

MATERIALS AND METHODS

Fish Samples

Sticklebacks were collected in minnow traps or seine nets from the locations summarized in Table S1. Laboratory populations were crossed using *in vitro* fertilization and grown in aquaria containing 0.35% seawater (~pH 7), at 16 to 18 °C, and a light cycle of 16 hours light, 8 hours dark. Skeletons were prepared by Alizarin red staining of ethanol fixed samples, and DNA preps were prepared from tail clips as described (1).

AFLP Screen

Arbitrary primers were used to amplify large numbers of stickleback genomic fragments from pooled DNA of 30 completely plated and 30 low-plated F2 progeny from the Japanese marine X Paxton Lake benthic mapping cross (Cross 1 in (2)) as previously described (3). Fragments that differed in abundance or size between pools were purified from 42% urea/5% polyacrylamide (Long Ranger gel solution, BioWhittaker Molecular Applications, Rockland, Maine) gels, cloned into TOPO-TA vectors (Invitrogen), and sequenced using primers that flanked the multiple cloning site. Locus-specific primers were then designed to amplify particular sequences from individual animals. Amplified fragments were tested for single strand conformation polymorphism (SSCP) using electrophoresis on polyacrylamide gels prepared with MDE or Long Ranger gel solution (BioWhittaker Molecular Applications, Rockland, Maine). Primers and PCR conditions for the newly developed *Stn345* and *Stn346* markers: *Stn345*, AATTCTGCCAAACCATCC, TATCCAATCTGCGATTTACC; *Stn346*, TAAGATGACAGCTTGAACGC, AATTCTCACTGATGTACTGTTTACC;

amplification assays; 5ng genomic DNA, 0.5 to 1 uM each primer, 0.25-0.125 mM of each dNTP, 1 uCi radiolabeled deoxycytidine 5' triphosphate [α - 32 P], 1.5 mM MgCl₂, 1X PCR buffer and 0.25U AmpliTaq polymerase/10 microliters reaction (Applied Biosystems); cycling conditions (referred to as PFC8): 93 °C 3 min, 95 °C 30 s, 56 °C 30 s, 72 °C 30 s, 5 cycles of 94 °C 30 s, 50 °C (*Stn 345*) or 56 °C (*Stn346*) 30 s, 72 °C 30 s, 35 cycles of 90 C 30 s, 50 °C (*Stn 345*) or 56 °C (*Stn346*) 30 s, 72 °C 30 s, followed by 72 °C 10 min, then cooled to 4 °C. The default annealing temperature for the PFC8 protocol is 56 °C.

Chromosome Walk

The CHORI-213 library (4) was screened by hybridization with overgo oligonucleotides designed to sequences in *Stn345* (AATTCTGCCAAACCATCCTGAGA, GCACCTCCCAGATCATTCTCAGGA), *Stn346* (CTTGAACGCTTCAGAGCAGCATCT, CGGAGGATTTGAGTTTAGATGCTG), and the ends of BAC clones 2D11 (AATTCAGTATCATCATCGAGTGC, GCATCCTCTCCTCTGGCACTCGA) and 128O17 (TGCTTGTTGTGCAGTGCAGGTTCT, AAACAATGACGGGGTGAGAACCTG). Sequences from the ends of individual clones were determined by cycle sequencing using primers from the vector backbone (SP6 and T7). Primers designed to BAC end sequences were used to map new positions on the chromosome walk in recombinant F2 animals from Cross 1. *Stn347* was scored using an SSCP assay with primers ACCAGCTTGTAACCGGTAGG and ATTGATATCCGGCCTCTGG and the PFC8 PCR protocol. The library screens and the

physical mapping data from a large BAC fingerprinting project (4) shows that the 0.68 cM interval from *Stn345* to *Stn346* can be spanned by 6 BAC clones (CHORI213-2D11, 128O17, 283M13, 131C10, 49K04, and 129C17) and corresponds to a physical interval of approximately 900 kb.

BAC clones 2D11, 128O17, and 129C17 were sequenced to completion (GenBank AC144489, AC145509, AC146539) as described (5). In order to identify genes in the candidate interval, the assembled stickleback sequence was analyzed by GENSCAN (6) and each predicted gene was compared to known genes in other species using BLASTP (7). Gene names in Figure 1 are based on official HGNC human names (8) where possible.

The CHORI-215 stickleback BAC library (4) was screened using two overgos specific for the last exon of *Eda*: CAGAGGATCTCCATACGCATAGTG and TGAGCGTGTCCTCATACTATGC and intron 6 of *Eda*: ACTTGCCAGAGCTGAACCCTCG and CAACCTGCTAACATTCGAGGGTT. Positive BAC clones 201K11 and 4D15 were isolated, and the *Eda* to *Gja3* interval was resequenced from the Paxton Benthic clones (GenBank ACxxx).

Linkage Disequilibrium Analysis

Thirty-two new sets of primers were designed flanking microsatellite sequences present at approximately 12 kb intervals throughout the assembled stickleback sequence (*Stn348* to *Stn379*): *Stn348F* CGTGTCTGTCAAGCTTGTTCC, *Stn348R* AATCCGTCGCTGCTCTGC, *Stn349F* GGCAAAGAAAGGTGGACAAC, *Stn349R* CTTTCAATGGGCCACCAAG, *Stn350F* TGTCTCAGTGCCATTTCCAG, *Stn350R*

