Supplementary material for **Rapid genetic adaptation to a novel ecosystem despite a large founder event**

Morgan M. Sparks[1*], Claire E. Schraidt[2], Xiaoshen Yin[3], Lisa W. Seeb[4], Mark R. Christie[1,2]

[1]Department of Biological Sciences, Purdue University; 915 W. State St., West Lafayette, IN

[2]Department of Forestry and Natural Resources, Purdue University; 715 W. State St., West Lafayette, IN

[3]MOE Key Laboratory of Marine Genetics and Breeding, College of Marine Life Sciences, Ocean University of China, Qingdao 266003, China

[4] School of Aquatic and Fishery Sciences, University of Washington, 1122 NE Boat Street, Box 355020, Seattle, WA 98195-5020, USA

*Corresponding author: Morgan M. Sparks; Department of Biological Sciences, Purdue University; 915 W. State St., West Lafayette, Indiana 47907; sparks35@purdue.edu

Keywords: agent-based models, genetic adaptation, introduced species, genetic drift, pink salmon, rapid evolution

Running title: Rapid evolution in an introduced fish

Supplementary methods

VCF filtering

We used GATK's recommended hard filtering thresholds to mark variants that did not pass the following filters quality by depth (QD) < 2.0, strand odds ratio (SOR) > 3.0, fisher strand (FS) > 60.0, root mean square mapping quality (MQ) < 40.0, mapping quality rank sum test (MQRankSum) < -12.5, and read position rank sum test (ReadPosRankSum) < -8.0., these filters mark loci that substantially deviate from thresholds for depth and quality to be removed by filtering software.

Sample relatedness

We used VCFtools' relatedness2 algorithm to determine pairwise sample relatedness in our joint vcf file (Danecek et al., 2011). One sample, from the British Columbia sample group (LAE_059) showed a pattern of relatedness to multiple sample groups that was indicative of contamination, likely during library preparation, and was removed from all downstream analyses. No other samples showed contamination or close relatedness to samples within their sample group (Fig. S1).

Contemporary effective population size estimates

We estimated contemporary effective population size to validate our GONE results using NeEstimator. First, we used VCFs which were individually filtered (by sample group) to a minor allele frequency of 0.05. For Great Lakes samples, we used all odd, even, and 3yr-old samples as a pool because our results indicated they were a single population and to generate

more accurate results from the program. We then randomly subsampled 1000 SNPs from each VCF to use in the program and estimated values with the three random subsets of 1000 SNPs.

Linkage disequilibrium pruned genome scan

To verify our genome scan results using ZF_{ST} and eigenGWAS methods, we also used a linkage disequilibrium pruned set of SNPs (R^2 >/= 0.2) and the program pcaadapt to look for outlier SNPs. We considered SNPs to be outliers if they had a false discovery rate corrected (Benjaminin-Hochberg) value </= 0.05. This value for each SNP is presented in the third panel of each plot in Figure S6.

Agent-based model

We created an age and stage structured model that allowed for age 0 and age 1 juveniles and only age 2 adults. All adults were removed after reproduction to mimic the semelparity of pink salmon, and the age and stage structure allowed us to recreate spawning events that only occurred in even or odd years. The sequence of events in the model were as follows: (1) creation of adults and juveniles, (2) creation of a single neutral locus at the desired starting allele frequency and in accordance with expected Hardy Weinberg proportions, (3) aging, where all individuals were aged by 1 year; the stage of individuals was changed from juvenile to adult if they became 2 years old, (4) reproduction of adults including the transmission of alleles in accordance with Mendelian inheritance, (5) mortality of all 2-year-olds after reproduction, and (6) juvenile mortality, where juvenile population sizes were approximately 7x that of adults and where there were twice as many age 0 juveniles as age 1 juveniles. Juvenile mortality included a stochastic component (set as a random deviate from a normal distribution) to mimic natural

fluctuations in population sizes. The sequence of events from steps 3-6 were repeated for 127 generations (254 years).

Our analyses next proceeded in two sequential steps. As a reminder our goal was to conservatively estimate the allele frequency at the time of the introduction event and then determine whether drift alone could explain the current allele frequency in the introduced population (Great Lakes). We first wanted to estimate the allele frequency at the time of the introduction event. Because the effective population size in the native range has declined since the introduction event (Figure 3b) and because we only sampled 30 individuals from the native range odd year, the estimates of allele frequencies from our sample may not reflect the range of allele frequencies present in GL odd at the time of introduction. For each *per2* allele frequency estimate from BC odd, we ran the model using the steps described above and the N_e estimated for BC odd via GONE using the oldest generation estimates as the first years in the model and the most recent generation estimates as the last years in the model (i.e., the model ran forward in time from 200 years in the past to the present). After the simulations were complete, we randomly sampled 30 individuals and calculated the ending allele frequency from that sample. By recording the allele frequency before and after sampling, we could disentangle the relative contributions of allele frequency changes due to genetic drift and those due to sampling. For each starting allele frequency, we performed 1000 replicate simulations that proceeded through all 127 generations (254 years, see methods below) of the model and were followed by one single sampling event.

We next wanted to determine whether the native-range allele frequencies could drift to a frequency equal to that found in our contemporary Great Lakes samples. Using the BC odd allele frequencies in *per2*, we again used the agent-based model to simulate changes in allele

frequencies this time using the N_e values for BC odd for the first 75 (generations 100-26 in figure 3B) and then switching to the GL odd N_e estimates for the next 52 generations (generations 1-52 in figure 3C) using the BC odd empirical estimates as our starting allele frequencies (See figure S2 for modeled population sizes through time). We selected these generations to capture the most likely effective population sizes at the time of introduction and the smallest effective population sizes during the introduction event. At the end of each simulation, we again sampled 30 individuals and calculated the final allele frequency. This process was replicated 15,999 times where a few runs (298) resulted in extinctions and were excluded from analyses. For the runs that did not result in extinction, we next calculated how many samples had allele frequency as high or higher than that found in our sample to calculate the probability of the final allele frequency occurring by chance alone. We also plotted the changes in allele frequencies through time. We selected these generations to capture the most likely effective population sizes at the time of introduction and the smallest effective population sizes during the introduction event. At the end of each simulation, we again sampled 30 individuals and calculated the final allele frequency.

Supplementary results

Contemporary effective population size estimates

Contemporary estimates from NeEstimator showed largely concordant values with those of GONE. Both programs indicated BC Even as having the largest contemporary sample size, while BC Odd and the Great Lakes samples were less than 4,000. The NeEstimator mean estimate with 95% confidence intervals for BC Even was 46,240 (9,862-34,766), however these were particularly biased by one replicate estimate which much higher than the other samples (118,696

vs. 10,984 and 9,040). Estimates for BC Odd and the GL pooled samples were 2,334 (2,081-2656) and 3,549 (3,285-3,859), respectively.

SNP annotation

It is worth noting that though non-exonic variants do not result in amino acid sequence change if these variants are upstream or downstream of genes, they may result in regulatory effects (reviewed in (Shabalina et al., 2013)). Additionally, snpEff quantifies the multiple effect impacts for a SNP (e.g., in the intronic region of one gene and upstream of another gene) based on the ENSEMBL system (see ENSEMBL glossary). Thirteen effects were characterized as having a moderate impact, while 14 were characterized as having low impact (SNPs could have multiple effects under the classification system of snpEff), the remaining (2,612) were calculated as modifier impact.

References

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., & Sherry, S. T. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158.

Liu, S., Ferchaud, A.-L., Grønkjær, P., Nygaard, R., & Hansen, M. M. (2018). Genomic parallelism and lack thereof in contrasting systems of three-spined sticklebacks. *Molecular Ecology*, *27*(23), 4725–4743.

Liu, X., & Fu, Y.-X. (2020). Stairway Plot 2: Demographic history inference with folded SNP frequency spectra. *Genome Biology*, *21*(1), 1–9.

Shabalina, S. A., Spiridonov, N. A., & Kashina, A. (2013). Sounds of silence: Synonymous nucleotides as a key to biological regulation and complexity. *Nucleic Acids Research*, *41*(4), 2073–2094.

Table S1. Sample information including sample group, ID, total reads across the two sequencing runs, average length of those reads, and average depth.

