Cell, Volume *131*

Supplemental Data

Cis-Regulatory Changes in *Kit Ligand* Expression and Parallel Evolution of Pigmentation in Sticklebacks and Humans

Craig T. Miller, Sandra Beleza, Alex A. Pollen, Dolph Schluter, Rick A. Kittles, Mark D. Shriver, and David M. Kingsley

Supplemental Experimental Procedures

Genotyping

Microsatellite genotyping was performed essentially as described (Peichel et al., 2001; Colosimo et al., 2004), except with the following PCR profile: 3 min of 94ºC, then 35 cycles of 94ºC for 10 sec, 58 ºC for 10 sec, 72 ºC for 20 sec. From five additional F2 families, 1182 fish were genotyped for *Stn194* and *Stn398*. Recombinant animals were typed with additional intervening markers, with some genotypes inferred by tightly linked microsatellites.

Marine panel genotyping for presence of PAXB-like *Kit ligand* variant was done by PCR using a 5' 6-FAM fluorescently labeled primer 5'-TGCACCTCCACTAGACACTGA-3' and primer 5'-AACGGAATATGATTTTTATTTTATG-3', which amplify a 168 bp product from the freshwater variant allele and a 174 bp product from marine alleles. Monomorphic fish were additionally typed with *Idh* primers (Peichel et al., 2004) to distinguish males from homozygous females. Allele frequencies are estimated assuming an equal ratio of male and female fish in non-carrier fish.

Kit ligand sequencing

Kitlg coding sequence was determined using 5' RACE with a SMART RACE cDNA Amplification Kit (Clontech) with reverse primer 5'-CAGGGAGGAGGGTCGCAGTAATCTG-3' and RT-PCR. The entire *Kitlg* coding region was amplified by RT-PCR from LITC and FTC gill tissue using primers 5'-TTGATTTCACGTTATTTGCAG-3' and 5'-GAGGAGGGTCGCAGTAATCT-3'; and 5'-CTTGGGGTACATTCGAGCAA-3' and 5'-TGCGATAGCGCGATGGAACT-3'. RT-PCR products were sequenced using ABI BigDye 3.1 on an ABI 3730xl capillary sequencer. *Kitlg* exons were PCR amplified from JAMA and PAXB genomic DNA using the following primers: exon 1: 5'-TCTCGGGAGCTTTGTTTCTC'-3' and 5'-CCACTTAAATGTCCGGTTCG-3'; exons 2 and 3: 5'-GCATTGAGTTGTGCGGAGTA-3' and 5'-TCCCTGAAAGAGGAAGAGCA-3'; exon 4: 5'-GGTAAGACATGGTGGCATCC-3' and 5'-GCCTCCTGCAAAGACACATT-3'; exon 5: 5'-GATGGACCCTTTGCAGTGAG-3' and 5'-

ACCGAAGGGCTGCAAGTAAT-3'; exon 6: 5'-GGCAGCAGCATGAAAAAGGCA-3' and 5'-CCTCTAACCATCTCAAAGGAT-3'; exons 7 and 8: 5'-CCTCATGATCTCCAACCTCCT-3' and 5'-GGTATGAAGCGTTGTGTTCG-3'; exon 9: 5'-TCCACGGCCAGTTAAACCTT-3' and 5'-GAGGGTTTACGCTCCGCTTAA-3'.

The tree in Figure 4B was made from 561 bp including exons 7 and 8 using primers listed above. Alignment and neighbor joining tree were done with ClustalX, and alignment displayed using the output from the BOXSHADE server (http://www.ch.embnet.org/software/BOX_form.html).

Genbank accession numbers for *Kitlg* cDNA and genomic sequences used for Figures 4 and S3 are EU218896-EU218903.