GCCGTGCAGTTGAAAGAATC, Stn351F CAGTCTGCAAGGTGACAAATG,
Stn351R TCTTGTCCTTTGGGTGTGTG, Stn352F CCTCCTCTTGCAGAAACCAG,
Stn352R GCACCCTTGCAAAGAGAAAG, Stn353F CACGGCTCAATACTGTGTGC,
Stn353R TCCATCTCGTGCTGTCTTTG, Stn354F CTCAGCCTTGAGCACAAATG,
Stn354R ATGCTGCCAGATTGTGATTG, Stn355F AGGCATCGACATCGAGAGAG,
Stn355R TCTTCCCTTCGCTCTGTAC, Stn356F ACTCCACCTACCCTGATGTCC,
Stn356R AGGCATACTACCGTTACCTCTG, Stn357F ATGCTCCGACAGACATTGG,
Stn357R CGGCTCGGAGAGACATTG, Stn358F CAGAAGAATGGTGCAGTGATTG,
Stn358R TGTGCTTAGGGTTGCAAATG, Stn359F TGACATTGGTCCGCTTCC,
Stn359R ACCTCCACATTTGACAAGC, Stn360F CTGTCCGAAACGATATGTGG,
Stn360R TGCAGATGATGTTTCAGTGC, Stn361F AGCCTTTCTCAGCCTTTGC,
Stn361R TGCTGACCTTCTCCCTTAGC, Stn362F
AAAGGAACTTGTTAGCACCAC, Stn362R ATCAGGAGTCCGCCAGTTC,
Stn363F CAGAGGAGATGTGGAGACACC, Stn363R
GAGGAGGATATGTTGGAGTTGC, Stn364F TCCCAGGTCTCTCCTCTGC, Stn364R
TAACGAGCAGCCCTATCTGG, Stn365F ACTTGGCCAGAGCTGAACC, Stn365R
AGGGTTTAACTACAACCAGAGC, Stn366F TGAAGTGATCGTCTCATTGCAT,
Stn366R AAAGGCGACCATGTTCCA, Stn367F TTGTGTACCACCACAGCTGAA,
Stn367R TGACAATAGTTGAGGCAATG, Stn368F GTGGATCATGTTCCCTGAGC,
Stn368R CCACGTGGACACAGACAGG, Stn369F
AGCAAAGGTCAGGTTGATCC, Stn369R TGAGGCAGAGTCCAGATTGC,
Stn370F ATTGCCAAGAAGAGCGAACC, Stn370R
GCTCTGTAATTGTGACCTGACC, Stn371F TACACAGCGGGTGTCTCTGC,

Stn371R CGACAGTTTCCAACATCATGG, Stn372F
CCTTCCCACCTCCTAATCTCC, Stn372R AGAAACACTTGCGGCTATGG,
Stn373F AGCCACGCGATAATTGAGG, Stn373R CCGCGGGTAACATAAACC,
Stn374F AAACGAAACCGACAACAAGG, Stn374R
GTTGAAGCCTGGTTCTTTTCG, Stn375F TCCAATATGCTTGCTCTCTGC,
Stn375R TGAGCCATGAAAGGATTTGC, Stn376F
CCAGGTCCGCTGTGTTGC, Stn376R TGGAACAGAAAGGGTACAGG,
Stn377F AGTTGGGAGTGACACCAAGG
Stn377R GCCAAAGACTCTGGATGAGG, Stn 378F CGGATCCGTCGCCATAAA
Stn 378R GGTCAAACTTTGTGTGAATGTG, Stn379F
CTCAGCAGCAGTTGAGTTCC, Stn379R GATTCCTGATGCAGCTACGG.

Alleles present at each microsatellite locus were determined in 46 completely and 45 low-plated individuals from the FRI stickleback population using PCR amplification and analysis on an ABI3730xl sequencer with GeneMapper v3.0 software. PCR conditions followed the PFC8 protocol except that for marker *Stn378*, 2 mM MgCl₂ was used, and for markers *Stn349-352*, *354,358,360,364* the annealing temperature was 60 C. The distribution of allele frequencies in completely and low-plated fish was compared using CLUMP, which performs a modified chi-square analysis to determine significance between populational differences (9). Figure 1B shows the log p-values of the chi-squared value (T4) from the 2X2 contingency table generated by CLUMP for each marker.

Gene Sequencing

Exons and introns of *Eda*, *Tnfsf13b*, *Garp*, and *Gjb1* were sequenced by amplification of genomic DNA from 1-2 animals in each population. Primers: *Eda* exon 1, TCCTCACGTGAATGAGATGC and ACAACCTGACACCGTGATCC; *Eda* exon 2/intron2, GAGCGTCTAGGTGCATAAAGC and GGATTTGGGTTTCTTTCTTGC; *Eda* exons 3-4, CGTTTCAAGCTTAATCACACG and TGAACTCGTGGGTGTTCTCC; *Eda* exons 5-6, ACTCCCTTCACTGGGAGAGC and GACTTTGCGTCAAAGACAGC; *Eda* exons 7-8, AGTTAAACCCTGCTGAGAGC and AGGCCCGGTGGTTGTGTAGTT; *Tnfsf13B* exon 1-intron2, CTGCGTCTTCCTTCTCACG and AACTGCAGGAATGATGTTGG; *Tnfsf13B* intron5-exon6, CATTGAGCGCTGGGCAA and GCGACAAGCTGCGTGATTAT; *Garp* exon 2 5'end, ACCAGCTCGACCTCATCTCT and TGCAGATACTTCAAACGGTTC *Garp* exon 2-3'end, TCTGCTACCAAGCAGTTTGA and AGTTTGCTATTCAGTTTCTCAC; *Gjb1* exon1-5' end, GAGATGTAGCTGAGCCTCCAC and TCATGTACGTCCACCACAGC and *Gjb1* exon1-3'end, TGATCAGCATCGTCTTCAGG and GATGGTTTAAACGTGCCTGT. Most PCR reactions were 20 μ l and used 20 ng of genomic DNA and the PFC8 cycling conditions. Amplifications for *Eda* exon 3-4 and *Gjb1*-exon1-5' included 10% DMSO in the reaction. Extension time for *Eda* exons 7-8

was modified to 60 sec. *Eda* exon 1 PCR conditions were modified to be 95 C 2 min, followed by 35 cycles of 95 C 30 sec, 56 C 30 sec, 68 C 30 sec, followed by 68 C 5 min.

