SUPPLEMENTAL INFORMATION

EXTENDED EXPERIMENTAL PROCEDURES

QTL Mapping of Armor Plate Size

To map stickleback armor plate size traits, we used 2638 F2 progeny from a cross between a wild-caught, completely plated marine female from Onnechikappu stream on the east coast of Hokkaido Island, Japan, and a wild-caught, low-plated benthic male from Paxton Lake, British Columbia (Colosimo et al., 2004). Fish grown to adulthood were stained with 0.008% Alizarin red in 1% Potassium hydroxide solution to reveal the armor plates, and the height and width of the ones in positions that correspond to plates five to eight in fully plated marine sticklebacks were measured using a reticle and a Nikon SMZ1500 microscope. We next calculated the residuals for the plate height and width measurements using Minitab Release 13.31 (Minitab, Inc.) following correction for fish standard depth and length, respectively, and for fish sex. The F2 fish were genotyped with thirty new microsatellite markers (Table S1) by analyzing PCR products with GeneScan 500 LIZ (Applied Biosystems) internal size standards on a 96-capillary array on an ABI3730xl (Applied Biosystems) and GeneMapper v3.0 software. The microsatellites were then added to the chromosome XX meiotic map using JoinMap 3.0 (Kyazma BV) (Van Ooijen and Voorrips, 2001) and the traits mapped with MapQTL 4.0 (Kyazma BV) using Interval Mapping method (Van Ooijen et al., 2000). High significance cut off levels ($p < 0.001$) were calculated using permutation test analysis in MapQTL 4.0 with default settings (1000 permutations). Highest LOD scores and percent variance explained (PVE) were obtained for the eighth plate on the right flank and seventh plate on the left flank for plate height and width traits, respectively, and they are shown in Figure 1A.

Geographic Sequence and Morphology Patterns

To look for repeated patterns of sequence divergence in the armor plate size QTL interval, we first organized the 21 completely sequenced sticklebacks that came from distinct populations from around the globe (Jones et al., 2012) into eleven pairs, each composed of a freshwater and a marine fish that are from geographically proximate locations. For each pair we then used the previously identified single nucleotide polymorphisms (SNPs) to calculate the fraction of nucleotides that are divergent in a sliding window with size 2500 bp and step 500 bp. The identities of the populations and the GPS coordinates of their collection sites are listed in Table S2.

To examine whether there are geographic patterns of armor plate size, we measured the armor plate height and width from many more fish (Table S2) from the same 21 populations used in the sequence analysis. Plate height and width were corrected for fish depth and length, respectively, and the average value for each trait for each of the populations was then calculated. The significance of the difference in average plate height and width values distributions between groups of freshwater and marine fish populations (Figures 1C and 1D) were calculated using two-sample Wilcoxon test.

Identification of Genes and Conserved Non-Coding Elements (CNEs)

To annotate all the predicted genes in the plate size intervals we examined an 800 kb region (chrXX:3,250,000-4,050,000 in gasAcu1 assembly) containing the entire plate height and width QTLs with multiple independent approaches. We used GENSCAN (Burge and Karlin, 1997) to predict all the proteins in the interval, which were then verified using blastp against both the human RefSeq protein database and against the non-redundant protein sequences (nr). In addition, we analyzed all consecutive 10 kb windows in the 800 kb region using blastx against the non-redundant protein sequences (nr). We also used the stickleback UCSC Genome Browser (gasAcu1 assembly) to retrieve all the Ensembl annotated genes and stickleback ESTs that map to the interval (Kingsley et al., 2004), and the ESTs were then verified using blastx against the non-redundant protein sequences (nr). The predictions from all the independent methods were combined and transcripts matching repetitive elements such as transposable elements were eliminated. Predictions not supported by at least two independent methods were also eliminated. The remaining predicted genes (Table S3) were then analyzed for their synteny in fish and other vertebrates using Genomicus (Muffato et al., 2010) (Figure S1B).

To estimate the number of conserved non-coding elements (CNEs) for each of the gene loci in the plate size interval we first defined their start and end coordinates, the end of the coding region of the upstream gene, and the start of the coding region of the downstream gene, respectively. We next used the Table Browser tool of the UCSC Genome Browser to recover the total number of regions conserved to

other vertebrates for each of the predicted gene loci from the "Most conserved" track from the "Comparative Genomics" group and subtracted the number that overlap coding sequences from the "Ensembl Genes" track. These values were then divided by the sizes of the loci in kb to obtain the number of CNEs per kb for each of the genes in the plate size interval (Table S3).

The conservation plot in the plate size interval (Figure S1A) was obtained using the stickleback genome (gasAcu1 assembly) UCSC Genome Browser "8 Species Multiz Alignment and Conservation" track in the "Comparative Genomics" group. This track uses the stickleback, medaka, tetraodon, fugu, zebrafish, chicken, mouse and human genomes to determine sequence conservation.