Sample Group	Sample ID	Total Reads (Millions)	Average Read Length (bp)	Depth (X)
BC_Even	LAE_006	294.66	134.99	14.73
BC_Even	LAE_012	276.55	135.2	13.85
BC_Even	LAE_024	314.93	133.15	15.53
BC_Even	LAE_030	249.17	133.4	12.31
BC_Even	LAE_036	314.4	136.01	15.84
BC_Even	LAE_042	270.13	133.66	13.37
BC_Even	LAE_053	332.07	135.77	16.7
BC_Even	LAE_056	797.53	145.14	42.87
BC_Even	LAE_057	238.93	130.59	11.56
BC_Even	LAE_058	268.38	126.24	12.55
BC_Even	LAE_059	532.74	133.03	26.25
BC_Even	LAE_062	250.82	129.04	11.99
BC_Even	LAE_064	221.1	132.28	10.83
BC_Even	LAE_065	301.15	133.28	14.87
BC_Even	LAE_071	304.27	133.38	15.03
BC_Even	LAE_076	212.07	131.75	10.35
BC_Even	LAE_077	399.27	132.14	19.54
BC_Even	LAE_080	632.66	148.73	34.85
BC_Even	LAE_081	299.92	131.47	14.6
BC_Even	LAE_082	281.83	134.81	14.07
BC_Even	LAE_083	88.86	127	4.18
BC_Even	LAE_086	298.18	133.16	14.71
BC_Even	LAE_087	344.85	134.57	17.19
BC_Even	LAE_088	318.66	134.33	15.85
BC_Even	LAE_089	224.9	132.08	11
BC_Even	LAE_093	245.43	131.82	11.98
BC_Even	LAE_095	229.31	129.82	11.03
BC_Even	LAE_098	261.24	130.57	12.63
BC_Even	LAE_099	89.83	122.18	4.06
BC_Even	LAE_100	387.56	133.11	19.11
BC_Odd	LAO_161	269.42	131.26	13.1
BC_Odd	LAO_162	240	134.47	11.95
BC_Odd	LAO_163	232.02	133	11.43
BC_Odd	LAO_164	298.79	132.01	14.61
BC_Odd	LAO_165	442.82	134.81	22.11
BC_Odd	LAO_170	241.97	128.14	11.48

BC_Odd	LAO_171	301.32	132.29	14.76
BC_Odd	LAO_173	375.46	131.01	18.22
BC_Odd	LAO_177	261.88	132.54	12.86
BC_Odd	LAO_182	284.34	132.14	13.92
BC_Odd	LAO_183	266.92	132.64	13.11
BC_Odd	LAO_185	201.85	131.47	9.83
BC_Odd	LAO_186	347.14	131.75	16.94
BC_Odd	LAO_187	266.85	132.83	13.13
BC_Odd	LAO_188	250.13	134.28	12.44
BC_Odd	LAO_189	232.93	129.03	11.13
BC_Odd	LAO_191	282.04	130.97	13.68
BC_Odd	LAO_192	325.44	126.93	15.3
BC_Odd	LAO_193	249.16	132.5	12.23
BC_Odd	LAO_194	204.99	131.78	10.01
BC_Odd	LAO_195	550.18	136.25	27.76
BC_Odd	LAO_198	109.09	125.59	5.07
BC_Odd	LAO_199	272.16	133.06	13.41
BC_Odd	LAO_201	235.81	132.4	11.56
BC_Odd	LAO_203	155.99	130.66	7.55
BC_Odd	LAO_204	250.95	131.46	12.22
BC_Odd	LAO_205	243.34	131.77	11.88
BC_Odd	LAO_206	413.15	134.75	20.62
BC_Odd	LAO_207	549.25	149.27	30.37
GL_Even	STL_003	193.7	124.4	8.92
GL_Even	STL_006	228.4	128.61	10.88
GL_Even	STL_007	281.23	133.32	13.89
GL_Even	STL_008	263.48	134.8	13.15
GL_Even	STL_009	313.48	134.09	15.57
GL_Even	STL_011	230.24	127.98	10.91
GL_Even	STL_012	223.01	127.07	10.49
GL_Even	STL_013	198.57	129.38	9.52
GL_Even	STL_014	181.68	129.13	8.69
GL_Even	STL_015	247.95	126.63	11.63
GL_Even	STL_016	236.63	133.51	11.7
GL_Even	STL_017	151.34	129.97	7.28
GL_Even	STL_018	334.74	136.58	16.93
GL_Even	STL_019	182.81	127.85	8.66
GL_Even	STL_020	189.56	121.73	8.55
GL_Even	STL_021	280.2	135.47	14.06
GL_Even	STL_022	201.26	124.06	9.25

GL_Even	STL_024	265.74	135.08	13.29
GL_Even	STL_025	184.61	123.3	8.43
GL_Even	STL_026	263.51	131.92	12.87
GL_Even	STL_028	234.17	127.23	11.04
GL_Even	STL_030	238.09	124.46	10.97
GL_Even	STL_031	294.33	132.78	14.47
GL_Even	STL_033	1078	134.33	53.63
GL_Even	STL_035	358.58	123.69	16.43
GL_Even	STL_036	257.85	132.09	12.61
GL_Even	STL_038	269.56	134.35	13.41
GL_Even	STL_039	306.8	134.34	15.27
GL_Even	STL_046	204.2	130.02	9.83
GL_Even	STL_048	174.9	126.03	8.16
GL_Odd	STL_101	242.04	133.2	11.94
GL_Odd	STL_105	291.59	133.82	14.45
GL_Odd	STL_112	282.43	134.84	14.11
GL_Odd	STL_113	264.15	134.68	13.18
GL_Odd	STL_125	252.1	131.75	12.3
GL_Odd	STL_136	229.57	133.56	11.36
GL_Odd	STL_159	180.8	126.41	8.47
GL_Odd	STL_160	306.39	133.93	15.2
GL_Odd	STL_187	239.57	133.87	11.88
GL_Odd	STL_189	574.97	135.98	28.96
GL_Odd	STL_191	177.04	132.06	8.66
GL_Odd	STL_201	290.93	133.17	14.35
GL_Odd	STL_205	266.34	133.43	13.16
GL_Odd	STL_206	288.22	134.98	14.41
GL_Odd	STL_210	279.13	135.17	13.97
GL_Odd	STL_226	454.45	146.55	24.67
GL_Odd	STL_228	556.82	135.63	27.97
GL_Odd	STL_244	223.26	133.41	11.03
GL_Odd	STL_247	284.11	133.62	14.06
GL_Odd	STL_259	370.44	135.01	18.52
GL_Odd	STL_264	260.28	132.71	12.79
GL_Odd	STL_270	243.79	134.23	12.12
GL_Odd	STL_271	406.94	136.65	20.6
GL_Odd	STL_290	241.24	132.32	11.82
GL_Odd	STL_304	210.51	133.44	10.4
GL_Odd	STL_315	200.44	133.33	9.9
GL_Odd	STL_316	264.56	132.73	13.01

GL_Odd	STL_318	228.44	132.78	11.23
GL_Odd	STL_328	260.18	133.42	12.86
GL_Odd	STL_331	224.59	134.01	11.15
GL_Odd_3yr	STL_104	335.31	148.14	18.4
GL_Odd_3yr	STL_110	194	133.34	9.58
GL_Odd_3yr	STL_111	498.35	136.68	25.23
GL_Odd_3yr	STL_134	241.21	134.1	11.98
GL_Odd_3yr	STL_139	144.89	131.12	7.04
GL_Odd_3yr	STL_153	225.78	134.48	11.25
GL_Odd_3yr	STL_175	295.2	133.37	14.58
GL_Odd_3yr	STL_179	197.46	131.99	9.65
GL_Odd_3yr	STL_194	128.85	135.15	6.45
GL_Odd_3yr	STL_203	341.5	131.15	16.59
GL_Odd_3yr	STL_252	205.66	130.9	9.97
GL_Odd_3yr	STL_302	456.95	131.55	22.26
GL_Odd_3yr	STL_308	229.22	133.7	11.35
GL_Odd_3yr	STL_310	171.89	132.86	8.46
GL_Odd_3yr	STL_319	280.08	135.64	14.07

Table S2. Analyses and the Variant Call Format file (vcf) used for their analysis. Note all vcfs listed are filtered. Joint vcf refers to files that went through all filtering as a single vcf, whereas sample group-specific refers to files filtered for minor allele frequency (MAF) in vcfs for a given sample group.

Analysis	**VCF used**
Windowed SNP counts, Total number SNPs	Sample group-specific, no MAF filter
GONE, NeEstimator	Sample group-specific
Tajima's *D*	Joint vcf, no MAF filter
Observed heterozygosity, $F_{ST,}$ eigenGWAS, fastSTRUCTURE, snpEff/snpSift, pcadapt	Joint vcf

Table S3: Chromosome, position, alleles, and allele frequencies for the GL Odd and BC Odd sample groups and modeled *p*-value at each SNP from our individual based models for the *per2* locus.

Chrom.	Pos.	A1	A2	A1.freq.GL	A2.freq.GL	A1.freq.BC	A2.freq.BC	pval
22	14723245	C	T	0.9333	0.0667	0.2321	0.7679	0.01529052
22	14723696	C	G	0.9667	0.0333	0.0345	0.9655	0
22	14724875	A	G	0.1000	0.9000	0.9655	0.0345	0.00102249
22	14725276	G	T	0.0667	0.9333	0.9655	0.0345	0.00102249
22	14725308	A	C	0.0667	0.9333	0.7931	0.2069	0.00303337
22	14725618	C	T	0.0833	0.9167	0.8036	0.1964	0.00917431
22	14726815	T	G	0.0833	0.9167	0.9138	0.0862	0.00203252
22	14727317	G	A	0.9333	0.0667	0.4074	0.5926	0.04805726
22	14728513	T	G	0.0667	0.9333	0.8103	0.1897	0.00506586
22	14728576	C	T	0.0833	0.9167	0.9310	0.0690	0.00101937
22	14728581	A	G	0.0833	0.9167	0.9310	0.0690	0.00101937
22	14729846	A	G	0.0667	0.9333	0.7759	0.2241	0.00812183
22	14730633	T	G	0.1000	0.9000	0.8103	0.1897	0.01013171
22	14731040	C	G	0.0833	0.9167	0.9655	0.0345	0.00102249
22	14731933	C	T	0.0667	0.9333	0.9483	0.0517	0.0010142
22	14732046	C	A	0.0667	0.9333	0.9310	0.0690	0
22	14732334	A	G	0.0667	0.9333	0.9138	0.0862	0.00203252
22	14732350	G	T	0.0667	0.9333	0.9138	0.0862	0.00203252
22	14732442	G	C	0.0833	0.9167	0.9138	0.0862	0.00203252
22	14735282	C	A	0.0833	0.9167	0.8214	0.1786	0.01022495
22	14737840	C	A	0.0833	0.9167	0.9655	0.0345	0.00102249
22	14739153	A	G	1.0000	0.0000	0.8448	0.1552	0.50823045
22	14741789	C	T	0.0667	0.9333	0.9655	0.0345	0.00102249
22	14742595	G	T	0.0667	0.9333	0.8103	0.1897	0.00506586
22	14742646	A	C	0.0833	0.9167	0.9828	0.0172	0.00102459