Amplified products were purified over a Sephacryl S-300 High Resolution Resin (Amersham Biosciences) column and directly sequenced using 20 ul reactions containing: 2-6 ul of column eluate, 3.2 pmol of sequencing primer, 3.2 ul of 5X sequencing buffer, and 1.5 ul of BigDye 3.1 sequencing premix (Applied Biosystems). Sequencing primers were the same as those used for PCR amplification, or additional primers designed for *Eda* exon 1, (AACATTGTATCCCGCTTCT), *Eda* exons3-4 (AATACAACAACCACGTTGAC, TTTAATACCTGCTGGACCT, and AGGAACAATGTGGCTGACGGA), and *Eda* exons 7-8, (GCATGCTTAGATCTTTCTGA and AGATGGAGCCTCTCCAAG).

RT-PCR

RNA from entire juvenile fish was isolated from crosses between FRI low-plated and FRI low-plated, FRI completely plated and FRI completely plated, FRI low-plated and PAXB low-plated, and LITC completely plated and LITC completely plated sticklebacks using the RNeasy Fibrous Tissue Midi Kit (QIAGEN). cDNA was made using oligo dT primer using the Retroscript Kit (Ambion). PCR amplification of the *Eda* transcripts were done in 25 μ l reactions with 2D11_195507 (TGGGATTATTCCTGCTGTCG) and 2D11_205407 (ATCCACCATCACCTCGTAGC) primers (0.50 uM) in *Eda* exons 1 and 8 respectively, 1.5 mM MgSO₄, 310 uM dNTPs, 0.75 units of AmpliTaq (Applied Biosystems), 1X PCRx amplification buffer and 0.5X PCRx enhancer solution

(Invitrogen). PCR cycling conditions were: 94 °C 2 min, followed by 35 cycles of 94 °C 30 sec, 55 °C 30 sec, 72 °C 60 sec, followed by 72 °C 5 min, followed by 4 °C hold. The expected 800 bp amplicons were gel-purified, cloned into TOPO-TA vectors (Invitrogen), and sequenced with primers surrounding the multiple cloning site.

Phylogenetic Analysis

A total of 1328 bp of sequence from *Eda* exons 2,3,4,7 and 8 were amplified and sequenced from 4 individuals per population. Sequences were aligned using ClustalX v1.803 (10), and a single composite sequence was constructed for each population using standard DNA ambiguity codes where necessary to accommodate polymorphisms.

For the nuclear phylogeny (hereafter, control tree), 193 SNPs from 25 nuclear loci were genotyped in 4 fish from each of 20 of the same populations previously used for EDA sequencing. Loci were chosen randomly from genes previously identified in a large EST sequencing project (Kingsley, 2004). Primers designed to 3 prime end sequences were initially screened for those that could successfully amplify 400 to 600 bp fragments from genomic DNA of both Japanese Marine and Paxton Benthic fish. Primer sets and PCR conditions for each locus are available upon request.

Each amplified fragment was sequenced with Applied Biosystems (ABI) Big-Dye Terminators from both the 5' and 3' directions. ABI 3730 sequence traces were basecalled with Phred and assembled with Phrap from U. of Washington (11). Polyphred v3 (12) was used to identify potential SNP sites and the resulting potential SNPs were reviewed and recalled in each individual by an experienced technician. SNP data were

concatenated into a single sequence file for each population, using DNA ambiguity codes where necessary to accommodate polymorphism.

EDA and nuclear trees were constructed using Bayesian analysis, maximum likelihood, and maximum parsimony approaches implemented in the MrBayes v3.0b4 (13), PHYLIP v3.62 (14), and PAUP* v. 4.0b10 (15) software packages. No molecular clock was enforced in any analysis, and the trees and branch lengths shown in Figure 3 depict topology rather than distance. MrBayes: we used a two parameter substitution model and a gamma distribution of evolutionary rates across sites, with the value of the shape parameter estimated from the data. The MCMC analysis used 4 chains, 2,000,000 generations sampled every 200 generations, and a burn-in period of 2500. Maximum likelihood: we used a two parameter substitution model and a gamma distribution of evolutionary rates across sites, with 6 rate parameters in the discrete approximation to the gamma distribution. The value of the shape parameter was estimated from the data. Bootstrap analysis was based on 100 replicates. Maximum parsimony: we used at least 5 random addition sequences and up to 10,000 rearrangements per addition sequence to find the most parsimonious trees. Bootstrap analysis was based on 100 replicates, with the results summarized in extended majority rule consensus trees.

We compared goodness of fit of phylogenetic trees to the SNP sequence data to test discrepancy between the EDA gene phylogeny and overall population histories. First, we compared each EDA tree to its control tree (estimated from random nuclear sequences) using the RELL option ($n = 1000$ replicates) of the Kishino-Hasegawa (KH) test (Bayesian and maximum likelihood trees) and using the Wilcoxon signed-rank test (parsimony trees). Next, we used the KH test and the Wilcoxon test to compare each

control tree to the best possible tree calculated from random nuclear sequences in which all freshwater populations except NAKA were constrained to form a monophyletic clade. This analysis was possible for maximum likelihood and parsimony trees only. Constrained trees were generated in PAUP*. *P*-values from the KH tests were similar to those produced by the Shimodaira-Hasegawa multiple-tree tests. KH tests used one-sided *p*-values, as recommended by Felsenstein (16).