Skin pigmentation

Skin pigmentation was initially scored in eight different body regions of F2 progeny from the Japanese marine x Paxton benthic cross. Pigmentation in the posterior ventrum region (posterior to the cloaca, and ventral to the posterior lateral line) was estimated by scoring the overall percent of the skin field covered in melanin, as well as by estimating the percentage of a line drawn from the cloaca to the posterior lateral line (Figure 3A). Both quantification methods gave similar results. To quantify melanocyte number, we counted melanocytes present within a 1 mm x 1 mm square centered in this posterior ventrum region in F2 fish over 35 mm standard length. Two-tailed t-tests were done in Microsoft Excel.

Stickleback populations

Sticklebacks were collected from Fishtrap Creek in Washington using minnow traps. Collection from other populations has been previously described (Colosimo et al., 2005).

In situ hybridization

A *Kitlg* riboprobe was generated by RT-PCR amplification of a 730bp fragment of *Kitlg* with primers 5'- CTTGGGGTACATTCGAGCAA-3' and 5'-GATTTCAATGACGTTCAACATGTT-3', cloning this into pCR4-TOPO (Invitrogen), and linearizing and transcribing with SpeI/T7 for antisense probe and NotI/T3 for sense probe. Gill arches were dissected from adult LITC and FTC fish, fixed overnight in 4% PFA, and stored at -20°C in methanol. Signal was detected using BM purple (Roche).

Allele-specific expression

For allele-specific expression quantification using pyrosequencing (Figure S4), RT-PCR with forward primer 5'-GATGACATCTCTAGGCTCTCCATT-3' and reverse primer (5'-biotinylated, HPLC-purified) 5'-TATCTTTTCGGTTGGAGGAAATGTT-3' were used to amplify *Kitlg* using a PCR program of 94ºC for 3 min, then 36 cycles of 94ºC for 15 sec, 58ºC for 15 sec, 72ºC for 20 sec followed by a 3 minute extension at 72ºC. PCR reactions contained 0.05 units/uL of AmpliTaq polymerase (ABI). These RT-PCR products were sent to EpigenDx (Worcester, MA), where pyrosequencing was done using primer 5'-CTCTCCATTCTCAGACAA-3' (designed to minimize mispriming sites) and analyzing a SNP at nucleotide position 156 of the *Kitlg* coding sequence.

One-way ANOVA and post hoc tests were done in SPSS 11.

Admixture mapping

A 231bp fragment containing the rs642742 SNP was amplified with primers 5'-CTGGTTCTTATCTGAAGGACT-3' (forward) and 5'-AAATCCATTCTGTGATAATAAC<u>C</u>ATG-3' (reverse). The underlined base of the reverse primer is a mismatched base, intentionally inserted to create a *BseG*I restriction site. Digestion with *BseG*I generates 205- and 26bp fragments in the presence of the rs642742*G allele. PCR reactions contained 0.3 µM of each primer, 0.2 mM of each deoxynucleotide triphosphate (dNTP), 10 mM Tris–HCl (pH 8.8), 50 mM KCl, 0.08% Nonidet, 1.5 mM MgCl2 and 1 U Taq polymerase. Samples were denatured for 3 min at 95ºC, followed by 20 cycles of 94ºC for 30s, 56ºC for 30s, and 72ºC for 30s, by 15 cycles of 94ºC for 30s, 54ºC for 30s, and 72ºC for 30s, followed by a 10min-extension at 72ºC. Digestions (1 U/µl) were performed at 55ºC for 1 h and DNA fragments were visualized by silver staining after non-denaturing electrophoresis separation in 9% polyacrylamide gels.

For ANCOVA, we used a general linear model where M index was the dependent variable and the rs642742 genotype was the independent fixed factor. To control for spurious associations due to differences in the ancestral proportions of the individuals (admixture structure) or due to differences in skin pigmentation between sexes, individual ancestry and sex were included as covariates. To determine the amount of M-index variation explained by rs642742 genotype, a multiple regression analysis was also performed, using rs642742 genotypes, ancestry and sex as independent variables. All these analyses were computed in SPSS 14.0.