Identification of Coding Changes

For identification of coding changes between freshwater and marine fish in the entire plate size interval, the following bacterial artificial chromosomes (BACs) with genomic DNA inserts from Salmon River Pacific basin marine fish (Kingsley et al., 2004) were completely sequenced (HudsonAlpha Institute for Biotechnology): CH213-4L05 (chrXX:3279878-3497783), CH213-101I14 (chrXX:3398293-3618126), CH213-124E09 (chrXX:3570385-3775338), CH213-72F04 (chrXX:3677176-3892221), CH213-288F23 (chrXX:3807740-3972457), and CH213-74J19 (chrXX:3951837-4156790). The sequences were aligned to the Bear Paw Lake Pacific basin freshwater reference genome, and SNPs were identified and intersected with the SNPs identified in the analysis of the 21 stickleback genomes (Jones et al., 2012) to obtain a list of sequence changes in coding regions that are observed in multiple independent stickleback populations (Table S4). The reading frames of the transcripts were verified by aligning their translated protein sequence to their human protein homologs using MUSCLE (Edgar, 2004). We next identified the sequence changes that lead to non-synonymous substitutions, and we used PolyPhen-2 (Adzhubei et al., 2010) to predict whether they are likely to be damaging for the function of the protein (Table S5).

Allele-Specific Expression Studies

To test whether there are regulatory changes in *GDF6*, we crossed a large-plated Little Campbell Pacific basin marine fish and small-plated Matadero Creek Pacific basin freshwater fish. We sequenced the coding regions of the *GDF6* and the nearby *BCKDHB* genes from both parents of the cross, identified SNPs at chrXX:3669475 and chrXX: 3633149 for *GDF6* and *BCKDHB*, respectively, and designed allele-specific expression assays (EpigenDx, Woburn, MA). At 83 days post fertilization, dermal flank tissue samples from the F1 hybrid fish containing armor plates from the areas corresponding to the fifth through eighth plate were dissected from the left and right flanks and RNA was extracted using Tri reagent (Invitrogen, Grand Island, NY). cDNA for each sample was then synthesized using SuperScript III (Invitrogen). Next, the regions containing the selected SNPs were PCR-amplified with the primers designed for each allele-specific assay, and pyrosequencing of two technical duplicates per each flank per each of three fish per each SNP was done by EpigenDx. The PCR-amplification primers were 5'–AAGGCGCTGCACGTGAAT–3' and 5'–AGGGCGCGATGATCCAGT–3' for *GDF6*, and 5'–TATCGCTCCAAACTTCCCATC–3' and 5'–CAAAGATGACTGCTGTTGG–3' for *BCKDHB*. The sequencing primer was 5'–GTGAATTTCAAGGAGCTGGG–3' for *GDF6* and 5'–CCAAACTTCACATCTCT–3' for *BCKDHB*. For each assay, we also used the same primer sets used in the above assays to amplify and clone the fragments from the parental fish, and pyrosequencing was done on mixtures where we made the fraction of the freshwater allele to be 0.3, 0.5, or 0.7. The observed freshwater allele abundance values were then used to obtain a linear regression equation, which we used to normalize the observed allele abundance values of the F1 hybrid fish samples. The significance of the deviation of the data from the expected freshwater allele abundance fraction of 0.5 in the total F1 hybrid fish mRNA was then determined using one sample t-test in R.

Fish Transgenic Studies

To test for armor plate enhancer activity differences between freshwater and marine fish, we PCR-amplified chrXX:3,610,110-3,614,309 region from Paxton Lake benthic freshwater fish and Little Campbell marine fish with the following primers:
5'–GGATTGCGGCCGCTACAGAGCCGAGCCGCCCTGCTAGCGG–3' and
5'–GGATTGCGGCCGCTAGGCACAGGAAGAAAGTGGCCAGAGG–3'. The primers have linkers with *NotI* sites, and the PCR products were digested with *NotI* and then cloned into the *PspOMI* site of the *pT2HE* vector, which has an *hsp70* basal promoter driving *EGFP* flanked by sequences with recognition sites for the *Tol2* transposase (Kawakami et al., 2004; Rada-Iglesias et al., 2011). The freshwater and

marine allele constructs were then co-injected with *Tol2* transposon mRNA into fertilized one-cell stage eggs (Hosemann et al., 2004) from Little Campbell marine population, and the fry were raised under standard aquarium conditions and checked daily for GFP expression.

To examine the significance of the transposon insertion, we modified the above transgenic constructs to contain either the freshwater enhancer with the L2 LINE transposon (Kapitonov et al., 2009) removed, or the marine enhancer with the same transposon sequence inserted in the appropriate location. These constructs also either added or removed the adjacent repetitive sequence not found in the marine enhancer, which is of high complexity (Price et al., 2005) and exists at 604 other locations in the stickleback genome (see Figure 2C).

The transposon was amplified from the freshwater construct with the following primers: 5'–ACCTAGAGAGTCAGTTATTTATTTACCATAGAC–3' and
5'–TTTTGTCGAAATATAACCAAATTCTTACC–3'. The marine enhancer and vector backbone were amplified and prepared to receive the transposon insertion with the following primers: 5'–TATGGTAAATAAATAACTGACTCTCTAGG–3' and
5'–TAGGGTAAGAATTTGGTTATATTTCGACAAAAGC–3'. These two amplicons were then joined through Gibson assembly to create a reporter construct with the transposon sequence inserted into the marine enhancer.

