File S1

Supplemental Materials and Methods

Grandparental population phenotyping

Skeletal morphologies of Japanese marine and Paxton benthic adult wild fish were compared by micro-computerized tomography using a Scanco uCT 40 scanned at 55kvp, 145 uA, at high resolution averaging four frames.

Phenotyping skeletal traits in F2 fish

We phenotyped 110 skeletal traits using a variety of methods described below. All traits were quantified on the left side, except for (1) premaxilla height and length which were quantified on the right side, and (2) premaxilla width, frontal width, supraoccipital traits, median fin and vertebral position traits, which are bilateral or midline measurements (see Figure 2). All linear measurements were quantified using an eye reticule on a Nikon SMZ1500 dissecting microscope unless noted otherwise.

Dissection method for branchial trait phenotyping

We developed a method to dissect out the entire branchial skeleton and mount it flat on a coverslip (Figure 2A). Briefly, under a Zeiss STEMI 2000 dissecting microscope with watchmaker's forceps, eyes were removed, and four cuts were made with iris or vannas scissors: two bilateral cuts dorsal to the opercle and hyomandibula through to the eye sockets, a cut across the frontal bone through the eye sockets, and a cut through the midline parasphenoid bone. Ventrally, the ceratohyals were disarticulated from the basihyal, and the urohyal removed. Next, the entire facial skeleton was removed, exposing the branchial skeleton. The epibranchials were detached from the neurocranium and the branchial skeleton removed by pulling the gut tube away from the rest of the fish. Soft tissue including the gut was removed and a single midline incision was made between the dorsal tooth plates to allow mounting the branchial skeletons flat on bridged coverslips as in Figure 2A. This method enables visualization of the entire branchial skeletal pattern from a dorsal view, as well as previously described variation in the pigmentation of the gill filaments from a ventral view (Miller *et al.* 2007).

Gill raker traits

Along the anterior/posterior axis, gill rakers are distributed across nine rows projecting from both the anterior and posterior faces of all five branchial segments, except for the fifth branchial segment, which has only an anterior row (Figure 2A). Using the edge of Alizarin-positive branchial bone staining, we defined four dorsal-ventral raker domains as follows: (1) hypo (all rakers medial to the ceratobranchial), (2) cerato (bounded by the edges of the ceratobranchial bones), (3) joint (between

epibranchial and ceratobranchial), and (4) epi (dorsal to the epibranchial) (Figure 2B-E). If a raker spanned these bone landmarks, the center of the raker base was used to assign each raker to a domain. We recorded raker number in each of these 25 anterior-posterior and dorsal-ventral domains using a Zeiss STEMI 2000 dissecting microscope. We also combined the individual domain phenotypes into 19 composite phenotypes in the following possible developmental modules: rows, segments (branchial arches), odd rows, even rows, all, and dorsal/ventral domains (hypo, cerato, joint, and epi).

In addition to gill raker number, we directly measured the inter-raker spacing distance at three positions (lateral, middle, and medial, Figure 2F) along row 2 rakers. Lateral spacing was measured between the second and third raker from the ceratobranchial-epibranchial joint, middle spacing was measured between two rakers in the middle of the ceratobranchial, and medial spacing was measured between the second and third raker from the hypobranchial-ceratobranchial joint. All three spacing measurements were made between the center of the base of the two rakers being measured. For all three spacing measurements, if an atypical raker spacing was present following the above landmarks, an adjacent raker space was recorded if it appeared more typical of the spacing within the row.

We phenotyped gill raker length (from raker tip to ceratobranchial bone of the third raker from the ceratobranchial/epibranchial joint in rows 1 and 2), but after no significant genetic effect was detected after scoring 92 F2 males, we did not pursue this trait further.

Pharyngeal dentition

We quantified pharyngeal tooth number on all three pharyngeal toothplates: the two dorsal toothplates (DTP1 and DTP2) attached to the pharyngobranchials (ANKER 1974), and the one ventral pharyngeal toothplate (VTP) attached to the fifth ceratobranchial (Figure 2G). Teeth were counted using a Zeiss Axiophot compound microscope with DIC optics. Baby teeth that were visible under DIC but did not stain with Alizarin red were not counted. In addition, we measured the lengths and widths of all three toothplates (Figure 2H) by recording the longest and widest point-to-point measurements between Alizarin-positive toothplate bone.

