

The Genetic Consequences of Spatially Varying Selection in the Panmictic American Eel (*Anguilla rostrata*)

Pierre-Alexandre Gagnaire,^{*1} Eric Normandeau,^{*} Caroline Côté,^{*} Michael Møller Hansen,[†] and Louis Bernatchez^{*}

^{*}Institut de Biologie Intégrative et des Systèmes, Département de Biologie, Université Laval, Quebec City, Quebec G1V 0A6, Canada, and [†]Department of Biological Sciences, Aarhus University, DK-8000 Aarhus C, Denmark

ABSTRACT Our understanding of the genetic basis of local adaptation has recently benefited from the increased power to identify functional variants associated with environmental variables at the genome scale. However, it often remains challenging to determine whether locally adaptive alleles are actively maintained at intermediate frequencies by spatially varying selection. Here, we evaluate the extent to which this particular type of balancing selection explains the retention of adaptive genetic variation in the extreme situation of perfect panmixia, using the American eel (*Anguilla rostrata*) as a model. We first conducted a genome scan between two samples from opposite ends of a latitudinal environmental gradient using 454 sequencing of individually tagged cDNA libraries. Candidate SNPs were then genotyped in 992 individuals from 16 sampling sites at different life stages of the same cohort (including larvae from the Sargasso Sea, glass eels, and 1-year-old individuals) as well as in glass eels of the following cohort. Evidence for spatially varying selection was found at 13 loci showing correlations between allele frequencies and environmental variables across the entire species range. Simulations under a multiple-niche Levene's model using estimated relative fitness values among genotypes rarely predicted a stable polymorphic equilibrium at these loci. Our results suggest that some genetic-by-environment interactions detected in our study arise during the progress toward fixation of a globally advantageous allele with spatially variable effects on fitness.

VARIABLE environmental conditions across species' ranges provide a basis for differential selection at polymorphic loci involved in local adaptation. In consequence, the level of locally adaptive genetic variation may be potentially increased, through a particular type of balancing selection whereby protected polymorphisms result from selection for different alleles in different environments. Depending on population structure, the degree of habitat choice, and the strength of selection, this process can lead to habitat specialization and eventually to ecological speciation (Maynard Smith 1966). However, when both dispersal across habitats and mating are random processes, local adaptation is impossible and polymorphism may be either lost

by drift or, under special conditions, protected by selection (Yeaman and Otto 2011). This evolutionary mechanism was first investigated more than half a century ago (Levene 1953), through a local density regulation model integrating variation in fitness of genotypes across niches and differential contribution of the niches to a panmictic reproductive pool. Levene demonstrated that a sufficient condition for a locally adaptive polymorphism to be maintained by selection requires the harmonic mean fitness of the heterozygote genotype to be higher than that of each homozygote, a process called "harmonic mean overdominance".

There is an increasing body of empirical evidence for cases of polymorphisms maintained by environmental heterogeneity (reviewed by Hedrick *et al.* 1976; Hedrick 1986, 2006). The most famous examples come from studies of allozyme variation (*e.g.*, Kreitman 1983; Sezgin *et al.* 2004), color polymorphism (Nachman *et al.* 2003; Hoekstra *et al.* 2004), adaptation to climate (Hancock *et al.* 2008; Kolaczkowski *et al.* 2011), and soil type (Turner *et al.* 2010), as well as pathogen and insecticide resistance (Garrigan and Hedrick

Copyright © 2012 by the Genetics Society of America
doi: 10.1534/genetics.111.134825

Manuscript received September 19, 2011; accepted for publication November 20, 2011
Supporting information is available online at <http://www.genetics.org/content/suppl/2011/11/30/genetics.111.134825.DC1>.

¹Corresponding author: Institut de Biologie Intégrative et des Systèmes, Département de Biologie, Université Laval, Pavillon Charles-Eugène-Marchand, 1030 Ave. de la Médecine, Quebec City, Quebec G1V 0A6, Canada. E-mail: pierre-alexandre.gagnaire.1@ulaval.ca

2003; Weill *et al.* 2003; Pelz *et al.* 2005). Since habitat choice or reduced gene flow increases the opportunity for the maintenance of locally adaptive polymorphisms in subdivided populations (Felsenstein 1976), these cases are usually more fully understood under migration–selection models. However, the framework of Levene’s model remains highly relevant to the study of species for which random mating and dispersal exist over large (*e.g.*, marine fishes and invertebrates) or local spatial scales (*e.g.*, sympatric host races of insects).

To date, the principal limitation to evaluating the retention of locally adaptive alleles in such species has been the lack of genomic resources. Only two studies that focused on one or two selected genes have empirically tested the maintenance of polymorphism under Levene’s model, in the leafhopper (Prout and Savolainen 1996) and the acorn barnacle (Schmidt and Rand 2001). In practice, the discovery of locally adaptive polymorphisms in panmixia is not straightforward for several reasons. First, there might be a negative trade-off between the number of loci influenced by spatially varying selection and individual locus effects on fitness, such that the whole adaptive load due to selection on unlinked loci remains sustainable for the population. Second, environmental changes can shift the frequency of some protected variants out of their domain of stability, resulting in the loss of formerly stable polymorphisms. Third, recombination should rapidly erase the effects of selection on the chromosomal neighborhood of the selected sites (Charlesworth *et al.* 1997; Przeworski 2002). As such, partial selective sweeps provoked by the establishment of new protected variants should leave only transient genomic footprints, further reducing the chance to find them. Nevertheless, these difficulties can be partly overcome by tracking protected polymorphisms using a high-density genome-scan approach. Owing to the development of high-throughput sequencing techniques, this strategy is now achievable in most nonmodel species (Stapley *et al.* 2010).

The American eel *Anguilla rostrata* is one of the most appropriate organisms for studying the evolutionary effects of spatially varying selection within Levene’s model framework (Karlin 1977). Mating of the whole species occurs in the Sargasso Sea (Schmidt 1923), after which planktonic larvae are dispersed by the Antilles Current and the Gulf Stream to the eastern North American coast over a large continental habitat, extending from Florida to Quebec and Labrador (Tesch 2003). Studies based on neutral molecular markers have shown that in this textbook example of panmixia, random mating occurs at the species scale (Avisé *et al.* 1986; Wirth and Bernatchez 2003). Evidence from the literature also supports that leptocephali larvae passively drift with the currents (Bonhommeau *et al.* 2010) and that, following metamorphosis, newly transformed unpigmented glass eels use a selective tidal stream transport mechanism to move landward (McCleave and Kleckner 1982). Thus, genotype-dependent habitat choice is unlikely to occur over a large geographical scale due to oriented horizontal

swimming and newly recruited glass eels are exposed to highly unpredictable conditions with respect to environmental parameters (*e.g.*, temperature, salinity, pathogens, and pollutants) during their early life history. Consistent with these observations, clinal variation attributed to single-generation footprints of spatially varying selection was found at three allozyme loci (Williams *et al.* 1973; Koehn and Williams 1978). However, allozyme studies focused on only a few metabolic genes and did not assess the retention of locally adaptive polymorphisms by spatially varying selection. Population genomics now offers powerful tools to bring empirical data to bear on this fundamental question in ecological genetics. Here, we discovered and typed annotated single-nucleotide polymorphisms (SNPs) in transcribed regions of the American eel genome to identify candidate genes potentially associated with environmental variables. An extensive spatiotemporal set of samples was then used to further test for selection at candidate loci, estimate niche-specific relative fitness among genotypes, and investigate conditions for the maintenance of polymorphism under a finite-population Levene’s model.

Materials and Methods

Preparation of cDNA libraries, contig assembly, and SNP discovery

We prepared cDNA libraries for 454 sequencing following the protocol described in Pierron *et al.* (2011). Two samples of 20 glass eels were collected just prior to settlement in freshwater at two river mouths located near the extreme ends of the species’ latitudinal range: the Grande Rivière Blanche in the lower St. Lawrence estuary (RB, 48°78’N, 67°70’W) and Florida (FL, 30°00’N, 81°19’W). Briefly, Poly(A) RNAs were individually extracted from entire glass eels and used as a template for cDNA amplification. Amplified cDNAs were then fragmented by sonication, and fragments from 300 to 800 bp were ligated to the standard 454 B primer and the standard 454 A primer, holding a 10-bp barcode extension at its 3’ end. Therefore, each individual could be identified by its unique barcode. For each sampling site, the 20 individually tagged libraries were pooled in equal amounts and sequenced on a half-plate of Roche GS-FLX DNA Sequencer at Genome Quebec Innovation Center (McGill University, Montreal, QC, Canada).

Base calling was performed using PyroBayes (Quinlan *et al.* 2008) after trimming adapters. Each read was then renamed according to its individual barcode, which was subsequently removed together with potential primers used for cDNA amplification. We performed a *de novo* assembly of the total sequencing data using CLC Genomic Workbench 3.7 (CLC bio), with a minimal read length fraction of 0.5 and a similarity parameter of 0.95. The consensus sequence of each *de novo* built contig was then used as a template for a reference assembly under the same parameters. This second round of assembly aimed at screening for additional reads that were not included into contigs during the step of *de novo*

assembly and excluding poor-quality contigs that did not recruit any read during the reference assembly procedure.

SNP discovery was performed using the neighborhood quality standard (NQS) algorithm (Altshuler *et al.* 2000; Brockman *et al.* 2008) implemented in CLC Genomic Workbench 3.7 (CLC bio). This method takes into account the base quality values to distinguish sequencing errors from actual SNPs. We set a minimum coverage of 20× per SNP site and used either a frequency threshold of 5% or a count threshold of 5 for the rarest variant (when the coverage exceeded 100×) to avoid the detection of sequencing errors as SNPs. Only biallelic SNPs were considered.

