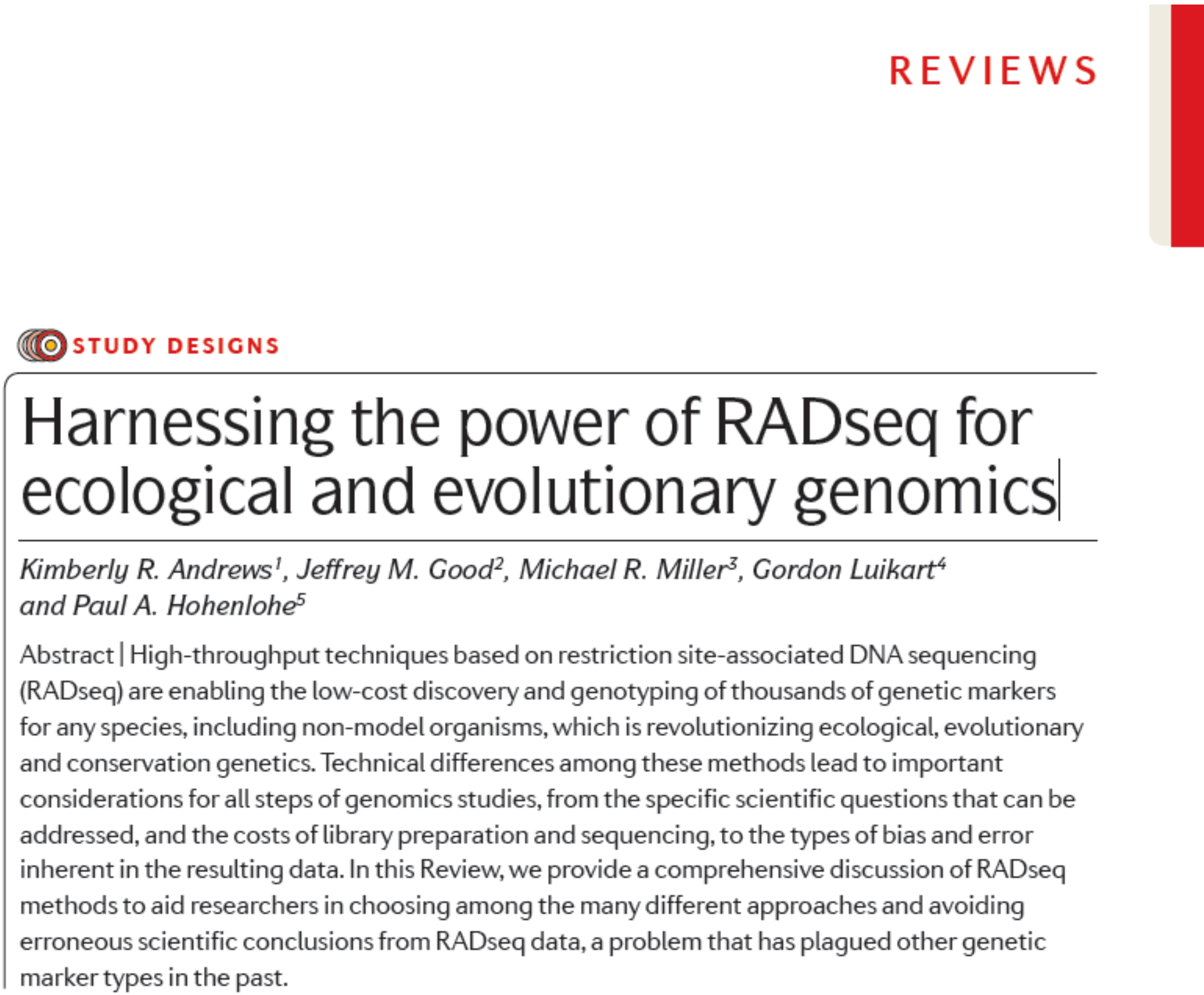


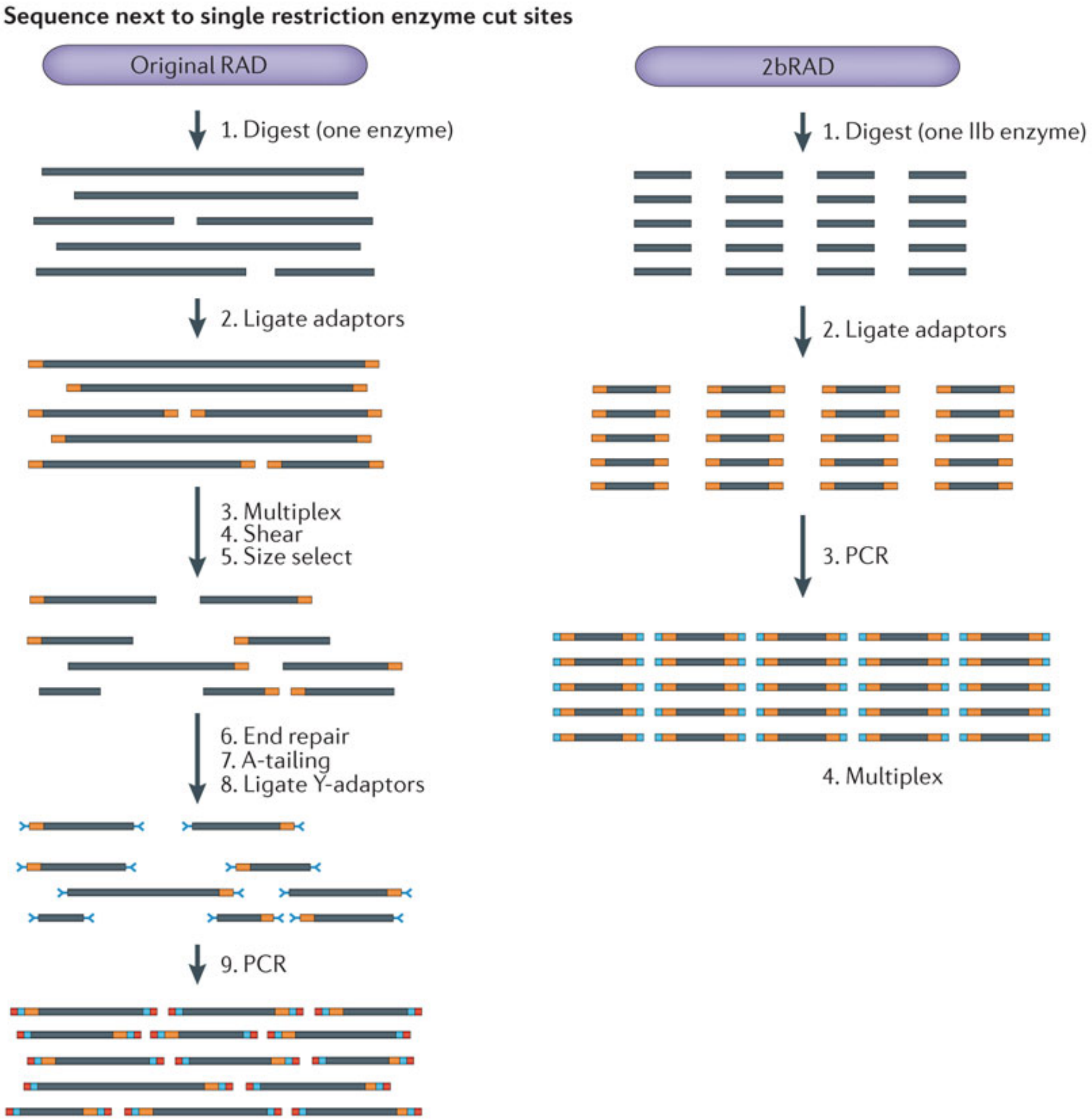
REVIEWS

STUDY DESIGNS

Harnessing the power of RADseq for ecological and evolutionary genomics

Kimberly R. Andrews[1], *Jeffrey M. Good*[2], *Michael R. Miller*[3], *Gordon Luikart*[4] *and Paul A. Hohenlohe*[5]

Abstract | High-throughput techniques based on restriction site-associated DNA sequencing (RADseq) are enabling the low-cost discovery and genotyping of thousands of genetic markers for any species, including non-model organisms, which is revolutionizing ecological, evolutionary and conservation genetics. Technical differences among these methods lead to important considerations for all steps of genomics studies, from the specific scientific questions that can be addressed, and the costs of library preparation and sequencing, to the types of bias and error inherent in the resulting data. In this Review, we provide a comprehensive discussion of RADseq methods to aid researchers in choosing among the many different approaches and avoiding erroneous scientific conclusions from RADseq data, a problem that has plagued other genetic marker types in the past.


Sequence next to single restriction enzyme cut sites
Original RAD
1. Digest (one enzyme)
2. Ligate adaptors
3. Multiplex
4. Shear
5. Size select
6. End repair
7. A-tailing
8. Ligate Y-adaptors
9. PCR
2bRAD
1. Digest (one IIb enzyme)
2. Ligate adaptors
3. PCR
4. Multiplex