Branchial bones

Along the dorsal-ventral axis, the branchial skeleton consists of: four epibranchials (EBs, dorsal bones in the roof of the buccal cavity); five ceratobranchials (CBs, long ventral bones in the floor of the buccal cavity), and three hypobranchials (HBs, short ventral bones in between the ceratobranchials and the midline). We measured the lengths of all five ceratobranchials and the first epibranchial using the two anterior corners of Alizarin-positive bone as landmarks (see Figure 2I). The lengths of the highly

three-dimensional epibranchials 2-4 and the widths of all ceratobranchials and epibranchials were not measured due to marked variation in mounting angles.

Jaw traits

Premaxillas were manually removed then soaked for several minutes in a dilute 2.5% bleach solution to remove soft tissue before measuring height, width, and length as in Figure 2J. Lower jaw measurements were quantified by dissecting out and separating the left dentary and articular as in Figure 2K, acquiring digitized images with an Evolution MP camera using ImageProPlus on a Leica MZFLIII microscope, then using ImageProPlus software to make linear measurements as in Figure 2K,L.

Skull and opercle traits

We quantified four skull traits: the linear measurement of frontal width or interorbital distance (Figure 2M), and three measurements of the supraoccipital crest (Figure 2N). Supraoccipital traits were quantified from digital images of the dorsal view of the skull taken with an Evolution MP camera on a Leica MZFLIII microscope and analyzed with ImageProPlus software. Three measurements of opercle size and shape were made: the length and width of the opercle, and a measurement of the width of the neck of the opercle (Figure 2O).

Median fin and vertebral traits

Spine serrations were scored from digital images of the second dorsal spine acquired with a Nikon D1X camera fixed to a Nikon SMZ-U microscope. Area of the anterior surface of the spine was calculated by counting the number of pixels in Photoshop (Adobe) and converting to square millimeters. The serration area (SRA) was calculated by subtracting a digitally smoothened dorsal spine area (i.e., a spine without serrations, SDSA) from the total spine area (SPA, Figure 2P). Pterygiophore and fin ray number, and anal spine lengths were quantified under a Leica S8APO microscope with an eye reticule. For all vertebral traits, animals were first X-rayed (Figure 2Q) at 5x magnification for 15-20 seconds at 20 kV in a Micro–50 cabinet specimen radiography machine (Faxitron). Positions of bones in the median skeleton were assigned a numerical value corresponding to the closest vertebra as described (AHN and GIBSON 1999). The position of the last dorsal and anal fin ray was determined based on the position of the pterygiophore that supported the fin ray. On occasion, the element was judged to be equidistant from two vertebrae and was assigned a value that was an average of the two vertebrae.

Genome-wide linkage map

Linkage map construction

A set of 275 microsatellites was genotyped in a single full-sibling family ("Family 4") of 370 fish from a Japanese marine (JAMA) by Paxton benthic (PAXB) freshwater F2 cross (Colosimo *et al.* 2004). These markers consisted of previously described sets of genome-wide microsatellites (Albert *et al.* 2008; Colosimo *et al.* 2004; Peichel *et al.* 2001) and markers near previously mapped genes (Colosimo *et al.* 2005; Knecht *et al.* 2007b; Miller *et al.* 2007; Shapiro *et al.* 2004). In addition, we added 16 new markers to the genetic map by genotyping new microsatellites near candidate genes with important roles in pharyngeal arch patterning in other vertebrates (*Dlx1/2*, *Dlx5/6*, *Dlx3*, *Msx1*, *Edn1*) as well as new positional markers. New markers were identified using a variety of methods including degenerate PCR, bacterial artificial chromosome (BAC) screening by radioactively labeled overgo hybridization, BAC end sequencing, physical map information, and publicly available previously sequenced BAC ends, as described in Supplemental Tables S1 and below. A linkage map (Figure S1) was constructed with JoinMap 3.0 (Kyazma), using previously described settings (Peichel *et al.* 2001) but by accepting more conservative LOD 6 groupings. The total map length is 1287.8 cM over 21 linkage groups, resulting in an average marker spacing of 5.1 cM. For each linkage group, proper phase was determined from the grandparental genotypes.