Individual genotype inference

There is a significant risk to misscore a heterozygote genotype by repeatedly sampling the same allele when the individual coverage is <5× (see Supporting Information, Figure S1). We corrected such artifactual heterozygote deficiencies by supposing within-sample Hardy–Weinberg equilibrium (HWE) while taking into account the stochasticity induced by the binomial sampling process of homologous sequences at each locus for each individual. For each SNP having >10 individuals sequenced in each sample (*i.e.*, total coverage ≥20×), allele counts were used to determine the observed genotype of each individual (*AA*, *Aa*, *aa*, or *NA* when no sequence data were available) to calculate the observed allelic frequencies. We then supposed within-sample HWE to estimate the number of expected individuals within each genotypic class in each sample given the number of individuals sequenced and the observed allelic frequencies in the sample. When the observed number of heterozygotes was below HWE predictions, new genotypes that were consistent with observed individual data were randomly drawn from a trinomial distribution with event probabilities ($P(AA)_{i,j}$; $P(Aa)_{i,j}$; $P(aa)_{i,j}$) corresponding to the probabilities of each genotype, given the observed data for the *j*th individual in sample *i*. For each locus showing HW deficiency, a new array of individual genotypes was generated until HWE expectations were verified for the sample. The individual genotype probabilities used to parameterize the trinomial sampling process were obtained from the following equation giving the probabilities of real genotypes (G_R) knowing the observed data (G_O) at a given locus,

$$P(G_R|G_O)_{i,j} = \begin{pmatrix} P(AA)_{i,j}; P(Aa)_{i,j}; P(aa)_{i,j} \\ \frac{p_i^2}{p_i^2 + p_i(1-p_i)(1/2)^{(N_{i,j}-1)}} & 0 & 0 \\ \frac{p_i(1-p_i)(1/2)^{(N_{i,j}-1)}}{p_i^2 + p_i(1-p_i)(1/2)^{(N_{i,j}-1)}} & 1 & \frac{p_i(1-p_i)(1/2)^{(N_{i,j}-1)}}{(1-p_i)^2 + p_i(1-p_i)(1/2)^{(N_{i,j}-1)}} \\ 0 & 0 & \frac{(1-p_i)^2}{p_i^2 + p_i(1-p_i)(1/2)^{(N_{i,j}-1)}} \end{pmatrix} \times G_{O_{i,j}},$$

where $N_{i,j}$ is the number of reads (*i.e.*, individual coverage) of individual *j* in sample *i*, $G_{O_{i,j}}$ is its observed genotype

$$\left(\text{with } G_O(AA) = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, G_O(Aa) = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \text{ and } G_O(aa) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right),$$

and p_i is the frequency of the *A* allele in sample *i*. Under this procedure, the genotype of an observed heterozygote was never modified, whereas observed homozygotes could be probabilistically assigned to heterozygotes. Since the sequencing error rate was already taken into account by the SNP detection method, it was neglected at this step to simplify the approach. Methodological validation performed on simulated data sets showed that our correction efficiently restored up to 50% of the hidden heterozygotes (see Figure S1).

Outlier detection

Individual genotypes obtained after treating for the heterozygote deficiency bias were used to detect SNPs potentially affected by diversifying selection between the two samples RB and FL. The empirical distribution of pairwise F_{ST} as a function of within-samples heterozygosity was compared to a neutral distribution simulated under a symmetrical two-island model assuming near random mating (Beaumont and Nichols 1996), using ARLEQUIN ver. 3.5 (Excoffier and Lisher 2010). This approach is more conservative than drawing random samples from a single panmictic population to derive the neutral distribution. For each outlier locus (*i.e.*, F_{ST} value located above the 99.5% quantile of the simulated distribution), the contig's consensus sequence was blasted against the nonredundant NCBI protein database (*nr*), using BLASTX with an *E*-value threshold of 10^{-5} (Altschul *et al.* 1997).

SNP genotyping

Individual SNP assays were developed using the KBiosciences Competitive Allele-Specific PCR genotyping system (KASPar). For each candidate contig, we targeted the SNP showing the highest F_{ST} value when possible. We also developed assays for SNPs identified within contigs of allozyme coding genes showing clinal variation in Williams *et al.* (1973): the *Sorbitol dehydrogenase* gene (*SDH*), two *Phosphoglucose isomerase* isoforms (*PGI-1* and *PGI-2*), and the *Alcohol dehydrogenase* gene (*ADH-3*). Our validation panel was finally completed with nonoutlier SNPs to 100 markers. All assays were tested with 80 individuals and only successfully genotyped SNPs were retained for subsequent genotyping.

A total of 992 individuals belonging to four distinct sample categories were genotyped (Table 1): (i) A reference sample of the 2007 cohort (before selection) consisting of 48 young leptocephali larvae collected in the Sargasso Sea soon after hatching in March and April 2007 (*SAR7*) during the Galathea III expedition (Munk *et al.* 2010); (ii) the first wave of recruiting glass eels belonging to the 2007 cohort, collected between January and July 2008 at 16 river mouths distributed from Florida to Quebec (*GLASS8*); (iii) 1-year-old

individuals from the 2007 cohort, sampled between February and June 2009 from four localities previously sampled in 2008, ranging between South Carolina and Quebec (OYO9); and (iv) glass eels belonging to the 2008 cohort, collected in 2009 at 5 river mouths distributed from South Carolina to Quebec (GLASS9) and that were also sampled in 2008 for the 2007 cohort.

Statistical analyses

We tested for HWE at each diploid locus within each of the four eel sample categories, using ARLEQUIN ver. 3.5 (Excoffier and Lisher 2010). We corrected for multiple independent tests using the false discovery rate correction ($\alpha = 0.05$). Multilocus global F_{ST} values among localities within sample categories were estimated and tested through 10,000 permutations. Outlier SNPs were searched on the basis of their level of genetic differentiation among localities within categories as well as between pairs of localities, using coalescent simulations under a symmetrical island model assuming near random mating.

For each locus, statistical associations between allelic frequencies and a set of four explanatory variables (sample category, latitude, longitude, and temperature) were assessed through logistic regressions using the R package *glmulti* (Calcagno and De Mazancourt 2010). Temperature data were obtained from a National Oceanic and Atmospheric Administration (NOAA) database containing georeferenced sea-surface temperatures along North America's coastlines (SST14NA), with a nominal spatial resolution of 14 km and a 48-hr update frequency. More precisely, we took the sea-surface temperature at river mouth averaged across the 10 days preceding the sampling date in each locality, which corresponded to recruitment at river mouths for the GLASS8 category. Because the exact date of arrival at river mouths was not known for the two other categories of samples, we used different temperature criteria: the three winter months (December to February) average river mouth temperature was used for the OYO9 category, and the sampling month average river mouth temperature was used for the GLASS9 category (Table 1). All possible models involving the four explanatory variables (including pairwise interactions) were fitted using samples from the three continental categories (GLASS8, OYO9, and GLASS9), and the best model was identified using a Bayesian information criterion (BIC). Because the best geographical coverage was achieved for the 2008 glass eels, the same approach was also performed using samples from the GLASS8 category only. For each SNP found in association with explanatory variables, individual haplotype information was retrieved from 454 sequencing data and used to evaluate between-sites linkage disequilibrium (LD), using the method for partially phased haplotypes in Haploview v4.2 (Barrett *et al.* 2005).

The multilocus spatial component of genetic variability at loci inferred to be influenced by spatially varying selection was determined using the spatial principal

component analysis method (sPCA) (Jombart *et al.* 2008) implemented in the R package *adeigenet_1.2-2* (Jombart 2008). The sPCA includes spatial information in the analysis of genetic data, which helps to reveal subtle global spatial structures such as geographic clines. The spatial proximity network among localities was built using the neighborhood-by-distance method. An abrupt decrease of the eigenvalues obtained by decomposing the genetic diversity from the spatial autocorrelation was used as a criterion to choose the principal component to interpret.

Evolution under Levene's model

The classical one-locus-two-allele model of Levene (1953) was extended by the addition of a genetic drift component. At each generation, mating occurs in panmixia, followed by random dispersal of genotypes across niches. Selection is a niche-specific process in which the frequency of allele *A* before selection in the *i*th niche, noted q_i , passes to q'_i after selection following the equation

$$q'_i = \frac{W_i q_i^2 + q_i(1 - q_i)}{W_i q_i^2 + 2q_i(1 - q_i) + V_i(1 - q_i)^2},$$

where W_i and V_i , respectively, denote the fitness of the homozygote genotypes *AA* and *aa* relative to that of the heterozygote genotype in the *i*th niche. Genetic drift is then modeled by randomly drawing $N_e \times C_i$ genotypes in each niche from a trinomial distribution with event probabilities

$$\left(P(AA)_i = q_i'^2; P(Aa)_i = 2q_i'(1 - q_i'); P(aa)_i = (1 - q_i')^2 \right).$$

The new frequency of allele *A* after selection and drift in the *i*th niche is noted q_i'' , and since C_i corresponds to the relative contribution of niche *i* to the global reproductive pool of effective size N_e , the frequency of allele *A* equals $\sum_i C_i q_i''$ in the next mating pool.

