Current Biology, Volume *29*

Supplemental Information

A Single Interacting Species Leads to Widespread Parallel Evolution of the Stickleback Genome

Sara E. Miller, Marius Roesti, and Dolph Schluter

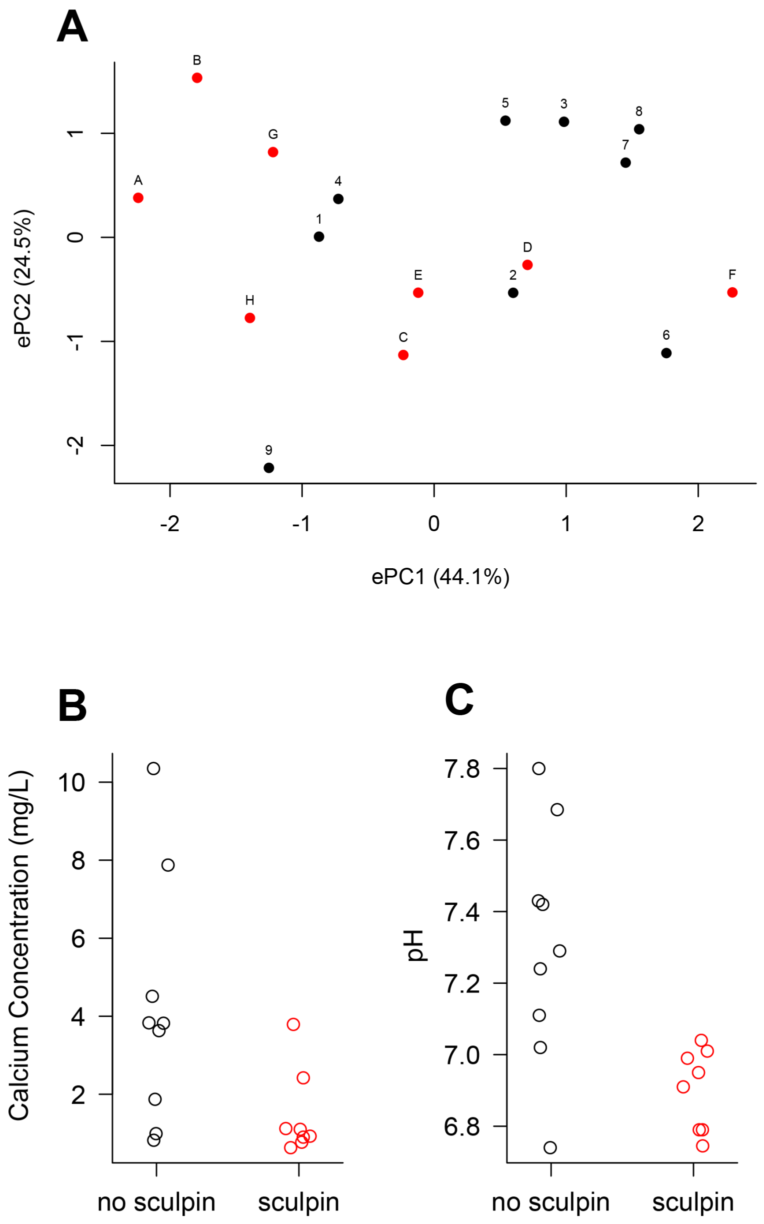

Figure S1: Little Environmental Difference Between Lakes with Sculpin and Lakes without Sculpin. Related to Figure 1B.

(A) Population positions along the first two environmental principal components (ePC1 and ePC2) from an analysis of nine environmental variables measured on freshwater lakes with sculpin (red A-H) and without sculpin (black 1-9). The two lake types broadly overlap in ePC1 and ePC2. **(B)** Calcium concentration [Ca] and **(C)** pH in study populations. Lakes with sculpin have relatively low calcium concentration and pH, whereas lakes without sculpin range from low [Ca] and low pH to high [Ca] and high pH. Points have been jittered to show overlap. Raw data are given in Table S1.