Nonsynonymous and synonymous substitution rates (K_a , K_s) were calculated using K-Estimator v6.1 (17).

Survey of Marine Populations for *Eda* Low Allele:

Samples of completely plated fish were collected from the mouth of the Little Campbell River, British Columbia in June and July of 2000 to 2004; and from the mouth of the Navarro River, California in May of 2003. All 302 Little Campbell fish were completely plated, consistent with extensive previous studies of the anadromous marine fish at this location (18). 106 of 109 Navarro fish were completely plated and 3 individuals were partially plated, consistent with a higher prevalence of partial morphs in the more southerly California population (19). DNA samples were prepared from ethanol fixed fish and individual samples were screened by PCR amplification with primers flanking 2 markers within introns 2 and 6 of *Eda*. *Stn380*, which is within intron 2, (GGCATTCCAATGTCTGTGG, CACTGAGAGCAGAGGAGAGACC) amplifies a 184 bp allele in completely plated animals and a 172 bp or a 189 bp allele in low-plated animals, and *Stn381*, which is within intron 6, (CACGGACTTACACCACAACG, ATTCGAGGGTTCAGCTCTGG), amplifies a 175 and/or a 165 bp allele in completely

plated animals and a 193 bp allele in low-plated animals. PCRs were visualized on an ABI 3730xl using GeneMapper v3 software.

Transgenics

The pCI neo-Eda A1 construct (20) was digested with BglI and MamI to release a 3.75 kb fragment containing the CMV promoter, the entire coding region and 1162 bp of the 3'UTR of the Eda-A1 cDNA, and the SV40 poly A site. This fragment was cloned into the pISceI-pBSIISK+ vector (21) that was cut with BamHI and EcoRV. 6-15 ng/ul of this construct was co-injected with 1 unit/ul of I-SceI meganuclease (Roche) into 1-cell stage embryos from a cross between low-plated fish (Friant, CA (FRIL) females and San Francisquito Creek, CA (SFC) males). Injections were carried out as described (22). Non-injected control siblings and injected animals were raised to at least 28 mm, when plate development is considered complete (23).

Genotypes at the endogenous *Eda* locus were verified using primers that flank an indel polymorphism in intron 1 of the *Eda* gene (*Stn382*, CCCTTAGAGAATTTCTAGCAG, CTGTCCCGGATCATAACGC) generating characteristic band sizes of 150 bp from the low-plated allele in FRI and SFC, and a 218 bp fragment from fish carrying the completely plated allele. Presence or absence of the injected transgene was confirmed using PCR with primers that amplify a 211 bp product from the coding region of the mouse EDA-A1 cDNA (CACCTCTGGCACCTAAGC, GCGAACACGCCTACTTTCC). PCR conditions followed the PFC8 protocol, except a total of 30 cycles of PCR and an annealing temperature of 56 °C were used.

Figure S1: Multiple sequence alignments of TNFSF13B, GARP, and GJB1.

Alignments are shown of the predicted protein sequences from marine (SRMA) and freshwater (PAXB) sticklebacks, *Fugu*, mice, and humans. (A) TNFSF13B (B) GARP (C) GJB1. Identical amino acids are shaded black, and conservative substitutions are shaded grey. Red dots indicate amino acid differences and blue dots indicate silent changes between SRMA and PAXB fish.

Table S1: Stickleback populations used in this study.

Mean plate counts are from the referenced papers or from five individuals per population.

| Pop. # in Fig.3 | Abbrev. | Longer Description | Collecting site reference and further information | Mean plate # | Year | Collectors |
|-----------------|---------|---|--|--------------|-------------|---------------------------------------|
| 1 | NAKA | Nakagawa Creek, Japan | (24) (25) | 5 | 2000 | J. McKinnon |
| 2 | JASE | Japan Sea | Akkeshi Harbor, Hokkaido, Japan | 33.8 | 2003 | C. Peichel & J.Kitano |
| 3 | JAMA | Japan Marine | Harutori population belonging to the Japan Pacific, rather than Japan Sea clade (24) (25) | 30 | 2000 | J. McKinnon |
| 4 | WALL | Wallace Lake, Alaska | (26) | 3.8 | 2003 | R. Bernhardt & F. vonHippel |
| 5 | AKST | Alaska Stream | (24) | 6 | 2001 | J. McKinnon |
| 6 | AKMA | Alaska Marine | (24) | 32.2 | | J. McKinnon |
| 7 | PAXB | Paxton Lake benthic, Texada Island, Canada | (27) | 0.35 | 2002 | D. Schluter Lab |
| 8 | PAXL | Paxton Lake limnetic, Texada Island, Canada | (27) | 5.5 | 2003 | D. Schluter Lab |
| 9 | SRST | Salmon River Stream | (24) | 5 | 2000 | J. McKinnon |
| 10 | LITC | Little Campbell River, Vancouver, Canada | (18) | 32.7 | 2000-2004 | D. Kingsley Lab, J. Kitano |
| 11 | SRMA | Salmon River Marine | (24) (25) | 27 | 2001 | D. Schluter Lab |
| 12 | COND | Conner Creek, Washington (Site D) | (28) | 7.2 | 2002 | B. Blackman & C. Peichel |
| 13 | NAVR | Navarro River, California | (29) | 31.6 | 2003 | D. Kingsley Lab |
| 14 | SFC | San Francisquito Creek, California | Creek along horse trail at Webb Ranch on Alpine Road, Portola Valley, CA | 5.8 | 2003 | K. Hosemann & A. Launer |
| 15 | FRIL | Friant, California | Dimorphic population of high and low plated fish (29). Sequence analysis and plate counts based on low morph. 36° 59' 7" N, 119° 43' 53.1" W | 7.5 | 1999-2004 | D. Kingsley Lab |
| 16 | WMSO | <i>G. aculeatus williamsoni</i> , Santa Clara River, CA | (30) | 0.13 | 2001 & 2002 | C. Peichel & D. Kingsley |
| 17 | LLOY | Demarest Lloyd State Park, Massachusetts | Tidal stream in Demarest Lloyd State Park on Barney's Joy Road in Dartmouth, Massachusetts | 32.6 | 2003 | P. Colosimo, M. Shapiro, & K. Olivera |
| 18 | WHEA | <i>G. wheatlandi</i> , Massachusetts | Tidal stream in Demarest Lloyd State Park on Barney's Joy Road in Dartmouth, Massachusetts | | 2003 | P. Colosimo, M. Shapiro, & K. Olivera |
| 19 | OMPL | Olmstead Park, Jamaica Plain, Massachusetts | Dimorphic population, as in Friant, CA (31). Sequence analysis and plate counts based on low morph. | 12.8 | 2000 | C. Peichel |
| 20 | NHR | New Harbor River, Nova Scotia, Canada | (32) | 32.2 | 2003 | B. Blackman |
| 21 | GJOG | Gjogur, Iceland | 65° 58' 60" N, 21° 21' 0" W | 33.8 | 2003 | B. Jónsson |
| 22 | BLAU | Blautaver, Iceland | 64° 2' 6.2" N, 18° 59' 2.1" W | 6.4 | 2002 | B. Jónsson |
| 23 | FADA | Loch Fada, Scotland | (33) (34) | 0 | 2001 | C. Peichel & D. Kingsley |