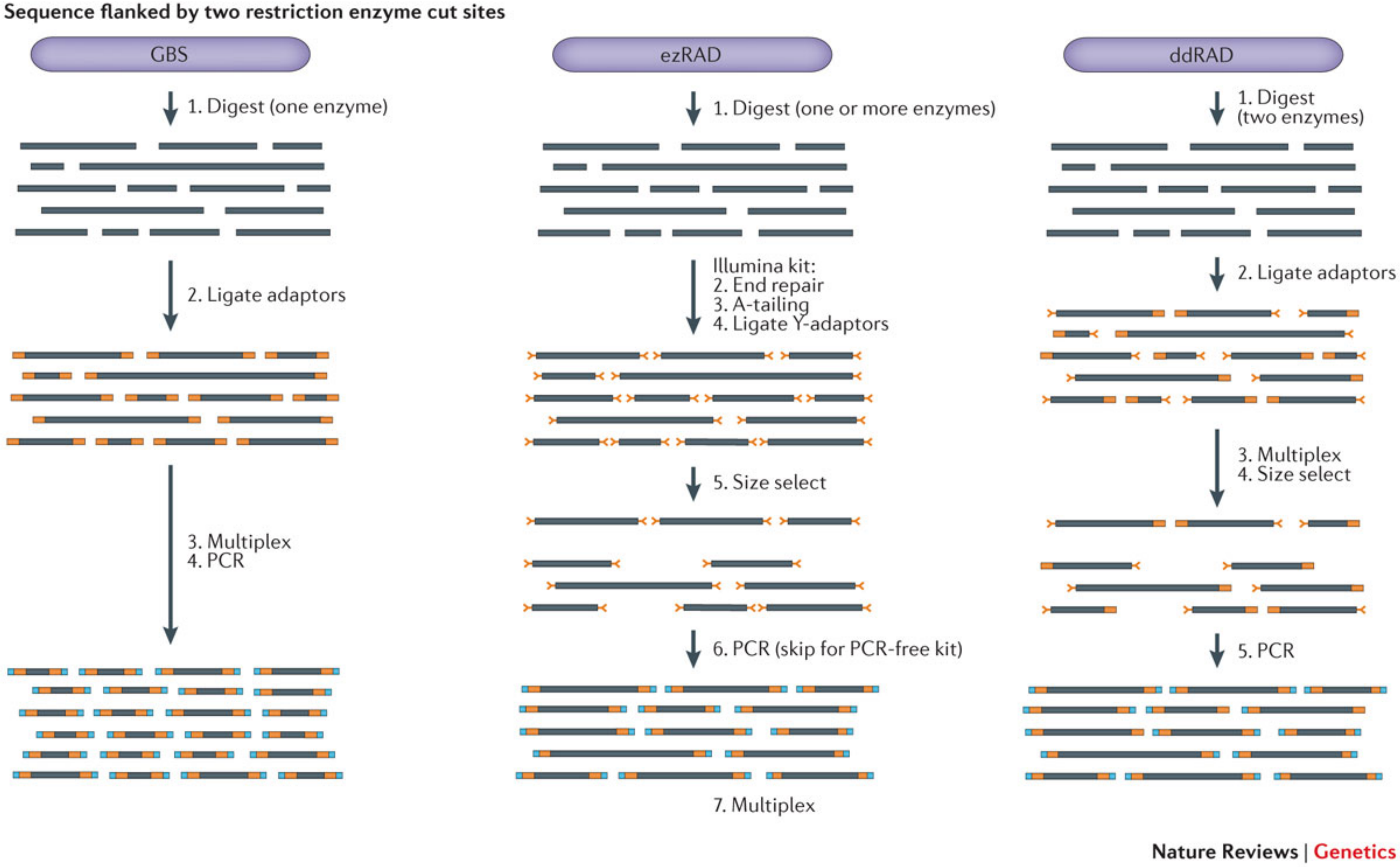

Sequence flanked by two restriction enzyme cut sites
GBS
1. Digest (one enzyme)
2. Ligate adaptors
3. Multiplex
4. PCR
ezRAD
1. Digest (one or more enzymes)
Illumina kit:
2. End repair
3. A-tailing
4. Ligate Y-adaptors
5. Size select
6. PCR (skip for PCR-free kit)
7. Multiplex
ddRAD
1. Digest
(two enzymes)
2. Ligate adaptors
3. Multiplex
4. Size select
5. PCR
Nature Reviews | Genetics

Table 1 | **Summary of trade-offs among five RADseq methods**

	Original RAD	**2bRAD**	**GBS**	**ddRAD**	**ezRAD**
Options for tailoring number of loci	Change restriction enzyme	Change restriction enzyme	Change restriction enzyme	Change restriction enzyme or size selection window	Change restriction enzyme or size selection window
Number of loci per 1 Mb of genome size*	30–500	50–1,000	5–40	0.3–200	10–800
Length of loci	≤1 kb if building contigs; otherwise ≤300 bp[‡]	33–36 bp	<300 bp[‡]	≤300 bp[‡]	≤300 bp[‡]
Cost per barcoded or indexed sample	Low	Low	Low	Low	High
Effort per barcoded or indexed sample[§]	Medium	Low	Low	Low	High
Use of proprietary kit	No	No	No	No	Yes
Identification of PCR duplicates	With paired-end sequencing	No	With degenerate barcodes	With degenerate barcodes	No
Specialized equipment needed	Sonicator	None	None	Pippin Prep[‖]	Pippin Prep[‖]
Suitability for large or complex genomes[¶]	Good	Poor	Moderate	Good	Good
Suitability for *de novo* locus identification (no reference genome)[#]	Good	Poor	Moderate	Moderate	Moderate
Available from commercial companies	Yes	No	Yes	Yes	No

*Estimated as follows: original restriction site-associated DNA sequencing (RADseq), assuming either a 6-cutter or an 8-cutter; 2bRAD, assuming type IIB enzymes with recognition sites containing 5–7 specific nucleotides; genotyping by sequencing (GBS), data from Elshire *et al.*[6]; double digest RAD (ddRAD), data from Table 1 in Peterson *et al.*[17] and allowing for up to double the size range; ezRAD, data from Toonen *et al.*[16] for species with reference genomes. [‡]Based on current single-end read-length limits in sequencing technology. [§]Assumes individual barcoding of many samples. [‖]Can alternatively be used with standard gel equipment. [¶]Based on the ability to reduce the total number of loci and lengths of loci. [#]Based on the lengths of loci to distinguish paralogues and duplicate sequence.