22	14742671	T	C	0.0667	0.9333	0.7931	0.2069	0.00303337
22	14742688	A	G	1.0000	0.0000	0.8276	0.1724	0.55442523
22	14743145	T	C	0.0833	0.9167	0.7759	0.2241	0.01116751
22	14743158	T	C	0.0667	0.9333	0.7759	0.2241	0.00812183
22	14746303	G	A	0.0667	0.9333	0.9655	0.0345	0.00102249
22	14746772	C	T	0.0833	0.9167	0.7931	0.2069	0.00606673
22	14747209	A	C	0.0667	0.9333	0.7931	0.2069	0.00303337
22	14747224	A	G	0.0667	0.9333	0.7931	0.2069	0.00303337
22	14748791	T	A	0.0517	0.9483	0.9655	0.0345	0.00102249
22	14748820	G	C	0.0000	1.0000	0.0000	1.0000	0
22	14750586	T	C	0.0000	1.0000	0.2321	0.7679	0.67074414
22	14750780	G	A	0.9000	0.1000	0.9483	0.0517	0.14198783
22	14750791	G	A	0.9167	0.0833	0.9655	0.0345	0.10429448

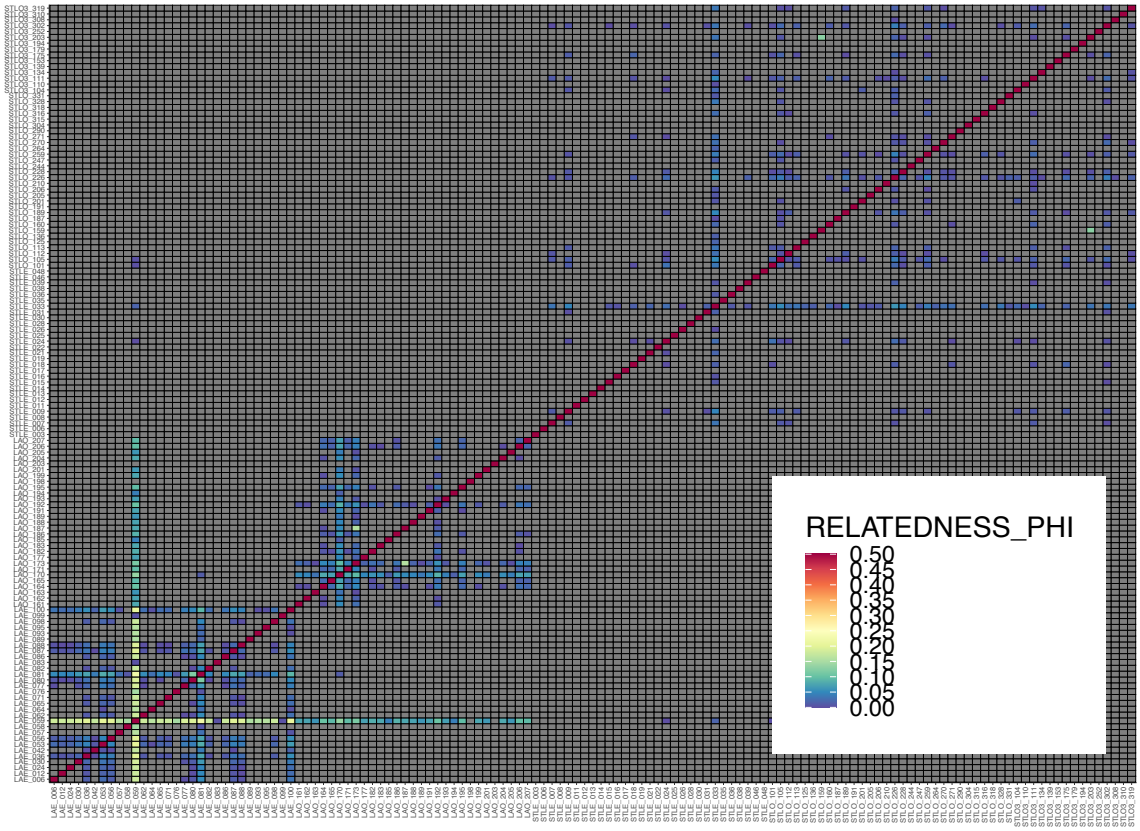


Figure S1. Plot of relatedness using the relatedness2 algorithm in VCFtools for individuals sequenced in this study. Note sample LAO_059, which has high relatedness across multiple sample groups was removed from the study. LAE and LAO refer to the even and odd sample groups from native range, while STLO, STLO3, and STLE refer to the two-year old odd, three-year old odd, and two-year old even sample groups in the introduced range (Table S1).

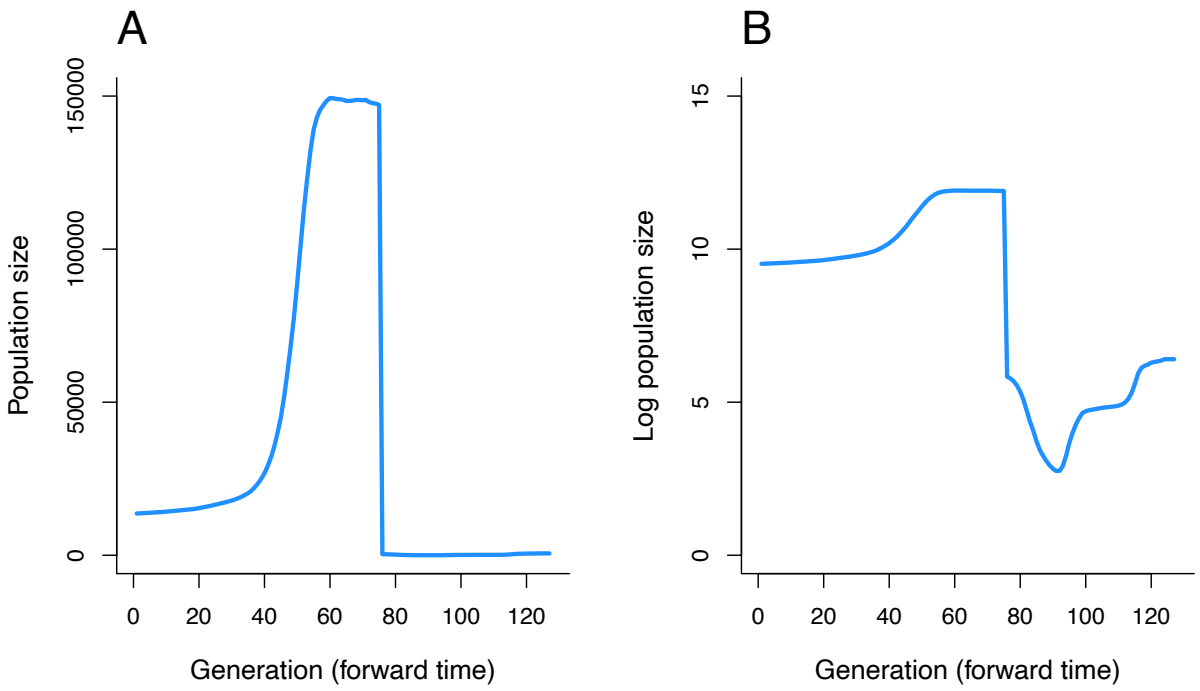


Figure S2. Population sizes used in the agent-based model. Values represent combined results of GONE estimates for BC odd and GL odd, where N_e values for BC odd were used for the first 75 generations (generations 100 to 26) and then switched to the GL odd N_e estimates for the next 52 generations (generations 52 to 1). Panel A shows raw values and panel B shows log-transformed values. N_e values for BC odd and GL odd can be found in figure 3. These population values are conservative because they use the entire bottleneck period estimated by GONE in the Great Lakes (generations 31-36), which extends further back in time than the actual bottleneck, and because here we used Ne as Nc in the model even though N_e:N_c ratios can be small (e.g., 0.2).