To test for equilibrium under Levene's model, empirical values of W_i and V_i were estimated from the observed genotypic frequencies in the SAR7 larval pool ($f_{AA}; f_{Aa}; f_{aa}$) and the modeled niche-specific genotypic frequencies after selection ($f_{AA_i}; f_{Aa_i}; f_{aa_i}$) following

$$W_i = \frac{f_{AA_i} f_{Aa}}{f_{AA} f_{Aa_i}} \quad \text{and} \quad V_i = \frac{f_{aa_i} f_{Aa}}{f_{aa} f_{Aa_i}},$$

where the ratios f_{AA_i}/f_{Aa_i} and f_{aa_i}/f_{Aa_i} were predicted by the regression models of $f_{AA'}/(f_{AA'} + f_{Aa'})$ and $f_{aa'}/(f_{aa'} + f_{Aa'})$, using the observed genotypic frequencies in the GLASS8 samples and the explanatory variables previously selected for this category (see *Results*). For each locus inferred to be influenced by spatially varying selection, the 16 estimated pairs of ($W_i; V_i$) were used to parameterize a 16-niche Levene's model in which the allelic frequencies observed in SAR7 were used as starting values. Different distributions of the C_i were explored, from uniform to normally distributed

Table 1 Sampling location and date, sample size, developmental stage, and river mouth temperature for each analyzed sample

Category	Sampling locality	Code	Development stage	N	Latitude	Longitude	Day	Month	Year	Temperature (°)
SAR7	Sargasso Sea	SAR_7	Leptocephali	48	—	—	—	—	2007	—
GLASS8	Gaspésie, Grande Rivière Blanche	GAS_G8	Glass eel	40	48.78	-67.70	14	June	2008	14.21 ^a
	Newfoundland, Codroy Bay	NF_G8	Glass eel	40	47.85	-59.26	16	July	2008	15.86 ^a
	Prince Edward Island, Rustico Bay	PEI_G8	Glass eel	40	46.43	-63.24	3	July	2008	19.83 ^a
	New Scotia, St. John's River	NS_G8	Glass eel	40	45.54	-64.70	25	April	2008	3.41 ^a
	Maine, Boothbay Harbor	MAI_G8	Glass eel	40	43.85	-69.65	1	May	2008	6.22 ^a
	New Hampshire, Taylor River	NH_G8	Glass eel	40	42.91	-70.84	23	April	2008	6.78 ^a
	Massachusetts, Parker River	MA_G8	Glass eel	40	42.69	-70.79	16	April	2008	7.04 ^a
	Connecticut, Taylor River	CO_G8	Glass eel	40	41.41	-70.55	5	May	2008	10.44 ^a
	New Jersey, Patcong Creek	NJ_G8	Glass eel	40	41.17	-72.23	4	April	2008	7.79 ^a
	Pensylvania, Delaware River	PEN_G8	Glass eel	40	39.89	-75.26	1 to 5	May	2008	9.50 ^a
	Delaware, Millsboro Pond Spillway	DEL_G8	Glass eel	40	38.35	-75.17	5	February	2008	7.00 ^a
	Virginia, Wormley Creek	VIR_G8	Glass eel	40	37.21	-76.49	28	March	2008	10.01 ^a
	North Carolina, Black Creek	NC_G8	Glass eel	40	34.46	-76.48	5 to 7	February	2008	10.44 ^a
	South Carolina, Cooper River	SC_G8	Glass eel	40	32.55	-80.00	13	February	2008	17.19 ^a
	Georgia, Altamaha River	GEO_G8	Glass eel	22	31.18	-81.28	8 to 23	January	2008	15.46 ^a
	Florida, Guana River	FLO_G8	Glass eel	40	30.00	-81.19	28	January	2008	18.09 ^a
OY09	Gaspésie, Grande Rivière Blanche	GAS_E9	Elver 1+	35	48.79	-67.70	—	June	2009	0.16 ^b
	Massachusetts, Parker River	MA_E9	Elver 1+	39	42.69	-70.79	—	April	2009	5.62 ^b
	Pensylvania, Crum Creek	PEN_E9	Elver 1+	40	39.86	-75.32	—	May	2009	7.22 ^b
GLASS9	South Carolina, Cooper River	SC_E9	Elver 1+	29	32.55	-80.00	—	February	2009	14.68 ^b
	Gaspésie, Grande Rivière Blanche	GAS_G9	Glass eel	40	48.79	-67.70	—	June	2009	10.57 ^c
	Nova Scotia, Caledonia	NS_G9	Glass eel	40	46.04	-59.96	—	June	2009	9.55 ^c
	Massachusetts, Parker River	MA_G9	Glass eel	39	42.69	-70.79	—	April	2009	5.65 ^c
	Pensylvania, Crum Creek	PEN_G9	Glass eel	40	39.86	-75.32	—	May	2009	13.57 ^c
South Carolina, Cooper River	SC_G9	Glass eel	20	32.55	-80.00	—	February	2009	12.31 ^c	

^a Ten-day average sea-surface temperature before sampling date (source NOAA: SST14NA).

^b Three winter months (December through February) average sea-surface temperature (source NOAA: SST14NA).

^c Sampling month average sea-surface temperature (source NOAA: SST14NA).

outputs among niches, and the population effective size parameter was set between 10^4 and 10^6 to assess genetic drift effects.

Results

Sequence assembly and SNP discovery

A total of 292.6 Mb of sequences were obtained from the two half-runs of 454 GS-FLX pyrosequencing, among which were 482,322 reads from the St. Lawrence estuary sample (RB, mean read length of 296 bp) and 495,482 reads from the Florida sample (FL, mean read length of 303 bp). These sequences were deposited in the NCBI sequence read archive SRA045712. Trimming adapters and individual barcodes and then filtering for sequence quality removed 5.3% of the reads from the RB data set and 5.1% from the FL data set. Processed reads were assembled into 22,093 contigs with an average length of 464 bp. *In silico* SNP detection allowed identifying 70,912 putative SNPs, 13,293 of which were retained after filtering for a minimal coverage of 10 reads from at least 10 different individuals in both samples RB and FL (*i.e.*, total coverage $\geq 20\times$). This filtering step

allowed inferring 78.1% of the 265,860 genotypes (20 individuals \times 13,293 SNPs) in sample RB and 79.7% in sample FL.

Candidate SNP detection and genotyping

A total of 163 outlier SNPs with estimated F_{ST} values ranging from 0.167 to 0.637 were detected (see Figure S2). However, for most candidate SNPs exhibiting the highest F_{ST} values, BLAST searches revealed the presence of reads matching alternative copies of duplicated genes within contigs. These false SNPs, which probably reflected differential expression patterns of paralogous genes (principally myosin isoforms) between samples RB and FL, were removed from subsequent analyses.

After these filtering procedures, our validation panel included 57 outlier SNPs and was completed to 100 markers with nonoutlier SNPs selected across the full range of heterozygosity. Successful genotyping was obtained for 73 of these 100 SNPs (see File S1), 70 of which were functionally annotated using BLASTX (see Table S1), and 44 were outliers from the initial screen. The genotyping success rate across all samples and loci was $>98\%$; KASPar primers used for genotyping are provided Table S2.

Table 2 Models selected for 13 loci associated with explanatory variables, for both the *GLASS8* data set and the three continental categories

Locus	Gene	<i>GLASS8</i>	Slope <i>P</i> -value	Three continental categories	Slope <i>P</i> -value
ACP_13914	<i>Acyl carrier protein</i>	TEMP	0.0013	TEMP	0.0001
ANX_2_249	<i>Annexin A2-A</i>	TEMP:LAT	0.0385	TEMP+TEMP:LONG+TEMP:LAT+LAT:LONG	0.0017 ^a
CST_21113	<i>Cystatin</i>	Null	—	TEMP:COHORT	0.0021
EIF_3F_341	<i>Translation initiation factor 3 subunit F</i>	Null	—	TEMP	0.0075
GPX_4_19607	<i>Glutathione peroxidase 4</i>	TEMP:LONG	0.0013	TEMP:LONG	0.0008
HSP_90A_15666	<i>Heat-shock protein 90 alpha</i>	TEMP:LONG	0.0355	TEMP+LONG+TEMP:LONG	0.0058 ^a
MDH_1393	<i>Malate dehydrogenase</i>	TEMP:LONG	0.0403	Null	0.0874 ^b
NCP_2_15547	<i>Nucleolar complex protein 2</i>	Null	—	TEMP	0.0539
NRAP_1541	<i>Nebulin-related anchoring protein</i>	TEMP:LAT	0.0425	TEMP:LONG	0.0590
PRP_40_16504	<i>Pre-mRNA-processing factor 40 homolog A</i>	LAT	0.0044	TEMP+LAT+TEMP:LAT	0.0202 ^a
SN4_TDR_374	<i>Staphylococcal nuclease domain-containing protein 1</i>	Null	—	TEMP:LONG	0.0281
TENT_02_11046	—	Null	—	TEMP	0.0082
UGP_2_2128	<i>UDP-glucose pyrophosphorylase 2</i>	TEMP:LONG	0.0177	TEMP+COHORT+TEMP:COHORT	0.0318 ^c

LAT, latitude; LONG, longitude; TEMP, temperature.

^a *P*-value associated to the term identified as best model for *GLASS8*, in the best model for all three continental categories.

^b *P*-value associated to the best model for *GLASS8*.

^c *P*-value associated to the term TEMP, in the best model for all 3 continental categories.

SNP variation patterns

Only one locus departed significantly from HWE expectations within the *SAR7* category, whereas 15, 7, and 10 loci showed significant HW disequilibrium within the continental categories *GLASS8*, *OYO9*, and *GLASS9*, respectively (see Table S3). Overall multilocus F_{ST} values calculated among locality samples were not significantly different from zero within each of the three continental categories (*GLASS8*, $F_{ST} = 0.0003$, $P = 0.318$; *OYO9*, $F_{ST} = 0.0015$, $P = 0.171$; *GLASS9*, $F_{ST} = 0.0022$, $P = 0.060$).

Logistic regressions between allelic frequencies and explanatory variables based on the data set containing the three continental categories (*GLASS8*, *OYO9*, and *GLASS9*) revealed contrasting patterns across loci. For 61 of 73 SNPs, all possible models involving the explanatory variables and their pairwise interactions were rejected. However, significant associations were detected for 10 loci and marginally significant associations for 2 loci (Table 2). The same approach performed within the *GLASS8* category alone revealed statistical associations for 8 loci, all but *MDH* being already detected in the analysis including all continental samples (Table 2). Among the 13 loci for which significant associations were found, 11 were also detected in outlier tests due to atypically high F_{ST} values ranging from 0.052 to 0.175 between some pairs of localities. After removing these 13 loci from the data set, the global multilocus F_{ST} values calculated among locality samples became null in two continental categories (*GLASS8*, $F_{ST} = -0.0015$, $P = 0.981$; *OYO9*, $F_{ST} = -0.0007$, $P = 0.647$) and was reduced in the third one (*GLASS9*, $F_{ST} = 0.0016$, $P = 0.151$).

