SNPs

Omy23 Outlier Region: 41 SNPs

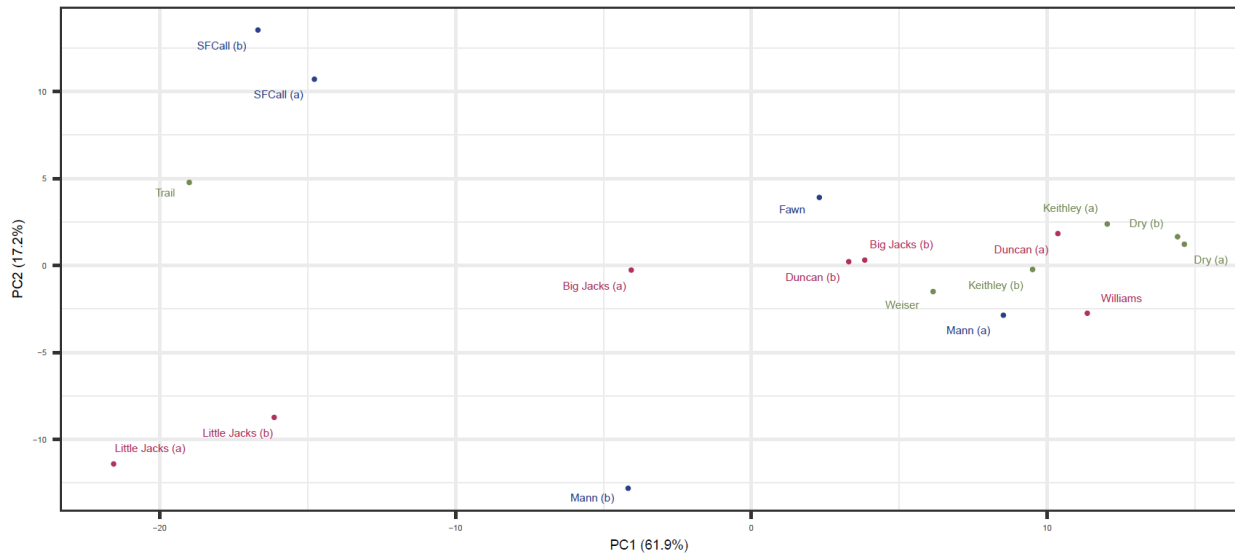Omy25 Outlier Region: 63 SNPs

Omy28 Outlier Region: 252 SNPs

Omy29 Outlier Region: 321 SNPs
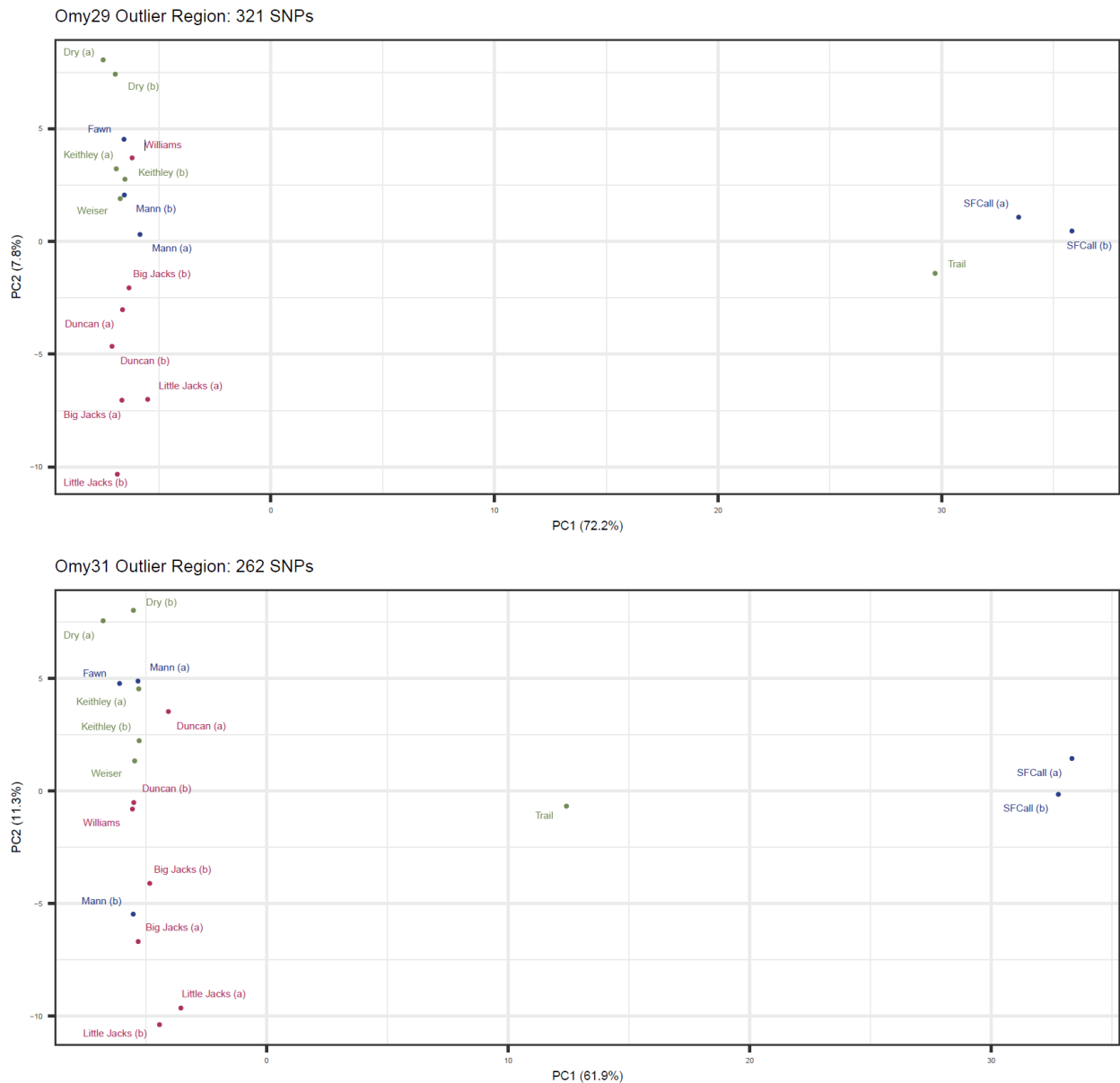

Omy31 Outlier Region: 262 SNPs

Figure S8. PCAs treating each temporal replicate independently for SNPs occurring in each of the 12 Local Score outlier regions identified for between-habitat comparisons. Habitat types cluster similarly here as in the PCAs that combined temporal replicates (see Figures S4 and S6), indicating that temporal variation in adaptive pressures likely had little influence on the results obtained when temporal replicates were combined. Sample sites are color-coded by ecotype: Blue = cold montane forest, Green = cool montane forest, Red = high elevation desert. "Weiser" = "Little Weiser."