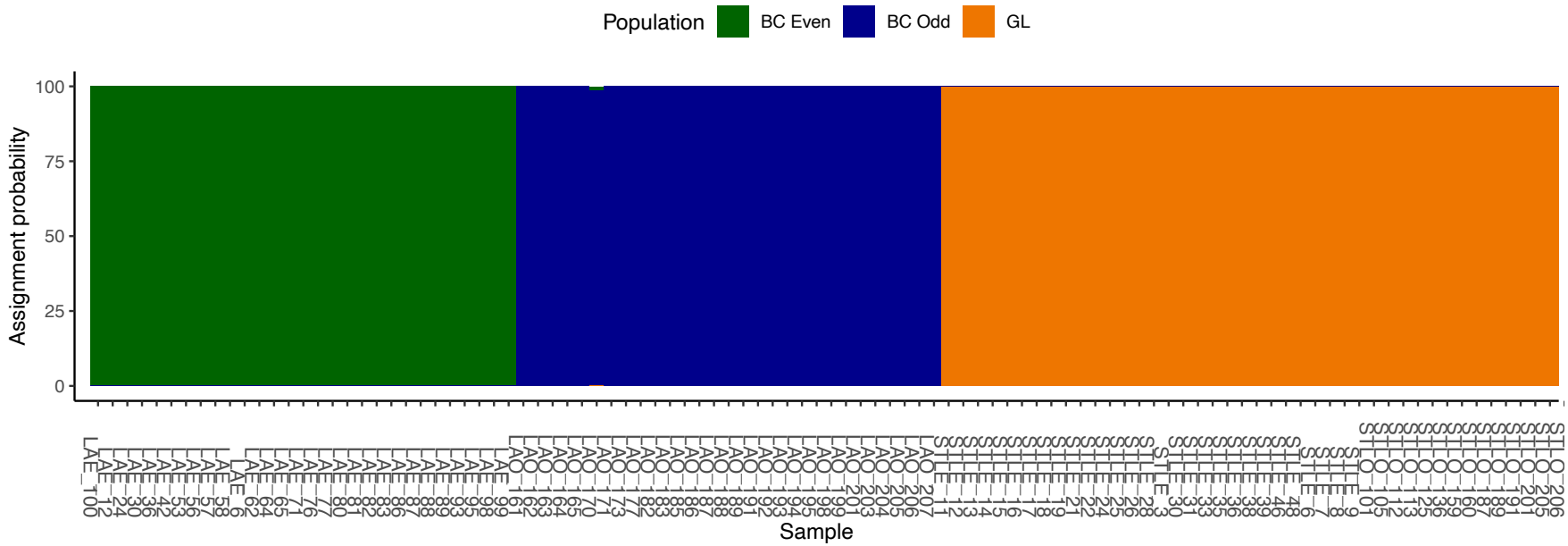


Figure S3. Population assignment determined by the program fastSTRUCTURE. Values of K= 1-5 were all evaluated and K=3 (shown here) was the best supported by the program. Note that all samples grouped according to their respective sample groups (LAE = BC Even, LAO = BC Odd, STLO/STLO3/STLE = Great Lakes, Table S1) with near perfect assignment probability.

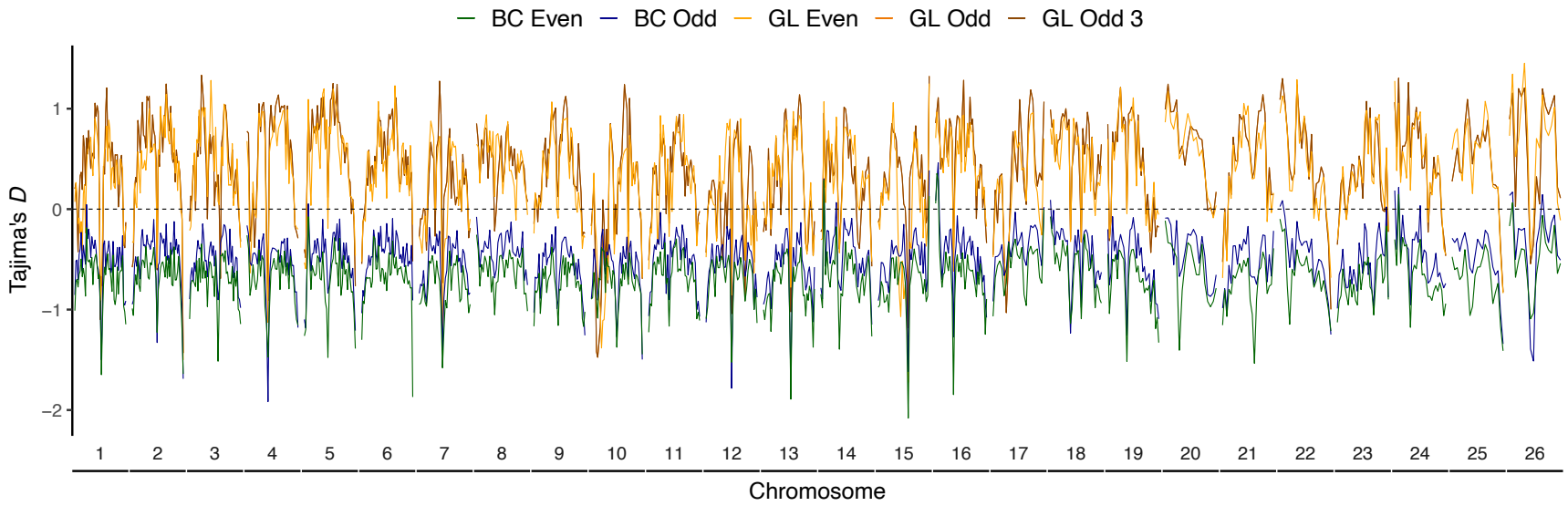


Figure S4. Windowed genome-wide Tajima's *D* (2.5 Mbp) for each sample group present in the study. Tajima's *D* was calculated using VCFtools and before the joint vcf was filtered for minor allele frequency to include low frequency SNPs.