Separately, the transposon was removed from the freshwater construct with the following primers: 5'–TTTTCTCGAAATAAACTGACTCTCTAGG–3' and
5'–TAGAGAGTCAGTTTATTTCGAGAAAAGC–3'. The resultant amplicon was circularized through Gibson assembly to create a reporter construct with the transposon sequence removed from the freshwater enhancer.

These modified constructs were then co-injected with *Tol2* transposase mRNA into fertilized one-cell stage embryos from the Matadero Creek, CA freshwater population. Fish were raised under standard aquarium conditions and checked regularly for GFP expression. Additional Matadero Creek embryos were also injected with unmodified freshwater and marine constructs to rule out any population-dependent reporter activity.

For stickleback transgenic studies testing the effect of overexpressing *GDF6*, constructs were made in *pT2HE* vector that contains synthesized *GDF6* cDNA fused to *EGFP* via a *2A* peptide (Szymczak et al., 2004). The *GDF6* cDNA was then placed under either the control of the same enhancers used in the GFP transgenic studies above, or the constitutive *pTK* or *pCMV* promoters. The constructs were then injected into one-cell stage marine embryos, as described above. In total, 2196 marine stickleback eggs were fertilized from 24 clutches, and 1657 one-cell stage embryos were injected with the *GDF6* constructs, with the rest kept as controls. At between 204 and 267 days post fertilization, the surviving 357 *GDF6*-injected fish and 203 control fish were collected. All fish were examined for GFP expression, typically seen in the eyes due to the *hsp* basal promoter activity. DNA was also extracted from the caudal and both pectoral fins of each fish, and was tested by PCR for the presence of the transgenic construct using the following primers:
5'–CCACGAACTTCTCTCTGTTAAAGC–3' and
5'–CGAGTATGTCCTGTATATGGAGAGC–3'.
168 *GDF6*-injected fish out of 357 scored as transgenic, either because they had detectable GFP expression or genotyped positive for the ectopic *GDF6* constructs by PCR. None of the 203 control fish tested positive by either assay. Seven out of the 168 transgenic fish had a plate size phenotype, which in all cases led to missing armor plates. In comparison, none out of the 189 *GDF6*-injected but not transgenic fish (Fisher's Exact Test p=0.005) and none out of the 203 control fish (Fisher's Exact Test p=0.004) had either missing or smaller plates. Skeletal phenotypes were further evaluated by micro-computed tomography (μCT). Alizarin red stained fish were scanned in 37% isopropanol using a Scanco μCT40 operated at 70 kVp, 114 μA, at medium resolution, and with 2x-averaging. Voxel size was 12 mm^3 for all fish. Evaluations were performed on 1975–2769 slices (~ 24–33 mm) beginning at the anterior tip of the mouth. Volumetric reconstruction images were imported into Adobe Photoshop, the plates were selected and overlayed with red color.

Mouse Transgenic Studies
The coordinates of hCONDEL.305 and hCONDEL.306 in the chimpanzee panTro2 genome assembly are chr8:94,538,975-94,539,466 and chr8:94,894,271-94,899,828, respectively. We confirmed the hCONDELs

in a panel of 23 human individuals from around the globe (Table S7) using PCR tests with the following primers:
5'–GGATTGCGGCCGCTAGATAGAGTTGTAAATTATAAGGCAGTTAG–3' and
5'–GGATTGCGGCCGCTATTACCCGCATAGAATCTCATGTTCC–3' for hCONDEL.305; and
5'–GGATTGCGGCCGCTAAGGGCAGCTGTGCTGAGCCTTGACC–3' and
5'–GGATTGCGGCCGCTAAAAAGCTGAGACTCAGTTTACATAAC–3' for hCONDEL.306.

For PCR amplifying and cloning hCONDEL.305, the following primers with linkers containing *SfiI* sites were used:
5'–ATGTGGCCAAACAGGCCTATTGATAGAGTTGTAAATTATAAGGCAGTTAG–3' and
5'–TGTGGCCTGTTTGGCCTATTTTACCCGCATAGAATCTCATGTTCC–3'. The resulting PCR product is chimpanzee chr8:94,538,442-94,539,803 (panTro2 genome assembly), which contains the homologous sequence of the entire 492 bp hCONDEL.305 plus an additional 870 bp of surrounding sequences.

For PCR amplifying and cloning hCONDEL.306, the following primers with linkers containing *NotI* sites were used:
5'–GGATTGCGGCCGCTATCTGGTCAGCAGGAAGGGTACTGAC–3' and
5'–GGATTGCGGCCGCTAAGACTTTAGAGGGTTGCTGGGAGTG–3'. The resulting PCR product is chimpanzee chr8:94,899,638-94,894,864 (panTro2 genome assembly), which contains the homologous sequence of 4775 bp out of 5558 bp that are deleted in humans. The bases excluded from the PCR-amplified region are repeats-rich regions.