For Bayesian modeling, the ADMIXMAP program was used to model the distribution of admixture and to test for association between rs642742A>G and the melanin index (McKeigue et al., 2000; Hoggart et al., 2003). This program is based on a Bayesian full probability model for population admixture, in which the parental admixture, locus ancestry and any missing marker genotypes can be specified as "missing data". With this assumption, all observed and missing data can be treated as random variables following non-informative prior distributions, and the posterior distributions of the missing data can be generated by Markov chain simulation (McKeigue et al., 2000). To test for association between rs642742 genotypes with skin pigmentation, ADMIXMAP fits a linear regression of the M index on sex and individual admixture, and estimates a score test by averaging over the posterior distribution of missing data (Hoggart et al., 2004). This test allows for uncertainty in estimation of individual admixture proportions from marker data. The program also gives the maximum likelihood ratio of the regression coefficient, estimated as the score/√observed information.

Table S1. Microsatellite Primer Sequences

primer name	position, scaffold 3 (Mb)	sequence, 5' to 3'
stn398-f	4.0	CACTCTCACACTCACACAACC
stn398-r		ATGGAGGATTTGCTGAACG
stn399-f	3.1	CTACTCGGGCTGGAGGACTT
stn399-r		GACAAGGGCTGTGTGTGGTA
stn400-f	0.6	AGACGATGGAGGGATTCACC
stn400-r		CTTTAGTACGAGCAGTTCTTCC
stn401-f	8.1	ACACTTCCTGGACACAGAGG
stn401-r		GAGCCACATGAACTGTGTCG
stn402-f	7.5	TTCTTCAGGGTGATTGTGC
stn402-r		TGTACTGCTTGTGGTCTCAGG
stn403-f	6.88	AGTTGTGTGACGGTCTCTCG
stn403-r		GCTTAACTGGACCCTTGACC
stn404-f	6.87	AAACACTCCAGGGTGCAAAC
stn404-r		CTCGCCTATCAGCTGTTCATT
stn405-f	6.845	CCTGTGTGCAGGATAAGTGG
stn405-r		CCACGAGAGGTCAGAAAGGA
stn406-f	6.84	AACATGCGCAGGGACTTT
stn406-r		CGTGCTCACCTCAACCAAT
stn407-f	6.7	CTGTCGGCCTGATCAGTTTC
stn407-r		TGCAGTGTATCCTCCATGAATC
stn408-f	6.54	AGCGTCCTGCCTATTGAGG
stn408-r		GAATGCTCCTGTGTGCGTTA
stn409-f	6.53	TTTGAGATGGAGGGTTAGGG
stn409-r		AACACAAGGACGGTTTGACC
stn410-f	6.5	TCTCTGCCTCCAGACTTTCC
stn410-r		AGCTTCAGGCCCTACATGG
stn411-f	6.4	CGATGTCCTGTTCACTGTGG
stn411-r		GTAACCAGGTCAGCGTTTCC/
stn412-f	6.3	CTCTTGACCTCCGATTCGTC
stn412-r		TGCCTCCTCGTCTCTTCAAT
stn413-f	5.8	CACCACACACGTCTCACAGC
stn413-r		CAAGTACGCCGAGTCCTACG
stn414-f	4.9	CAGCATGGTTCAAGGTTGC
stn414-r		CATCACAGCTCCCAGTCTGC
stn415-f	4.6	CCCAGACCTGCAGCACTCTA
stn415-r		AAACACAACGGCATCACAGA
stn416-f	4.3	ATCCTCGGTCCCTAACGAGT
stn416-r		GGGTCCAATTACTCCCAGAA
stn417-f	4.1	ATCAATCGGTCGACCTCTCC
stn417-r		AGCTCCATGAGCCACTCG

All forward primers were fluorescently labeled with 6-FAM at their 5' end. In Figure 2A, the markers at positions 6.6 and 5.6 are *Stn256* and *Stn192*, respectively.