MOLECULAR ECOLOGY

Molecular Ecology (2013) **22**, 2864–2883 doi: 10.1111/mec.12330

The population structure and recent colonization history of Oregon threespine stickleback determined using restriction-site associated DNA-sequencing

JULIAN CATCHEN,[1] SUSAN BASSHAM,[1] TAYLOR WILSON, MARK CURREY, CONOR O' BRIEN, QUICK YEATES and WILLIAM A. CRESKO
Institute of Ecology and Evolution, University of Oregon, Eugene, OR 97403, USA

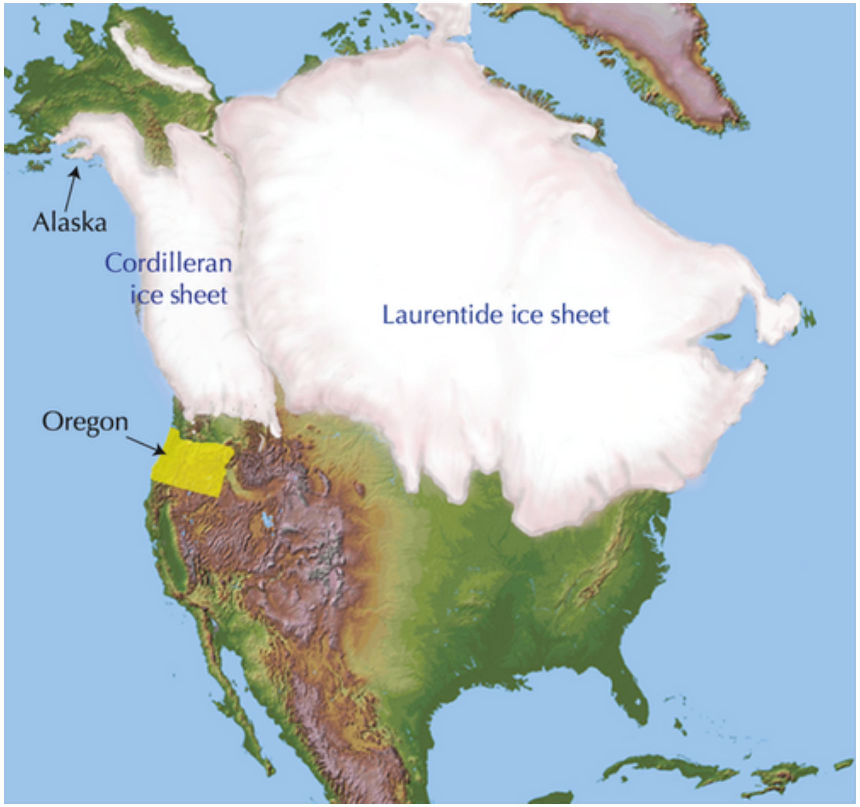


Figure 1. Extent of the last glacial maximum approximately 18 000 years ago, showing how the present-day State of Oregon outlined in yellow was not glaciated during this period.

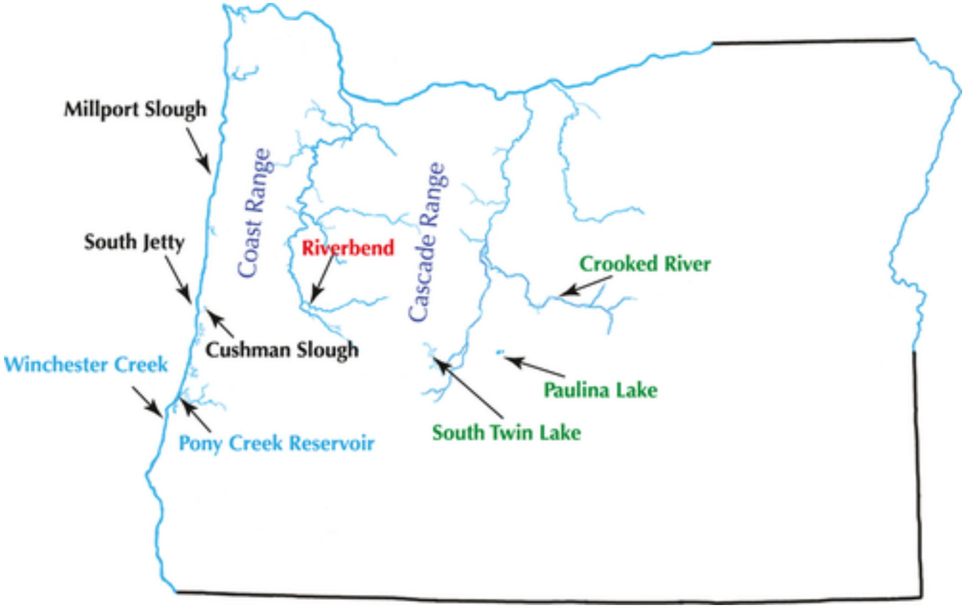


Figure 2. Map of Oregon sample locations. Populations analysed in this study come from the Pacific coast (Millport Slough, South Jetty, Cushman Slough, Winchester Creek, Pony Creek Reservoir), the Willamette Basin (Riverbend in the McKenzie River) and central Oregon (South Twin Lake, Paulina Lake and Crooked River).