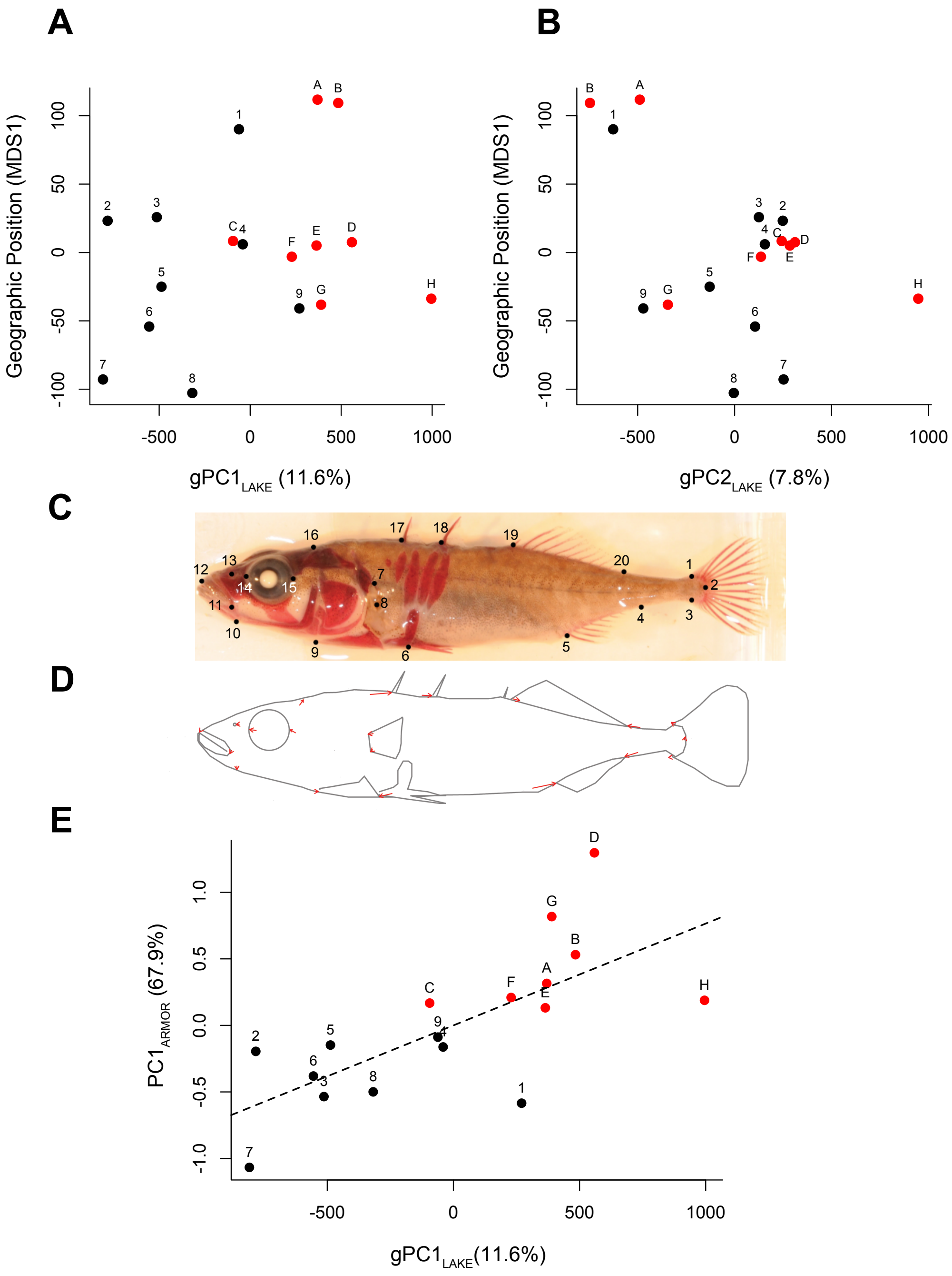

Figure S2: Genomic Variation is Associated with Stickleback Morphology not Spatial Distribution of Study Lakes. Related to Figure 2.
(A) Spatial distribution of lakes was summarized using the first multi-dimensional scaling axis of pairwise geographic distance among lakes, as in Figure 1B. The major axis of geographic position of populations (MDS1) does not correlate with the first principal component of genetic variation among all 17 freshwater lake populations of stickleback (gPC1$_{LAKE}$, as in Figure 2A, see main text) (t=1.2, df=15, P=0.25). **(B)** the second principal component of genetic variation (gPC2$_{LAKE}$) is associated with geographic position of populations (MDS1) (t=-2.1, df=15, P=0.05). **(C)** Twenty morphometric landmarks outlining body shape and fin insertion points. Landmarks were used to assess shape differences between stickleback populations from lakes with and without sculpin using stained stickleback samples. See supplementary methods for details. **(D)** Stickleback from lakes with and without sculpin differ predictably in body shape.