| | | | | | |
|----|------|-------------------------------------|------------------------------------|------|-------------|
| 24 | NOST | Norway Stream | (24) 60° 32.955' N, 5° 2.866' E | 7.2 | J. McKinnon |
| 25 | NEU | Neustädter Binnenwasser, Germany | 54° 6' 41" N, 10° 48' 45" E | 29.2 | M. Kalbe |
| 26 | SCX | Schwale, Germany | 54° 4' 14" N, 10° 6' 00" E | 6.4 | M. Kalbe |

Table S2. Sequence differences between marine (SRMA) and freshwater (PAXB)

BACs across the minimal conserved low morph interval.

There are a total of 269 single base pair changes and 65 length polymorphisms (indels and changes in microsatellite or homopolymer repeats) in the minimum low morph haplotype (Table 1), whose total length is 15,883 bp in the SRMA BAC, and 14,967 bp in the PAXB BAC. These changes are distributed in many different exons, introns, and intergenic regions, and could affect either the regulation or protein functions of genes in the low morph region. No evidence was seen for accelerated protein sequence evolution ($K_a/K_s > 1$) for any of the genes in the interval (Table S3). Multiple noncoding changes were also observed in NAKA fish (0.32 % divergence with marine sequence found in SRMA, based on an initial survey of 1882 noncoding bp in the candidate interval).

| Gene Region (exons, introns, or intergenic) | Length | # SNPs | | | # Indels | | | Length changes in µsatellites or poly(N) | | |
|--|---------|---------|------|----------|-------------|---------------|----------------|---|---------------|----------------|
| | | Total # | Syn. | Non-Syn. | 1- 10 bp | 11- 100 bp | 101- 1000bp | 1- 10 bp | 11- 100 bp | 101- 1000bp |
| EDA (Exons 2-8) | 798 bp | 4 | 3 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| TNFSF13B (Exons 1-6) | 705 bp | 5 | 3 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| GARP (Exons 1-2) | 1953 bp | 24 | 9 | 15 | 0 | 0 | 0 | 0 | 0 | 0 |
| EDA (Introns 1-7) | 9311 bp | 185 | N/A | N/A | 20 | 8 | 2 | 17 | 5 | 1 |
| TNFSF13B- (Introns 1-5) | 1141 bp | 19 | N/A | N/A | 0 | 0 | 0 | 1 | 1 | 0 |
| GARP (Intron 1) | 109 bp | 0 | N/A | N/A | 0 | 1 | 0 | 0 | 0 | 0 |
| EDA to TNFSF13B intergenic | 1178 bp | 15 | N/A | N/A | 4 | 0 | 0 | 2 | 0 | 0 |
| TNFSF13B to GARP intergenic | 955 bp | 17 | N/A | N/A | 1 | 0 | 0 | 1 | 1 | 0 |

Table S3. Ka/Ks ratios for the coding regions of genes in the conserved low morph haplotype.

Nonsynonymous changes per nonsynonymous site (K_a), synonymous changes per synonymous site (K_s), and the ratio of K_a to K_s (35, 36) were calculated for the indicated genes in the marine (SRMA) and freshwater (PAXB) sequences. No evidence was seen for accelerated amino acid divergence in any of the genes ($K_a/K_s > 1$). This test cannot rule out important functional changes in particular amino acid residues, or possible changes in non-coding regulatory sequences surrounding the coding exons.

| Gene | K_a | K_s | K_a/K_s |
|-----------------|-------------------------|-------------------------|-----------------------------|
| <i>Eda</i> | 0.0055 | 0.016 | 0.35 |
| <i>Tnfsf13b</i> | 0.0052 | 0.0086 | 0.60 |
| <i>Garp</i> | 0.011 | 0.017 | 0.61 |
| <i>Gjb1</i> | 0.0054 | 0.015 | 0.35 |