Most selected regression models revealed significant interactions between spatial variables and river mouth tem-

perature (Table 2), which was measured at different timescales for the three continental categories (Table 1). River mouth temperature at recruitment was selected as the best model for one locus in the *GLASS8* category (*Acyl-Carrier Protein*, *ACP*; Figure 1A). To identify more precisely when this parameter was the most biologically relevant for this locus, we used a sliding-window analysis to test whether the correlation could be improved, using temperature data from different time periods surrounding glass eels recruitment. We found that the correlation was rapidly lost when the 10-days window used to calculate the average sea-surface temperature was shifted around the period corresponding to recruitment (Figure 1B). Furthermore, because the timing of glass eels' recruitment varied considerably across sampling locations, river mouth temperatures were neither correlated with latitude ($R^2 = 0.113$, $P = 0.203$) nor correlated with longitude ($R^2 = 0.036$, $P = 0.484$) during the recruitment period (Table 1).

Multivariate analysis of the eight loci significantly associated with explanatory variables in the *GLASS8* category showed that most of the variability was explained by the first principal component, since the first eigenvalue of the sPCA was highly positive (Figure 2, top right). The global structure illustrated by individual lagged scores on the first principal component showed a synthetic latitudinal cline (Figure 2), corresponding to the multilocus spatial component of genetic variation at the eight loci inferred to be under spatially varying selection. A logistic regression of locality scores against latitude ($R^2 = 0.76$, $P < 0.0001$) showed that the center of the cline coincides with the coastal zone where the latitudinal gradient of nearshore sea-surface temperature is the strongest over the sampling period. Moreover, river mouth temperature averaged across the whole sampling period was a better predictor of

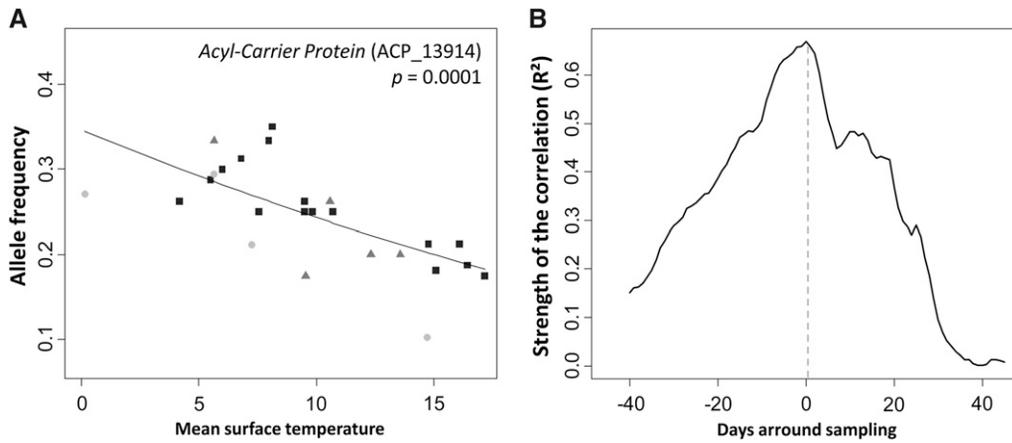


Figure 1 Correlation between river mouth temperature and allele frequencies at locus *ACP*. Logistic regression is based on all three continental categories. (A) Allele frequencies in the *GLASS8* category are represented by solid squares, *OYO9* by circles with light shading, and *GLASS9* by triangles with dark shading. (B) Sliding-window analysis of the coefficient of determination (R^2) between allele frequencies at locus *ACP* in the *GLASS8* category and the values predicted using river mouth temperature data. For

each day within a 3-month period centered on the sampling date (which also corresponded approximately to the date of arrival at river mouths), surface temperature was taken as the mean value across the 10 previous days.

locality scores than latitude ($R^2 = 0.88$, $P < 0.0001$). A highly similar synthetic latitudinal cline was obtained when analyzing the three continental categories together (see Figure S3), supporting the temporal stability of the observed pattern.

One of the 13 SNPs associated with explanatory variables was a nonsynonymous polymorphism (*Nucleolar Complex Protein 2*, *NCP-2*), whereas for 9 of the 12 other SNPs, at least one nonsynonymous segregating site was identified within a 1-kb region (see Figure S4). Heterozygosity within the 13 contigs usually followed a monotonical trend and substantial levels of linkage disequilibrium ($r^2 > 0.5$) were sometimes found between remote SNPs. These results suggest that the indirect influence of selection through linkage with a nearby functional mutation was more likely than direct selection at the focal SNPs. However, this should not introduce any bias in the estimation of the relative fitness values used in the following simulations.

Assessment of polymorphism stability under Levene's model

Simulating the evolution of allelic frequencies for the eight loci significantly associated with explanatory variables in the *GLASS8* category led to two different predictions. Under the hypothesis of uniform contribution among niches and an effective population size (N_e) of 10^5 , simulations predicted polymorphism stability for two SNPs and allele fixation for the remaining six loci (Table 3). The simulated evolution of allelic frequencies over generations as generated by the model can be found for each of the eight loci in Figure S5. Frequency at equilibrium was fairly close to that measured in the *SAR7* sample for the two SNPs predicted to be protected by spatially varying selection. Concerning the six transient polymorphisms, allele fixation was generally reached within <80 generations. Most importantly, the invading allele was always the derived state after identifying the ancestral allele through BLASTN search.

The results obtained under different assumptions on the relative contributions among niches and population size

did not radically change these predictions, as the same two protected polymorphisms were repeatedly inferred across scenarios. However, estimating the niche-specific relative fitness directly from the observed genotype frequencies in the *GLASS8* category (*i.e.*, instead of using values predicted by the regression models) increased the number of protected polymorphisms from two to four (see Table S4).

Discussion

Evidence for single-generation footprints of spatially varying selection

Our results provide strong indications that young glass eels colonizing different areas of the species range are exposed to differential patterns of selection, resulting in significant shifts in allele frequencies within a short timescale. The alternative hypothesis of a subtle neutral population genetic structure in the American eel was not supported by previous works, since panmixia has never been rejected using neutral markers (Avise *et al.* 1986; Wirth and Bernatchez 2003). This conclusion was reiterated here on the basis of 60 neutral SNPs genotyped over 944 individuals distributed across three temporal categories. Moreover, a neutral pattern imposed by spatially restricted gene flow is not consistent with the finding that 85% (11 of 13) of the loci associated with explanatory variables were also detected as outliers and that different regression models were selected across these markers. Consequently, our data do not support the existence of a spatial population structure due to deviation from panmixia in *A. rostrata*. Alternatively, passive genotype-specific habitat choice could possibly occur if the genes associated with environmental variables underlie differences in leptocephalus stage duration. However, the observation that early-metamorphosing leptocephali preferentially recruit to the center of the continental distribution range, whereas late-metamorphosing larvae mostly settle in northern and southern locations (Wang and Tzeng 1998), is inconsistent with the clinal multilocus spatial component detected at

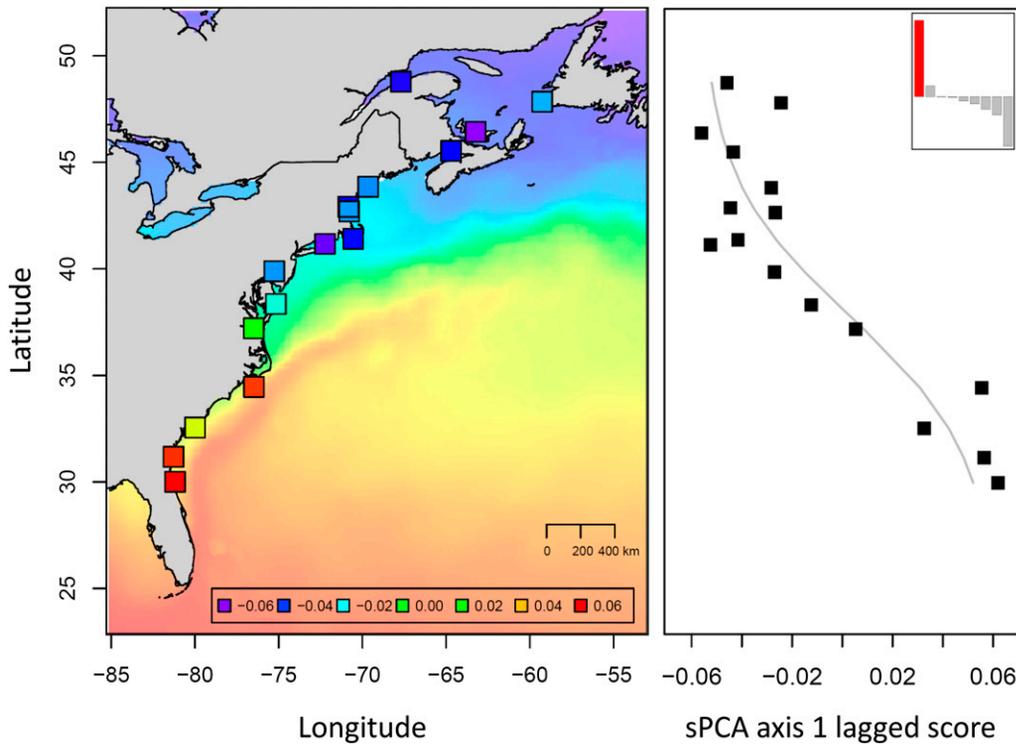


Figure 2 Synthetic multilocus spatial variation component in the 2008 glass eels. The spatial component analysis was based on genetic variation at the eight loci significantly associated with explanatory variables in the *GLASS8* category. The 16 sampling sites are represented on the map by squares colored according to each locality's lagged score on the first principal component, as indicated in the inset. Sea-surface temperatures averaged across the whole sampling period (from January 8 to July 16, 2008) are represented on the same color scale for indication (purple, 0.2°; red, 27.3°). The plot on the right shows the shape of the synthetic multilocus cline, as well as the decomposition of the product of the variance and the spatial autocorrelation into positive, null, and negative components (top right corner). The dinal structure corresponds to the highly positive eigenvalue in red.

these loci (Figure 2). Moreover, half of these loci displayed HWE deviations in continental samples but not in the larval sample, which is incompatible with the habitat choice hypothesis. Therefore, we propose that spatially varying selection is the most parsimonious mechanism underlying the observed patterns of genetic variation.