Arrows show the differences in the mean landmark positions for lakes without sculpin (arrow head) compared to the change in landmark positions for lakes with sculpin. Arrow lengths have been multiplied by four to increase visibility. **(E)** Comparison between the first genetic principal component (based on all SNPs in all lake populations, except those on the sex chromosome; $gPC1_{LAKE}$) and the first principal component axis of armor traits ($PC1_{ARMOR}$). A greater value of $PC1_{ARMOR}$ indicates increased defensive armor. Greater armor is associated with more extreme genomic differentiation along $gPC1_{LAKE}$ (overall: R=0.68, N=17, P=0.002). Each point represents a single individual from a unique lake population. Lakes with sculpin are red (A-H) and lakes without sculpin are black (1-9).

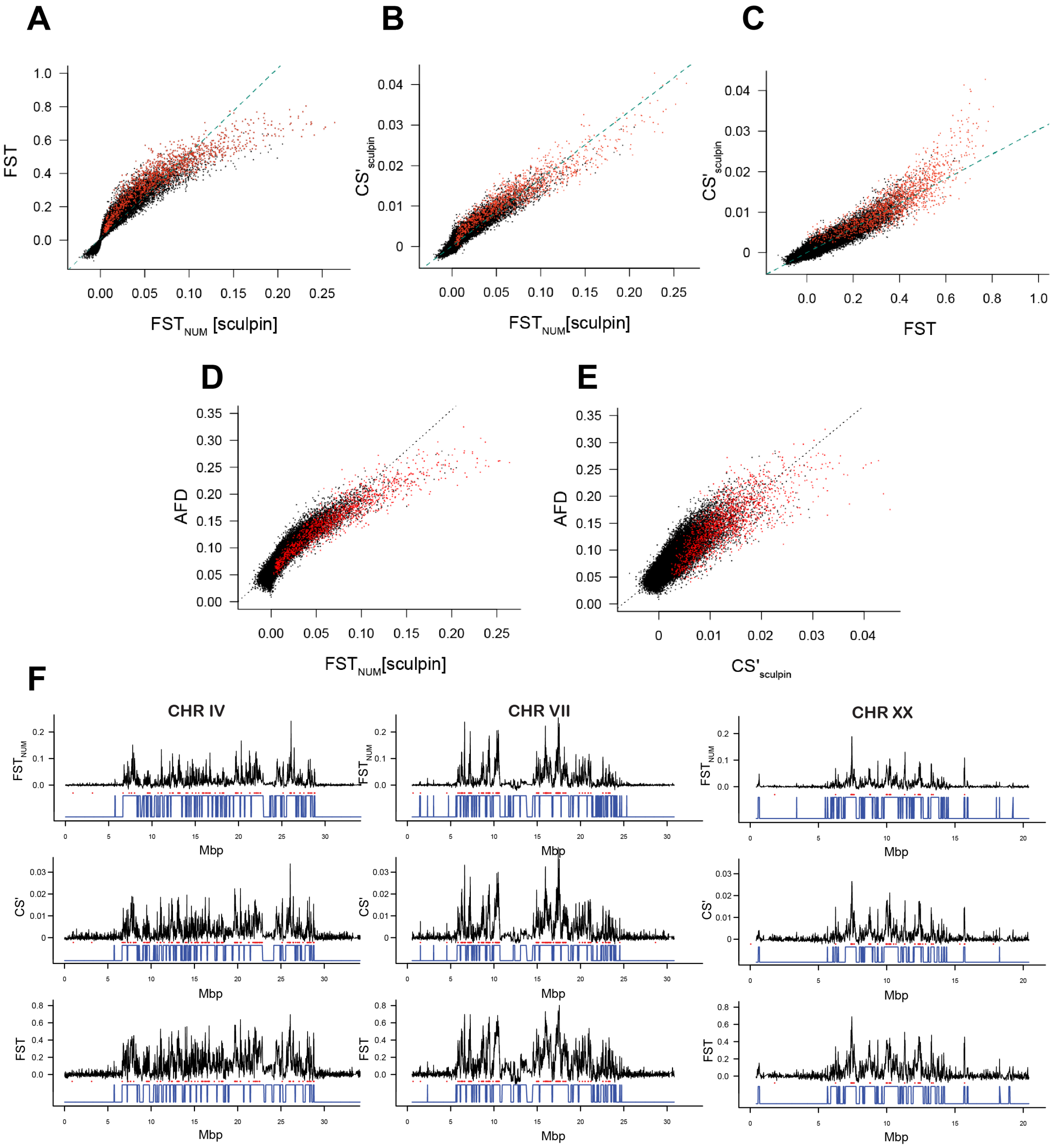

Figure S3: Correlation Between Different Metrics for Measuring Genetic Differentiation. Related to Figure 2 and STAR Methods.

Comparison of differentiation metrics between stickleback from lakes with and without sculpin across the genome. All scores are measured in 10,000-bp windows (5,000-bp step size). **(A)** Plot of FST and $\mathrm{FST}_{\mathrm{NUM}}$. Outlier windows for $\mathrm{FST}_{\mathrm{NUM}}$ are shown in