Supporting Information

Protas et al. 10.1073/pnas.1013850108

SI Materials and Methods

Degenerate PCR. Degenerate primers were designed by using alignments of *Drosophila melanogaster, Tribolium castaneum, Apis mellifera, and Daphnia pulex* sequences. Degenerate PCR was performed by using the following protocol: 94 °C for 3 min, 35 cycles of 94 °C for 30 s, 50 °C for 30 s, 72 °C for 1 min, and finally 72 °C for 10 min, and then chilled to 4 °C (the red primers in Dataset S1). Some degenerate primers used were designed by using the primer design program CODEHOP (http://blocks.fhcrc.org/codehop.html) and an annealing temperature of 60 °C (the blue primers in Dataset S1). The first PCR was performed with the forward and reverse primer. A second PCR was performed by adding 1 μL of the first PCR as template and either the same two primers or nested primers (if listed in Dataset S1). Previously published primers were used to amplify the *Asellus aquaticus* ortholog of *hedgehog* (1). The gene-associated markers not listed in Dataset S1 were isolated from gene-specific sequences generated by Floragenex (see below) or were amplified by incorrect priming of degenerate primers.

Floragenex Generated SNPs by Using RAD LongRead Technology. Genomic DNA was purified from thorax tissue by using the QIAamp DNA micro kit (Qiagen). Ten micrograms of DNA from a cave individual, a surface individual, and a hybrid individual were digested with the restriction endonuclease SbfI, and RAD tag libraries were produced (2). RAD tag libraries were sequenced by using paired end (2 × 60 bp) sequencing chemistry and an Illumina Genome Analyzer IIx at the University of Oregon High Throughput Sequencing Facility in Eugene, OR. Similar to published methods of genome assembly using short-read Illumina/Solexa data (3), the RAD LongRead method allows for the assembly of DNA sequences flanking restriction enzyme sites. Sequence data were categorized by "barcode" and was assembled into a reference for SNP detection. Poor sequences, with a converted phred score of Q10 or lower, were eliminated and remaining sequences were compiled into RAD sequence clusters, identical at the single end Illumina read. Sequences with a minimum of 20× and a maximum of 400× sequence coverage at the RAD single end reads were used. The variable paired end sequences were assembled by the Velvet sequence assembler into contigs (4). Floragenex software using the Needleman–Wunsch algorithm aligned raw paired-end sequence reads to the "reference" contig sequences from the cave thorax sample. Raw paired end reads from the cave, surface, and hybrid samples were then aligned with the reference cave thorax contig set; SNPs were called where the cave and surface sequences differed and the hybrid sequence contained both alleles.

Bias. We calculated the odds ratio for each binary trait (Tables S3 and S4). We added 0.5 to each value (Table S3) to address the issue of having zero values in the denominator. For the one quantitative trait examined, eye size, we used multiple imputation in R/qtl, which calculated the phenotypic variance responsible. Regarding the Noor effect, we used a Z-test to investigate whether the marker density flanking the areas of the genome containing the large effect loci was different from the average marker density across the genome (Table S4).

1. Vargas-Vila MA, Hannibal RL, Parchem RJ, Liu PZ, Patel NH (2010) A prominent requirement for single-minded and the ventral midline in patterning the dorsoventral axis of the crustacean *Parhyale hawaiensis*. *Development* 137:3469–3476.
2. Baird NA, et al. (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* 3:e3376.
3. Hiatt JB, Patwardhan RP, Turner EH, Lee C, Shendure J (2010) Parallel, tag-directed assembly of locally derived short sequence reads. *Nat Methods* 7:119–122.
4. Zerbino DR, Birney E (2008) Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18:821–829.