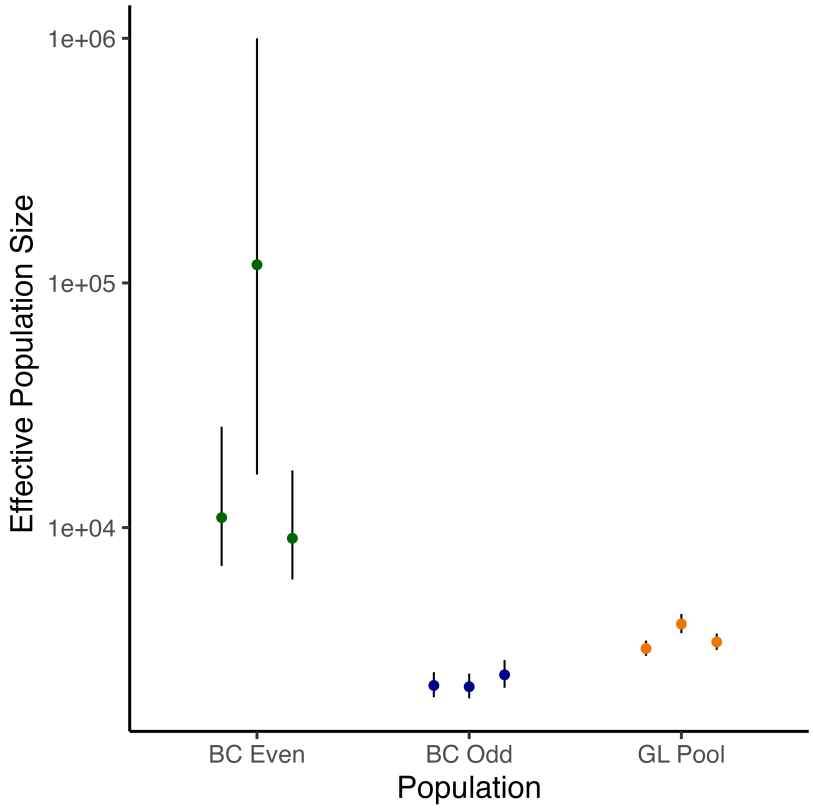
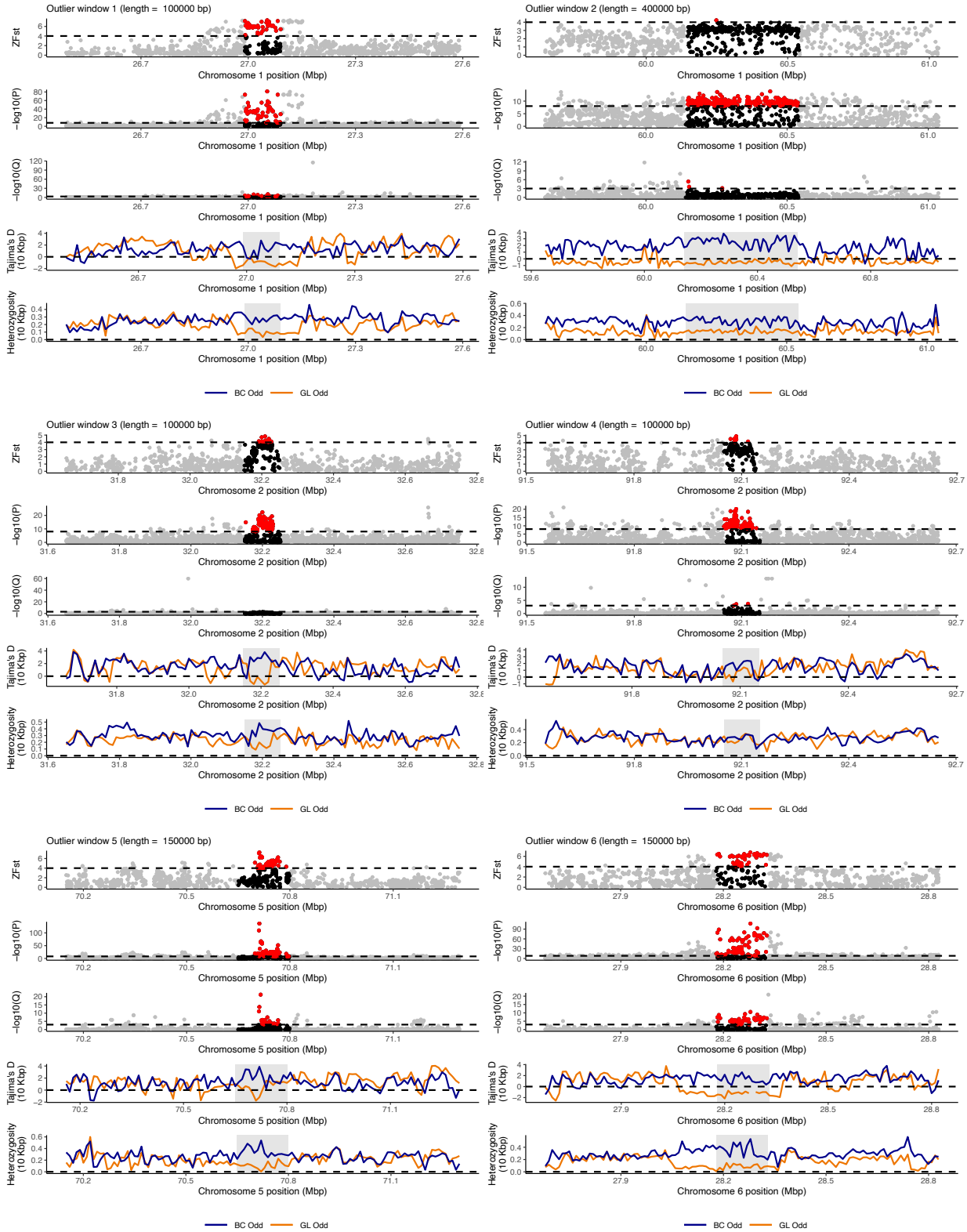
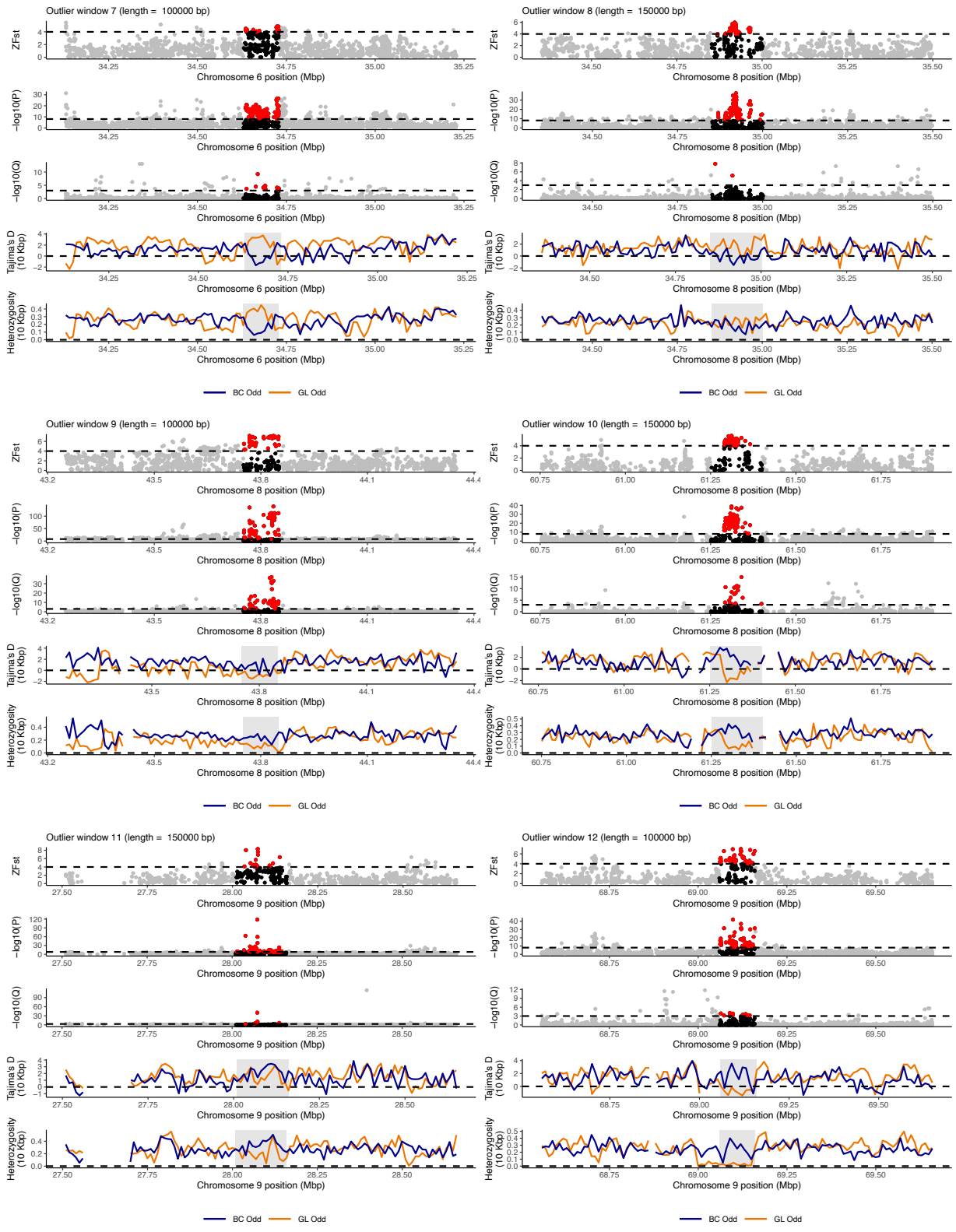
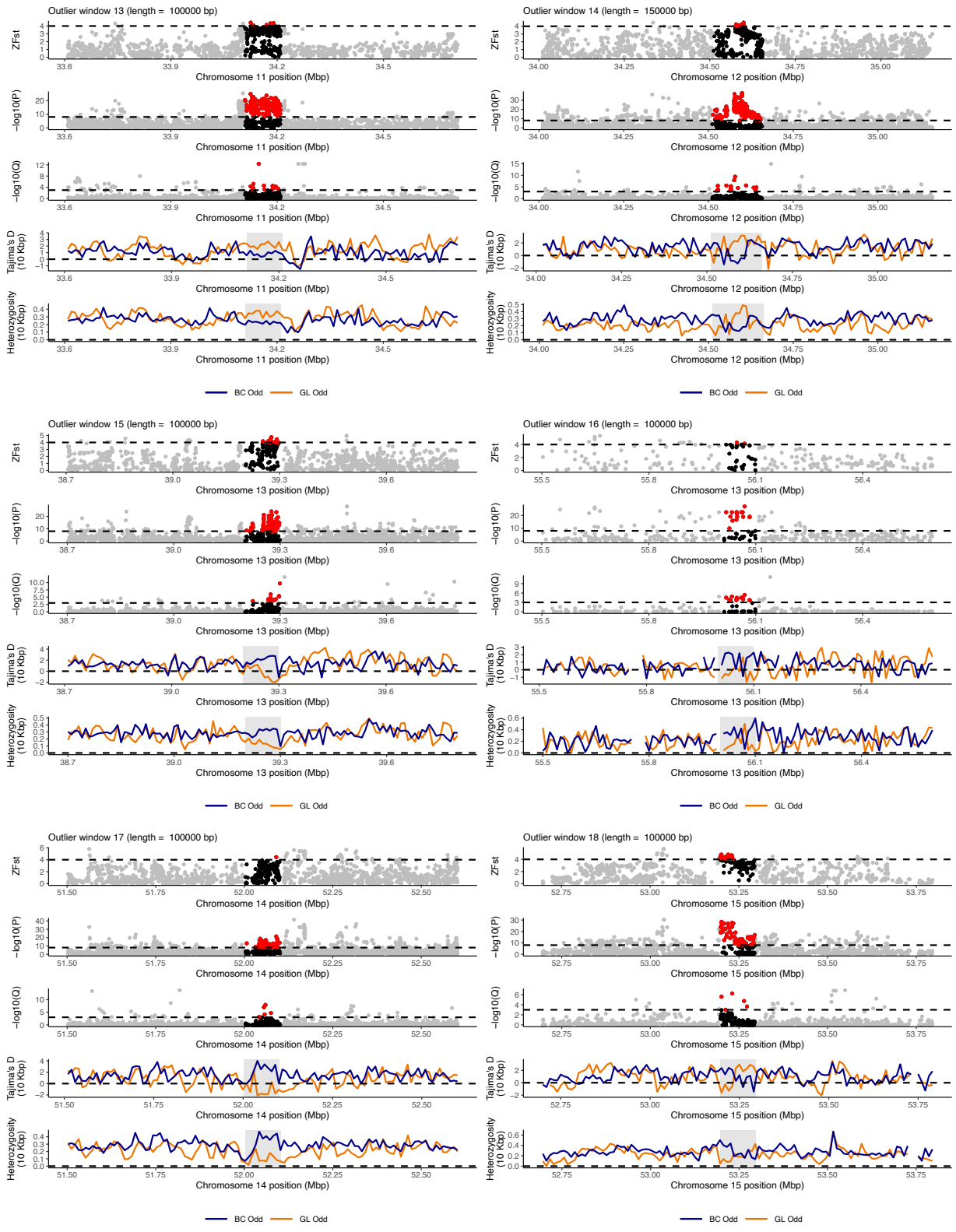
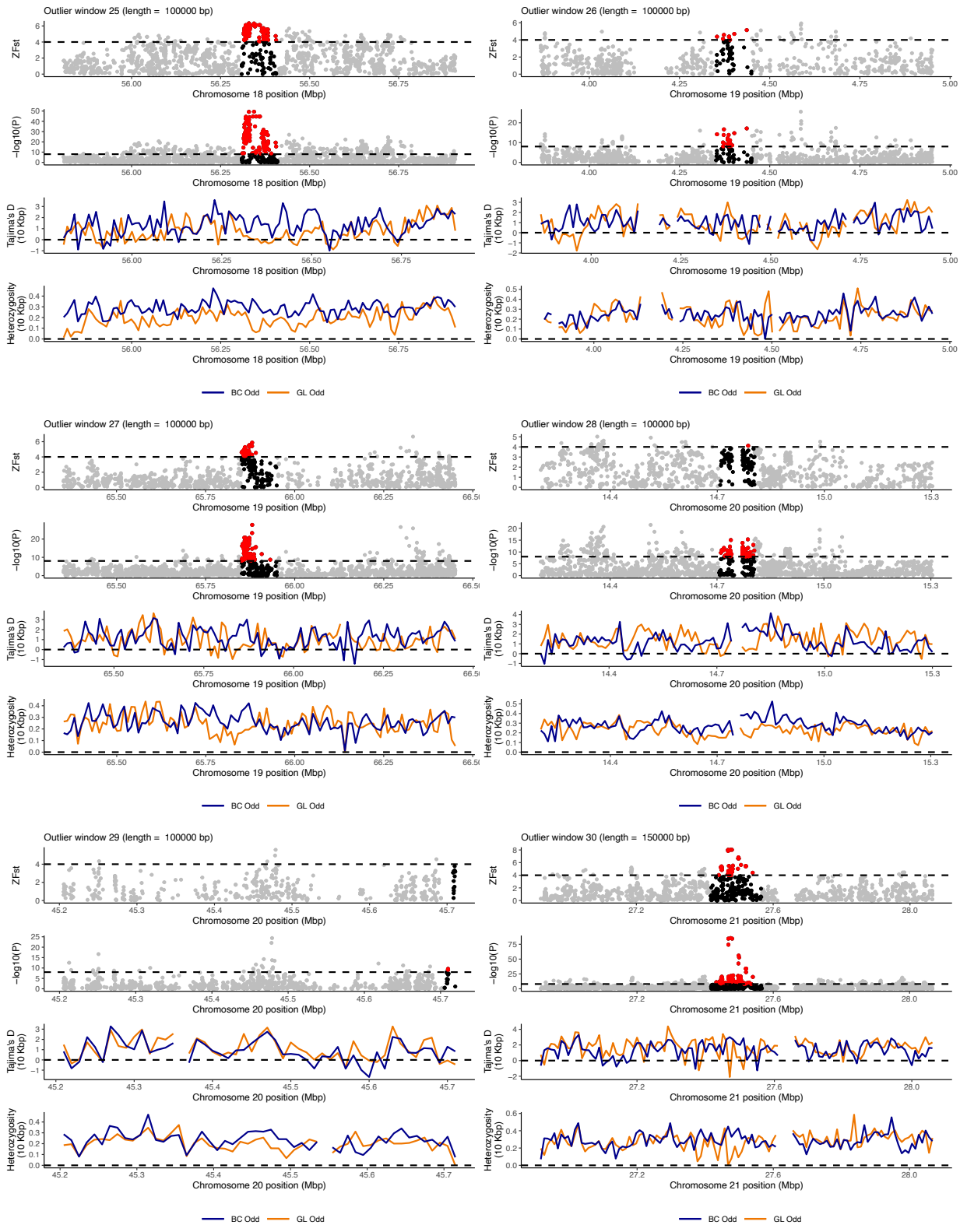