red. These values are highly correlated (R^2= 0.88, P<2e16) **(B)** Plot of $\mathrm{CS'}_{\mathrm{sculpin}}$ and $\mathrm{FST}_{\mathrm{NUM}}$. Outlier windows for $\mathrm{CS'}_{\mathrm{sculpin}}$ are shown in red. The two variables are highly correlated (R^2= 0.91, P<2e16). **(C)** Plot of $\mathrm{CS'}_{\mathrm{sculpin}}$ and F_{ST}. Outlier windows for $\mathrm{CS'}_{\mathrm{sculpin}}$ are shown in red. The two variables are highly correlated (R^2= 0.86, P<2e16). **(D)** Plot of $\mathrm{FST}_{\mathrm{NUM}}$ and Allele Frequency Differences (AFD). Outlier windows from the $\mathrm{CS'}_{\mathrm{sculpin}}$ analysis are highlighted in red. The two variables are highly correlated (R^2= 0.85, P<2e16). **(E)** Plot of $\mathrm{CS'}_{\mathrm{sculpin}}$ and Allele Frequency Differences (AFD). Outlier windows from the $\mathrm{CS'}_{\mathrm{sculpin}}$ analysis are highlighted in red. The two variables are highly correlated (R^2= 0.77, P<2e16). **(F)** $\mathrm{FST}_{\mathrm{NUM}}$, CS', and F_{ST} values between stickleback from lakes with and without sculpin plotted across three highly differentiated chromosomes. Outlier windows are indicated in red below each metric. State changes as predicted by a hidden markov model (HMM) are show for each metric with a blue line.