Supplementary References and Notes

1. C. L. Peichel *et al.*, *Nature* **414**, 901 (2001).
2. P. F. Colosimo *et al.*, *PLOS Biology* **2**, 635 (2004).
3. D. G. Ransom, L. I. Zon, *Methods Cell Biol.* **60**, 195 (1999).
4. D. M. Kingsley *et al.*, *Behavior*, (in press).
5. C. L. Peichel *et al.*, *Curr. Biol.* **14**, 1416 (2004).
6. C. Burge, S. Karlin, *J. Mol. Biol.* **268**, 78 (1997).
7. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, *J. Mol. Biol.* **215**, 403 (1990).
8. H. M. Wain, M. J. Lush, F. Ducluzeau, V. K. Khodiyar, S. Povey, *Nucleic Acids Res.* **32 Database issue**, D255 (2004).
9. P. C. Sham, D. Curtis, *Ann. Hum. Genet.* **59**, 97 (1995).
10. J. D. Thompson, D. G. Higgins, T. J. Gibson, *Nucleic Acids Res.* **22**, 4673 (1994).
11. B. Ewing, P. Green, *Genome Res.* **8**, 186 (1998).
12. D. A. Nickerson, V. O. Tobe, S. L. Taylor, *Nucleic Acids Res.* **25**, 2745 (1997).
13. J. P. Huelsenbeck, F. Ronquist, *Bioinformatics* **17**, 754 (2001).
14. J. Felsenstein, *Cladistics* **5**, 164 (1989).
15. D. L. Snofford, *PAUP*. Phylogenetic analysis using parsimony (* and other methods)*. (Sinauer Associates, Sunderland, MA, 2002).
16. J. Felsenstein, *Inferring phylogenies* (Sinauer, Sunderland, MA, 2004).
17. J. M. Comeron, *J. Mol. Evol.* **41**, 1152 (1995).
18. D. W. Hagen, *J. Fish. Res. Board Can.* **24**, 1637 (1967).
19. J. V. Baumgartner, M. A. Bell, *Evolution* **38**, 665 (1984).
20. A. K. Srivastava *et al.*, *Hum. Mol. Genet.* **10**, 2973 (2001).
21. V. Thermes *et al.*, *Mech. Dev.* **118**, 91 (2002).
22. K. E. Hosemann, P. F. Colosimo, B. R. Summers, D. M. Kingsley, *Behavior*, (in press).
23. K. Igarashi, *Annot. Zool. Jpn.* **43**, 34 (1970).
24. J. S. McKinnon *et al.*, *Nature* **429**, 294 (2004).
25. D. Schluter, E. A. Clifford, M. Nemethy, J. S. McKinnon, *Am. Naturalist* **163**, 809 (2004).
26. M. A. Bell, G. Orti, *Copeia*, 314 (1994).
27. J. D. McPhail, *Can. J. Zool.* **70**, 361 (1992).
28. J. D. McPhail, *J. Fish. Res. Bd. Can.* **26**, 3183 (1969).
29. J. C. Avise, *Genet. Res.* **27**, 33 (1976).
30. R. R. Miller, C. L. Hubbs, *Copeia*, 52 (1969).
31. M. A. Bell, J. V. Baumgartner, *Copeia* **1984**, 258 (1984).
32. T. R. Haglund, D. G. Buth, D. M. Blouw, *Biochem. Syst. Ecol.* **18**, 559 (1990).
33. R. N. Campbell, R. B. Williamson, *Proc. Royal. Soc. Edinburgh* **77B**, 377 (1979).
34. N. Giles, *J. Zool.* **199**, 535 (1983).
35. W. H. Li, C. I. Wu, C. C. Luo, *Mol. Biol. Evol.* **2**, 150 (1985).
36. W. H. Li, *J. Mol. Evol.* **36**, 96 (1993).