Figure S5. Contemporary effective population size estimates for sample groups in the native and introduced ranges of pink salmon. Lines show the extent of the upper and lower 95% confidence intervals. Estimates were generated using the program NeEstimator. Note that samples from the Great Lakes were pooled into an individual group and that the y-axis is on a log scale.


Outlier window 1 (length = 100000 bp)
Outlier window 2 (length = 400000 bp)
Outlier window 3 (length = 100000 bp)
Outlier window 4 (length = 100000 bp)
Outlier window 5 (length = 150000 bp)
Outlier window 6 (length = 150000 bp)
ZFst
−log10(P)
−log10(Q)
Tajima's D (10 Kbp)
Heterozygosity (10 Kbp)
Chromosome 1 position (Mbp)
Chromosome 2 position (Mbp)
Chromosome 5 position (Mbp)
Chromosome 6 position (Mbp)
BC Odd
GL Odd


Outlier window 7 (length = 100000 bp)
Outlier window 8 (length = 150000 bp)
Outlier window 9 (length = 100000 bp)
Outlier window 10 (length = 150000 bp)
Outlier window 11 (length = 150000 bp)
Outlier window 12 (length = 100000 bp)
ZFst
−log10(P)
−log10(Q)
Tajima's D (10 Kbp)
Heterozygosity (10 Kbp)
Chromosome 6 position (Mbp)
Chromosome 8 position (Mbp)
Chromosome 9 position (Mbp)
BC Odd
GL Odd


Outlier window 13 (length = 100000 bp)
Outlier window 14 (length = 150000 bp)
Outlier window 15 (length = 100000 bp)
Outlier window 16 (length = 100000 bp)
Outlier window 17 (length = 100000 bp)
Outlier window 18 (length = 100000 bp)
ZFst
−log10(P)
−log10(Q)
Tajima's D (10 Kbp)
Heterozygosity (10 Kbp)
Chromosome 11 position (Mbp)
Chromosome 12 position (Mbp)
Chromosome 13 position (Mbp)
Chromosome 14 position (Mbp)
Chromosome 15 position (Mbp)
BC Odd
GL Odd


Outlier window 25 (length = 100000 bp)
Outlier window 26 (length = 100000 bp)
Outlier window 27 (length = 100000 bp)
Outlier window 28 (length = 100000 bp)
Outlier window 29 (length = 100000 bp)
Outlier window 30 (length = 150000 bp)
ZFst
−log10(P)
Tajima's D (10 Kbp)
Heterozygosity (10 Kbp)
Chromosome 18 position (Mbp)
Chromosome 19 position (Mbp)
Chromosome 20 position (Mbp)
Chromosome 21 position (Mbp)
BC Odd
GL Odd

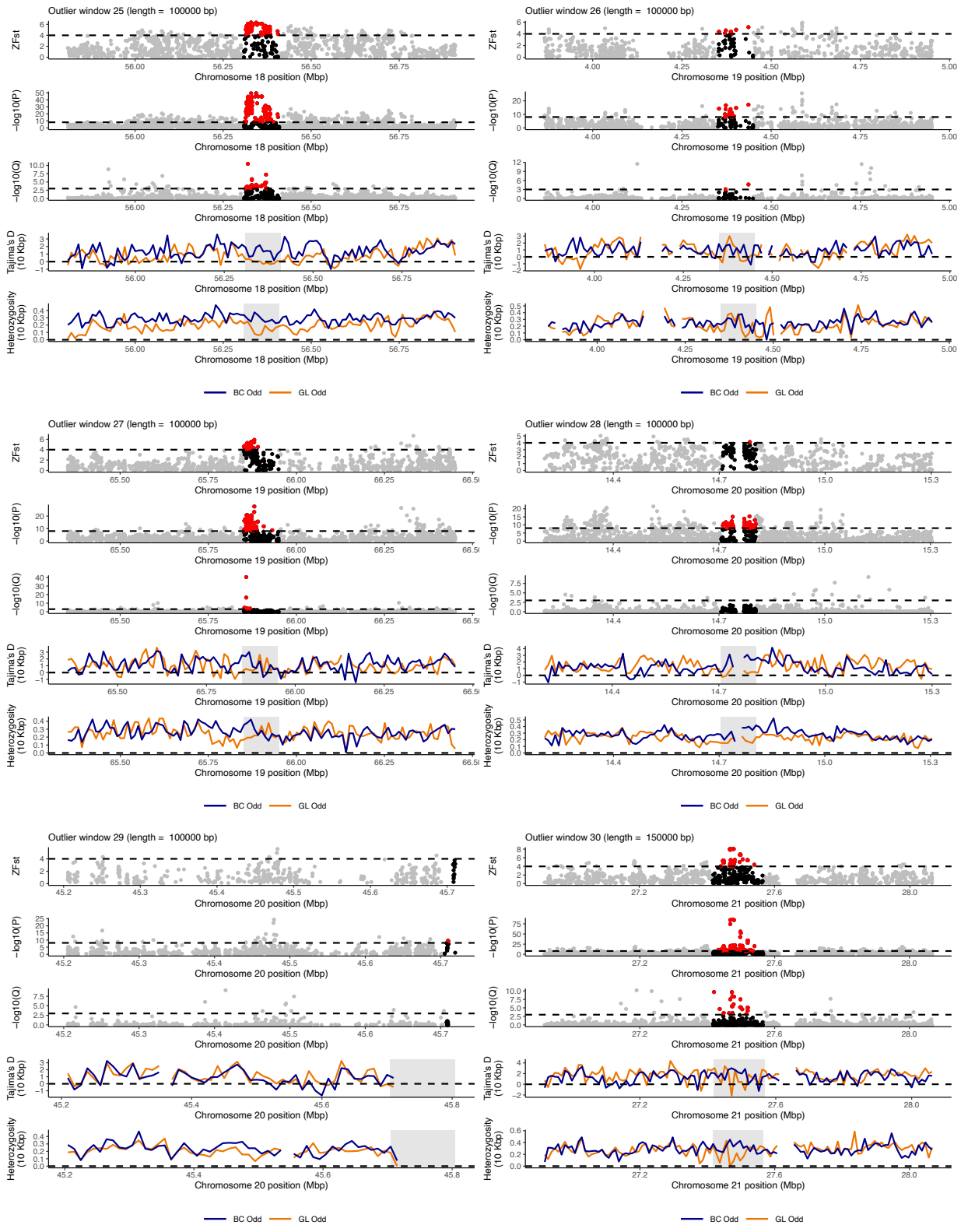


Figure S6. Plots of sliding window outlier regions (34 total) determined using a threshold of ≥ 5 ZF_{ST}. These plots represent combined windows if they were overlapping (47 total windows)