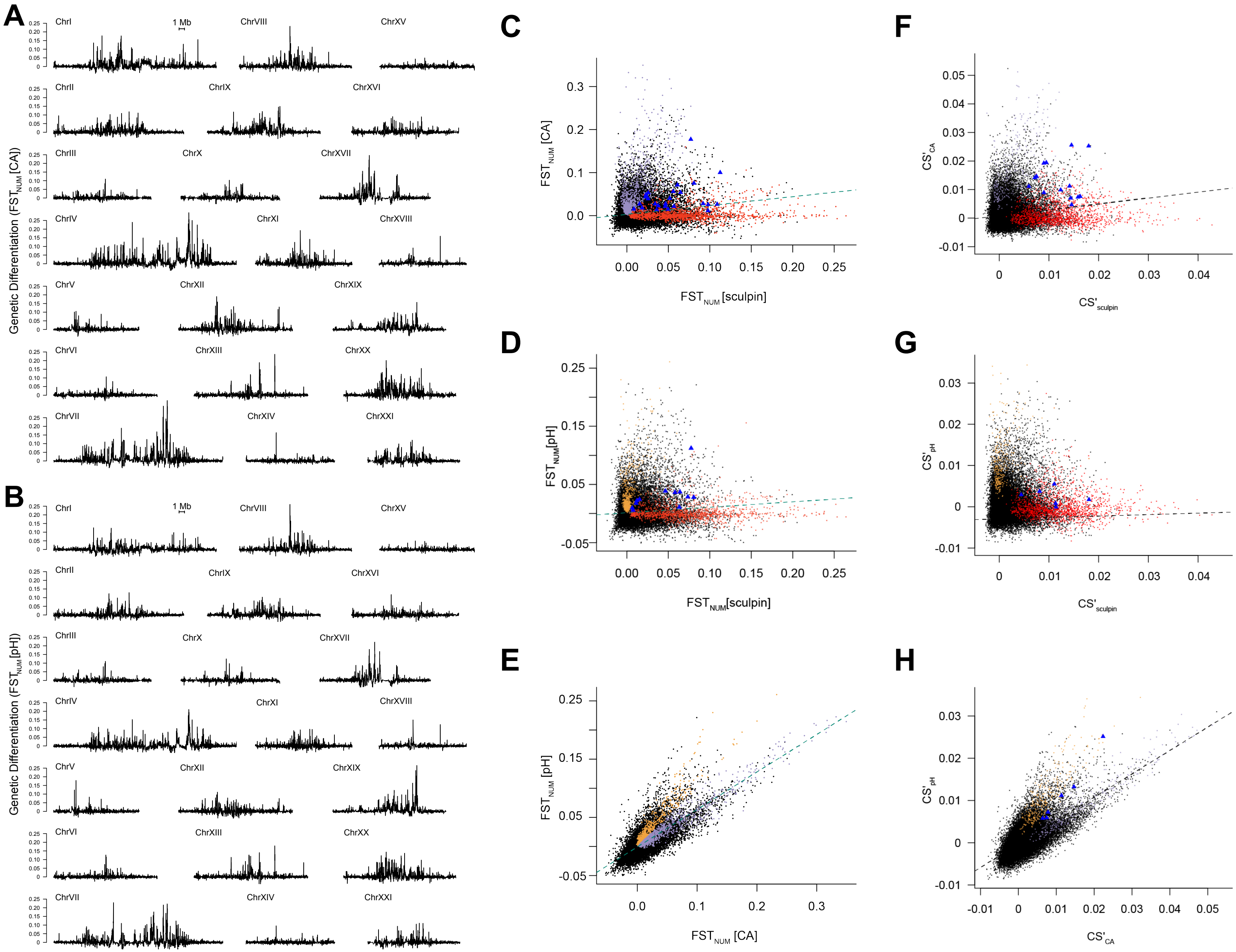
A
ChrI
1 Mb
ChrVIII
ChrXV
ChrII
ChrIX
ChrXVI
ChrIII
ChrX
ChrXVII
ChrIV
ChrXI
ChrXVIII
ChrV
ChrXII
ChrXIX
ChrVI
ChrXIII
ChrXX
ChrVII
ChrXIV
ChrXXI
0.25
0.20
0.15
0.10
0.05
0
Genetic Differentiation (FST_{NUM}[CA])
B
ChrI
1 Mb
ChrVIII
ChrXV
ChrII
ChrIX
ChrXVI
ChrIII
ChrX
ChrXVII
ChrIV
ChrXI
ChrXVIII
ChrV
ChrXII
ChrXIX
ChrVI
ChrXIII
ChrXX
ChrVII
ChrXIV
ChrXXI
Genetic Differentiation (FST_{NUM}[pH])
C
FST_{NUM} [CA]
FST_{NUM} [sculpin]
0.3
0.2
0.1
0.0
0.00
0.05
0.10
0.15
0.20
0.25
D
FST_{NUM}[pH]
FST_{NUM}[sculpin]
0.25
0.15
0.05
-0.05
E
FST_{NUM} [pH]
FST_{NUM} [CA]
0.0
0.1
0.2
0.3
F
CS'_{CA}
$CS'_{sculpin}$
0.05
0.04
0.03
0.02
0.01
0
-0.01
G
CS'_{pH}
$CS'_{sculpin}$
0.03
0.02
0.01
0
-0.01
0.04
H
CS'_{pH}
CS'_{CA}
0.05