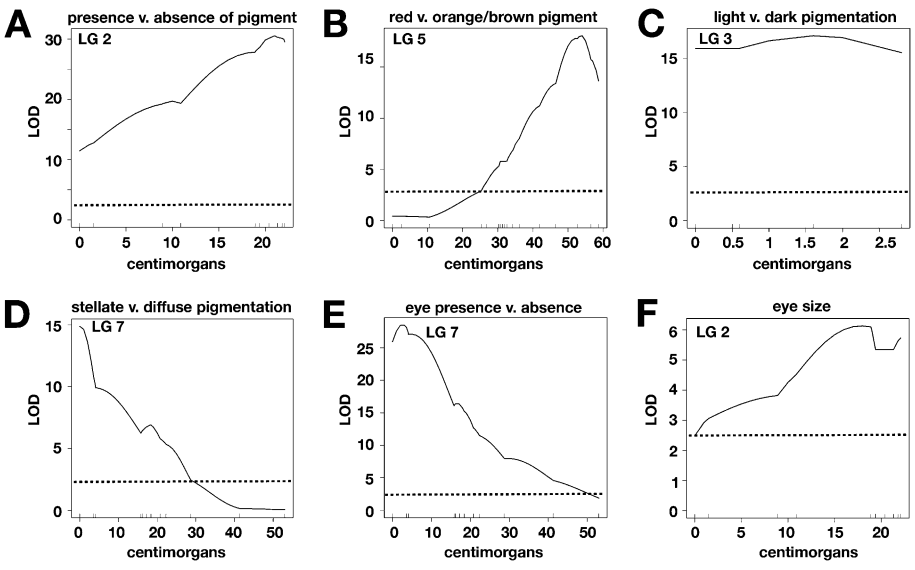

Fig. S1. Loci/QTL responsible for eye and pigmentation traits. Using the nonparametric interval mapping method in R/qtl, the following significant loci and QTL were found. (*A*) Presence vs. absence of pigment. (*B*) Red vs. orange/brown pigment. (*C*) Light vs. dark pigment. (*D*) Stellate vs. diffuse pigment. (*E*) Eye presence vs. absence. (*F*) Eye size. Dotted lines show the genome-wide significance levels ($\alpha < 0.05$) by permutation test, LOD = 2.32.

Table S1. Backcross populations

Backcross no.	Female	Male	No. of progeny	No. of broods
1	V	W′	5	1
2	W	W′	20	1
3	V	X′	19	1
4	X	X′	7	1
5	Y	Y′	92	4
6	Z	Y′	32	3
7	W	Z′	19	1

We used seven backcross families generated by five different females (named V, W, X, Y, and Z) and four different males (named W′, X′, Y′, and Z′). The number of progeny per backcross family ranged from 5 to 92 individuals. Two of the families were composed of multiple broods where the same cave male and female mated and produced offspring multiple times.