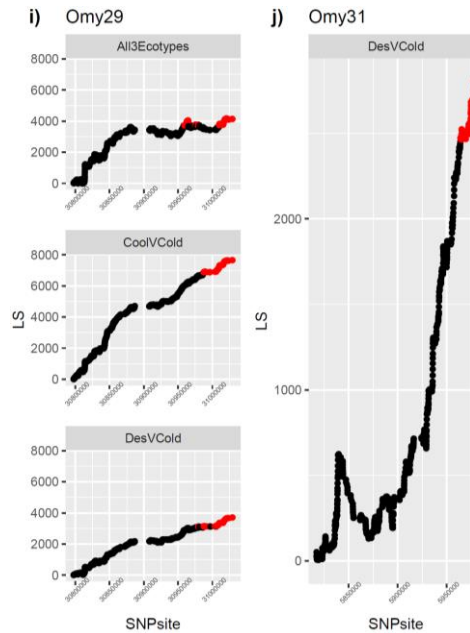
15
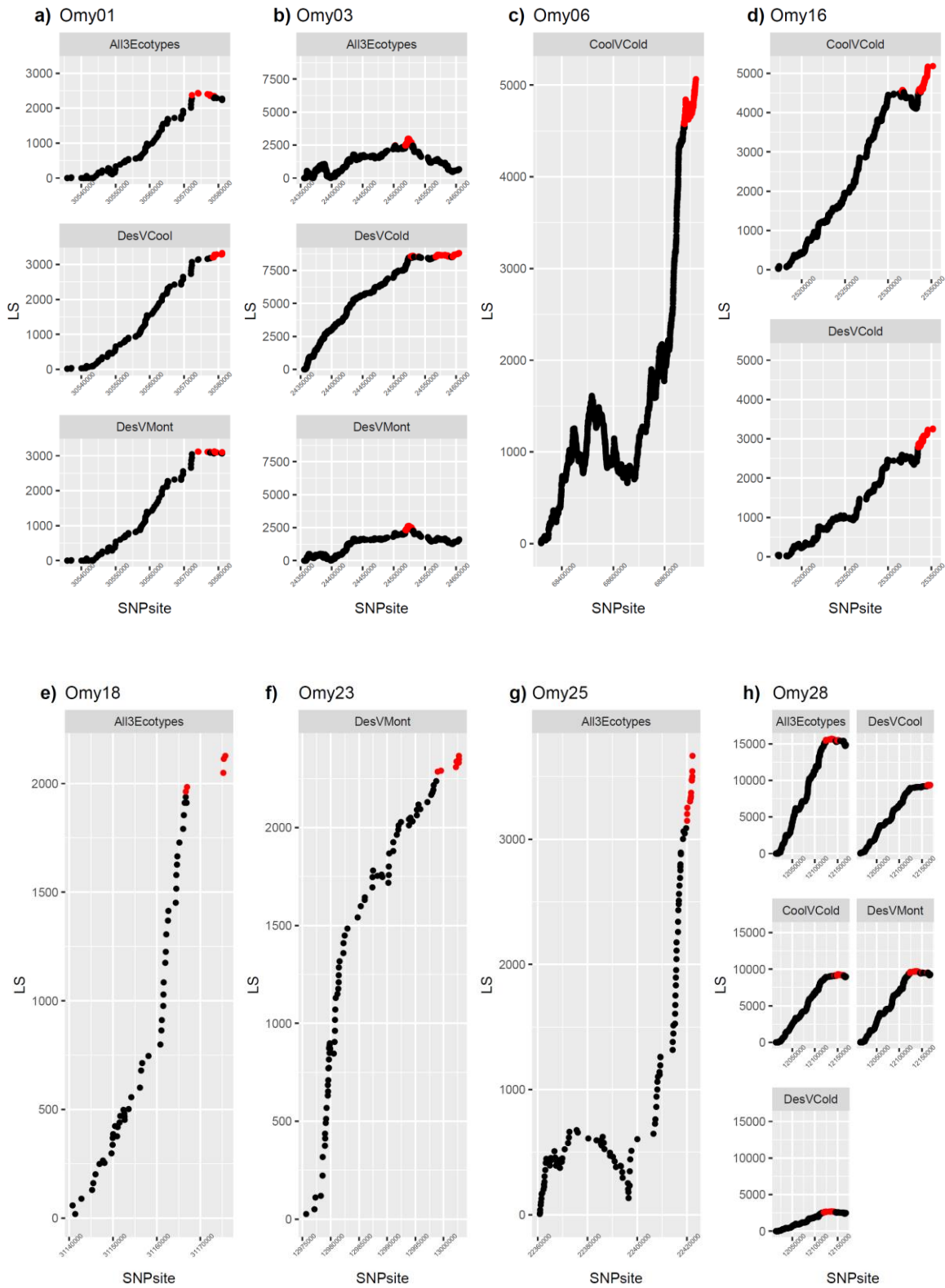
Figure S9. Local score (LS) values across all SNPs in each of the ten outlier regions identified by between-habitat comparisons with the full dataset (including Snake River and Kootenai River populations). Results are shown for each pairwise habitat comparison that was significant for each outlier region. The highest 10% of local score values are highlighted in red. "All 3" = Desert vs. Cool vs. Cold, "Des" = Desert, "Mont" = montane (i.e., both cool and cold montane).