Figure S1 (B) GARP

| | | |
|-------|-----|---|
| SRMA | 1 | -MVRHMFPEHLLLSWISINDLYITGLTYDPPKEHWSNQNHSYVPLDLDVRLRRLLDLSNNFIROLHPALPN--LQQLDSSNQDLISGCAFEDLALBEELNLS |
| PAXB | 1 | -MVRHMFPEHLLLSWISINDLYITGLTYDPPKEHWSNQNHSYVPLDLDVRLRRLLDLSNNFIROLHPALPN--LQQLDSSNQDLISGAFENLALBEELNLS |
| Fugu | 1 | -MVTHMFPEHLLLSFS--KGFCLAKEQSNHSHSLPDLDDVRLRRLLDLSNFIROLHMLAPY--LQQLDSSNQDLISGAFENLALBEELNLS |
| Mouse | 1 | -----MRPAEPPGAGAVDGDRCGNLHLSVPSLPHSRMLILDAMPKLDLWNHSCQAPRLENLHSHCHLDRISHYAFREQGHENLVLVA |
| Human | 1 | MRPQIILLLALLTGLAAHQDKVPCMKMVKKVSQCVLGLQVPSVPPDTEITLDSLQNLRSILASPEGCFYALRHLHDLSTNEISFLQPGAFQALTHEHLSLA |
| SRMA | 102 | RNLSKDIQNNKKALQSHIRLKSLDISKMDLGDDAVELYLRNKPSSLDQKMTGNALTLSHNLFFKESKGRRAITIDDNHSVIEEGTFEFLRQLEMLNARNNLA |
| PAXB | 102 | RNLSKDIQNNKKALQSHIRLKSLDISKMDLGDDAVELYLRNKPSSLDQKMTGNALTLSHNLFFKESKGRRAITIDDNHSVIEEGTFEFLRQLEMLNARNNLA |
| Fugu | 96 | RNELGNLCSNKTALREGLKTLDSMGLDGDAAEALYONKSEFLHAKMTGNVLRYSKLFKQSKMKSISIDDDLISVIOQGTFFPLRHLLESANLAQNNLV |
| Mouse | 90 | DNRLSENKESAALHLLGLRRRLDLSNLSLTDMAALMQLNSLSLEVSLARNTLRRDDDDIFEGLEHVELDLQRNYPFEIEGGAFDLTLRRLRNLAYNNLP |
| Human | 106 | HNRLAMATALSAGGLGPRVTSLDLSGNSLYSGLLERLIGEAPSLHTELSAENSLTRTRHTRTFRDMFALEQLDLHNSVIMDIEDGAFEGPLRLTHLNSRNSLT |
| SRMA | 207 | HICDFKLEQVKYLNLSRNSVEFFVTR-DDOLYRLEIIDLSHNKLLYFPKMNRLKYHLQNNLVGAIN-----SEAMVSEANALYSEIMSEKIVRKNLH |
| PAXB | 207 | HICDFKLEQVKYLNLSRNSVEFFVTR-DDOLYRLEIIDLSHNKLLYFPKMNRLKYHLQNNLVGAIN-----SEAMVSEANALYSEIMSEKIVRKNLH |
| Fugu | 201 | YICDFKLEQVKYLNLSRNSVEFFVTHE-DEOLYMLEIIDLSHNKLLYFPKMNRLKYHLQNNLVGAIN-----SEAMVSEANALYSEIMSEKIVRKNLH |
| Mouse | 195 | CIVDFSLQLRFLNVSXNLIEMFLAAR-EVAFLEIIDLSHNQLLFFPLPQCCKLFTLLOQNNMGFYRELYNITSQEMVAQFLLDVGNVTNITTWNLWEEF |
| Human | 211 | CISDFSLQLRVLDLSCNSIEAFOTASQPAEQFQITLWIDLRENKLLHFFEDLAALPRLLYINLNSNLIIRLP-----TGPPQDSKGIHAPSEG-WSALPLGAPSGNA |
| SRMA | 305 | SNWRMLPILYIDL SYNHFSPPLETSLTSLTLNLFSPNCLQNIAMNIRHDESGISEQLFFPPLSRKYENLQOS-----NGLAFISPLFANALTO |
| PAXB | 305 | SNWRMLPILYIDL SYNHFSPPLETSLTSLTLNLFSPNCLQNIAMNIRHDESGISEQLFFPPLSRKYENLQOS-----NGLAFISPLFANALTO |
| Fugu | 299 | SSWRQMPVHFDLSENHEPSPVETLSLSLTLNLFSENCLRKLMMVVK-DNSGREQLYFHALKNDMOS-----NGLMGLSPLFANALIR |
| Mouse | 299 | SSDLSARFLDMSONEFHLPDGFKKTPSLSHLNQONCLKMLHIREHEPPGALTEDDLSHNQDAEHDAPGLTGSURNLRFVNLSSNQLLGVPTGLPFDNASS |
| Human | 310 | SGRPLSQNLNLDLSYNETIELIPDSFLEHTLSLCFNLNERNCLRTFEARRIGSLPCLMLLDLSHNALETHEL GARALGSLR--TLTLLQGNALRDLPEYTFANLAS |
| SRMA | 394 | IETLNLQNSVOPCASIDRS---TQQIHLNNTSCVVIHQHRTLKHNLKKNENIKMLQPNFPLKLTALVSNLQAGNY-HMVVGRDLEGVQETLQSLIIS--EMNMT |
| PAXB | 394 | IETLNLQNSVOPCASIDRS---TQQIHLNNTSCVVIHQHRTLKHNLKKNENIKMLQPNFPLKLTALVSNLQAGNY-HMVVGRDLEGVQETLQSLIIS--EMNMT |
| Fugu | 387 | LESNLQNAVQPCDPACLONPQSRWQSLNASCVAVSKLTLKRLNKENHIQIIEHNTFKETSLLENLAGNS-HLTQEGALEGQNTLQSLVIS-EVYLS |
| Mouse | 404 | ITTIIDMSHNTISLCPQMVVDWE-----GPPSCVDFRNGSRLSLSDDCGKALQDCFPQGTSTLTHELSSNWGVNGSISPLWAVAPLQVLSLSDVYGLGSG |
| Human | 412 | LQRLNLQGNRVSPCG---GPDEP-----GPPSCVAVFSGITSLRSHLSVAVNEIEELLRAGAFLLHTPLTEITLSSNP-GEFVATGALGGLFALEVLALQ--GNGLM |
| SRMA | 492 | SDDLSPCMPALTQLNISGNRLDLPSSISCSPLTEVDIRNNHFESLNSLILDLSVR-HNIMYISGNHFNCCDSMWLIHIDGLKMKLEPDANDAVCFHSDRSIAM |
| PAXB | 492 | SDDLSPCMPALTQLNISGNRLDLPSSISCSPLTEVDIRNNHFESLNSLILDLSVR-HNIMYISGNHFNCCDSMWLIHIDGLKMKLEPDANDAVCFHSDRSIAM |
| Fugu | 489 | S-NVSLPEMSALTYLNIANNVLDLSPGAFSCSPLKEIDLRNRFVSLNREMTLALSAE-LELMYISGNVENCCDSMWLITVHERKIKVDDINSTLCFTENNVVM |
| Mouse | 503 | AAEMDFSAFGNLRALDLSGNSLTSFPKPKGSLALRTLDRRNSLTALPQKVVSEQLRGTITTYLSQNPYDCCGVEGWALQQHFKTWADLSMVTICNLSKIVRV |
| Human | 505 | VLQVDLPCFICLKRLLNLAENRLSHLPAWTQAVLEVLDRNNSFSLPQSAMGGLETS-HRRLYLQGNPLSCCG-NGWFAAQLHQGRVDVYDATQDLICRFSQEE |
| SRMA | 596 | REYLRNPSGYCLFHFKAQEINFGTMIVLVFVTVVTLVLIIFTRKVFV-----TQISFIV----- |
| PAXB | 596 | REYLRNPSGYCLFHFKAQEINFGTMIVLVFVTVVTLVLIIFTRKVFV-----TQISFIV----- |
| Fugu | 592 | ADLLRPTLDDCTFHPEAEVHFGQLVAVLFGTVLTVLVLVGRKVVWFYAVTSPSLTQTFVVFQHKCSAMTK |
| Mouse | 608 | VELPEGLPQCKWEQVDTGLFYLILPSCLLLVACTVVFLLFKKPL-----ICVIKSRCHWSSIY----- |
| Human | 608 | VLSLHVREDCEKGGGLKNILNIIILITFLVSAILLTFLAACCCVRRQK-----FNQOYKA----- |