Table 3 Summary genetic statistics for all populations split into those calculated for only nucleotide positions that are polymorphic in at least one Oregon population (top, 'Variant positions'), as well as all nucleotide positions across all restriction-site associated DNA (RAD) sites regardless of whether they are polymorphic or fixed (bottom, 'All positions'). These statistics include the average number of individuals genotyped at each locus (*N*), the number of variable sites unique to each population (Private), the number of polymorphic (top) or total (bottom) nucleotide sites across the data set (Sites), percentage of polymorphic loci (% poly), the average frequency of the major allele (P), the average observed heterozygosity per locus (H_{obs}), the average nucleotide diversity (π), and the average Wright's inbreeding coefficient (F_{IS})

	N	Private	Sites	% poly	P	H_{obs}	π	F_{IS}
Variant positions								
Millport Slough	65.70	7195	91 466	45.62	0.958	0.0567	0.0631	0.0341
Cushman Slough	95.61	13 110	114 056	60.24	0.952	0.0667	0.0734	0.0395
South Jetty	85.40	13 813	114 036	61.68	0.949	0.0690	0.0766	0.0435
Pony Creek Reservoir	67.66	2139	114 651	32.06	0.951	0.0664	0.0710	0.0190
Winchester Creek	21.78	3420	114 553	38.70	0.948	0.0714	0.0786	0.0258
Riverbend	134.12	3412	91 726	26.33	0.961	0.0540	0.0584	0.0176
Paulina Lake	20.55	392	114 727	6.55	0.983	0.0229	0.0230	0.0011
Crooked River	22.63	1409	114 743	8.91	0.980	0.0282	0.0284	0.0013
South Twin Lake	49.63	271	114 742	10.31	0.978	0.0294	0.0296	0.0015
All positions								
Millport Slough	66.61	7195	1 897 050	2.20	0.998	0.0027	0.0030	0.0016
Cushman Slough	96.36	13 110	2 433 350	2.82	0.998	0.0031	0.0034	0.0018
South Jetty	86.01	13 813	2 433 310	2.89	0.998	0.0032	0.0036	0.0020
Pony Creek Reservoir	68.06	2139	2 434 000	1.51	0.998	0.0031	0.0033	0.0009
Winchester Creek	21.88	3420	2 433 900	1.82	0.998	0.0034	0.0037	0.0012
Riverbend	136.32	3412	1 897 210	1.27	0.998	0.0026	0.0028	0.0009
Paulina Lake	20.62	271	2 434 030	0.31	0.999	0.0011	0.0011	0.0001
Crooked River	22.71	392	2 434 140	0.42	0.999	0.0013	0.0013	0.0001
South Twin Lake	49.81	1409	2 434 170	0.49	0.999	0.0014	0.0014	0.0001

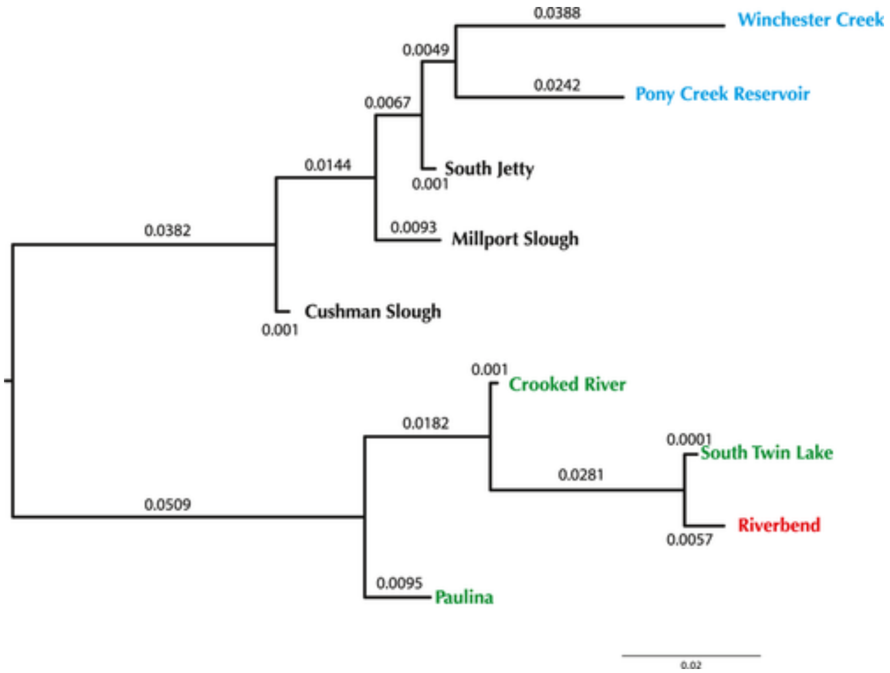


Figure 4. Neighbour-joining tree created using the pairwise FST values as a distance metric. The tree shows a clear split between the coastal populations (top; oceanic in black and freshwater in blue) and the inland populations (bottom). The central Oregon populations (green) clearly cluster together with the Willamette Basin population from the McKenzie River site at Riverbend (red), supporting the hypothesis of an introduction into central Oregon of stickleback from the Willamette Basin.