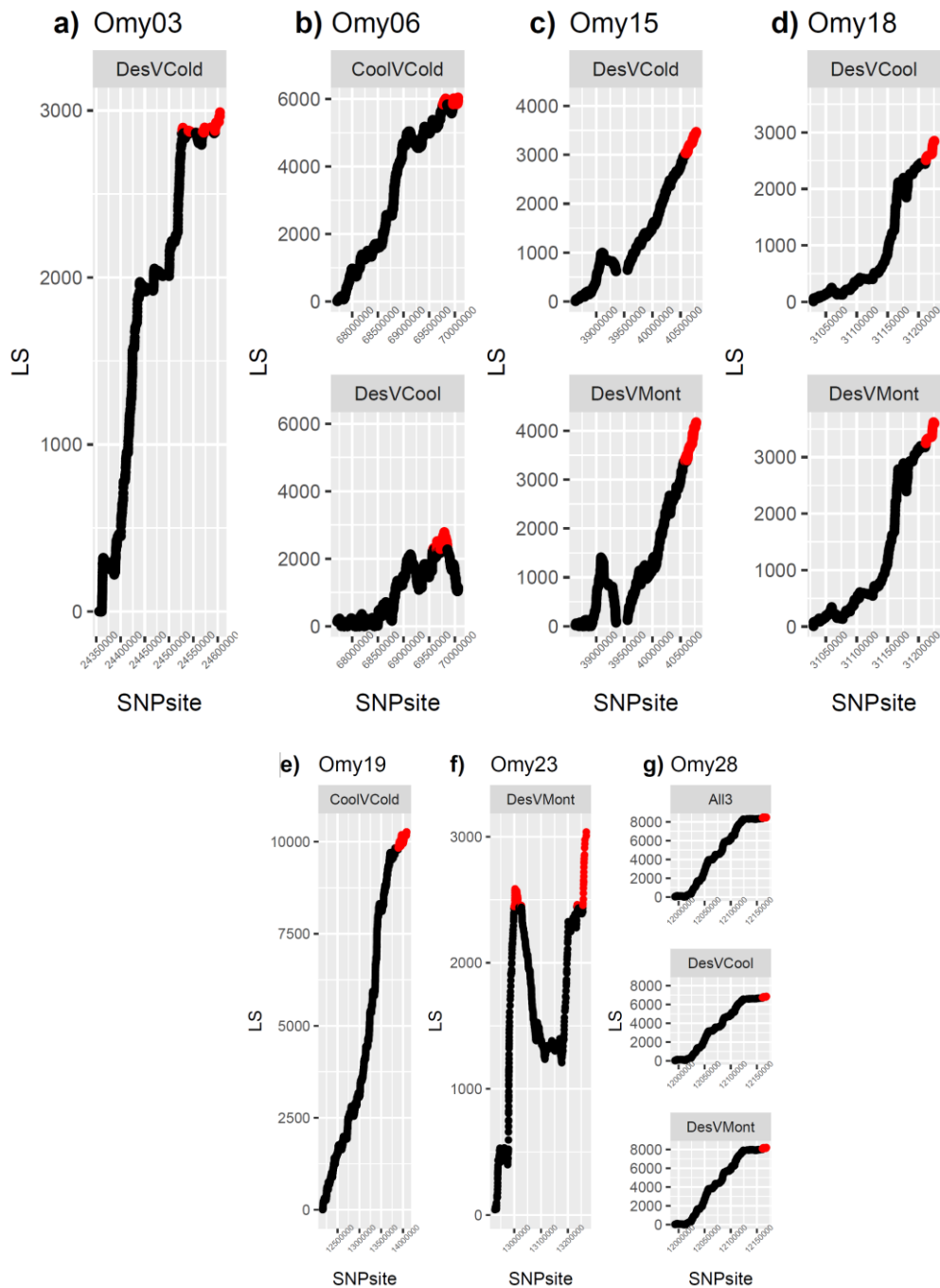
Figure S10. Local score (LS) values across all SNPs in each of the seven outlier regions identified by between-habitat comparisons with the Snake River data subset (excluding Kootenai River populations). Results are shown for each pairwise habitat comparison that was significant for each outlier region. The highest 10% of local score values are highlighted in red. "All 3" = Desert vs. Cool vs. Cold, "Des" = Desert, "Mont" = montane (i.e., both cool and cold montane).
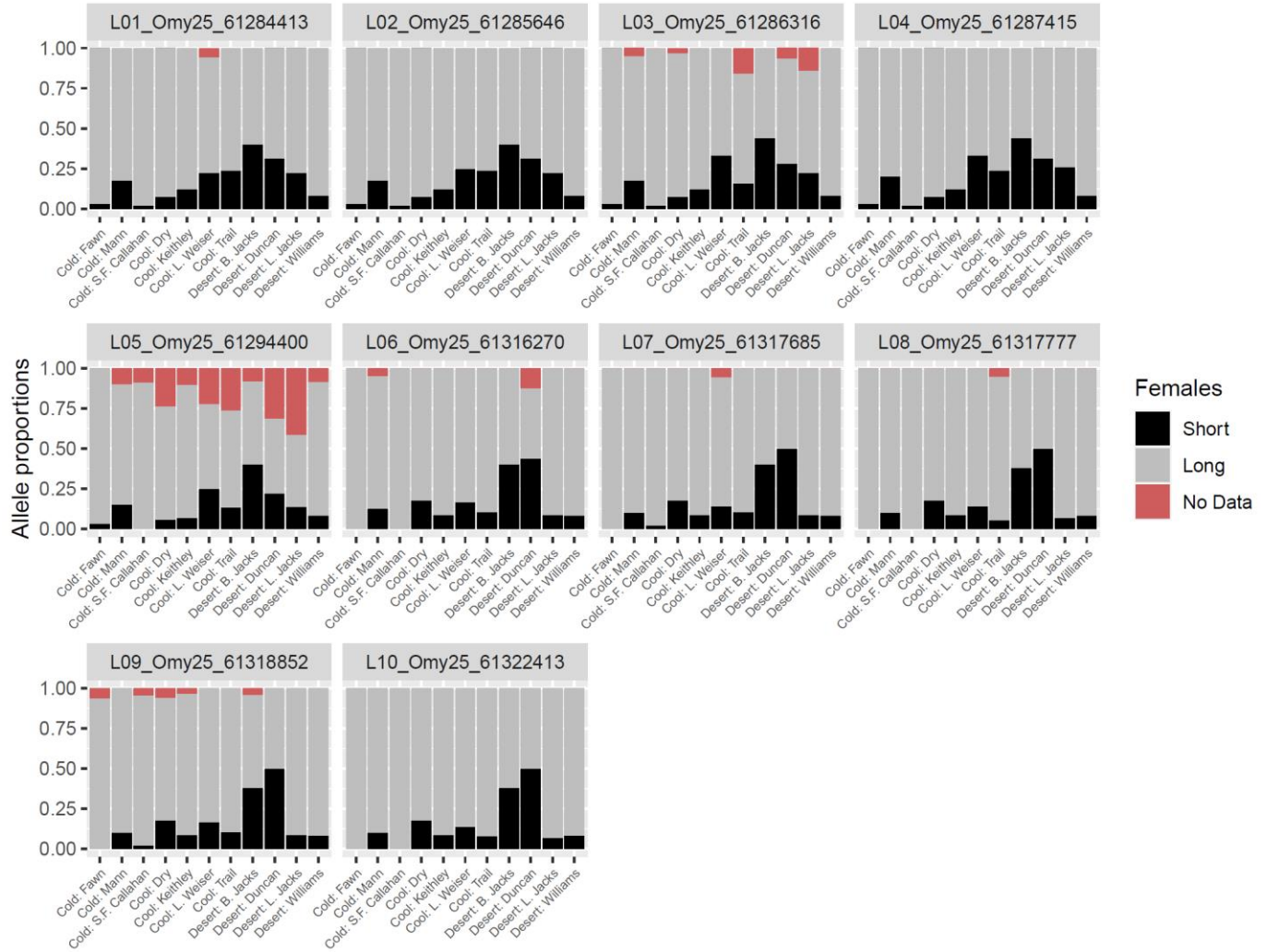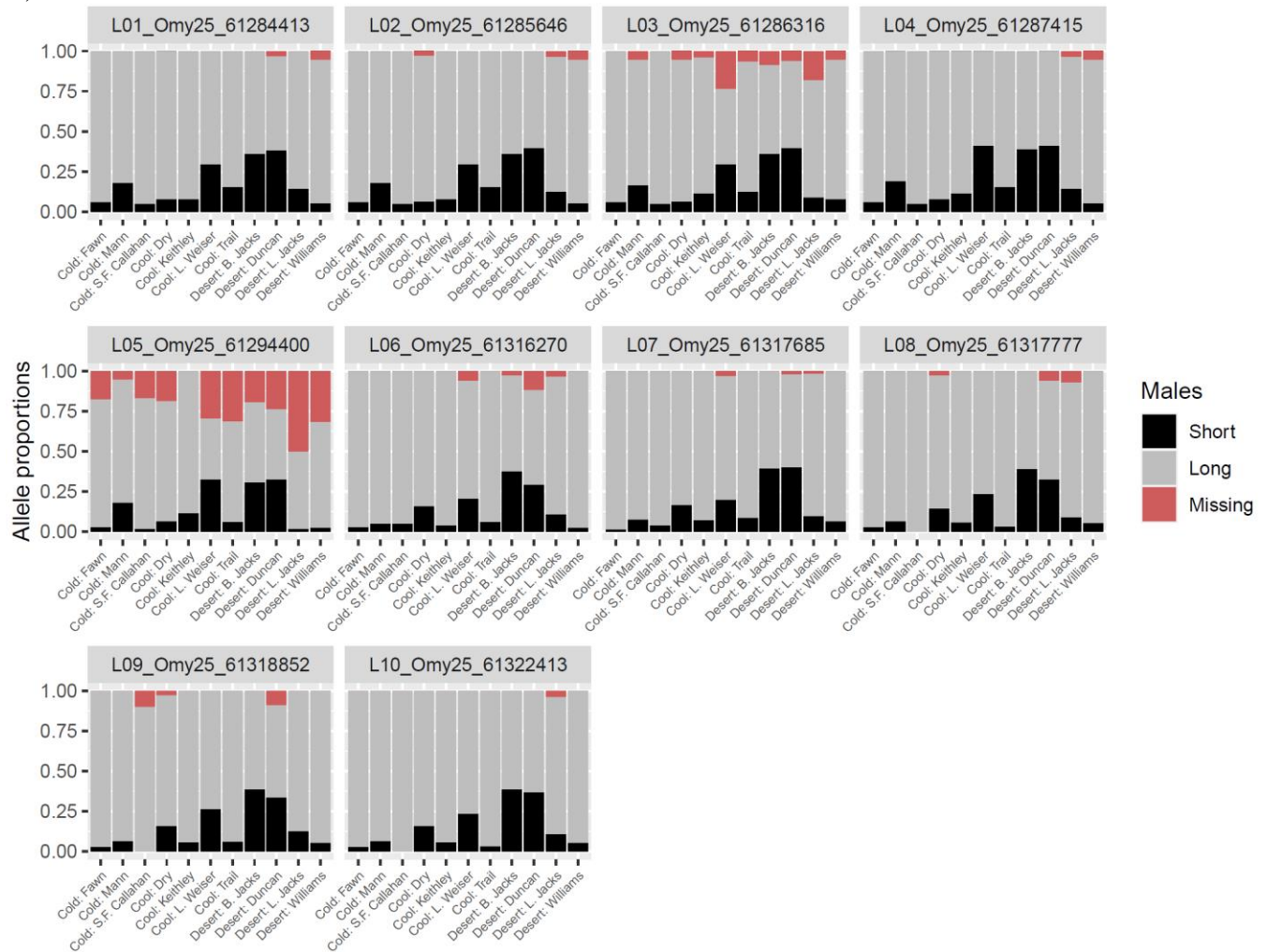
18

A)
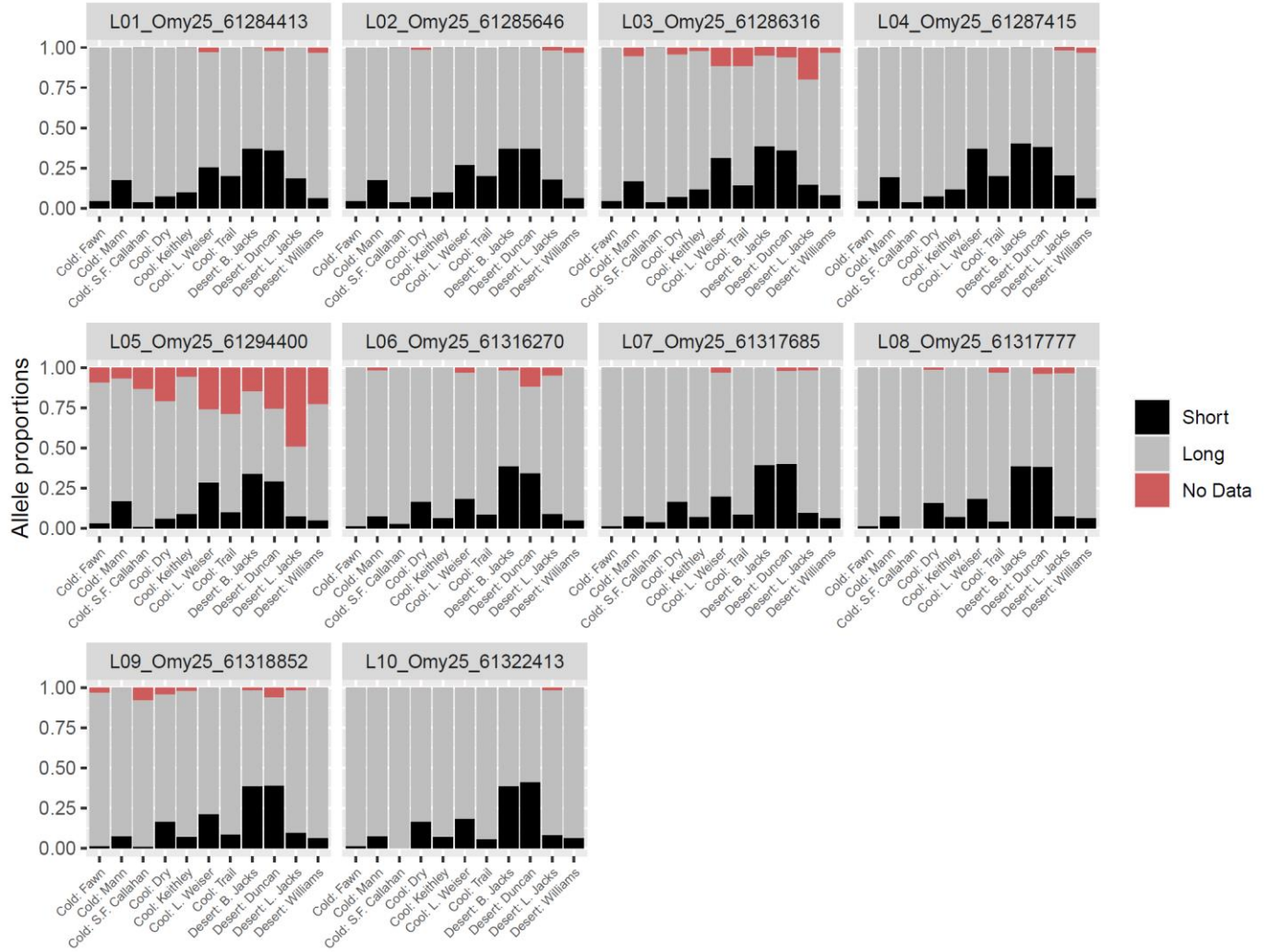
B)

C)

Figure S11. Allele proportions for each of the ten GTseq SNPs occurring within 100kb of the Omy25 outlier region. Plot headers show each SNP locus name. The