The diversity of the statistical models retained to explain genetic variation across selected loci may suggest the implication of different locus-specific selective factors. Without detailed information on such agents, the choice of spatial and temperature variables as proxies for the ecological conditions experienced by eels was justified, as, for instance, several environmental factors covary with latitude along the North Atlantic coasts (Schmidt *et al.* 2008). Covariation between latitude and spatially varying selective factors has previously been used to illustrate multilocus spatial patterns attributed to selection in heterogeneous envi-

ronments in *Drosophila melanogaster* (Sezgin *et al.* 2004). Here, we additionally used this synthetic multilocus signal to demonstrate the overall temporal stability of the observed patterns (Figure S3). Owing to the apparent panmixia and the lack of evidence for large-scale habitat choice in the American eel, any genetic pattern left by spatially varying selection at a given generation will be inevitably erased at the next generation. Consequently, the temporal stability of the observed genetic patterns between the *GLASS8* and *GLASS9* categories probably reflects the repeated action of similar natural selection pressures in the two consecutive year cohorts covered by this study. Given that only a very small proportion (<0.5%) of the larvae survive until glass eels reach the coasts and that the glass eel survival rate is ~10% (Bonhommeau *et al.* 2009), our sampling scheme was designed with the intent to detect changes in allelic frequencies occurring during the early stages of eels' life

Table 3 Simulated evolution of allelic frequencies under Levene's model for the eight loci statistically associated with explanatory variables in the *GLASS8* category

Locus	Predictive model	$1/\sum(C_i/W_i)$	$1/\sum(C_i/V_i)$	Frequency Sargasso	Levene's model prediction
ACP_13914	TEMP	0.7661	1.3688	0.6875	Fixation of the derived allele
ANX_2_249	TEMP:LAT	0.8325	0.2900	0.9565	Equilibrium at 0.81
GPX_4_19607	TEMP:LONG	1.0031	1.8379	0.9149	Fixation of the derived allele
HSP_90A_15666	TEMP:LONG	2.1926	1.0759	0.1170	Fixation of the derived allele
MDH_1393	TEMP:LONG	0.6397	0.7761	0.3913	Equilibrium at 0.375
NRAP_1541	TEMP:LAT	0.6065	1.1201	0.5426	Fixation of the derived allele
PRP_40_16504	LAT	4.7856	1.1736	0.1277	Fixation of the derived allele
UGP_2_2128	TEMP:LONG	0.7568	1.4254	0.4468	Fixation of the derived allele

Uniform contribution among niches and a population effective size of 10^5 were assumed in these simulations.

cycle. Moreover, we sampled the first wave of early-recruiting glass eels before potential settlement cues may affect up-estuary migration depending on individual condition and temperature (Sullivan *et al.* 2009).

Although we used variables that mirror continental factors better than open-ocean factors, a decoupling between river mouth and nearshore continental shelf sea-surface temperatures in the northeastern part of the species range [localities Gaspésie (GAS), Newfoundland (NF), and Prince Edward Island (PEI); Table 1] may have influenced our results. Glass eels recruiting in this region during early summer first face cold water temperatures while crossing the continental shelf before entering warmer estuary waters influenced by river outflows. Differential mortality during the cross-shelf transport may thus have resulted in the selection of explanatory models involving interactions between river mouth temperature and latitude or longitude, as observed for six loci of eight in the *GLASS8* category. This interpretation is further supported by the close correspondence between the multilocus spatial variation component and the nearshore averaged sea-surface temperature pattern in Figure 2. Although it suggests that the continental shelf sea-surface temperature should have been used in the regression analyses, this variable remains too difficult to measure without knowing the trajectories of eels during the cross-shelf transport. Using hydrodynamic models for backtracking larval transport may thus help in selecting additional meaningful variables in future studies. Admittedly, as in any other study of this type, our approach cannot fully capture the signal of all spatially varying selection pressures and probably underestimates the number of genes under spatially varying selection. On the other hand, the strong association found at locus *Acyl-Carrier Protein (ACP)* between allele frequencies and temperature at recruitment (Figure 1) shows that the river mouth temperature is a relevant variable. Indeed, settlement in estuaries is a critical period during which glass eels do not feed (Sullivan *et al.* 2009) and probably live on their fatty acids reserves.

Three of the five allozymes previously studied (Williams *et al.* 1973) were included in our analysis to assess whether clines observed at the protein level could be detected at the DNA level. The SNP developed for the *Malate dehydrogenase* gene (*MDH*), which was also detected as an outlier in our initial 454 transcriptome scan, was the only one to show a significant association with environmental variables. In the allozyme study, however, genetic heterogeneity at this locus was observed only among samples of adults and not at the glass eel stage. The lack of a significant pattern for the SNPs developed at the *ADH* and *PGI* loci may be due to problems of paralogy or to a lack of LD with the SNPs under selection.

The fate of selected polymorphisms under Levene's model

Covariation between environmental variables such as temperature and the direction and strength of selection has

been suspected for a long time to actively maintain polymorphisms in heterogeneous environments (reviewed by Hedrick *et al.* 1976; Karlin 1977). Depending on the overall sum of local selective effects, spatially varying selection can, however, lead to two different outcomes: (i) balanced selection for different alleles in different environments can maintain polymorphism over generations, while (ii) globally unbalanced local effects of directional selection may lead to allelic fixation (Levene 1953). Recent theoretical developments have shown that substantial multilocus polymorphism can be maintained under Levene's model, in particular when locally advantageous alleles are partially dominant (Bürger 2010). This includes cases of local dominance that specifically arise with enzymes when fitness is a concave function of the activity level while the heterozygote's enzymatic activity is intermediate to that of both homozygotes (Gillespie and Langley 1974).

Here, equilibrium was tested through simulations under Levene's model to account for combinations of parameters leading to nontrivial evolution of allelic frequencies within a finite population. This approach is relevant since the underlying conditions of Levene's model perfectly fit the American eel's life cycle, which is characterized by random mating and dispersal, and a local density regulation (Vollestad and Jonsson 1988). While exploring a realistic range of parameter values, stable equilibrium was predicted at only two of eight tested loci. This relatively low proportion may be partly explained by uncertainties due to the methodological approach. For instance, a lack of precision in the estimation of fitness parameters, but also in the relative outputs among niches, will obviously affect the realism of the simulations (Schmidt and Rand 2001). Here, we considered only 16 river mouths among a much greater number of existing rivers harboring the American eel along the North Atlantic coast. Yet, those sampling locations are distributed evenly across the species range and are therefore representative of the variation in selection direction and intensity potentially encountered by glass eels. By considering only the earliest life stages, we may fail to catch differential selection acting later in the life cycle. While we cannot rule it out, the fact that most of the mortality occurs before entering freshwater (Bonhommeau *et al.* 2009) reduces this possibility. Moreover, it is likely that selection acting at later life stages plays on different sets of genes. Finally, successive waves of recruiting glass eels can face different conditions depending on their date of arrival (*i.e.*, sea-surface temperature), and interannual global variations in the selective parameters may also exist. The natural settings are thus likely more complex than considered in the model. However, it has been shown that temporal variation in selection is less efficient in maintaining locally adaptive polymorphisms compared to spatially varying selection (Ewing 1979).

Thus, it appears plausible that the low proportion of protected polymorphisms truly reflects the relatively restrictive conditions required for equilibrium (Levene 1953). When a locally advantageous allele appears by mutation and

successfully escapes random loss when rare, it has more chance to invade the panmictic gene pool and to become fixed than to stabilize at an intermediate, stable frequency. Because the transitory phase to fixation will often last for a few hundred generations, loci that are undergoing incomplete selective sweeps may not be easily discovered unless they are frequent enough to be detected with a genome-scan approach. In populations with a large census size, however, new adaptive mutations can frequently occur (Karasov *et al.* 2010) and may result in selective sweeps if the overall effects of spatially varying selection are unbalanced. The finding that, in our simulations, the invading allele was always a derived state for each of the six predicted unstable SNPs supports the hypothesis of linkage with such a globally advantageous mutation that has not already reached fixation.

Incomplete sweeps are expected to leave a specific pattern in the haplotype structure. Because the derived allele increasing in frequency has an atypically long-range LD compared to neutral ancestral variants segregating at the same frequency (Sabeti 2006; Voight *et al.* 2006), the measure of LD can be used to detect ongoing directional selection. Here, sequence information retrieved from 454 sequencing data was insufficient to perform such tests on the basis of the haplotype structure, which require phased haplotypes extending outside the selected gene. However, measuring LD on the basis of available information within contigs revealed the existence of substantial linkage ($r^2 > 0.5$) between some sites that are likely separated by a few kilobases if the presence of introns is taken into account.

Implications for adaptation and conservation of American eel

The Gene Ontology (GO) molecular functions of the genes inferred to be involved in $G \times E$ interactions mostly encompassed major metabolic functions, among which are lipid metabolism (*ANX-2*, inhibition of phospholipase A2; *ACP*, acyl carrier activity; *GPX-4*, phospholipid-hydroperoxide glutathione peroxidase activity), saccharide metabolism (*MDH*, malate dehydrogenase activity; *UGP-2*, UDP-glucose pyrophosphorylase activity), and protein biosynthesis (*EIF-3F*, translation initiation factor; *PRP-40*, pre-mRNA-processing activity). The best predictive models of all these genes included temperature, a factor known to have a strong influence on the level of metabolism in the American eel (Walsh *et al.* 1983). Moreover, the center of the synthetic multilocus latitudinal cline coincided with the region where the warm waters of the Gulf Stream drift away from the coasts. Although the American eel occupies a wide latitudinal range, its thermal preferendum is rather elevated for the temperate zone, since glass eels have a highly reduced swimming ability below 7° (Wuenschel and Able 2008), elvers optimally grow at 28° (Tzeng *et al.* 1998), and yellow eels stop feeding and become metabolically depressed below 10° (Walsh *et al.* 1983). Therefore, selective effects are logically

expected at the relatively low temperatures locally encountered between metamorphosis and recruitment to estuaries, although phenotypic plasticity may also account for the wide range of temperature tolerance in *A. rostrata* (Daverat *et al.* 2006).