Two-copy concatenation in head-to-tail orientation of the chimpanzee hCONDEL.305 sequences (NS03659, Coriell Institute for Medical Research) and a single copy of the chimpanzee hCONDEL.306 sequences (NS03659) were cloned into the *SfiI* or the *NotI* site, respectively, of an *hsp-LacZ* minimal promoter expression vector (Mortlock et al., 2003). The chimpanzee hCONDEL.306 sequences were also subcloned into the *NotI* site of *hsp-CreER-T2* minimal promoter expression vector (Indra et al., 1999). The three constructs were linearized with *SalI* and injected into the pronuclei of fertilized FVB embryos (Xenogen Biosciences and Stanford Transgenic Facility). For the *hsp-LacZ* constructs, embryos were collected at embryonic day 14.5 and 16.5 for hCONDEL.305, and embryonic day 12.5 and 14.5 for hCONDEL.306, and were assayed for lacZ activity. For the *hsp-CreER-T2* construct, two independent stable lines (*pVBI347L2* and *pVBI347L3)* were obtained and animals were bred to floxed-*ROSA26* reporter mice (Soriano, 1999). Tamoxifen (3 mg of tamoxifen dissolved in corn oil per 30 g female weight) was administered by oral gavage at different developmental time points and embryos were then assayed for lacZ activity two to eleven days later at stages specified in the figures. For more detailed analysis, lacZ stained hindlimbs were embedded in gelatin, sectioned in transverse orientation using a Leica CM3050 S, and counterstained with Nuclear Fast Red (Vector Laboratories, CA).

Phenotypes in *Gdf6* Knock Out Mice

Skull and digit length measurements were carried out on Alizarin red stained *Gdf6* mutant and wild type mice. Frontal and parietal skull plate length and width were measured from 12 wild type and 16 *Gdf6* null mice and were normalized to nasal plates length and width, respectively. Relative length of hindlimb digits was determined for eight wild type and nine *Gdf6* null mice. We measured the length of each of the hindlimb phalanges on digits I, III, and V, normalized the measurements to femur length, took the sum of the values of the two phalanges for digit I and the three phalanges for digits III and V to get the relative total length for each digit, and then averaged the values of the left and right hindlimbs. We then used two-sample t-tests in R determine whether the differences in skull and digit sizes are significant between the wild type and *Gdf6-/-* mice.

Animal Procedures

All mouse and fish procedures used in this study were carried out in accordance with protocols approved by the Stanford University Institutional Animal Care and Use Committee.

SUPPLEMENTAL TABLES

Table S1. Microsatellite Markers Used in this Study, Related to Figure 1

Marker	Position on chrXX	Forward primer	Reverse primer
VBI007	2276966-2277202	TGATGCTCCATGTTGAGTAAGC	GCCCAGTCTCCCTGTTATGG
VBI009	3263592-3263774	CCTCTGCCAATCAATCTCG	GACGACCATGGAGAGTTTGC
VBI019	3384263-3384372	AGTTTCTCTCACAGTGCAACC	TCCAGCTTTGCAGATTAGGG
VBI020	3553520-3553679	GTTGCAACAGACCAATCACC	ACCTCAGTTAAGCCGATTGC
VBI010	3658358-3658469	AACCCTTTGCAGACAAGAGC	GCTTCTTGTGCCATCTGTACC
VBI032	3675224-3675350	ACAGATTCATGGTGCCTTCC	GAAGTGAGGGCAGGTCAGG
VBI033	3692001-3692104	GGTGTGCATGCCTTGATATG	ATACCTCGTGGCAGCAACTC
VBI034	3705052-3705201	AGGAAGGGAGGGAGGTATGG	TGCTTCTCACCGTAAATCTGG
VBI035	3712769-3712851	CATTCCTCTGGTTGCTCTGG	GGCTGTGAGGACTCAGTGG
VBI021	3722590-3722696	CTTCACAGGCTTCCTATGTGG	GACTCCTGCAGCACTTAAAGC
VBI036	3750153-3750224	CATTAACAAGAAGGGCAGTCG	ACCAGCTGCATTATGTGTGC
VBI037	3790203-3790341	ACTGGAGGAGGACATTGAGG	CTTGAGTGCTCTCATCCTTCC
VBI038	3839454-3839552	CTTGGCCCACACATGATACC	CATGTGGGCTCAGTTCAAGC
VBI022	3864228-3864447	GGCTTTCGGCTGCTGAGG	CGCATTCTAATCCCAACTCTGG
VBI039	3888423-3888476	TGAGCCTTGTTCTCCTCTCG	CACCGACACCTCTCACACC
VBI040	3931383-3931507	CAGAGCAACACAATCATCATCC	GCAGAGTGCGAACTCAAAGC
VBI042	3979804-3979923	TCCAGGGTTCAGACAACAGC	AAAGGGATTTCCTTCAGACG
VBI015	4018336-4018552	CTCTGACGACGGCTACAACC	CACCTGAACAATGACCAACC
VBI043	4098183-4098281	GGTCCTCTGTGTGTTGATGC	TCAGGAACGACCTCTCAAGC
VBI044	4131151-4131233	ACTCTGCTTCTCCACCAACG	ATCTGGCTTGTGCTTCTTGG
VBI045	4165087-4165169	ATCCCTGAGCGAGTAACAGC	TCTCTGCCAATCTCTCACTCC
VBI047	4260955-4261171	TAACGGTTGAGAAAGTGACC	CAGACGTTCCTCTCTGATGC
VBI016	4451784-4451955	CAACGGGTAGACACTAATTTGC	AACGGAGCATCTGACAATGG
VBI026	4643210-4643314	CCTCACACAAGACAGGAAACC	GCTCCAAGCTGTGTCATTACC
VBI028	4886207-4886344	TGGAGATAGTGGGTCTGTGG	TGTTACCATGGTCGATGTCC
VBI017	4968704-4968872	CCACAGTCAGACACTGTGACG	TGTGGTCAGCAGATGTAAACG
VBI029	5096646-5096850	CTGAAAGCAAGATAACGTGTGG	GTTGTTTCCTCGTGTGATTC
VBI030	5242011-5241202	CACTGACGCATTGAAGAACG	TTCACTGATGCAAACCATCG
VBI031	5335734-5335879	ATCTCTCCCTTCGCACACG	CAGAGCGAGGAAACTACAAAGG
VBI004	5512098-5512227	CAGTCACACATGCGAAGTCC	CTGCACTGCAAGTGAAAAGG