Table S2. Genes in Pigmentation QTL Interval

locus abbreviation	**Gene name**
EIF4G2	eukaryotic translation initiation factor 4 gamma, 2
XM_001341198	hypothetical locus 100001178
LYVE1	lymphatic vessel endothelial hyaluronan receptor 1
C12orf29	chromosome 12 open reading frame 29
CEP290	centrosomal protein 290 kDa
TMTC3	transmembrane and tetratricopeptide repeat containing 3
Kitlg	kit ligand, stem cell factor, steel factor
SLC23A1	solute carrier family 23, member 1 (sodium-dependent vitamin C transporter)
DUSP6	dual specific phosphatase 6
WDR51B	WD repeat domain 51B
Tsga14	testis specific, 14
Cpa1	carboxypeptidase A1, pancreatic
Tes	testis derived transcript (3 LIM domains)
Cav2	caveolin 2
Cav1	caveolin 1

The fifteen genes in the 315 kb pigmentation QTL interval defined by high resolution mapping (*Stn405* to *Stn409*, see Figure 2). Genic content was determined using a combination of GENSCAN, nucleotide BLAST (blastn) and translated BLAST (blastx) searches to the NCBI non-redundant database. All non-repetitive GENSCAN predictions with support from homology to other organisms are listed. Gene abbreviation names listed are the HUGO genome nomenclature committee designations for the fourteen genes with human homologs, and the NCBI accession number for XM_001341198. Exon 2 of Cav1 is meiotically excluded, as *Stn409* is located in intron 1 of Cav1. SLC23A1 has no homology to SLC24A5, as the SLC designation denotes function, not homology.

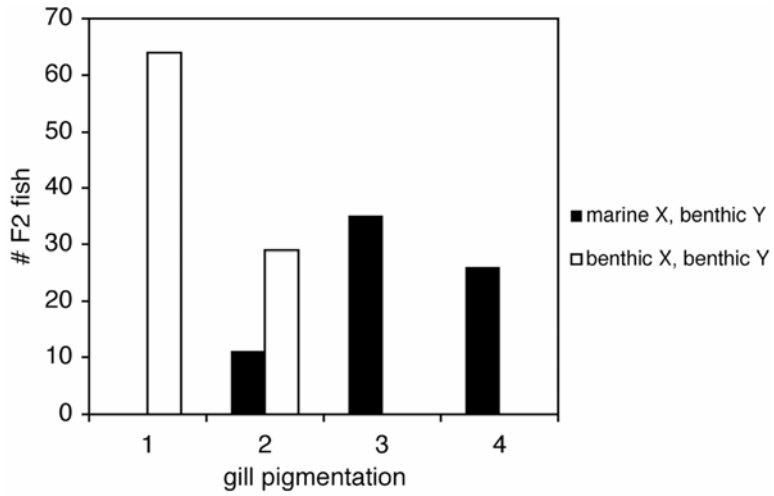

Figure S1. In males, extreme gill pigmentation phenotypes are concordant with LG19 pigmentation QTL genotype. Histograms of gill pigmentation scores for males with a marine X chromosome (black bars) or a Paxton benthic X chromosome (white bars). Classes 1, 3, and 4 are fully predictive of genotype at *Stn191*, the peak marker for the LG19 pigment QTL.

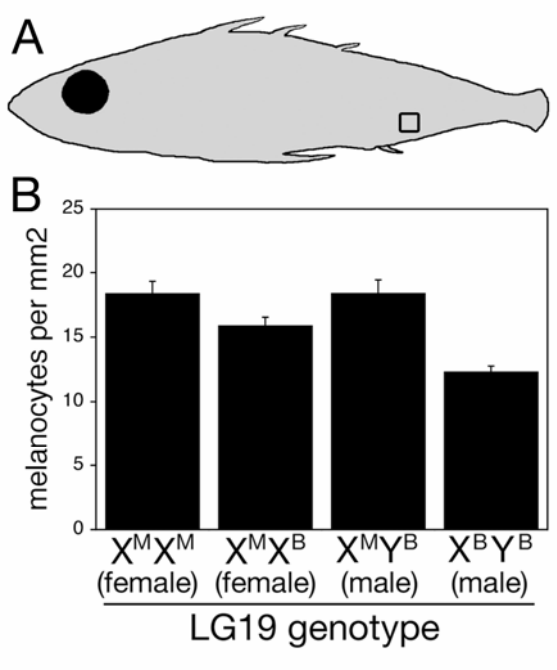

Figure S2. The LG19 pigmentation QTL controls melanocyte number in ventral skin.
(A) Schematic with black box denoting the position of the 1 mm x 1 mm square in the ventral flank in which melanocytes were counted.
(B) Melanocytes per millimeter^2 of ventral skin in F2 fish of different LG19 genotypes. Shown are means +/- SEM, with 22 to 27 F2 animals per class.