Two genes involved in defense response were also detected (*CST*, cysteine endopeptidase inhibitor activity; *SN4-TDR*, nuclease activity). Since selective factors related to pathogen exposure do not always correlate with temperature or geographic coordinates, other genes whose variation patterns could not be explained with our set of explanatory variables may also play a role in resistance to pathogens in *A. rostrata*. For instance, the innate immune response gene *TRIM-35* showed strong departure from HWE in glass eels and atypically high levels of genetic differentiation between some localities (F_{ST} values up to 0.174). In parallel, simulating the evolution of allelic frequencies at this locus on the basis of observed genotype frequencies predicted a stable equilibrium (results not shown). This observation warrants further investigation, especially since the *TRIM-35* gene cluster, which is located in a region of significantly elevated nucleotide diversity in the threespine stickleback (*Gasterosteus aculeatus*), is also a candidate target of balancing selection in this species (Hohenlohe *et al.* 2010).

In conclusion, we have screened >13,000 SNPs in transcribed regions of the American eel genome and identified several genes undergoing spatially varying selection associated with the highly heterogeneous habitat used by this species. Due to our methodological approach, however, the number of genes involved in $G \times E$ interactions has likely been underestimated, and the causative agents of selection remain partially unknown. Nevertheless, the higher proportion of transient vs. stable polymorphisms suggests that locally adaptive polymorphisms are not easily maintained by spatially varying selection when local adaptation is impossible. Under such conditions, theory predicts that phenotypic plasticity, by broadening the environmental tolerance of individual genotypes, provides a more functionally adaptive response to spatial environmental variation (Sultan and Spencer 2002). Indeed, the costs induced by selection on locally adaptive traits are particularly severe in the case of random mating and in the absence of habitat choice (Lenormand 2002). For eels, as for other highly fecund marine species facing huge mortality rates during larval stages, phenotypic plasticity may represent the main mechanism for coping with habitat heterogeneity (Edeline 2007), and our results suggest that differential expression of paralogous genes may be involved in this regulation. Nevertheless, the finding of locally selected mutations spreading to fixation in *A. rostrata* suggests that this high census size species may be regularly subject to new locally adaptive mutations. How the recent population decline of Atlantic eels (Wirth and Bernatchez 2003) affects their adaptability to changing environments is still poorly understood and will be a matter of further investigations.

Acknowledgments

We are grateful to M. Castonguay for his help in sample collection and to R. MacGregor as well as anonymous referees and Associate Editor D. Begun for their constructive and helpful comments on the manuscript. We also thank F. Pierron, M. H. Perreault, V. Bourret, and G. Côté for their precious help with molecular experiments and C. Sauvage for help with bioinformatic analyses. L.B. and P.A.G. designed research; P.A.G. performed research; C.C., M.M.H., E.N., and P.A.G. contributed reagents, materials, and analysis tools; P.A.G. analyzed data; and P.A.G., E.N., M.M.H. and L.B. wrote the paper. This research was funded by a grant from Natural Science and Engineering Research Council of Canada (Discovery grant program) as well as by a Canadian Research Chair in Genomics and Conservation of Aquatic Resources (to L.B.). P.A.G. was supported by a Government of Canada Postdoctoral Fellowship. M.M.H. was supported by the Danish Council for Independent Research|Natural Sciences (grant 09-072120).

Literature Cited

- Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang *et al.*, 1997 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25: 3389–3402.
- Altshuler, D., V. Pollara, C. Cowles, W. Van Etten, J. Baldwin *et al.*, 2000 An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature* 407: 513–516.
- Avise, J. C., G. S. Helfmant, N. C. Saunders, and L. S. Hales, 1986 Mitochondrial DNA differentiation in North Atlantic eels: population genetic consequences of an unusual life history pattern. *Proc. Natl. Acad. Sci. USA* 83: 4350–4354.
- Barrett, J. C., B. Fry, J. Maller, and M. J. Daly, 2005 Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21: 263–265.
- Beaumont, M. A., and R. A. Nichols, 1996 Evaluating loci for use in the genetic analysis of population structure. *Proc. R. Soc. Lond. B Biol. Sci.* 263: 1619–1626.
- Bonhommeau, S., O. Le Pape, D. Gascuel, B. Blanke, A. M. Tréguier *et al.*, 2009 Estimates of the mortality and the duration of the trans-Atlantic migration of European eel *Anguilla anguilla* leptocephali using a particle tracking model. *J. Fish Biol.* 74: 1891–1914.
- Bonhommeau, S., M. Castonguay, E. Rivot, R. Sabatié, and O. Le Pape, 2010 The duration of migration of Atlantic *Anguilla* larvae. *Fish Fish.* 11: 289–306.
- Brockman, W., P. Alvarez, S. Young, M. Garber, G. Giannoukos *et al.*, 2008 Quality scores and SNP detection in sequencing-by-synthesis systems. *Genome Res.* 18: 763–770.
- Bürger, R., 2010 Evolution and polymorphism in the multilocus Levene model with no or weak epistasis. *Theor. Popul. Biol.* 78: 123–138.
- Calcagno, V., and C. de Mazancourt, 2010 glmulti: an R package for easy automated model selection with (generalized) linear models. *J. Stat. Softw.* 34: i12.
- Charlesworth, B., M. Nordborg, and D. Charlesworth, 1997 The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genet. Res.* 70: 155–174.
- Daverat, F., K. E. Limburg, I. Thibault, J. C. Shiao, J. J. Dodson *et al.*, 2006 Phenotypic plasticity of habitat use by three temperate eel species *Anguilla anguilla*, *A. japonica* and *A. rostrata*. *Mar. Ecol. Prog. Ser.* 308: 231–241.
- Edeline, E., 2007 Adaptive phenotypic plasticity of eel diadromy. *Mar. Ecol. Prog. Ser.* 341: 229–232.
- Ewing, E. P., 1979 Genetic variation in heterogeneous environments VII. Temporal and spatial heterogeneity in infinite populations. *Am. Nat.* 114: 197–212.
- Excoffier, L., and H. E. L. Lisher, 2010 Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Res.* 10: 564–567.
- Felsenstein, J., 1976 The theoretical population genetics of variable selection and migration. *Annu. Rev. Genet.* 10: 253–280.
- Garrigan, D., and P. W. Hedrick, 2003 Perspective: detecting adaptive molecular evolution, lessons from the MHC. *Evolution* 57: 1707–1722.
- Gillespie, J. H., and C. H. Langley, 1974 A general model to account for enzyme variation in natural populations. *Genetics* 76: 837–884.
- Hancock, A. M., D. B. Witonsky, A. S. Gordon, G. Eshel, J. K. Pritchard *et al.*, 2008 Adaptations to climate in candidate genes for common metabolic disorders. *PLoS Genet.* 4: e32.
- Hedrick, P. W., 1986 Genetic polymorphism in heterogeneous environments: a decade later. *Annu. Rev. Ecol. Syst.* 17: 535–566.
- Hedrick, P. W., 2006 Genetic polymorphism in heterogeneous environments: the age of genomics. *Annu. Rev. Ecol. Syst.* 37: 67–93.
- Hedrick, P. W., M. E. Ginevan, and E. P. Ewing, 1976 Genetic polymorphism in heterogeneous environments. *Annu. Rev. Ecol. Syst.* 7: 1–32.
- Hoekstra, H. E., K. E. Drumm, and M. W. Nachman, 2004 Ecological genetics of adaptive color polymorphism in pocket mice: geographic variation in selected and neutral genes. *Evolution* 58: 1329–1341.
- Hohenlohe, P. A., S. Bassham, P. D. Etter, N. Stiffler, E. A. Johnson *et al.*, 2010 Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genet.* 6: e1000862.
- Jombart, T., 2008 adegenet: an R package for the multivariate analysis of genetic markers. *Bioinformatics* 24: 1403–1405.
- Jombart, T., S. Devillard, A. B. Dufour, and D. Pontier, 2008 Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity* 101: 92–103.
- Karasov, T., S. W. Messer, and D. A. Petrov, 2010 Evidence that adaptation in *Drosophila* is not limited by mutation at single sites. *PLoS Genet.* 6: e1000924.
- Karlin, S., 1977 Gene frequency patterns in the Levene subdivided population model. *Theor. Popul. Biol.* 11: 356–385.
- Koehn, R. K., and G. C. Williams, 1978 Genetic differentiation without isolation in the American eel, *Anguilla rostrata*. II. Temporal stability of geographic patterns. *Evolution* 32: 624–637.
- Kolaczowski, B., A. D. Kern, A. K. Holloway, and D. J. Begun, 2011 Genomic differentiation between temperate and tropical Australian populations of *Drosophila melanogaster*. *Genetics* 187: 245–260.
- Kreitman, M., 1983 Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. *Nature* 304: 411–417.
- Lenormand, T., 2002 Gene flow and the limits to natural selection. *Trends Ecol. Evol.* 17: 183–189.
- Levene, H., 1953 Genetic equilibrium when more than one ecological niche is available. *Am. Nat.* 87: 331–333.
- Maynard Smith, J., 1966 Sympatric speciation. *Am. Nat.* 100: 637–650.