Figure S4: Extensive Genomic Differentiation in Response to pH and Calcium Concentration is not Correlated with Genomic Differentiation Between Lakes with and Lakes Without Sculpin. Related to Figure 3.
(A) Genome-wide distribution of genetic differentiation (FST_{NUM} [CA]) calculated between stickleback in low and high calcium concentration lakes lacking sculpin. **(B)** Genome-wide distribution of FST_{NUM} [pH] calculated between stickleback in low and high pH lakes lacking sculpin. All chromosomes are plotted on the same scale. The sex chromosome (XIX) is included below but was not included in analyses reported in the manuscript. **(C)** FST_{NUM} [CA] scores calculated between stickleback from lakes with low or high calcium concentrations plotted against FST_{NUM} [sculpin] scores between lakes with and without sculpin. The two values were weakly correlated (R^2= 0.039). Outlier windows for the analysis between lakes with low and high calcium concentration are in purple, whereas outlier windows for FST_{NUM} [sculpin] are in red. **(D)** FST_{NUM} [pH] scores calculated between stickleback from lakes with low or high pH. The two scores are not correlated (R^2= 0.013). Outlier windows for FST_{NUM} [pH] are in orange whereas outlier windows for FST_{NUM} [sculpin] calculated between populations in lakes with and without sculpin are in red. **(E)** Scores for FST_{NUM} [CA] and FST_{NUM} [pH] were moderately correlated (R^2= 0.64). Outlier windows from the comparison between lakes with low and high pH are highlight in orange, whereas outlier windows for the comparison between lakes having low and high calcium concentration are highlighted with purple. There were no shared outlier windows. **(F)** CS'_{CA} scores calculated between stickleback from lakes with low or high calcium concentrations plotted against $CS'_{sculpin}$ scores between lakes with and without sculpin. The two scores are only weakly correlated (R^2= 0.024). Outlier windows for the analysis between populations in sculpin-absent lakes with low and high calcium concentration are in purple, whereas outlier windows for $CS'_{sculpin}$ between populations in lakes with and without sculpin are in red. The fourteen shared outlier windows for CS'_{CA} and $CS'_{sculpin}$ are identified with blue triangles. **(G)** CS'_{pH} scores calculated between populations from sculpin-absent lakes having low and high pH plotted against $CS'_{sculpin}$ scores between populations from lakes with and without sculpin across the same 10,000-bp windows (step size 5000-bp). The two scores are not correlated (R^2= 0.002). Outlier windows for CS'_{pH} are in orange whereas outlier windows for $CS'_{sculpin}$ calculated between populations in lakes with and without sculpin are in red. The six shared outlier windows for CS'_{pH} and $CS'_{sculpin}$ are identified with blue triangles. **(H)** Scores for CS'_{CA} and CS'_{pH} in 10,000-bp windows (step size 5,000-bp) calculated using only lakes without sculpin. These scores are moderately correlated (R^2= 0.49). Outlier windows from the comparison between lakes with low and high pH are highlight in orange, whereas outlier windows for the comparison between lakes having low and high calcium concentration are highlighted with purple. The six shared outlier windows for CS'_{CA} and CS'_{pH} are identified with blue triangles. All scores for CA and pH are calculated only between stickleback from sculpin-absent lakes. FST_{NUM} and CS' scores were calculated using the same 10,000-bp sliding windows (5000-bp step size).