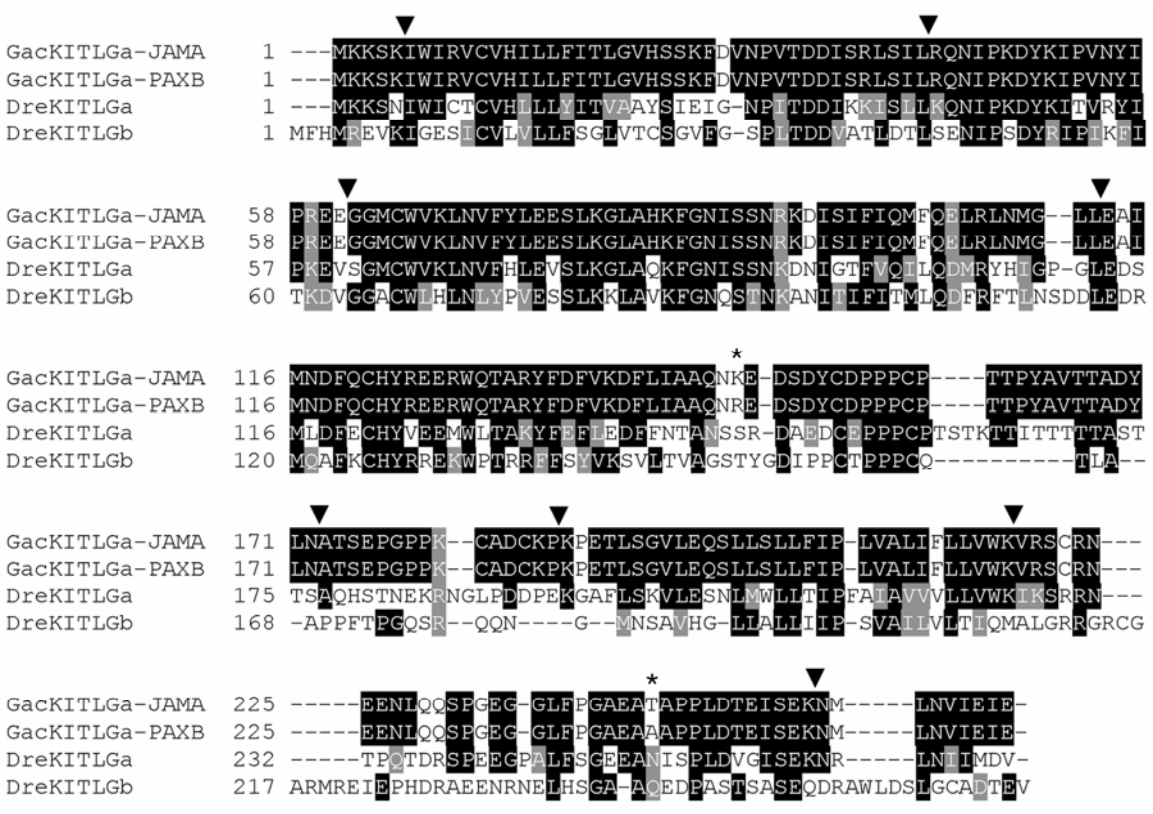

Figure S3. Sequence alignment of predicted KITLG amino acid sequences in fish. Sequences of stickleback (*Gasterosteus aculeatus)* Kitlg on LG19 (GacKITLGa) in Japanese marine (JAMA) and Paxton benthic (PAXB), and zebrafish (*Danio rerio*) KITLGa and KITLGb (Hultman et al., 2007). The stickleback *Kit ligand* we describe in this paper is orthologous to zebrafish *Kitlga*. Intron positions are marked with arrowheads. The two amino positions which vary between JAMA and PAXB, marked in asterisks, are at non-conserved residues. The predicted amino acid sequences of LITC and FTC are identical to JAMA and PAXB, respectively.