first six SNPs are inferred to occur within the upstream promoter region for the *SIX6* gene, the next three occur in the first transcribed region of the *SIX6* gene, and the last occurs in the downstream intergenic region. Previous studies have identified *O. mykiss* phenotypes associated with each allele, including an earlier ("Short") or later ("Long") age at maturity (Willis et al. 2020). "No Data" indicates the proportion of samples that failed to produce genotype data. A) All samples; B) Female samples only; C) Male samples only.
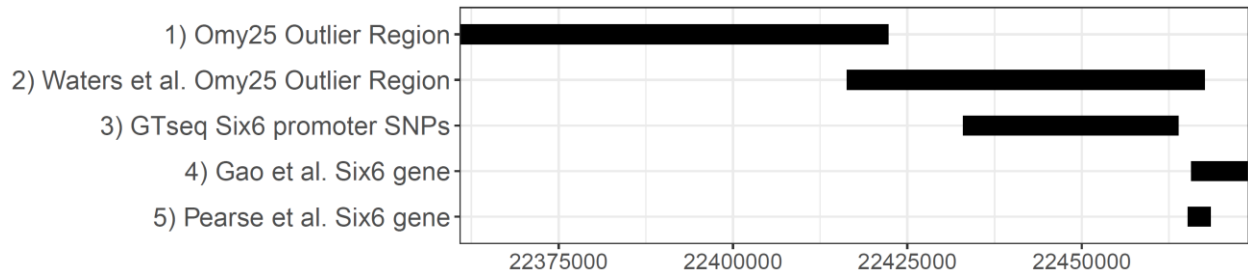
Figure S12. Map of the chromosomal positions of 1) the Omy25 outlier region identified in this study, 2) an outlier region identified in Waters et al. (2020) that was associated with age at maturity in rainbow trout, 