Cloning Dlx and Msx genes

Intergenic (*Dlx5/*6) or genic (*Msx1*, *Dlx3a*, and *Dlx3b*) regions of new genes added to map were amplified by PCR using the following primers (all sequences 5' to 3'). For *Dlx5/6,* PCR primers GGTGGGAAAGTGTTTGCACACC and CTGAGACAATCCGCATTCCTGTGG were designed to conserved intergenic sequences (Zerucha *et al.* 2000) which were found to flank Stn339 in intervening genomic sequence. For *Dlx3a* and *Dlx3b*, portions of two stickleback *Dlx3* genes were amplified and sequenced using a common forward degenerate primer (GGGTGAAGATHTGGTTYCARAA) and a reverse degenerate primer for either *Dlx3a* (CGGGCTGRTACCARTTYTGRTG) or *Dlx3b* (CGCCCTGYTGRTACCARTGRTT). The resulting *Dlx3* sequences were used to design two gene specific overgoes (see below) for BAC screening. Both *Dlx5/6* and *Dlx3* PCRs used Little Campbell marine genomic DNA as a template. For *Msx1,* degenerate RT-PCR primers CCGTTCAGCGTCGARGCNCTNATGGC and GGGGTGRTACATRCTRTANCC were used with oligo-dT reverse primed cDNA harvested from a 1 cM long Little Campbell marine fry. The resulting RT-PCR amplicon sequence was used to design overgoes (see below) for BAC screening.

Overgo screening, BAC end sequencing, and genotyping

Overgo screening was performed as described at www.chori.org/bacpac/overgohyb.htm. *Dlx1/2* overgoes were directly designed to conserved intergenic sequence (Ghanem *et al.* 2003). Forward and reverse overgo sequences (5' to 3') for each marker or gene were: *Msx1*:
CGGTAGTCTGGATACTTCAGTTCC and GCCCATCGATAAAGCAGGAACTGA; Stn207: TTTCAGCAGGTGCAACGTTTCCAC and AACTAAGAAGGCGAGCGTGGAAAC; *Dlx1/2:* ACCAAGATCTCGAGTGCACAATGT and CCTCATTACGCTGATGACATTGTG; *Dlx3a*: GGCGGCAGTATTAAGAGTAATGCG and CGGTGGGATCCACAAGCGCATTAC; *Dlx3b*: CCGACGCACAGCTCGTCGCCGCCA and TATAATCCTCCAGGTATGGCGGCG; Stn48: GTGCCAGAAAACTTGCATTCCAGG and ATCCCCTCACGTCACACCTGGAAT; EaccMgtg: GCAGGGTGATTGAATGTCTTCACT and GTCCTTAGGAAGATGCAGTGAAGA; 48B15.t7: AACAGTGTTGAGCGCTGAAATGCC and ACCTGTATGCACACACGGCATTTC; Stn292:
AAGATACGGGCTGATGAGCAGTGA and TTCTTACTACGCCTCCTCACTGCT; Stn222: TCGCACTTCAGACACTAAGCCTTG and TGAAGGGTGTCCAAACCAAGGCTT; Bmp6: TGTGACGTTGACCTCAGCTAGACT and GAGGATTTAAACCGGGAGTCTAGC.
For overgo screening, three pairs of labeled overgoes were combined in one hybridization bottle containing four filters, and positive BACs subsequently identified using a combination of the physical map (Kingsley *et al.* 2004) and PCR screening. BAC ends were sequenced using ABI3730xl manufacturer's suggestions, using 8 uL of ABI BigDye per 20uL sequencing reaction. Genotypes were generated essentially as previously described (Miller *et al.* 2007; Peichel *et al.* 2001).