Table S2. Gene-specific markers in *Asellus aquaticus*

Marker name	Likely gene	BLAST ID	Species	Gene name	Coverage, %	*e* value
aa1	bmp10	XM_002196574.1	*Taeniopygia guttata*	bone morphogenetic protein 10	96	8.00e-11
aa2	egfr	XM_002411848.1	*Ixodes scapularis*	epidermal growth factor receptor	54	4.00e-34
aa3	disconnected	NM_078638.2	*Drosophila melanogaster*	disconnected	45	1.00e-12
aa4	dachsund	XM_002427823.1	*Pediculus humanus corporis*	dachshund	25	2.00e-09
aa5	pointed	XM_002429051.1	*Pediculus humanus corporis*	c-ets-1-B	39	3.00e-13
aa6	hedgehog	AB125742.1	*Achaearanea tepidariorum*	hedgehog	73	1.00e-19
aa7	sineoculis	XM_001096585.2	*Macaca mulatta*	six homeobox 1	80	6.00e-69
aa8	singed	XM_002423695.1	*Pediculus humanus corporis*	singed	51	3.00e-12
aa9	barh1	XM_001861865.1	*Culex quinquefasciatus*	homeobox protein b	22	1.00e-05
aa10	rap	XM_623288.2	*Apis mellifera*	retina aberrant in pattern	34	3.00e-23
aa11	sox14	FJ613627.1	*Scylla paramamosain*	sox14	45	5.00e-36
aa12	h15	AM404112.1	*Tegenaria atrica*	h15-3	93	4.00e-24
aa13	cinnabar	DQ106550.1	*Limbodessus pinnaclesensis*	cinnabar	13	3.90e-02
aa14	vsx	U19995.1	*Caenorhabditis elegans*	ceh-10	57	2.00e-27
aa15	punch	GU953222.1	*Biston betularia*	GTP cyclohydrolase I	29	1.00e-04
aa16	schizo	XM_002422978.1	*Tribolium castaneum*	guanyl-nucleotide exchange factor	78	6.00e-32
aa17	scarlet	XM_001848914.1	*Culex quinquefasciatus*	scarlet	69	2.00e-12
aa18	white1	AF442747.1	*Tribolium castaneum*	white	67	3.00e-05
aa19	rotund	NM_001144549.1	*Drosophila melanogaster*	rotund	98	4.00e-37
aa20	neprilysin	XM_002408091.1	*Ixodes scapularis*	neprilysin	64	6.00e-28
aa21	coracle	XM_001943191.1	*Acyrthosiphon pisum*	coracle	18	7.00e-18
aa22	tfp	DQ848565.1	*Locusta migratoria*	tfp1	57	2.00e-14
aa23	karmoisin	XM_001857512.1	*Culex quinquefasciatus*	monocarboxylate transporter	37	4.00e-16
aa24	tensin	XM_001602507.1	*Nasonia vitripennis*	tensin	49	7.00e-34
aa25	pbx	XM_001120618.1	*Apis mellifera*	pbx/knotted 1 homeobox 1.1	57	4.00e-19
aa26	tramtrack	NM_001164138.1	*Tribolium castaneum*	tramtrack	30	6.00e-12
aa27	labial	NM_001114290.1	*Tribolium castaneum*	labial	22	7.00e-07
aa28	nanos	EU289288.1	*Parhyale hawaiensis*	nanos	62	6.00e-05
aa29	msx	AB302960.1	*Pandinus imperator*	msxA	38	5.00e-09
aa30	tef-1	XM_002187355.1	*Taeniopygia guttata*	tef-1 gamma	69	7.00e-44
aa31	myosin	AF474966.1	*Glyptonotus antarcticus*	myosin	32	8.00e-27
aa32	tiptop	XM_002422528.1	*Pediculus humanus corporis*	tiptop	50	4.00e-18
aa33	ptcr	EF442429.1	*Apis mellifera*	patched-related protein	64	1.00e-24
aa34	plexin	XM_001661601.1	*Aedes aegypti*	plexin b	77	2.00e-73
aa35	none	n.a.	n.a.	n.a.	n.a.	n.a.
aa36	scorched	XM_969086.2	*Tribolium castaneum*	F-actin-capping protein subunit alpha	20	5.00e-08
aa37	pipsqueak	XM_965219.2	*Tribolium castaneum*	bmp-induced factor	52	2.00e-13
aa38	vermilion	NM_078558.2	*Drosophila melanogaster*	vermilion	49	1.00e-15
aa39	fringe	XM_002128597.1	*Ciona intestinalis*	fringe	38	2.00e-07
aa40	echinus	XM_969086.2	*Tribolium castaneum*	F-actin-capping protein subunit alpha	20	2.00e-07
aa41	opa	XM_623410.2	*Apis mellifera*	optic atrophy 1-like	63	9.00e-17
aa42	catsup	XM_395560.3	*Apis mellifera*	solute casolute carrier family 39 (zinc transporter)	20	1.00e-06
aa43	AP-1	XR_015060.1	*Apis mellifera*	AP-1	67	9.00e-28
aa44	no ocelli	XM_969829.2	*Tribolium castaneum*	nocA no ocelli	44	1.00e-29
aa45	cytochrome P450	NM_001130436.1	*Tribolium castaneum*	cytochrome P450	64	1.00e-10
aa46	abrupt	XM_002424613.1	*Pediculus humanus corporis*	abrupt	10	6.00e-07
aa47	kekkon	XM_002431457.1	*Pediculus humanus corporis*	leucine-rich repeat-containing protein 24 precursor	83	2.00e-22
aa48	net	XM_962827.1	*Tribolium castaneum*	net	20	2.00e-05

Table S2. Cont.