3) the region encompassing the GTseq panel SNPs that are associated with age at maturity and are inferred to occur within the *SIX6* promoter region, 4) the *SIX6* gene in the Gao et al. (2021) genome assembly, 5) the *SIX6* gene in the Pearse et al. (2019) genome assembly. The x-axis shows chromosomal positions in reference to the Gao et al. (2021) assembly.
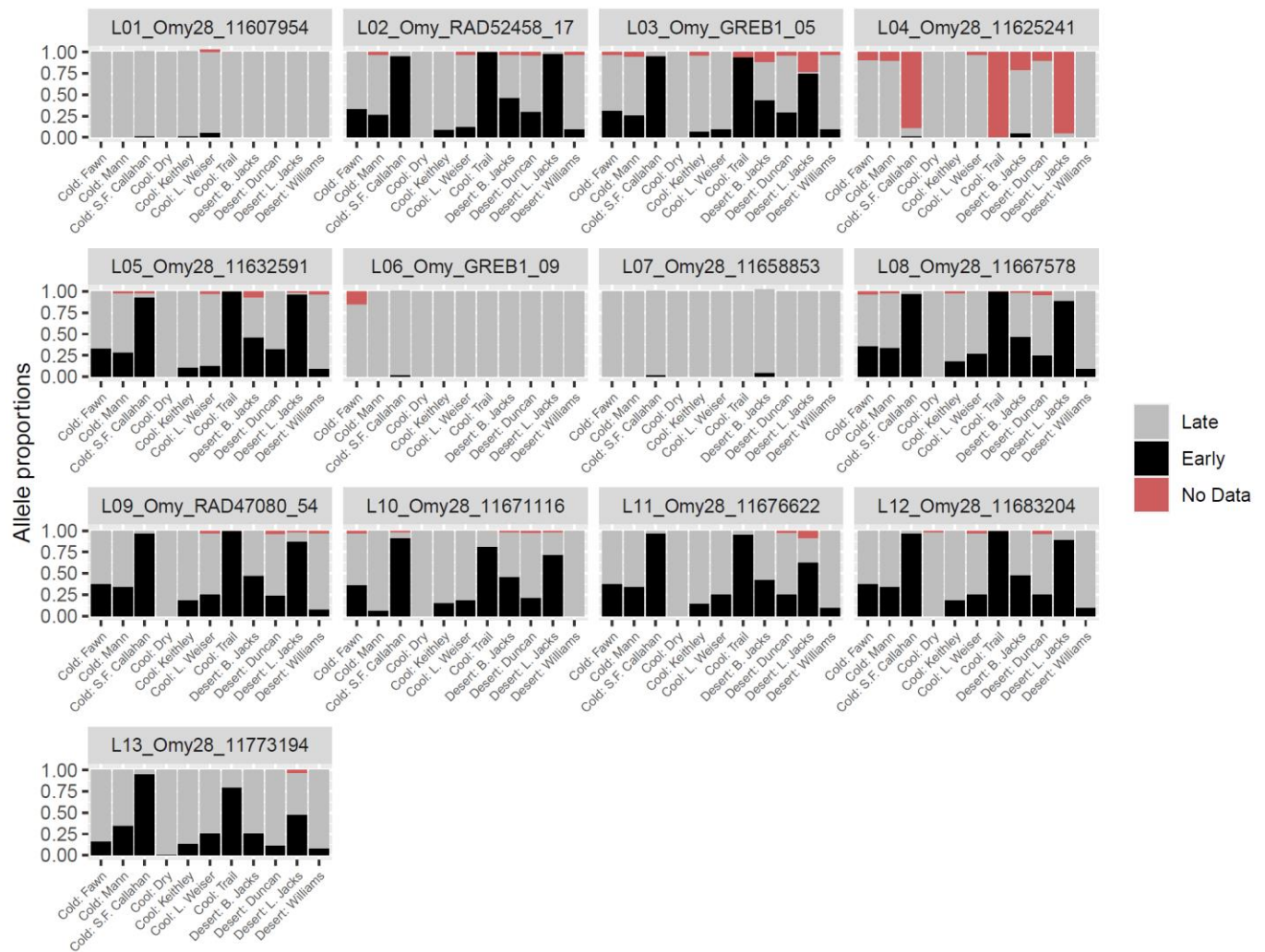
Figure S13. Allele proportions for each of the 13 GTseq SNPs occurring within the Omy28 outlier region. SNP locus names are in the plot headers. The first six SNPs occur within the exonic and intronic regions of the *GREB1L* gene, the next six SNPs occur within the intergenic region, and the last SNP occurs within an intron of the *ROCK1* gene. Previous studies have identified *O. mykiss* phenotypes