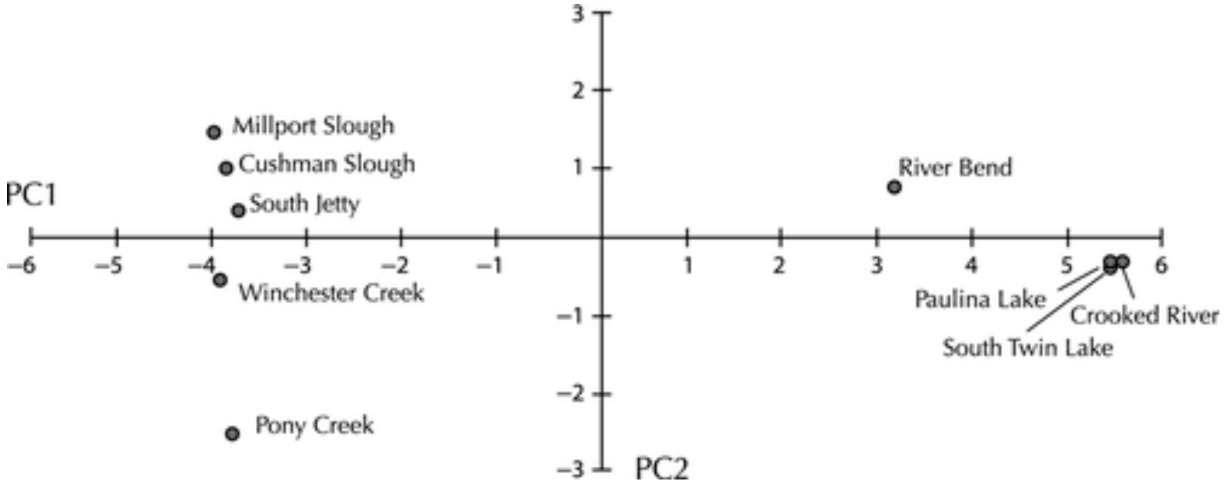


Figure 5. Distribution of average scores along PC1 and PC2 axes of genetic variation for each population. PC1 encompasses the vast majority of the genetic variation (~90%), and clearly differentiates the coastal Oregon stickleback populations from those in the Willamette Valley and central Oregon. Although PC2 explains less of the overall genetic variation, within each region it still has some explanatory power for partitioning populations.

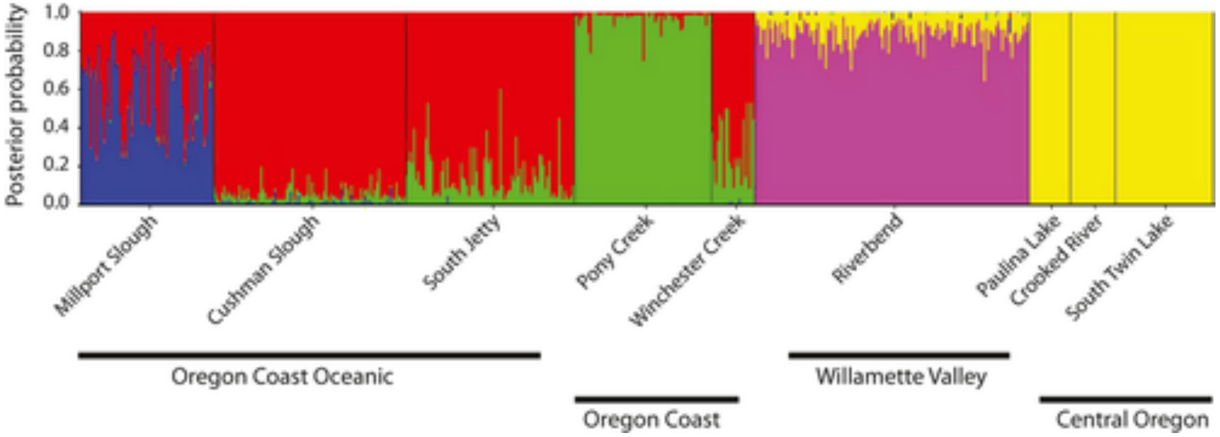


Figure 7. Plots of posterior probabilities of group assignment of each individual into five clusters based upon the results of a structure analysis. The results are grouped by population of origin for each individual. Each vertical bar represents a different individual from one of nine populations. The colour proportion for each bar represents the posterior probability of assignment of each individual to one of five clusters of genetic similarity.