Table S2. Fish Populations Used in this Study, Related to Figure 1

Name	Location	Ecotype	Latitude	Longitude	n[a]
BEPA	Bear Paw Lake, AL	Pacific freshwater	61.615	-149.757	10
BIGL	Big River Lagoon, CA	Pacific freshwater	39.300	-123.728	11
FTC	Fish Trap Creek, WA	Pacific freshwater	48.931	-122.487	11
HUTU	Humptulips Lake, WA	Pacific freshwater	47.231	-123.957	9
MATA	Matadero Creek, CA	Pacific freshwater	37.393	-122.162	11
MUDL	Mud Lake, AL	Pacific freshwater	61.593	-149.340	11
PAXB	Paxton Lake Benthic, BC	Pacific freshwater	49.703	-124.522	11
NOST	Norway Stream, Norway	Atlantic freshwater	59.765	5.712	11
SCX	Schwale River, Germany	Atlantic freshwater	54.080	10.083	10
SHEL	River Shiel, Scotland	Atlantic freshwater	56.748	-5.698	7
TYNE8	River Tyne Site 8, Scotland	Atlantic freshwater	55.943	-2.785	11
BDGB	Bodega Bay, CA	Pacific marine	38.325	-123.041	11
BIGR	Big River Downstream, CA	Pacific marine	39.289	-123.747	11
JAMA	Japanese Marine, Japan	Pacific marine	42.973	144.329	9
RABS	Rabbit Slough, AL	Pacific marine	61.556	149.249	11
SALR	Salmon River Anadromous, BC	Pacific marine	49.175	-122.594	8
ANTL	Antigonish Landing, Nova Scotia	Atlantic marine	45.698	-61.878	8
GJOG	Gjogur, Iceland	Atlantic marine	65.980	-21.440	11
GORT	Gorten Sands, Scotland	Atlantic marine	56.911	-5.888	10
NEU	Neustadt, Germany	Atlantic marine	54.058	10.877	11
TYNE1	River Tyne Site 1, Scotland	Atlantic marine	55.999	-2.520	N/A

[a] – The number of fish used for plate size measurements in Figures 1C and 1D. No additional fish were available for TYNE1.

Table S3. Genes and Conserved Non-Coding Elements (CNEs) in the Armor Plate Height and Width QTL Intervals, Related to Figure 1

Gene	Gene name	QTL interval	Coding region, chrXX	Entire locus, chrXX	Locus size, kb	#CNEs	#CNEs per 1 kb
USP45	*Ubiquitin Carboxyl-Terminal Hydrolase 45*	Height	3,278,433-3,301,766	3,273,202-3,303,524	30323	75	2.47
HSP67B2 (Copy 1)	*Drosophila Heat-Shock Protein 67B2-Like, C*opy 1	Height	3,303,525-3,305,197	3,301,767-3,306,097	4331	4	0.92
HSP67B2 (Copy 2)	*Drosophila Heat-Shock Protein 67B2-Like, Copy 2*	Height	3,306,098-3,307,836	3,305,198-3,319,388	14191	50	3.52
PRDM13	*PR Domain Zinc Finger Protein 13*	Height	3,319,389-3,324,634	3,307,837-3,358,063	50227	178	3.54
ME1	*Malic Enzyme 1*	Height	3,358,064-3,409,590	3,324,635-3,425,306	100672	213	2.12
PRSS35	*Serine/Arginine Protease 35*	Height	3,425,307-3,426,581	3,409,591-3,431,039	21449	43	2.00
SNAP91	*Clatrin Coat Assembly Protein AP180*	Height	3,431,040-3,467,939	3,426,582-3,469,997	43416	43	0.99
DOPEY1	*Protein Dopey-1*	Height	3,469,998-3,490,913	3,467,940-3,492,736	24797	18	0.73
UBE2CBP	*Ubiquitin-Conjugating Enzyme E2C-Binding Protein*	Height	3,492,737-3,507,559	3,490,914-3,524,055	33142	41	1.24
TPBG	*5T4 Oncofetal Trophoblast Glycoprotein*	Height	3,524,056-3,525,147	3,507,560-3,528,045	20486	37	1.81
IBTK	*Inhibitor of Bruton Agammaglobulinemia Tyrosine Kinase*	Height	3,528,046-3,544,218	3,525,148-3,549,863	24716	20	0.81
FAM46A	*Family 46 Protein A*	Height	3,549,864-3,553,161	3,544,219-3,596,586	52368	52	0.99
BCKDHB	*Branched Chain Alpha-Keto Acid Dehydrogenase*	Height	3,596,587-3,633,220	3,553,162-3,636,338	83177	83	1.00
TTK	*Dual Specificity Protein Kinase TTK*	Height	3,636,339-3,643,358	3,633,221-3,644,831	11611	5	0.43
EBAG9	*Estrogen Receptor-Binding Fragment-Associated Gene 9*	Height	3,644,832-3,648,639	3,643,359-3,653,038	9680	17	1.76
SYBU	*Syntabulin*	Height	3,653,039-3,662,115	3,648,640-3,664,497	15858	40	2.52