associated with each allele, including an "Early" or "Late" migration timing (Willis et al. 2020). "No Data" indicates the proportion of samples that failed to produce genotype data.
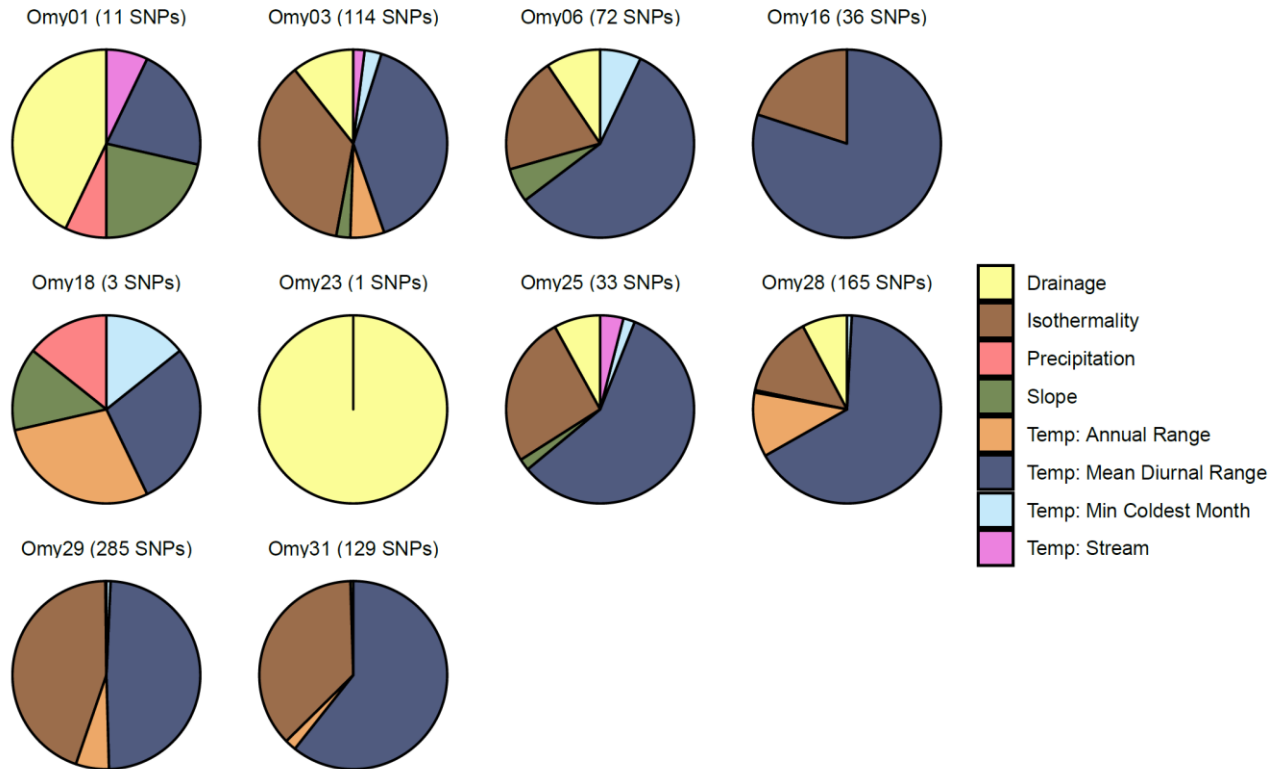
Figure S14. Proportions of environmental associations for LFMM outlier SNPs in each Local Score outlier region identified by the full dataset analysis. Plots are labeled based on the chromosome in which the outlier region occurred, and the number of SNPs with significant LFMM associations is shown in parentheses. Each environmental association is treated independently, and therefore SNPs associated with more than one environmental variable are represented more than once. Proportions should be interpreted with caution for the three outlier regions that have <15 SNPs.