within a chromosome. Shown in each plot in descending order is the ZF_{ST} for each SNP, where outlier SNPs that are above the 4 ZF_{ST} threshold are colored red. The window itself is denoted with the black and red SNPs while the 500 kbp flanking region in the upstream and downstream directions is shown in grey for comparison. Shown in the next lower plot are the same SNPs presented in the same style but with -$\log_{10}$ transformed *p*-values from eigenGWAS with a Bonferonni corrected *p*-value threshold. The next window uses the same format but presents data from pcaadapt and presents FDR correct *Q*-values for a dataset of linkage disequilibrium (R^2 >/= 0.2) pruned SNPs. In the subsequent two windows are Tajima's *D* and heterozygosity presented in non-overlapping sliding windows of 10 kbp. Occasionally missing values are presented which are indicated by gaps in the lines of those plots. The grey polygon in each window represents the same window as the black SNPs for the first three panels. The colors of the lines correspond to the populations compared in the F_{ST} and eigenGWAS methods, blue for source BC odd population and orange for the introduced GL odd population. At the top of each plot is the window number which corresponds to the table in the supplemental data, as well as the length of the window (e.g., if it was combined with another step or not). Genes in the above windows are included in the SI Data file.

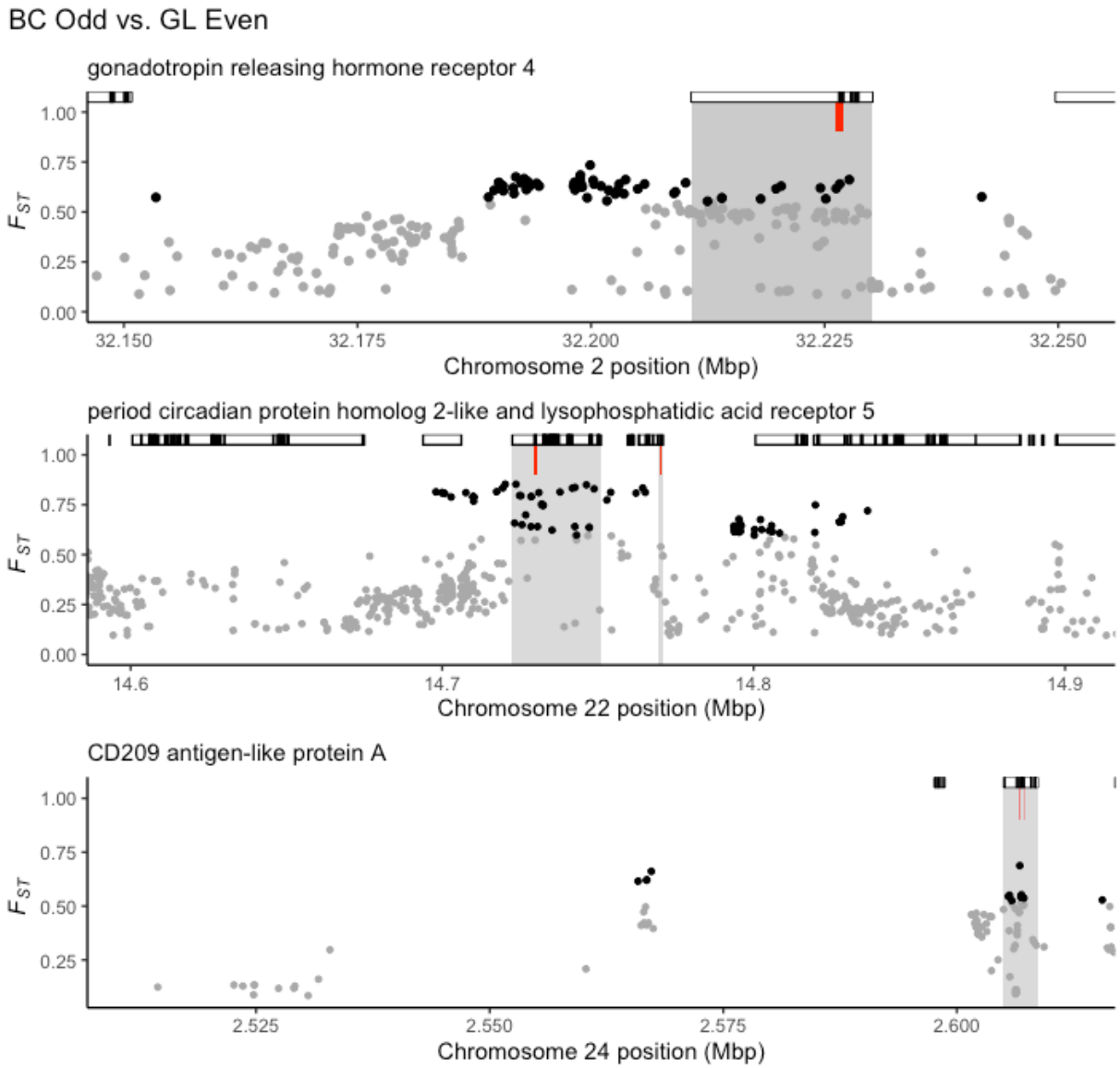


Figure S7. To validate our F_{ST} results from our archetypal sample group comparison (BC odd vs. GL odd) we also compared our BC odd (source) and GL even sample groups. Only 4 of the 9 windows with genes with missense variants were shared between the samples. Those windows and the genes with missense variants (red lines) are shown above. The raw F_{ST} values are presented here and those genes are indicated with grey polygons. Significantly differentiated SNPs are shown with black circles. Coding and non-coding regions of genes are represented by black filled and open rectangles, respectively.