GDF6	*Growth/Differentiation Factor 6*	Height and width	3,664,498-3,669,742	3,662,116-3,736,567	74452	305	4.10
ENY2	*Enhancer of Yellow 2 Homolog*	Width	3,736,568-3,738,629	3,669,743-3,776,983	107241	334	3.11
TRHR	*Thyrotropin-Releasing Hormone Receptor*	Width	3,776,984-3,781,307	3,738,630-3,788,627	49998	40	0.80
TMEM74	*Transmembrane Protein 74*	Width	3,788,628-3,790,084	3,781,308-3,827,002	45695	95	2.08
EMC2	*ER membrane Protein Complex Subunit 2*	Width	3,827,003-3,844,761	3,790,085-3,852,598	62514	157	2.51
EIF3E	*Eukaryotic Translation Initiation Factor 3 Subunit E*	Width	3,852,599-3,874,950	3,844,762-3,878,711	33950	109	3.21
RSPO2	*R-Spondin Family, Member 2*	Width	3,878,712-3,925,038	3,874,951-3,934,001	59051	229	3.88
HAMP (Copy 1)	*Hepcidin*, Copy 1	Width	3,934,002-3,934,520	3,925,039-3,937,142	12104	28	2.31
HAMP (Copy 2)	*Hepcidin*, Copy 2	Width	3,937,143-3,937,661	3,934,521-3,941,043	6523	11	1.69
HAMP (Copy 3)	*Hepcidin*, Copy 3	Width	3,941,044-3,941,558	3,937,662-3,999,223	61652	122	1.98
PLEC	*Plectin*	Width	3,999,224-4,049,483	3,941,559-4,063,500	121942	200	1.64

Table S4. SNPs in the Coding Regions of the Plate Height and Width QTL Intervals, Related to Results

Position (chrXX)	Gene	Ref[a]	Alt[b]	Codon change[c]
3303687	*HSP67B2 C1*	T	C	TCT (Ser) to TCC (Ser)
3304225	*HSP67B2 C1*	A	G	GAA (Glu) to GAG (Glu)
3306775	*HSP67B2 C2*	G	A	CAG (Gln) to CAA (Gln)
3306890	*HSP67B2 C2*	G	A	**GTC (Val) to ATC (Ile)**
3321207	*PRDM13*	T	A	**TTC (Phe) to TAC (Tyr)**
3323320	*PRDM13*	A	G	CCA (Pro) to CCG (Pro)
3323503	*PRDM13*	G	C	CGG (Arg) to CGC (Arg)
3323560	*PRDM13*	C	A	GGC (Gly) to GGA (Gly)
3358434	*ME1*	C	T	**ATG (Met) to ATA (Ile)**
3371306	*ME1*	A	T	GGT (Gly) to GGA (Gly)
3371327	*ME1*	C	A	GCG (Ala) to GCT (Ala)
3372640	*ME1*	C	T	**AGG (Arg) to AAG (Lys)**
3425525	*PRSS35*	T	C	CGT (Arg) to CGC (Arg)
3432273	*SNAP91*	T	C	GTA (Val) to GTC (Val)
3434468	*SNAP91*	C	G	GGG (Gly) to GGC (Gly)
3470190	*DOPEY1*	T	C	CTA (Leu) to CTG (Leu)
3470913	*DOPEY1*	G	A	TCC (Ser) to TCT (Ser)
3473329	*DOPEY1*	C	A	**GCC (Ala) to TCC (Ser)**
3474567	*DOPEY1*	G	A	TAC (Tyr) to TAT (Tyr)
3475084	*DOPEY1*	T	G	AGG (Arg) to CGG (Arg)
3475166	*DOPEY1*	T	C	CAA (Gln) to CAG (Gln)
3475657	*DOPEY1*	G	C	ACC (Thr) to ACG (Thr)
3475699	*DOPEY1*	C	T	GCG (Ala) to GCA (Ala)
3475801	*DOPEY1*	A	G	AGT (Ser) to AGC (Ser)
3475807	*DOPEY1*	C	T	CAG (Gln) to CAA (Gln)
3476926	*DOPEY1*	C	T	GCG (Ala) to GCA (Ala)
3477263	*DOPEY1*	G	T	TCC (Ser) to TCA (Ser)
3477983	*DOPEY1*	G	C	GGC (Gly) to GGG (Gly)
3478012	*DOPEY1*	G	A	CTG (Leu) to TTG (Leu)
3478196	*DOPEY1*	A	G	CGT (Arg) to CGC (Arg)
3478367	*DOPEY1*	G	A	GCC (Ala) to GCT (Ala)
3478415	*DOPEY1*	A	G	TCT (Ser) to TCC (Ser)
3478638	*DOPEY1*	C	T	**GGC (Gly) to GAC (Asp)**
3479156	*DOPEY1*	G	T	CTC (Leu) to CTA (Leu)
3486321	*DOPEY1*	A	G	TTT (Phe) to TTC (Phe)
3490147	*DOPEY1*	G	A	GGC (Gly) to GGT (Gly)
3494691	*UBE2CBP*	A	G	ACA (Thr) to ACG (Thr)
3524458	*TPBG*	C	G	CTG (Leu) to CTC (Leu)
3524514	*TPBG*	C	T	**GGC (Gly) to AGC (Ser)**
3534110	*IBTK*	C	T	**GCG (Ala) to GTG (Val)**
3535194	*IBTK*	G	C	CTG (Leu) to CTC (Leu)
3535504	*IBTK*	G	C	GTG (Val) to GTC (Val)
3535504	*IBTK*	G	C	GTG (Val) to GTC (Val)
3537180	*IBTK*	A	G	GAA (Glu) to GAG (Glu)
3538854	*IBTK*	A	G	CAA (Gln) to CAG (Gln)
3539865	*IBTK*	T	C	**TCC (Ser) to CCC (Pro)**
3540447	*IBTK*	A	G	**AGC (Ser) to GGC (Gly)**
3551874	*FAM46A*	C	A	ATC (Ile) to ATA (Ile)
3552873	*FAM46A*	T	G	CTT (Leu) to CTG (Leu)
3602405	*BCKDHB*	A	G	CGT (Arg) to CGC (Arg)
3620700	*BCKDHB*	C	G	CCG (Pro) to CCC (Pro)