Figure S1 (C) GJB1

| | | | | | | | | | | | |
|-------|-----|------------|-----------|----------|----------|---------|----------|-------|-------|----------|-----------|
| SRMA | 1 | MNWGSFYAV | ISGVNRHST | GIGRVWLS | VIFIFRIL | VLVVA | AESVWGDE | KSGFV | CNTQQ | PGC | |
| PAXB | 1 | MNWGSFYAV | ISGVNRHST | GIGRVWLS | VIFIFRIL | VLVVA | AESVWGDE | KSGFV | CNTQQ | PGC | |
| Fugu | 1 | MNWGTFYAL | ISGVNRHST | GIGRVWLS | VIFIFRIL | VLVVA | AESVWGDE | KSGFT | CNTQQ | PGC | |
| Mouse | 1 | MNWTGLYTL | ISGVNRHST | AIGRVWLS | VIFIFRIM | VLVVA | AESVWGDE | KSGFI | CNTLQ | PGC | |
| Human | 1 | MNWTGLYTL | ISGVNRHST | AIGRVWLS | VIFIFRIM | VLVVA | AESVWGDE | KSGFI | CNTLQ | PGC | |
| SRMA | 61 | NSVCYDQFFP | ISHIRLWAL | QLILVST | PALLVAM | HVAHRRH | VDKKV | LKKT | GRG | GPKLELI | |
| PAXB | 61 | NSVCYDQFFP | ISHIRLWAL | QLILVST | PALLVAM | HVAHRRH | VDKKI | LKKT | GRG | GPKLELI | |
| Fugu | 61 | NSVCYDQFFP | ISHIRLWAL | QLILVST | PALLVAM | HVAHRRH | VDKKI | LKRA | GRCT | PKDLEQL | |
| Mouse | 61 | NSVCYDHFPP | ISHVRLWS | LQLILVST | PALLVAM | HVAHQOH | IEKKM | LRLE | GHG | DPLHLEEV | |
| Human | 61 | NSVCYDQFFP | ISHVRLWS | LQLILVST | PALLVAM | HVAHQOH | IEKKM | LRLE | GHG | DPLHLEEV | |
| SRMA | 121 | KNQKFOITG | ALWWTYMI | SIVFR | IVLEVA | FLYIF | YLYPG | FKM | VRLV | KCASYP | CPNTVD |
| PAXB | 121 | KNQKFOITG | ALWWTYMI | SIVFR | IVLEVA | FLYIF | YLYPG | FKM | VRLV | KCASYP | CPNTVD |
| Fugu | 121 | KNQRFQITG | ALWWTYMI | SIFR | IVFEVA | FLYIF | YLYPG | FKM | VRLV | KCD | SYP |
| Mouse | 121 | KRHKVHISG | LWWTYV | ISVVFR | LLFEAV | FMYV | FYLYP | GYA | MVRLV | KCEAF | PCPNTVD |
| Human | 121 | KRHKVHISG | LWWTYV | ISVVFR | LLFEAV | FMYV | FYLYP | GYA | MVRLV | KCD | VYPCPNTVD |
| SRMA | 181 | VSRPTEKTI | FTVFMLA | VSGLC | VLLNLA | EVA | YLI | FRAC | KRCL | RGT | EEES |
| PAXB | 181 | VSRPTEKTI | FTVFMLA | VSGLC | VLLNLA | EVA | YLI | FRAC | KRCL | QGT | EEES |
| Fugu | 181 | VSRPTEKTI | FTVFMLG | VSGCV | VLLNLA | EVA | YLI | GRAC | RQ | CIR | GEETS |
| Mouse | 181 | VSRPTEKTV | FTVFMLA | ASGIC | ILNVA | EVVYLI | IRAC | XRR | AQR | RSN | PPSR |
| Human | 181 | VSRPTEKTV | FTVFMLA | ASGIC | ILNVA | EVVYLI | IRAC | ARR | AQR | RSN | PPSR |
| SRMA | 239 | -TYKONEIN | QLIAEQ | --AL | KSKFA | VSKK | SPT | --EK | GER | C | S |
| PAXB | 239 | -TYKONEIN | QLIAEQ | --AL | KSKFA | VSKK | SPA | --EK | GER | C | S |
| Fugu | 239 | -SYQONEIN | ELLDH | --P | RSK | F | G | V | T | K | K |
| Mouse | 241 | PEYKONEIN | KLSEQ | DGSL | KDIL | RRSP | GT | G | A | E | K |
| Human | 241 | PEYKONEIN | KLSEQ | DGSL | KDIL | RRSP | GT | G | A | E | K |