QTL mapping and analyses

Trait transformations

Trait processing and analysis was performed in R (http://www.R-project.org/). A custom pipeline was made to correct each trait for sex and/or size-dependence, log-transform if appropriate, and to remove phenotypic outliers as follows. First each trait was tested for size dependence by linear regression vs. standard length (SL), for sex dependence by a one-way ANOVA using sex as a factor, and for sex and size dependence using SL as a covariate and sex as a main effect in a General Linear Model (GLM) ANOVA. If the trait was neither sex nor size dependent, raw trait values were used for QTL mapping. If there was SL-dependence but no sex-dependence, traits were regressed against SL to obtain residuals. If there was sex-dependence but not size dependence, sex was corrected for using the residuals of a one-way ANOVA with sex as a factor. If traits were significantly dependent upon both sex and size, the residuals of the GLM ANOVA were used for QTL mapping. Outliers (defined as fish that had trait values greater than four standard deviations from the mean trait value) were removed and ANOVAs, regressions, and GLM ANOVAs were redone without outliers. Outliers were rare and consisted of only 35 values for 17 total traits (AH, DH, DL, DS1L, DS2L, FDP, IL1, IL2, OPL, OPN, PD4, PML, PMW, SDS2A, SOL, SPA, and SRA). Traits were log-transformed when the

transformation equalized variances (in sextiles ranked by standard length) by Levene's test for equality of variances, and/or normalized the residuals by an Anderson-Darling test of normality.

QTL mapping

QTL mapping was performed in R/qtl (Broman and Sen 2009). Initial QTL mapping was performed with *scanone* with Haley-Knott regressions (hk). For each phenotype, ten thousand permutations were performed to determine a LOD threshold at which alpha equals 0.05. The average of these trait-specific thresholds was 4.1; thus this value was used as the QTL significance threshold for all traits. All significant QTL by *scanone* were also identified by *stepwise* mapping, so the larger *stepwise* set of QTL are presented here. The *stepwise* algorithm was performed by an automated forward-backward stepwise search for QTL using *stepwiseqtl* with a main penalty of 4.1, which was the average penalty from 100 *scantwo* permutations for each trait. QTL peak markers and LOD scores were calculated using *refineqtl* and percent variances explained were calculated with *fitqtl*. For a small number of traits (n=11), the *stepwiseqtl* output included markers with a LOD less than 4.1. These markers were conservatively removed. In 10 cases, the *stepwiseqtl* output included two markers on the same chromosome. Only cases where both peak markers had LODs greater than 4.1 and also had non-overlapping 1.5 LOD intervals were considered as two QTL. In cases where the two linked markers had overlapping 1.5 LOD intervals, only the peak marker with the highest LOD was considered a QTL. LOD scores for QTL on chromosomes that did not have significant effects were determined with *addqtl*, adjusting for QTL that were identified from the stepwise search. Additional QTL were included from *addqtl* if they surpassed a 4.1 LOD score threshold and LODs were recalculated as above. This *addqtl/refineqtl/fitqtl* process was iteratively repeated for three rounds and all QTL that had a LOD score above 4.1 in the final *fitqtl* model were included in the final QTL set. LOD scores for phenotypes with no significant QTL were determined by *scanone* with Haley-Knott regressions. Heat maps in Figures 3 and S2-S6 use color schemes from http://colorbrewer2.org/.

Anatomical specificity of QTL

For investigating the anatomical specificity of QTL (Figure 4), the subset of QTL with clearly or likely serially homologous domains (QTL in the raker, teeth, branchial, jaw, and spine classes) were considered. QTL controlling raker spacing were excluded because this phenotype was only quantified on one segment, and QTL controlling toothplate size and tooth number were analyzed separately.

Principal Components Analysis

To determine the major axes of skeletal variation in the dataset, we performed Principal Components Analysis (PCA) using the *FactoMineR* package in R. Phenotypes were size/sex/log-adjusted as necessary (see Table S2) and Z-scored. Missing data were imputed using the *imputePCA* command, then weighted PCA was performed using phenotype weights such that the total weight for each phenotype class was equal. We performed PCA on all phenotypes, excluding composite phenotypes where the non-composite phenotypes comprising the composite phenotype were also present. The first five principal components explained 18.4, 9.4, 4.9, 4.6, and 4.4 percent of the phenotypic variance, respectively. The coordinates for each fish for the five largest principal components were extracted and QTL were mapped as described above.