3631141	*BCKDHB*	T	G	AGA (Arg) to CGA (Arg)
3632161	*BCKDHB*	C	T	ACG (Thr) to ACA (Thr)
3633149	*BCKDHB*	T	G	GGA (Gly) to GGC (Gly)
3636695	*TTK*	T	G	ACA (Thr) to ACG (Thr)
3637606	*TTK*	A	G	ATT (Ile) to ATC (Ile)
3637651	*TTK*	T	C	CCA (Pro) to CCG (Pro)
3639213	*TTK*	G	A	CTC (Leu) to CTT (Leu)
3639269	*TTK*	G	A	**CCC (Pro) to TCC (Ser)**
3639498	*TTK*	A	G	**CTT (Leu) to CCT (Pro)**
3646497	*EBAG9*	T	A	ACT (Thr) to ACA (Thr)
3653314	*SYBU*	G	T	CGA (Arg) to AGA (Arg)
3653435	*SYBU*	G	A	TCC (Ser) to TCT (Ser)
3653658	*SYBU*	C	G	**GGG (Gly) to GCG (Ala)**
3653699	*SYBU*	C	A	TCG (Ser) to TCT (Ser)
3654125	*SYBU*	T	A	GCC (Ala) to GCT (Ala)
3669475	*GDF6*	C	G	GGC (Gly) to GGG (Gly)
3737087	*ENY2*	T	C	ACA (Thr) to ACG (Thr)
3777003	*TRHR*	T	G	**CAC (His) to CCC (Pro)**
3789673	*TMEM74*	A	G	**AAC (Asn) to AGC (Ser)**
3834909	*TTC35*	T	C	GCA (Ala) to GCG (Ala)
3925002	*RSPO2*	A	G	CAA (Gln) to CAG (Gln)
4032267	*PLEC1*	G	C	GGG (Gly) to GGC (Gly)
4032907	*PLEC1*	C	G	CTC (Leu) to CTG (Leu)
4038685	*PLEC1*	G	C	GCG (Ala) to GCC (Ala)
4039507	*PLEC1*	C	T	AGC (Ser) to AGT (Ser)
4040746	*PLEC1*	G	T	ACG (Thr) TO ACT (Thr)
4041325	*PLEC1*	A	G	GAA (Glu) to GAG (Glu)
4041577	*PLEC1*	G	A	CAG (Gln) to CAA (Gln)
4043387	*PLEC1*	A	C	ACA (Thr) TO ACC (Thr)
4044065	*PLEC1*	T	C	ATT (Ile) to ATC (Ile)
4047284	*PLEC1*	G	A	AAG (Lys) to AAA (Lys)
4047917	*PLEC1*	C	T	TAC (Tyr) to TAT (Tyr)

[a] Nucleotide in the reference freshwater stickleback genome from Bear Paw Lake

[b] Alternative nucleotide present in the Salmon River BACs and in the sequenced 21 genomes

[c] Non-synonymous substitutions are shown in bold

Table S5. Non-Synonymous Substitutions in the Plate Size Interval and PolyPhen-2 Predictions, Related to Results