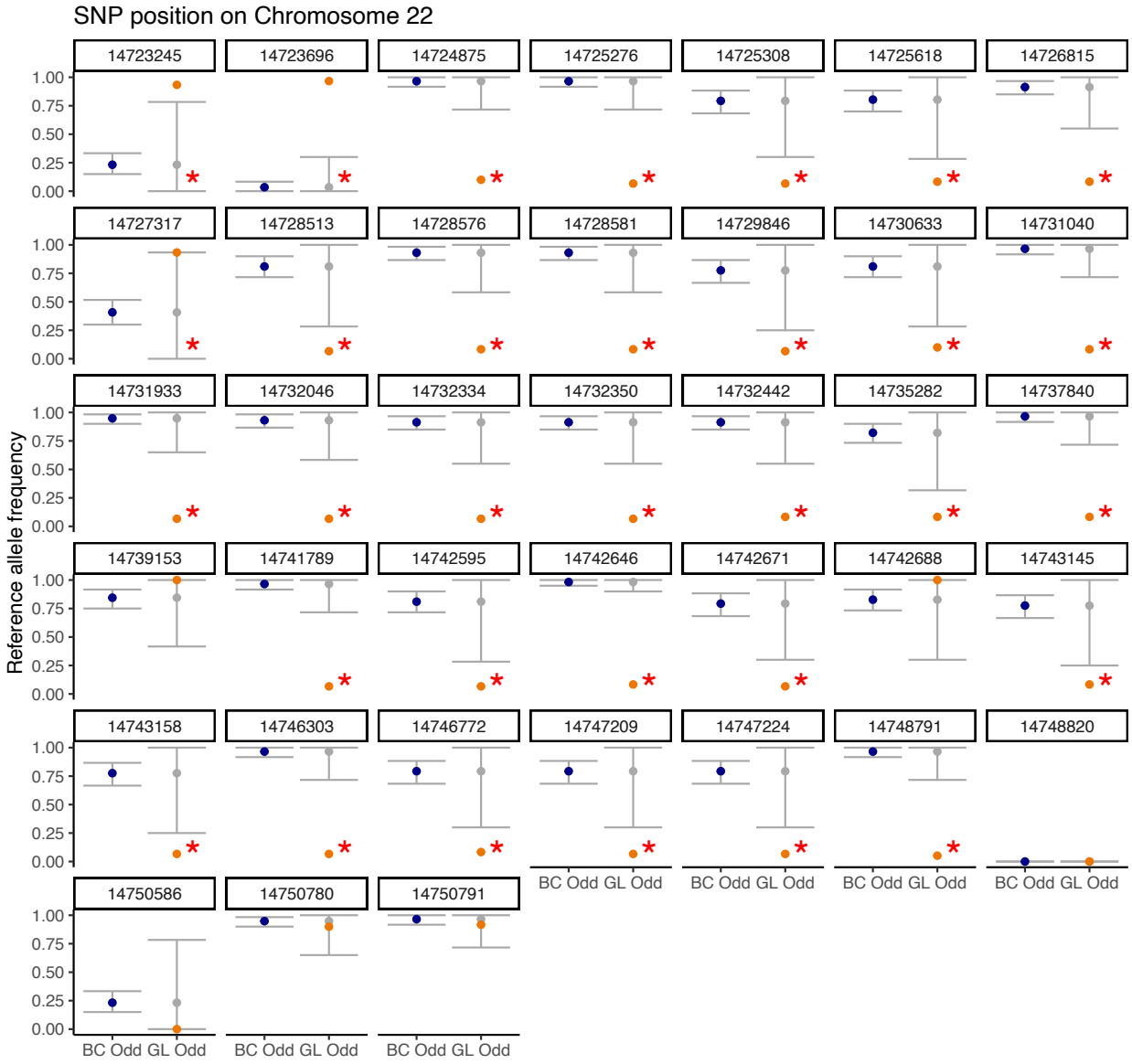


Figure S8. All SNPs present in *per2* locus (BC odd vs GL odd). Each individual figure shows the reference allele frequency of the SNP (position listed above) in the BC Odd (blue) and GL Odd (orange) sample groups. The grey point and error bars indicate the simulated allele frequency and 95% quantiles from our agent-based model using the BC odd allele frequency as the starting point. For each SNP, we also calculated the probability that the empirical estimate of the allele frequency in the introduced population could have occurred by chance alone (including the effects of both genetic drift and sample sizes). To calculate this probability, we counted the number of simulations that were equal to or more extreme than the empirically observed allele frequency and divided by 1000 (the number of simulations run for each of the 38 SNPs). For p-values less than 0.05, we considered the allele frequency change as being more likely to have been driven by selection; indicated by red asterisks in the right column (mean p-value = 0.0051; median = 0.0020). These values are indicated by solid lines in Figure 5b for the observed data. Notice, for significant SNPs, that small changes in the estimates of BC Odd allele frequencies would be unlikely to have a large effect given that CIs for the GL Odd are often far away from the empirical estimates (one exception being the SNP at location 14727317).

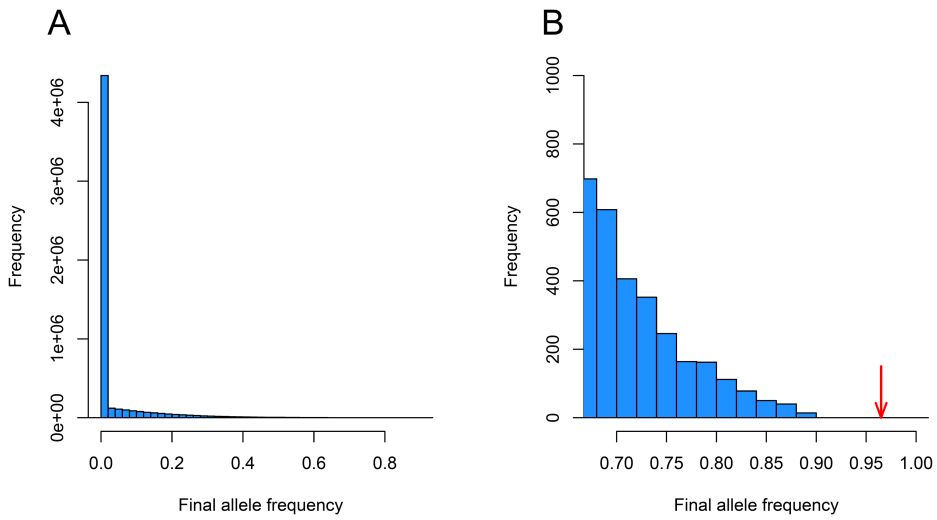


Figure S9. Results of 5,349,616 individual-based simulations for a *per2* allele with an empirically observed starting allele frequency of 0.0344 in the native-range, source population (BC) and a "final" allele frequency of 0.9655 in the introduced range (Great Lakes). We performed this large number of simulations to explore the possibility of chance alone (*e.g.*, drift) resulting in a large increase in allele frequency, especially when considering the observation that there were 5,080,474 SNPs identified throughout the genome. When examining the results of all 5,349,616 simulations (panel A), we see that in the vast majority of cases, the final allele frequency was equal to zero. When zooming in on the right side of the distribution of final, simulated allele frequencies (panel B), we see that not a single simulation resulted in a final allele frequency as large or larger than the allele frequency observed empirically (red arrow). This result suggests that the differences in allele frequencies for this particular SNP were highly unlikely to occur by chance. These analyses were computationally intensive, which is why they were performed for a single locus of large effect.

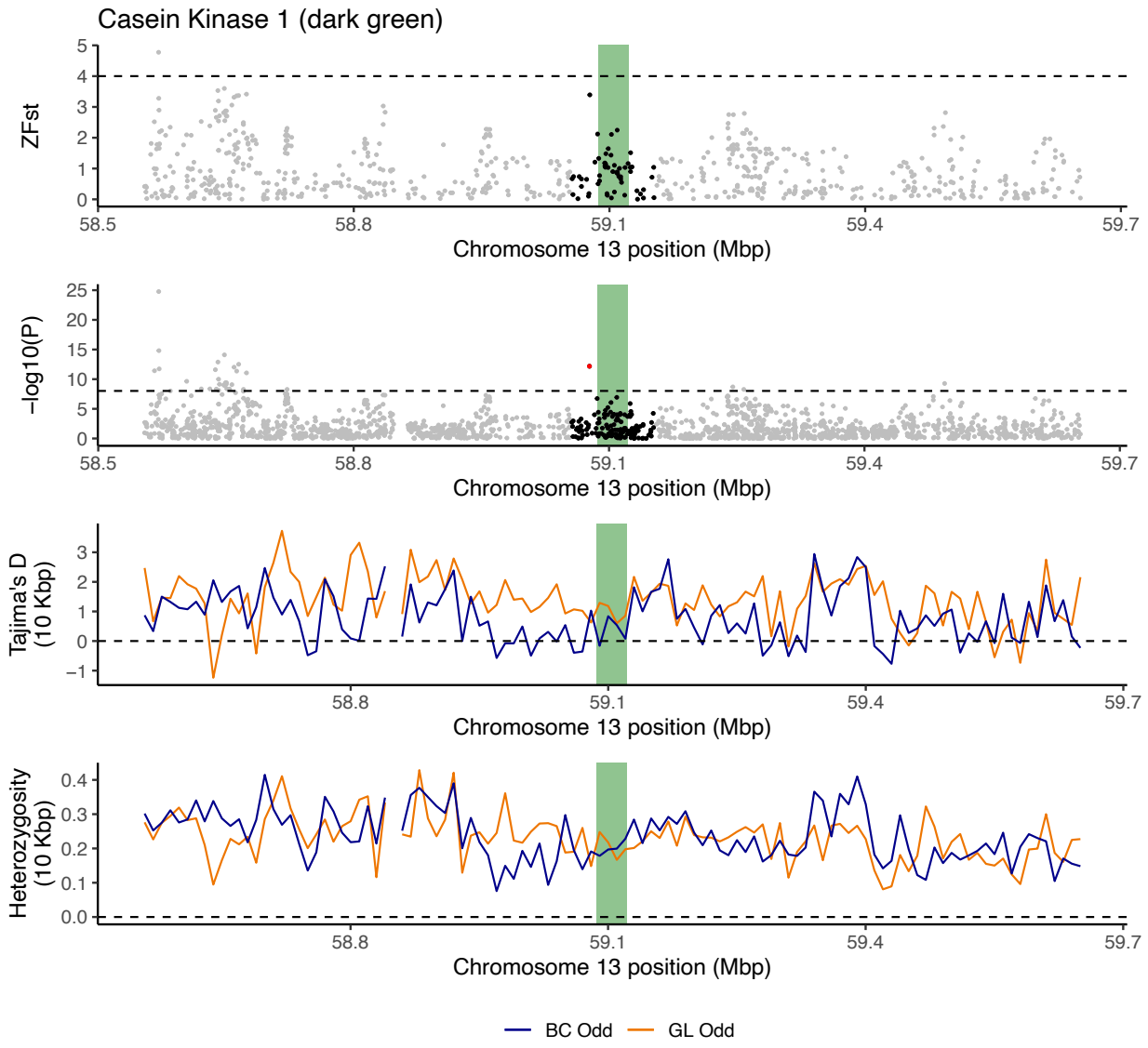


Figure S10. Population differentiation and genomic patterns within and flanking Casein Kinase 1 (green polygon), the priming kinase for *per2*. The top plot shows ZF_{ST} within the gene, the 100 kbp region around the gene (black points), and the flanking region upstream and downstream (gray points). The following plot shows the same setup but using eigenGWAS -$\log_{10}$ tranformed *p*-values. Significant SNPs in both are represented by a red point, which occur above the dashed lines. The final two plots show windowed Tajima's *D* and heterozygosity in 10 kbp windows for each sample group.