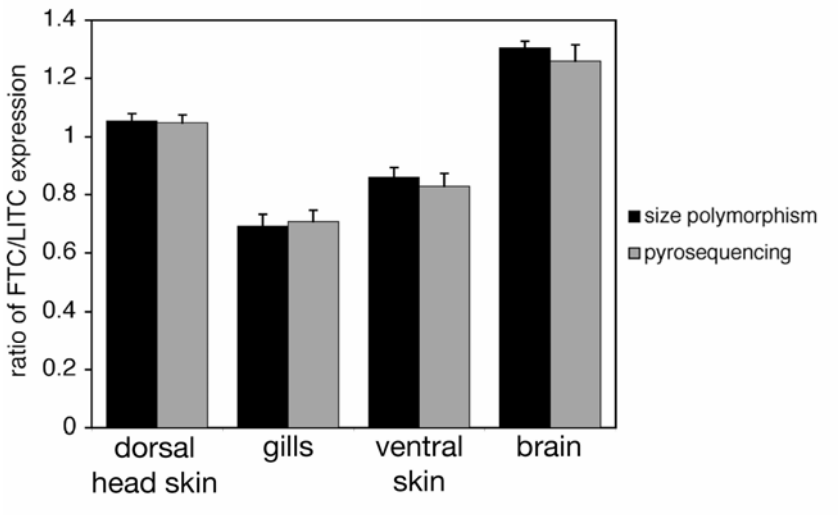

Figure S4. Two methods of quantifying allele-specific gene expression give similar results. Ratios of FTC/LITC *Kitlg* expression levels in dorsal head skin, gills, ventral skin, and brain for size polymorphism (black bars, see Figure 6) and pyrosequencing (gray bars). Shown are means +/- SEM from eight to fifteen F1 hybrids.

Supplemental References

Colosimo, P.F., Peichel, C.L., Nereng, K., Blackman, B.K., Shapiro, M.D., Schluter, D., and Kingsley, D.M. (2004). The genetic architecture of parallel armor plate reduction in threespine sticklebacks. PLoS Biol. *2*, E109.

Colosimo, P.F., Hosemann, K.E., Balabhadra, S., Villarreal, G., Jr., Dickson, M., Grimwood, J., Schmutz, J., Myers, R.M., Schluter, D., and Kingsley, D.M. (2005). Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. Science *307*, 1928-1933.

Hoggart, C.J., Parra, E.J., Shriver, M.D., Bonilla, C., Kittles, R.A., Clayton, D.G., and McKeigue, P.M. (2003). Control of confounding of genetic associations in stratified populations. Am. J. Hum. Genet. *72*, 1492-1504.

Hultman, K.A., Bahary, N., Zon, L.I., and Johnson, S.L. (2007). Gene Duplication of the zebrafish kit ligand and partitioning of melanocyte development functions to kit ligand a. PLoS Genet. *3*, e17.

McKeigue, P.M., Carpenter, J.R., Parra, E.J., and Shriver, M.D. (2000). Estimation of admixture and detection of linkage in admixed populations by a Bayesian approach: application to African-American populations. Ann. Hum. Genet. *64*, 171-186.

Peichel, C.L., Nereng, K.S., Ohgi, K.A., Cole, B.L., Colosimo, P.F., Buerkle, C.A., Schluter, D., and Kingsley, D.M. (2001). The genetic architecture of divergence between threespine stickleback species. Nature *414*, 901-905.

Peichel, C.L., Ross, J.A., Matson, C.K., Dickson, M., Grimwood, J., Schmutz, J., Myers, R.M., Mori, S., Schluter, D., and Kingsley, D.M. (2004). The master sex-determination locus in threespine sticklebacks is on a nascent Y chromosome. Curr. Biol. *14*, 1416-1424.