Marker name	Likely gene	BLAST ID	Species	Gene name	Coverage, %	*e* value
aa49	dll	NM_001131037.1	*Apis mellifera*	distal-less	27	1.00e-32
aa50	nckx30	XM_396230.2	*Apis mellifera*	sodium/potassium/calcium exchanger Nckx30C	27	2.00e-27
aa51	amos	XM_001121050.1	*Apis mellifera*	absent MD neurons and olfactory sensilla	16	6.00e-03
aa52	rhomboid	XM_001122653.1	*Apis mellifera*	stem cell tumor	29	4.00e-09
aa53	calcitonin	XM_001653601.1	*Aedes aegypti*	calcitonin receptor	45	9.00e-04
aa54	masquerade	Y11145.2	*Pacifastacusleniusculus*	masquerade-like	61	7.00e-04
aa55	suH	AJ717514.1	*Cupiennius salei*	suppressor of hairless	69	8.00e-93
aa56	Dscam	GQ154653.1	*Litopenaeus vannamei*	down syndrome cell adhesion molecule	98	3.00e-17
aa57	eyegone	NM_001114345.1	*Tribolium castaneum*	eyegone	18	2.00e-29
aa58	nubbin	HM063420.1	*Asellus aquaticus*	nubbin	57	2.00e-46
aa59	sal	XM_968136.2	*Tribolium castaneum*	spalt-like protein	14	7.00e-09
aa60	ia2	XM_969473.2	*Tribolium castaneum*	receptor-type tyrosine-protein phosphatase N2	47	2.00e-51
aa61	syt1	NM_001172394.1	*Nasonia vitripennis*	synaptotagmin 1	98	8.00e-06
aa62	ef-2	AF240816.1	*Armadillidium vulgare*	elongation factor-2	100	7.00e-35
aa63	GD20337	XM_002103257.1	*Drosophila simulans*	GD20337	31	3.00e-08
aa64	fgfr	XM_965738.2	*Tribolium castaneum*	fibroblast growth factor receptor	76	2.00e-21
aa65	agrin	XM_001339729.4	*Danio rerio*	EGF-like fibronectin and laminin domains (EGFLAM)	46	5.00e-05
aa66	stretchin-mlck	XM_001121724.1	*Apis mellifera*	stretchin-mlck	68	8.00e-29
aa67	notch	AB287421.1	*Achaearanea tepidariorum*	notch	48	5.00e-22
aa68	groucho	XM_002427207.1	*Pediculus humanus corporis*	groucho	10	5.00e-04
aa69	msp300	XM_623753.1	*Apis mellifera*	muscle-specific protein 300	56	5.00e-22
aa70	tango	FJ745381.1	*Chironomus tepperi*	tango	89	6.00e-15
aa71	carboxylesterase	EF675186.1	*Bemisia tabaci*	carboxylesterase 2	52	3.00e-22
aa72	dumpy	XM_001119920.1	*Apis mellifera*	dumpy	65	1.00e-09
aa73	rhodopsin	GQ221739.1	*Neogonodactylus oerstedii*	opsin	98	4.00e-33
aa74	broad	XM_001602311.1	*Nasonia vitripennis*	broad-complex core-protein	22	8.00e-05
aa75	rosy	XM_001501646.2	*Equus caballus*	xanthine dehydrogenase/oxidase	63	2.00e-33
aa76	sp	FN562992.1	*Parhyale hawaiensis*	sp6-9	69	9.00e-42
aa77	guanylate cyclase	EF988651.1	*Penaeus monodon*	guanylate cyclase	76	9.00e-19
aa78	optix	AB239695.1	*Anthopleura japonica*	Six-A	99	9.00e-56
aa79	none	n.a.	n.a.	n.a.	n.a.	n.a.
aa80	hth	XM_001951080.1	*Acyrthosiphon pisum*	homothorax	29	2.00e-19
aa81	pax2	XM_962948.2	*Tribolium castaneum*	shaven	45	4.00e-31
aa82	vps33b	XM_002736163.1	*Saccoglossus kowalevskii*	vacuolar protein sorting-associated protein 33B	36	1.00e-07
aa83	eya	NM_010165.2	*Mus musculus*	eyes absent 2	36	5.70e-02
aa84	klu	NM_057714.3	*Drosophila melanogaster*	klumpfuss	83	2.00e-28
aa85	repo	DQ263615.1	*Oncopeltus fasciatus*	repo	14	5.00e-05
aa86	fatfacets*	AY591387.1	*Macaca fascicularis*	ubiquitin specific peptidase 9	49	5.00e-10
aa87	pale*	DQ347833.1	*Acanthopagrus schlegelii*	tyrosine hydroxylase	38	1.00e-11
aa88	pax6*	AL929172.6	*Danio rerio*	pax6a	20	2.00e-13
aa89	ash*	DQ489559.1	*Panulirus argus*	achaete-scute complex protein	40	3.00e-16
aa90	burgundy*	XM_001604219.1	*Nasonia vitripennis*	gmp synthase	13	1.00e-12
aa91	lightoid*	XM_001599094.1	*Nasonia vitripennis*	rab-related protein	40	5.00e-52
aa92	lim*	XM_002106460.1	*Drosophila simulans*	lim1	33	1.00e-24
aa93	garnet*	XM_002666259.1	*Danio rerio*	adaptor-related protein complex 3	42	2.00e-19

Sequences flanking the SNPs used for genotyping were input into BLAST. Listed are the gene ID, species, query coverage, and *e* value of the highest BLAST hit. Often, the query coverage is low because a large portion of the sequence is intronic. We did not list the highest BLAST hit if the gene ID did not use a gene name. Because we are using often only a small piece of coding sequence and intronic sequence, the highest BLAST hit is suggestive of the identity of the gene but does not confirm the identity. Markers with an asterisk were genotyped by using PCR. n.a., not applicable.