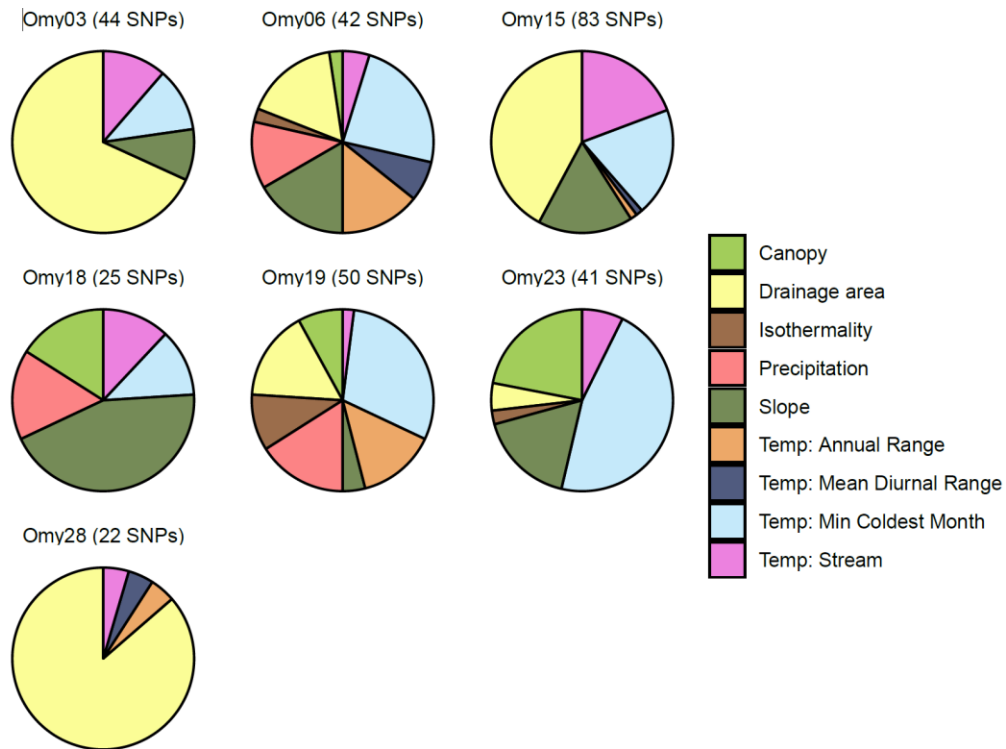
Figure S15. Proportions of environmental associations for LFMM outlier SNPs in each Local Score outlier region for the Snake River data subset analysis (excluding Kootenai River populations). Plots are labeled based on the chromosome in which the outlier region occurred, and the number of SNPs with significant LFMM associations is shown in parentheses. Each environmental association is treated independently, and therefore SNPs associated with more than one environmental variable are represented more than once.
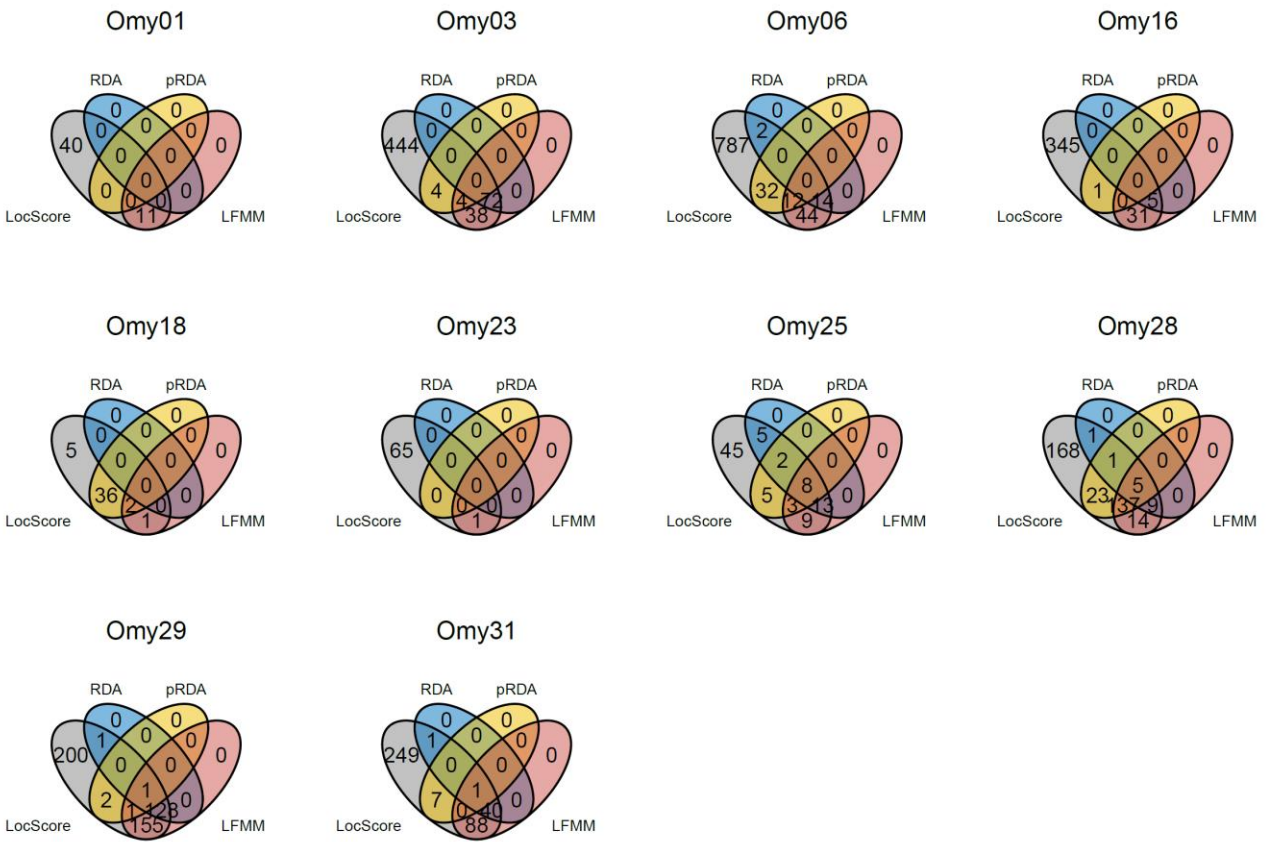
Figure S16. Overlap of outlier SNPs across detection methods for all SNPs occurring in each of the ten outlier regions identified by Local Score ("LocScore") analysis for the full dataset (including both Kootenai River and Snake River populations). Plots are labeled based on the chromosome in which the outlier region occurred. SNPs that were outliers for more than one environmental variable were counted only once.