	Population	Type	ID	Latitude	Longitude	Area (ha)	Perimeter (m)	Depth (m)	Elevation (m)	Distance to sea (m)	SRP (uM)	Na (µM)	Ca (mg/L)	pH	Env.PC1	Env.PC2
Lake																
	Ambrose	sculpin	C	49.733195	-124.02437	29.8	3200	13.3	56	940	0.9	76.40	1.12	6.79	-0.2331	-1.1315
	Black	no sculpin	9	48.773234	-125.096594	130.0	8400	11.1	111	1350	0	115.55	0.99	6.74	-1.2513	-2.2164
	Brown	sculpin	E	49.741765	-123.915103	18.8	1796	3.5	49	1120	0.23	54.07	0.90	6.79	-0.1204	-0.5325
	Bullocks	no sculpin	7	48.874112	-123.509031	9.4	1300	4	33	3310	0.97	424.75	3.82	7.43	1.4507	0.7180
	Cedar	sculpin	A	50.204813	-125.565719	31.0	4900	3	204	11800	0.31	54.80	0.77	6.99	-2.2409	0.3795
	Cranby	no sculpin	2	49.696084	-124.50654	44.6	3280	3.2	69	2530	0.34	159.76	7.88	7.69	0.5989	-0.5335
	Hoggan	no sculpin	6	49.152667	-123.828194	19.7	2219	3	63	310	0	278.76	3.83	7.02	1.7593	-1.1130
	Kirk	no sculpin	3	49.739441	-124.586289	8.3	1372	8.3	121	2543	0.73	126.94	10.35	7.80	0.9836	1.1108
	Klein	no sculpin	4	49.730087	-123.968612	13.5	2650	12	135	2660	1.01	47.62	1.87	7.29	-0.7252	0.3689
	North	sculpin	D	49.748275	-123.970265	12.8	1737	10.06	45	1040	0.29	100.66	2.42	7.01	0.7073	-0.2651
	Ormond	sculpin	B	50.180256	-125.52577	8.0	1753	6.57	230	10140	0.07	59.74	0.63	6.95	-1.7948	1.5333
	Pachena	sculpin	H	48.838338	-125.029038	58.7	4389	10.7	88	4730	0.295	87.76	1.10	6.91	-1.3951	-0.7751
	Paq	sculpin	F	49.613543	-124.0229	12.1	1785	2.2	12	880	0.67	308.70	3.79	7.04	2.2593	-0.5296
	Rosseau	sculpin	G	48.830345	-124.992733	13.1	2120	5.8	137	6405	0.28	74.10	0.93	6.75	-1.2199	0.8192
	Stowell	no sculpin	8	48.781861	-123.444086	5.6	983	4.6	77	1400	2.03	218.87	4.51	7.42	1.5530	1.0392
	Tom	no sculpin	1	50.256375	-124.923292	17.0	2600	1.7	198	1689	0.03	60.95	0.82	7.11	-0.8719	0.0063
	Trout	no sculpin	5	49.50843	-123.876057	7.6	1308	5.8	157	2420	0.92	162.34	3.63	7.24	0.5403	1.1214
	Mean	no sculpin				28.4 ± 13.3	2679 ± 761	6.0 ± 1.2	107 ± 17	2023 ± 304	0.67 ± 0.22	177.3 ± 39	4.19 ± 1.05	7.3 ± 0.11	0.45 ± 0.38	0.06 ± 0.38
		sculpin				23 ± 5.9	2710 ± 458	6.9 ± 1.4	102 ± 28	4632 ± 1569	0.38 ± 0.09	102 ± 30	1.45 ± .04	6.9 ± .01	-0.51 ± 0.06	-0.06 ± .04
	Mann-Whitney Test					U=30; P=0.61	U=29; P=0.54	U=32; P=0.74	U=40; P=0.74	U=32; P=0.74	U=45; P=0.41	U=53; P=0.11	U=59; **P=0.03**	U=63; **P=0.01**	U=51; P=0.17	U=40; P=0.74
Marine																
	Bamfield Inlet	marine	M6	48.824876	-125.136342											
	Little Cambell River	marine	M3	49.014982	-122.772487											
	Oyster Lagoon	marine	M2	49.613611	-124.029722											
	Salmon River	marine	M5	49.175163	-122.59434											
	Seyward Estuary	marine	M1	50.373634	-125.929494											
	West Creek	marine	M4	49.191545	-122.655709											

Table S1: Abiotic Characteristics of Study Populations. Related to Figure 1B.
Area, perimeter, and mean depth were obtained from HabitatWizard (http://www.env.gov.bc.ca/habwiz/). Elevation and the distance from the lake to the nearest ocean were calculated using Google Maps. Soluble reactive phosphorus (SRP), sodium concentration (Na), calcium concentration (Ca), and pH were directly measured from water samples. Mann-Whitney U tests were used to test for differences in abiotic variables between lakes with sculpin and without sculpin. Statistically significant differences are shown in bold.

	FSTNUM		FST		CS'	
Chromosome	**Number of outlier windows**	**Number of regions of high divergence**	**Number of outlier windows**	**Number of regions of high divergence**	**Number of outlier windows**	**Number of regions of high divergence**
I	101	17	102	14	120	16
II	49	8	49	6	50	9
III	9	0	9	0	6	3
IV	265	28	265	23	331	32
V	2	0	2	0	4	0
VI	11	2	11	2	8	4
VII	432	20	433	16	493	23
VIII	70	10	70	10	86	12
IX	62	10	62	9	71	12
X	8	2	8	2	9	2
XI	27	8	27	9	31	10
XII	142	9	141	8	164	11
XIII	6	2	6	1	11	3
XIV	5	1	5	0	5	1
XV	6	0	6	1	7	4
XVI	11	3	11	3	14	4
XVII	38	6	38	4	48	7
XVIII	17	3	17	2	16	2
XIX (Sex)	Not included	Not included	Not included	Not included	Not included	Not included
XX	110	8	111	8	120	9
XXI	21	4	21	4	45	7
unassembled	3	Not included	3	Not included	6	Not included
Total	1395	141	1397	122	1645	171

Table S2: Outlier Windows Between Stickleback from Lakes with and without Sculpin are Non-Randomly Distributed. Related to Figure 3.
The number of 10,000 bp outlier windows for $CS'_{sculpin}$ and the number of regions of high divergence identified by the hidden markov model (HMM). Chromosome refer to the genome alignment in [S1].