Investigating biases in dominance

For simulations investigating QTL detection biases for dominance, two cases of QTL were compared: dominant QTL (d/a=1; heterozygotes have the same mean phenotype as the benthic B_1B_2 genotype) and additive QTL (d/a=0; heterozygous mean phenotype equals the mean of the M_1M_2 and B_1B_2 homozygous phenotypes). 400 samples for each value of dominance were used. Effect sizes span the boundaries of detection. Quantities in the simulation were based on the results of the analysis of the trait "DTP2," using Haley-Knott regression and a step size of 5 and a LOD threshold of 4.5. The simulations model a normally distributed trait, with QTL effect sizes ranging from 0 to 5, and a constant residual variance of 30 within each genotypic class, and assume no genotyping errors and no missing values. After detecting four QTL, all QTL were entered into a linear model in R/qtl. Obtained effect sizes for the four QTL ranged from LODs of 5 to 8, with a residual variance of about 28. Exploration with these numbers showed that effect sizes between 0 and 5 led to probabilities of detection ranging from about 0 to about 1.

Overlap with marine-freshwater divergent regions

The number of marine-freshwater divergent genomic regions that show evidence of repeated selection [the HMM or CSS signals of selection from (JONES *et al.* 2012)] within the 1.5 LOD interval of each QTL was determined. To test for enrichment of signals of selection within various groups of QTL, the mean number of overlaps of the QTL group was divided by the mean number of overlaps from 1000 simulations of random placement of signals of selection across the genome. P values were calculated by comparing the mean number of signal of selection-QTL overlaps to a null distribution of simulated placements of signals of selection. For determining the number of signals of selection overlapping the three trait clusters, the following coordinates were used: 2.34-28.56 Mb, 1.71-14.68 Mb, and 0-8.94 Mb for chromosomes 4, 20, and 21, respectively. These physical coordinates correspond to the genetic range on each chromosome that spans all of the clustered QTL shown in Figure 7, based upon markers flanking the 1.5-LOD interval listed in File S3. These coordinates were also used to identify putative

developmental regulatory genes within the trait clusters with Gene Ontology (GO) terms of "multicellular organismal development," "growth factor activity," or "regulation of transcription, DNA-dependent."

Files S2-S4

Available for download as Excel files at http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.114.162420/-/DC1

File S2 Genotype and Phenotype data

All raw phenotype, adjusted phenotype (see Table S2) and genotype data used for QTL mapping. Raw genotypes were coded as follows: A and C are the marine grandparental alleles, B and D are the freshwater grandparental alleles. Genotypes of F2 fish were coded as follows: NA = missing, 1 = AC, 2 = BC, 3 = AD, 4 = BD, 5 = A = AC or AD, 6 = B = BC or BD, 7 = C = AC or BC, 8 = D = AD or BD, 10 = AD or BC.

File S3 Summary of all trait QTL

All detected significant QTL affecting skeletal traits are shown. QTL statistics are displayed in two ways: 1) allowing interpolated markers calculated every cM to be the peak markers or boundaries of the 1.5 LOD intervals or 2) allowing only real markers to be the peak markers or boundaries of the 1.5 LOD intervals. Mean phenotypes are displayed for Z scored phenotypes after size adjustment, sex adjustment, and log transformation, as necessary. For each QTL, one additive effect (half the difference between the two homozygous classes) and two dominance effects (difference between a heterozygous class and the mean of the two homozygous classes) is shown, as well as two dominance values (*d/a).* QTL in raker, opercle, median fin, and vertebrae classes that had expected directions of evolutionary change were determined to be concordant or antagonistic to the direction of evolutionary change.

File S4 Summary of principal component QTL

All detected significant QTL for the five largest principal components (PC1-5) are shown. QTL statistics are displayed in two ways: 1) allowing interpolated markers calculated every cM to be the peak markers or boundaries of the 1.5 LOD intervals or 2) allowing only real markers to be the peak markers or boundaries of the 1.5 LOD intervals.