- McCleave, J. D., and R. C. Kleckner, 1982 Selective tidal stream transport in the estuarine migration of glass eels of the American eel (*Anguilla rostrata*). *J. Conseil* 40: 262–271.
- Munk, P., M. M. Hansen, G. E. Maes, T. G. Nielsen, M. Castonguay *et al.*, 2010 Oceanic fronts in the Sargasso Sea control the early life and drift of Atlantic eels. *Proc. Biol. Sci.* 277: 3593–3599.
- Nachman, M. W., H. E. Hoekstra, and S. L. D'Agostino, 2003 The genetic basis of adaptive melanism in pocket mice. *Proc. Natl. Acad. Sci. USA* 100: 5268–5273.
- Pelz, H. J., S. Rost, M. Hünerberg, C. Fregin, A. C. Heiberg *et al.*, 2005 The genetic basis of resistance to anticoagulants in rodents. *Genetics* 170: 1839–1847.
- Pierron, F., E. Normandeau, M. Defo, P. Campbell, L. Bernatchez *et al.*, 2011 Effects of chronic metal exposure on wild fish populations revealed by high-throughput cDNA sequencing. *Ecotoxicology* 20: 1388–1399.
- Prout, T., and O. Savolainen, 1996 Genotype-by-environment interaction is not sufficient to maintain variation: levine and the leafhopper. *Am. Nat.* 148: 930–936.
- Przeworski, M., 2002 The signature of positive selection at randomly chosen loci. *Genetics* 160: 1179–1189.
- Quinlan, A. R., D. A. Stewart, M. P. Strömberg, and G. T. Marth, 2008 Pyrobayes: an improved base caller for SNP discovery in pyrosequences. *Nat. Methods* 5: 179–181.
- Sabeti, P. C., 2006 Positive natural selection in the human lineage. *Science* 312: 1614–1620.
- Schmidt, J., 1923 The breeding places of the eel. *Philos. Trans. R. Soc. B Biol. Sci.* 211: 179–208.
- Schmidt, P. S., and D. M. Rand, 2001 Adaptive maintenance of genetic polymorphism in an intertidal barnacle: habitat-and-life-stage specific survivorship of MPI genotypes. *Evolution* 55: 1336–1344.
- Schmidt, P. S., and E. A. Serrao, G. A. Pearson, C. Riginos, P. D. Rawson *et al.*, 2008 Ecological genetics in the North Atlantic: environmental gradients and adaptation at specific loci. *Ecology* 89: S91–S107.
- Sezgin, E., D. D. Duvernell, L. M. Matzkin, Y. Duan, C.-T. Zhu *et al.*, 2004 Single-locus latitudinal clines and their relationship to temperate adaptation in metabolic genes and derived alleles in *Drosophila melanogaster*. *Genetics* 168: 923–931.
- Stapley, J., J. Reger, P. G. D. Feulner, C. Smadja, J. Galindo *et al.*, 2010 Adaptation genomics: the next generation. *Trends Ecol. Evol.* 25: 705–712.
- Sullivan, M. C., M. J. Wuenschel, and K. W. Able, 2009 Inter and intra-estuary variability in ingress, condition and settlement of the American eel *Anguilla rostrata*: implications for estimating and understanding recruitment. *J. Fish Biol.* 74: 1949–1969.
- Sultan, S. E., and H. G. Spencer, 2002 Metapopulation structure favors plasticity over local adaptation. *Am. Nat.* 160: 271–283.
- Tesch, F., 2003 *The Eel*. Blackwell Science, Oxford.
- Turner, T. L., E. C. Bourne, E. J. Von Wettberg, T. T. Hu, and S. V. Nuzhdin, 2010 Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nat. Genet.* 42: 260–263.
- Tzeng, W. N., Y. T. Wang, and C. H. Wang, 1998 Optimal growth temperature of the American eel, *Anguilla rostrata* (Le Sueur). *J. Fish Soc. Taiwan* 25: 111–115.
- Voight, B. F., S. Kudravalli, X. Wen, and J. K. Pritchard, 2006 A map of recent positive selection in the human genome. *PLoS Biol.* 4: e72.
- Vollestad, L. A., and B. Jonsson, 1988 A 13-year study of the population dynamics and growth of the European eel *Anguilla anguilla* in a Norwegian river: evidence for density-dependent mortality, and development of a model for predicting yield. *J. Anim. Ecol.* 57: 983–997.
- Walsh, P. J., G. D. Foster, and T. W. Moon, 1983 The effects of temperature on metabolism of the American eel *Anguilla rostrata* (Le Sueur): compensation in the summer and torpor in the winter. *Physiol. Zool.* 56: 532–540.
- Wang, C.-H., and W.-N. Tzeng, 1998 Interpretation of geographic variation in size of American eel *Anguilla rostrata* elvers on the Atlantic coast of North America using their life history and otolith ageing. *Mar. Ecol. Prog. Ser.* 168: 35–43.
- Weill, R., G. Lutfalla, K. Mogensen, F. Chandre, A. Berthomieu *et al.*, 2003 Insecticide resistance in mosquito vectors. *Nature* 423: 136–137.
- Williams, G. C., R. K. Koehn, and J. B. Mitton, 1973 Genetic differentiation without isolation in the American Eel, *Anguilla rostrata*. *Evolution* 27: 192–204.
- Wirth, T., and L. Bernatchez, 2003 Decline of North Atlantic eels: A fatal synergy? *Proc. Biol. Sci.* 270: 681–688.
- Wuenschel, M., and K. Able, 2008 Swimming ability of eels (*Anguilla rostrata*, *Conger oceanicus*) at estuarine ingress: contrasting patterns of cross-shelf transport? *Mar. Biol.* 154: 775–786.
- Yeaman, S., and S. P. Otto, 2011 Establishment and maintenance of adaptive genetic divergence under migration, selection, and drift. *Evolution* 65: 2123–2129.

Communicating editor: D. Begun

GENETICS

Supporting Information

<http://www.genetics.org/content/suppl/2011/11/30/genetics.111.134825.DC1>

The Genetic Consequences of Spatially Varying Selection in the Panmictic American Eel (*Anguilla rostrata*)

Pierre-Alexandre Gagnaire, Eric Normandeau, Caroline Côté, Michael Møller Hansen, and Louis Bernatchez

File S1
Supporting Data

File S1 is available for download at <http://www.genetics.org/content/suppl/2011/11/30/genetics.111.134825.DC1> as an Excel file.

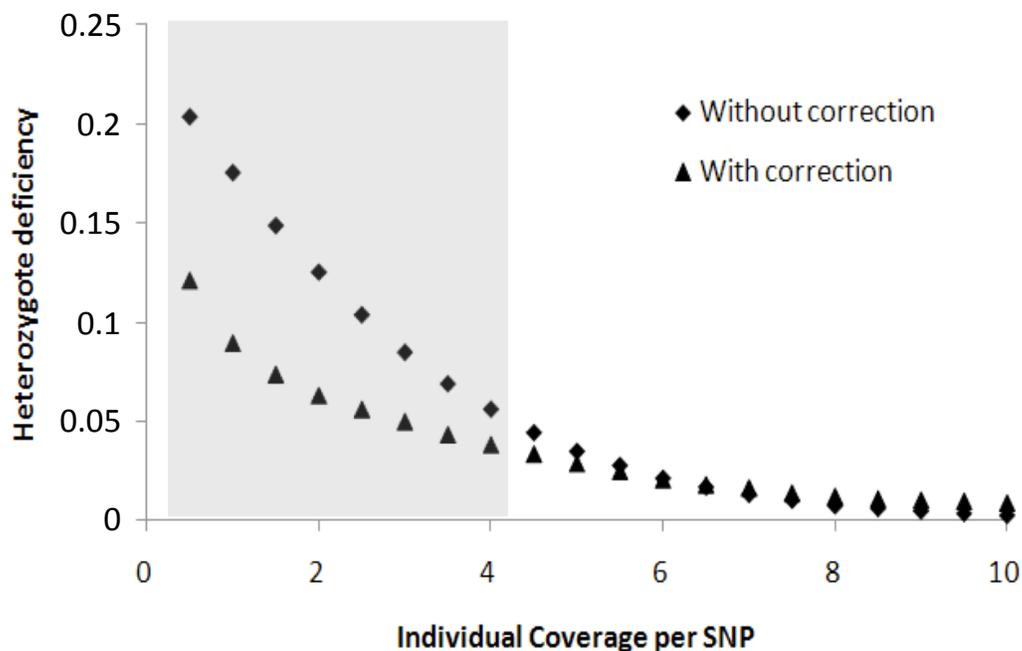


Figure S1 Correction for hidden heterozygotes: methodological validation on simulated datasets. Polymorphism data was simulated for 20 diploid individuals belonging to a panmictic population, at 10000 bi-allelic loci for which the allelic frequency spectrum was analytically derived from equation 51 in Tajima (1989). Individual genotypes were drawn from a trinomial distribution with event probabilities $(\rho^2; 2\rho(1-\rho); (1-\rho)^2)$, where ρ is the frequency of the A allele at a given locus. In order to reproduce the random sampling procedure of individual sequences in our high-throughput sequencing approach, uniform probabilities over individuals and loci were used to randomly draw a number of copies per locus per individual with a given targeted mean coverage. Knowing the number of sampled copies for each individual at each site, observed data were finally drawn using binomial sampling with event probabilities depending on individual genotype. Simulated data were first used to calculate the mean observed heterozygote deficiency per site. Our procedure to correct for heterozygote deficiency caused by limited coverage was then applied on simulated data to evaluate its efficiency. Results are presented for a series of simulated datasets characterized by a mean individual coverage value ranging from 0.5 to 10x. The grey domain represents the individual coverage interval in which most *in silico* SNPs lied in our study.