SNP, chrXX	Ref[a]	Alt[b]	Gene	Freshwater to marine allele change	Prediction for change to marine allele	Score	Marine to freshwater allele change	Prediction for change to freshwater allele	Score
3306890	G	A	*HSP67B2*	V146I	benign	0.002	I146V	benign	0.001
3321207	T	A	*PRDM13*	F119Y	benign	0.001	Y119F	benign	0.000
3358434	C	T	*ME1*	M529I	benign	0.000	I529M	benign	0.003
3372640	C	T	*ME1*	R149K	benign	0.001	K149R	benign	0.044
3473329	C	A	*DOPEY1*	A1977S	benign	0.035	S1977A	benign	0.001
3478638	C	T	*DOPEY1*	G1087D	benign	0.079	D1087G	benign	0.000
3524514	C	T	*TPBG*	G226S	benign	0.009	S226G	benign	0.102
3534110	C	T	*IBTK*	A530V	benign	0.001	V530A	benign	0.018
3539865	T	C	*IBTK*	S1119P	benign	0.000	**P1119S**	**probably damaging[c]**	**0.992**
3540447	A	G	*IBTK*	S1149G	benign	0.271	G1149S	benign	0.000
3639269	G	A	*TTK*	P617S	benign	0.002	S617P	benign	0.012
3639498	A	G	*TTK*	L580P	possibly damaging	0.948	P580L	benign	0.043
3653658	C	G	*SYBU*	G306A	benign	0.000	A306G	benign	0.000
3777003	T	G	*TRHR*	H369P	benign	0.015	P369H	benign	0.000
3789673	A	G	*TMEM74*	N116S	benign	0.400	S116N	benign	0.000

[a] Nucleotide in the reference freshwater stickleback genome from Bear Paw Lake

[b] Alternative nucleotide present in the Salmon River BACs and in the sequenced 21 genomes

[c] This change was not a consistent difference between additional marine and freshwater populations, with both predicted amino acid variants found in both ecotypes (Jones et al. 2012).

Table S7. Human DNA Samples Used for PCR Validation of hCONDEL.305 and hCONDEL.306, Related to Figure 5

ID	Population
NA10473	Pygmy
NA10539	Melanesian
NA10975	Mayan
NA11198	Peru
NA11373	Cambodian
NA11521	Druze
NA13597	Atayal
NA13619	Adygei
NA15199	Human Variation Panel - Hungarian
NA15886	Basque
NA16654	Chinese
NA17003	Human Variation Panel - Northern European
NA17021	Human Variation Panel - Indo Pakistani
NA17041	Human Variation Panel - Middle Easatern (Version 1)
NA17094	Iberians
NA17109	Human Variation Panel - African American Panel of 50
NA17342	Human Variation Panel - African South of the Sahara
NA17371	Human Variation Panel - Greek
NA17380	Human Variation Panel - Africans North of the Sahara
NA17385	Human Variation Panel - Pacific
NA19027	International Hapmap Project - Luhya from Webuye
NA19238	International Hapmap Project - Yoruba in Ibadan, Nigeria
NA20845	Gujarati Indians in Houston

SUPPLEMENTAL REFERENCES

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nat. Methods *7*, 248-249.

Burge, C., and Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. J. Mol. Biol. *268*, 78-94.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792-1797.

Hosemann, K.E., Colosimo, P.F., Summers, B.R., and Kingsley, D.M. (2004). A simple and efficient microinjection protocol for making transgenic sticklebacks. Behaviour *141*, 1345-1355.

Indra A.K., Warot X., Brocard J., Bornert J.M., Xiao J.H., Chambon P., and Metzger, D. (1999). Temporally-controlled site-specific mutagenesis in the basal layer of the epidermis: comparison of the recombinase activity of the tamoxifen-inducible Cre-ER(T) and Cre-ER(T2) recombinases. Nucleic Acids Res *27*, 4324-4327.

Kapitonov, V.V., Tempel, S., and Jurka, J. (2009). Simple and fast classification of non-LTR retrotransposons based on phylogeny of their RT domain protein sequences. Gene *448*, 207-13.

Kawakami, K., Takeda, H., Kawakami, N., Kobayashi, M., Matsuda, N., and Mishina, M. (2004). A transposon-mediated gene trap approach identifies developmentally regulated genes in zebrafish. Dev. Cell *7*, 133-144.

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

Muffato, M., Louis, A., Poisnel, C.E., and Roest Crollius, H. (2010). Genomicus: a database and a browser to study gene synteny in modern and ancestral genomes. Bioinformatics *26*, 1119-1121.

Price, A.L., Jones, N.C., and Pevzner, P.A. (2005). De novo identification of repeat families in large genomes. Bioinformatics *21*, i351-i358.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011) A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279-283.

Soriano, P. (1999). Generalized *lacZ* expression with the ROSA26 Cre reporter strain. Nat. Genet. *21*, 70-71.

Szymczak, A.L., Workman, C.J., Wang, Y., Vignali, K.M., Dilioglou, S., Vanin, E.F., and Vignali, D.A. (2004). Correction of multi-gene deficiency in vivo using a single 'self-cleaving' 2A peptide-based retroviral vector. Nature Biotech. *22*, 589-594.

Van Ooijen, J., Boer, M., Jansen, R., and Maliepaard, C. (2000). MapQTL 4.0: software for the calculation of QTL positions on genetic maps (user manual).

Van Ooijen, J.W., and Voorrips, R.E. (2001). JoinMap 3.0, Software for the calculation of genetic linkage maps. In Plant Research International (Wageningen, the Netherlands).