Table S3. Number of animals heterozygous or homozygous at the peak marker for the five binary traits

Trait	Absence vs. presence of pigmentation (aa75)		Red vs. orange and brown pigmentation (aa45)		Light vs. dark pigmentation (aa83)		Stellate vs. diffuse pigment pattern (aa48)		Eye absence vs. presence (aa92)	
Binary code	0	1	0	1	0	1	0	1	0	1
Corresponding phenotype	No pigment	Pigment	Red	Orange, brown	Light red, orange	Red, brown	Stellate	Diffuse	No eyes	Eyes
Heterozygous	12	90	2	54	0	66	0	44	1	81
Homozygous	83	1	34	1	21	2	35	4	51	0

Table S4. Trait name, LG (linkage group), marker name, location on the linkage group, LOD score, odds ratio (log_e odds ratio) (qualitative traits), percent variance (quantitative traits), and number of flanking markers within 10 cM of each marker

Trait	LG	Marker	cM	LOD	Odds ratio	% Variance	Flanking markers
Presence vs. absence of pigmentation	2	aa75	21.31	37.69	403, (6.0)	n.a.	9
Red vs. orange/brown pigmentation	5	aa45	52.7	21.77	501.4, (6.2)	n.a.	5
Light vs. dark pigmentation	3	aa83	1.6	18.28	1143.8, (7.0)	n.a.	5
Stellate vs. diffuse pigmentation	7	aa48	0	19.25	781, (6.7)	n.a.	3
Eye presence vs. absence	7	aa92	3.6	35.68	5595.8, (8.6)	n.a.	3
Eye size	2	aa56	18.9	6.35	n.a.	29.69	9

n.a., not applicable.

Table S5. Primers used to assay genotypes for markers aa86–aa93

Marker	Primers	Sequence
aa86	FATFACETSF	TGCTGGCAAGTTTTCTCTATGGCACT
	FATFACETSR	CAGAATGCAAAATGGATGATGATGA
aa87	PALEF	AAGATATTAAAAGGCGGTCGTAAAC
	PALER	GGAGGGTAAAACCAGTACGTCCTGCAA
aa88	PAX6F	CACACCGCCTAGTTGGTTAACACCAT
	PAX6R	GACGCAAGAAGTACATTAGTATAATTTCAG
aa89	ASHF	GATGTAGTCCACCGCCTGCTTCAGA
	ASHR	TAAACTTTGCCCCAAACAGCAACAA
aa90	BURGUNDYF	CAAAAGGGTGATTAGCATGCAGTTGA
	BURGUNDYR	TGGTGGAACGGTTATAAGAAAACCAG
aa91	LIGHTOIDF	AGTTCAGGCCAGGTAGCATTGACTTTGG
	LIGHTOIDR	GTCTTTATTGCCTTCGCGGAAGCTC
aa92	LIMF	TGGTTACTGTTATGCAATTTACCAAG
	LIMR	TGATGTGTCTGGTGGGTTTCGGAGT
aa93	GARNETF	TGGGTAACCCTGCATTGTATATTTTGTG
	GARNETR	TCTGTGGGGGTCCGTAGTCCGTTTA

The associated alleles could not be distinguished via Sequenom analysis, so either allele-specific PCRs were designed (that only amplified the surface allele) or size specific PCRs were designed (where the surface and cave alleles were different sizes).

Dataset S1. List of degenerate primers used to amplify candidate genes

Dataset S1 (XLS)

Markers not listed are those that were isolated from Floragenex sequences and, therefore, were not degeneratively amplified. Sequences in red are primers for which a 50 °C annealing temperature was used, and sequences in blue are primers for which a 60 °C annealing temperature was used.

Dataset S2. Sequences of SNPs used for Sequenom and PCR assays

Dataset S2 (XLS)

Sequences flanking each SNP are listed as well as the primers used for the Sequenom assay. SNPs used for the assays are in brackets. The surface allele is listed first and the cave allele is listed second. Other SNPs present in the sequences are listed as N except where it is a very large insertion and indicated by **. In many of the sequences, there were areas where one of the populations had a deletion or insertion and those are listed as a string of Ns. Some of the sequences might contain additional SNPs not demarcated by an N. Catsup and SuH sequences contain a fragment of the original degenerate PCR and, therefore, might be missing intronic sequences.