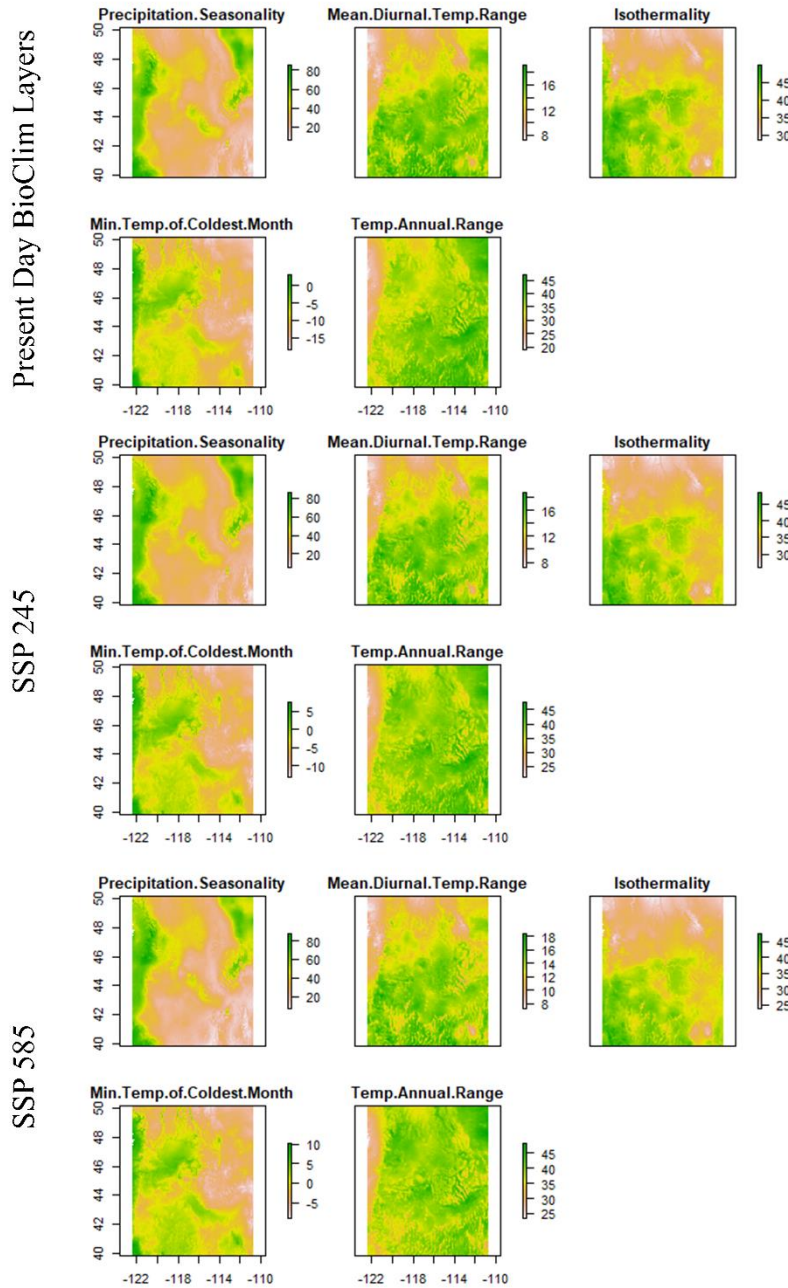
Figure S17. WorldClim climate variables for ecological niche models and genetic offset calculations.
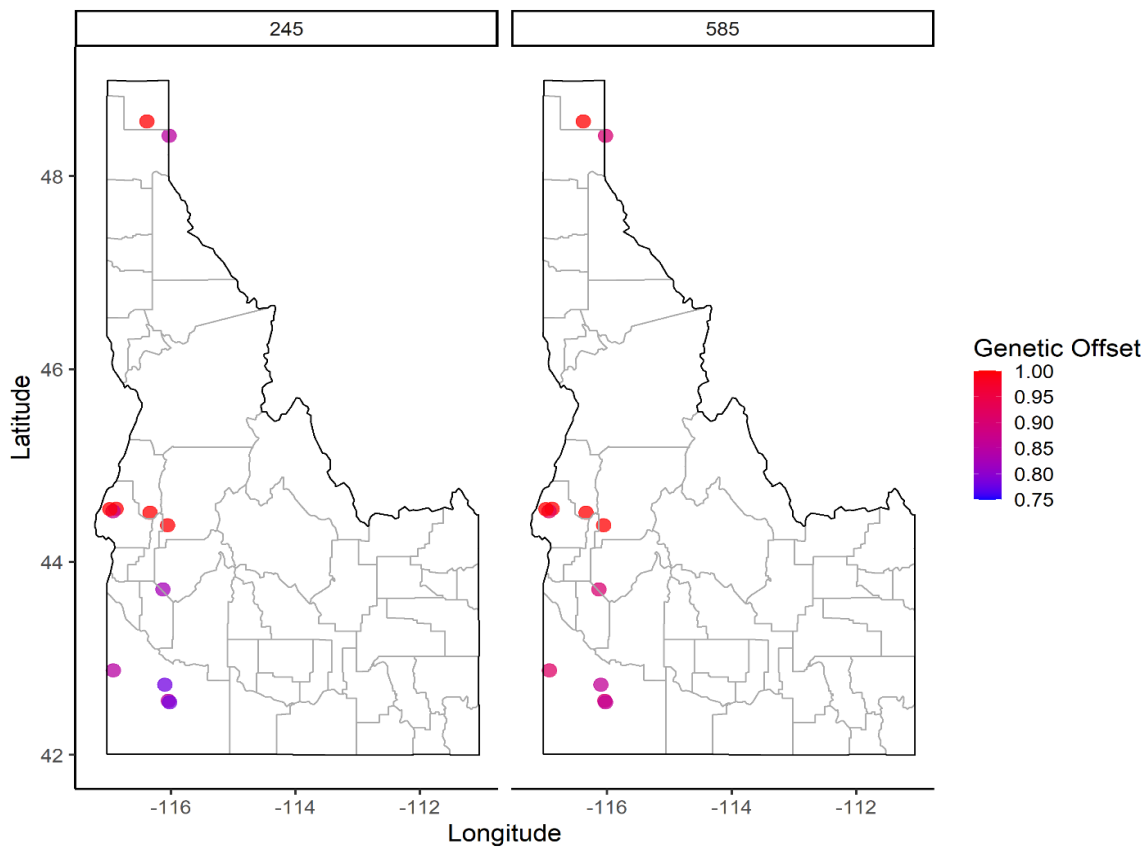
Figure S18. Genetic offset calculation for the period 2081-2100 across two Shared-Socioeconomic Pathways (SSP245 and SSP585).