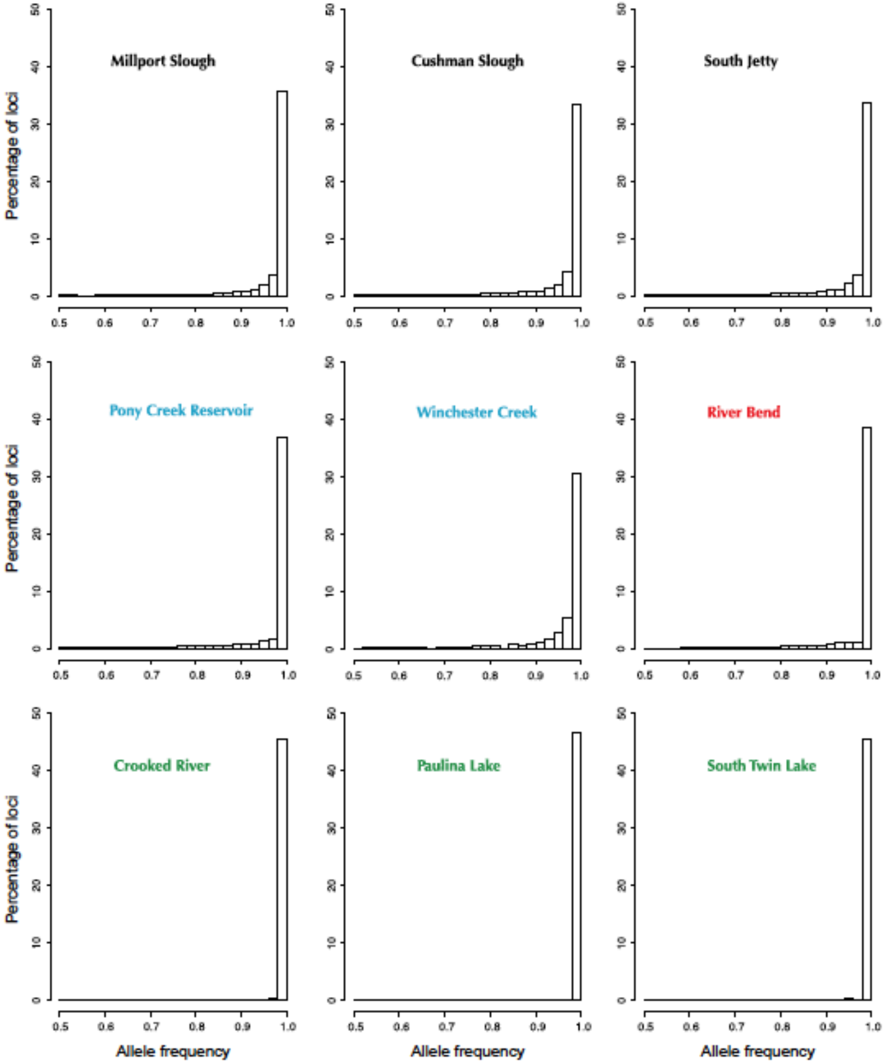


Figure 3. Allele frequency spectrum distribution for loci in each population that are SNPs in the Oregon system. The x-axis represents categories of the major allele frequency (p in each population) with the frequency distribution of the loci in each category on the y-axis. The majority of loci that are polymorphic across populations are fixed within each population (right-most category equal to 1.0). For coastal and Willamette Basin populations, a similar range of allele frequencies is seen, from 0.5 to 0.99 as expected for older populations at equilibrium. However, in central Oregon populations (bottom row), the distribution is clearly shifted to the right, as expected, if these populations were recently founded by a small number of individuals. (Coastal oceanic, black; coastal freshwater, blue; Willamette Basin, red; central Oregon, green).