GO.ID	GO Term Description	Annotated	Significant	Expected	P-value
GO:0006811	ion transport	486	19	11.61	0.00016
GO:0007271	synaptic transmission, cholinergic	4	2	0.1	0.0033
GO:0006897	endocytosis	15	3	0.36	0.00494
GO:0001878	response to yeast	19	3	0.45	0.00981
GO:0006506	GPI anchor biosynthetic process	8	3	0.19	0.01091
GO:0051260	protein homooligomerization	64	5	1.53	0.01806
GO:0016311	dephosphorylation	135	5	3.22	0.0183
GO:0010842	retina layer formation	24	3	0.57	0.01878
GO:0007423	sensory organ development	307	9	7.33	0.0238
GO:0003403	optic vesicle formation	1	1	0.02	0.02389
GO:0006809	nitric oxide biosynthetic process	1	1	0.02	0.02389
GO:0045602	negative regulation of endothelial cell differentiation	1	1	0.02	0.02389
GO:0045737	positive regulation of cyclin-dependent protein kinase activity	1	1	0.02	0.02389
GO:0001886	endothelial cell morphogenesis	1	1	0.02	0.02389
GO:0016444	somatic cell DNA recombination	1	1	0.02	0.02389
GO:0006622	protein targeting to lysosome	1	1	0.02	0.02389
GO:0006047	UDP-N-acetylglucosamine metabolism	1	1	0.02	0.02389
GO:0010749	regulation of nitric oxide mediated signal transduction	1	1	0.02	0.02389
GO:0051570	regulation of histone H3-K9 methylation	1	1	0.02	0.02389
GO:0000045	autophagosome assembly	1	1	0.02	0.02389
GO:0021905	forebrain-midbrain boundary formation	1	1	0.02	0.02389
GO:0016255	attachment of GPI anchor to protein	1	1	0.02	0.02389
GO:0016055	Wnt signaling pathway	83	5	1.98	0.02554
GO:0003007	heart morphogenesis	126	7	3.01	0.03042
GO:0060037	pharyngeal system development	14	2	0.33	0.04278
GO:0006189	'de novo' IMP biosynthetic process	2	1	0.05	0.0472
GO:0050890	cognition	2	1	0.05	0.0472
GO:0008057	eye pigment granule organization	2	1	0.05	0.0472
GO:0021884	forebrain neuron development	2	1	0.05	0.0472
GO:0019731	antibacterial humoral response	2	1	0.05	0.0472
GO:0048755	branching morphogenesis of a nerve	2	1	0.05	0.0472
GO:0003308	negative regulation of Wnt signaling pathway involved in heart development	2	1	0.05	0.0472
GO:0046654	tetrahydrofolate biosynthetic process	2	1	0.05	0.0472
GO:0048814	regulation of dendrite morphogenesis	2	1	0.05	0.0472

Table S3: Gene Ontology (GO) Terms Enriched in Outlier Windows. Related to Figure 3.
Enriched Biological Processes Gene Ontology (GO) terms calculated using topGO for windows that were identified as outliers using FST_{NUM}, $CS'_{sculpin,}$ and FST. The term annotated refers to the number of genes in the dataset that were annotated with that GO term. Significant refers to the number of genes with the given GO term in outlier windows. Expected is the number of genes with the GO term expected by chance. P-values were calculated from Fisher exact tests.

Supplemental Reference

S1. Glazer A.M., Killingbeck E.E., Mitros T., Rokhsar D.S., and Miller C.T. (2015). Genome Assembly Improvement and Mapping Convergently Evolved Skeletal Traits in Sticklebacks with Genotyping-by-Sequencing. G3-Genes Genom. Genet. *5*, 1463–72.