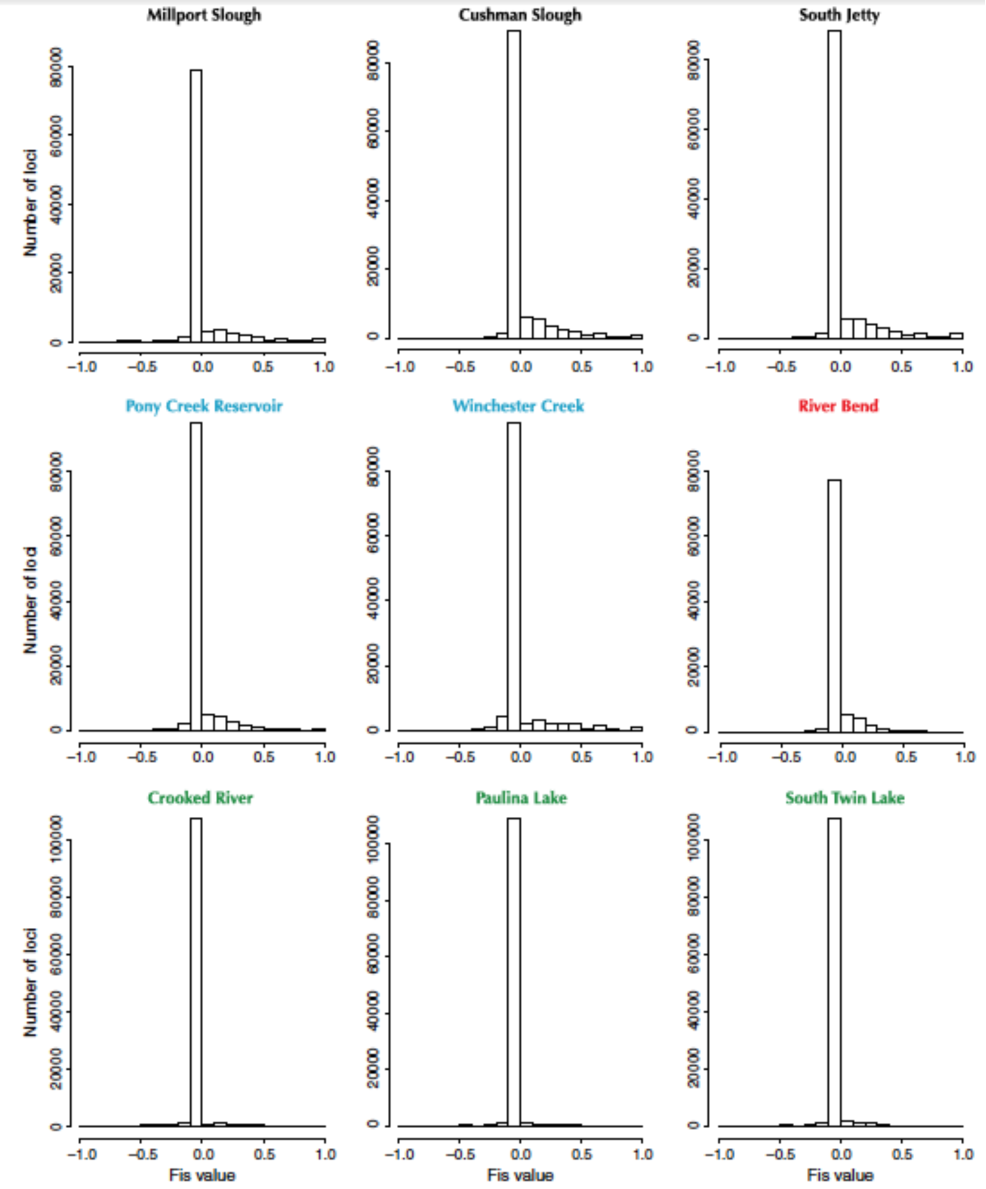
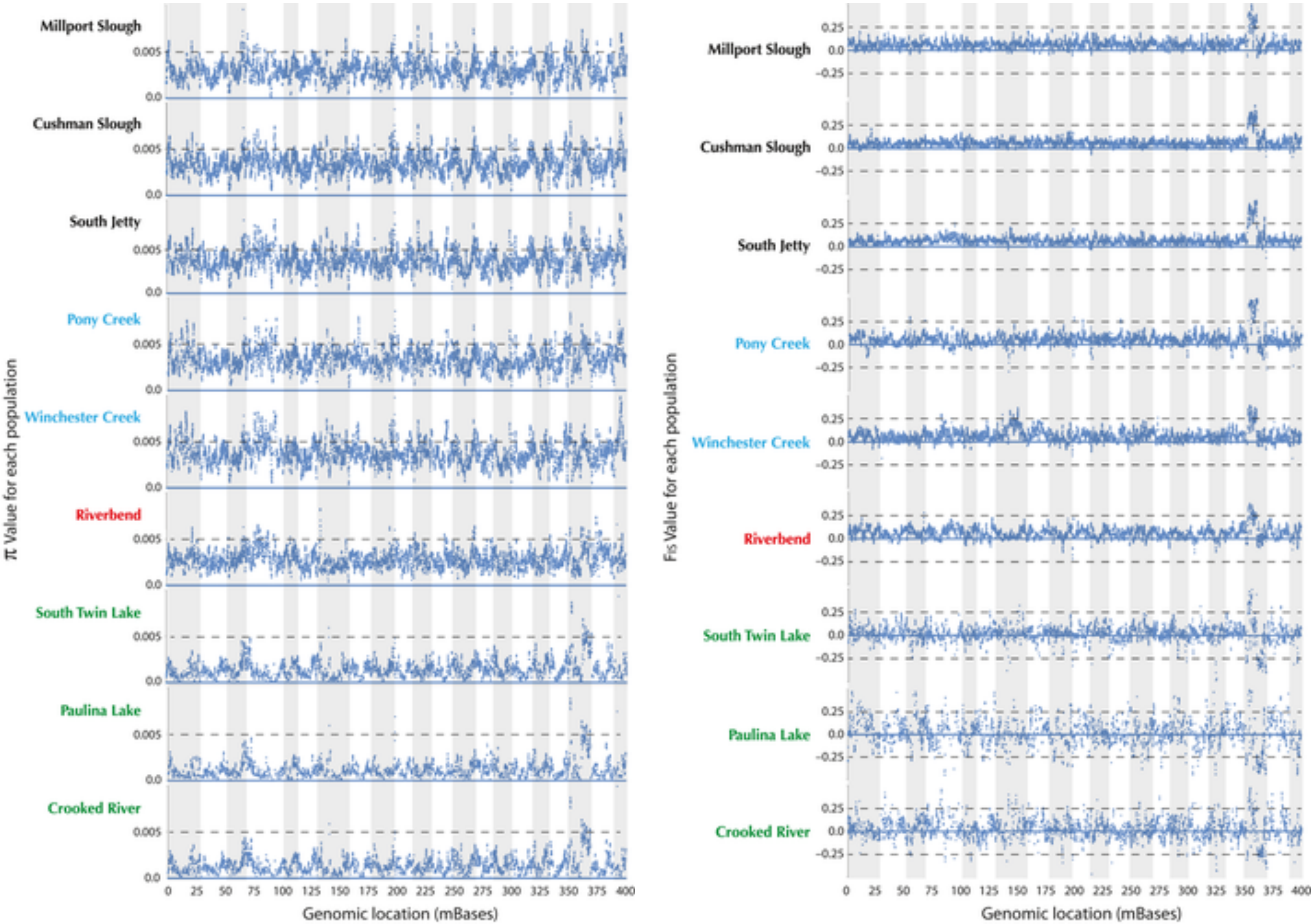


Figure 6. Frequency distribution of FIS values across loci within each population. The majority of loci within each population are zero or nearly so, indicating a lack of pervasive overall structure within each population. However, for several populations, including the coastal oceanic populations (top row), a noticeable fraction of loci exhibit values >1, including an appreciable number that are 1.0. These higher values are present in the coastal freshwater populations (Pony Creek and Winchester Creek), less in Riverbend, and noticeably depleted in the central Oregon populations (bottom row). (Coastal oceanic, black; coastal freshwater, blue; Willamette Basin, red; central Oregon, green).



π Value for each population
Millport Slough
Cushman Slough
South Jetty
Pony Creek
Winchester Creek
Riverbend
South Twin Lake
Paulina Lake
Crooked River
0.005
0.0
0 25 50 75 100 125 150 175 200 225 250 275 300 325 350 375 400
Genomic location (mBases)
Fis Value for each population
Millport Slough
Cushman Slough
South Jetty
Pony Creek
Winchester Creek
Riverbend
South Twin Lake
Paulina Lake
Crooked River
0.25
0.0
-0.25
0 25 50 75 100 125 150 175 200 225 250 275 300 325 350 375 400
Genomic location (mBases)