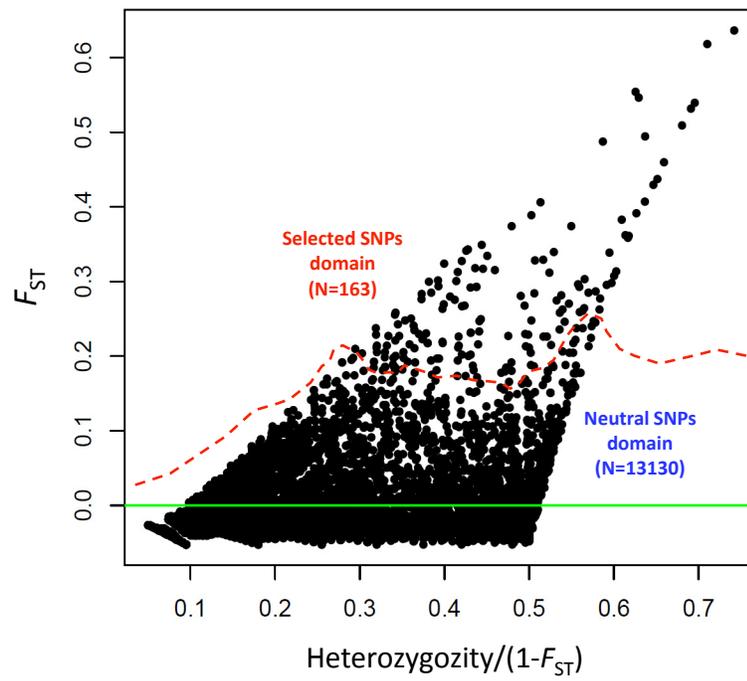


Figure S2 Joint distribution of F_{ST} values and (heterozygosity within samples)/($1-F_{ST}$) estimated between samples RB and FL for the 13293 *in silico* SNPs. The green solid line shows the expected F_{ST} value under panmixia and the red dashed line represents the 99.5% quantile of the distribution of simulated F_{ST} obtained under a neutral model assuming near panmixia.

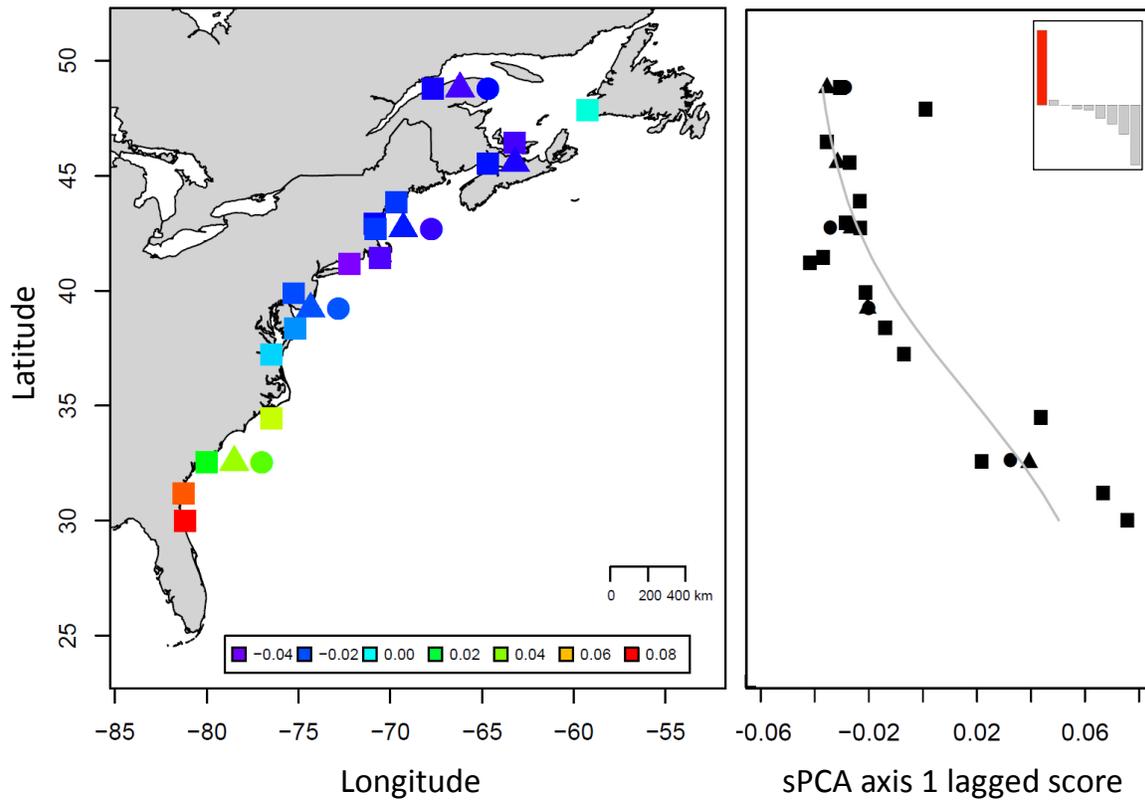


Figure S3 Temporal stability of the synthetic multilocus latitudinal cline over 3 continental categories. Squares represent locality samples from the *GLASS8* category, circles those from the *OYO9* category and triangles those from the *GLASS9* category. The analysis was based on genetic variation at the 8 loci significantly associated with explanatory variables in the *GLASS8* category. Colors are scaled to the lagged scores of the first principal component, as indicated in the caption. The plot on the right shows the shape of the synthetic multilocus cline over the three continental categories, as well as the decomposition of the product of the variance and the spatial autocorrelation into positive, null and negative components (upper right corner). The clinal structure corresponds to the highly positive eigenvalue in red.

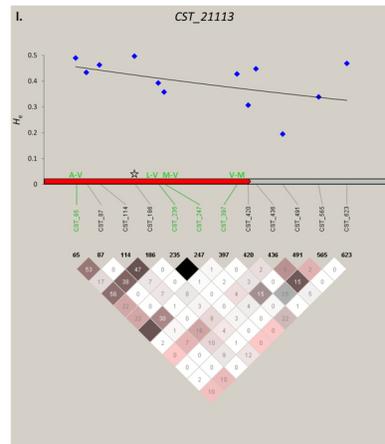
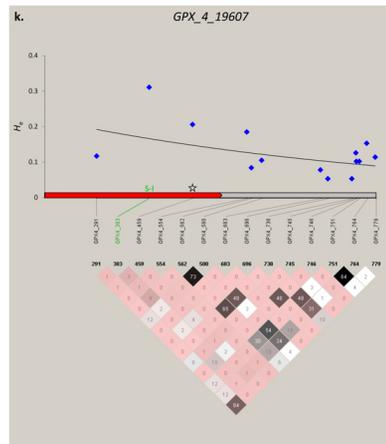
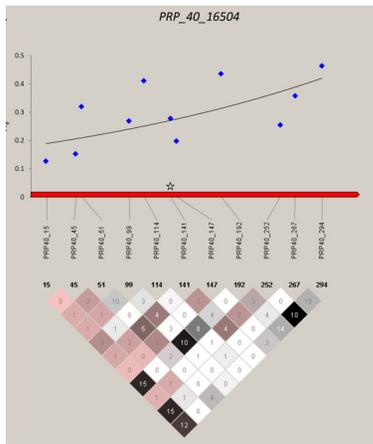
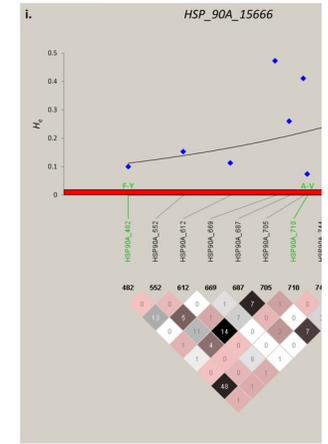
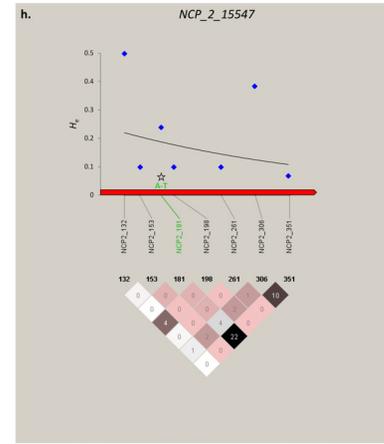
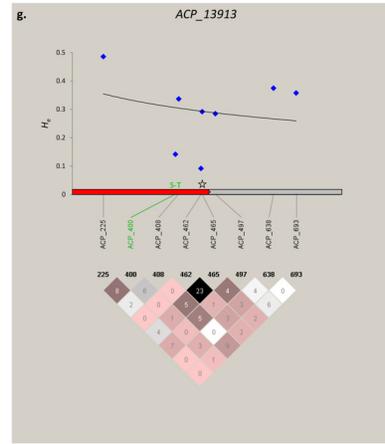
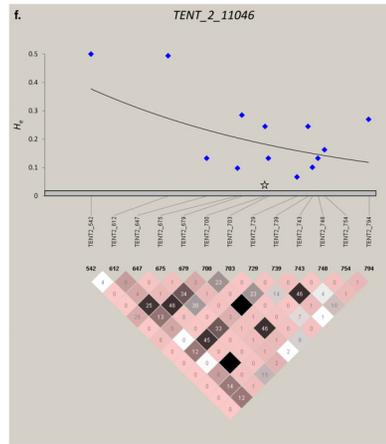
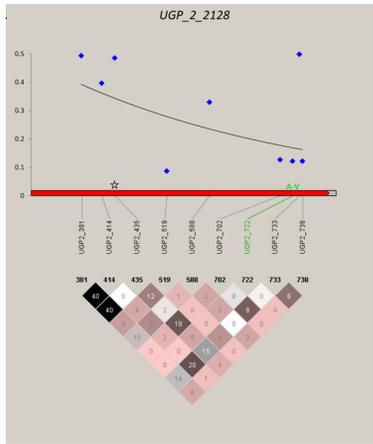
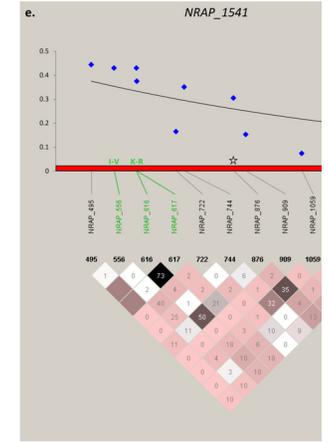
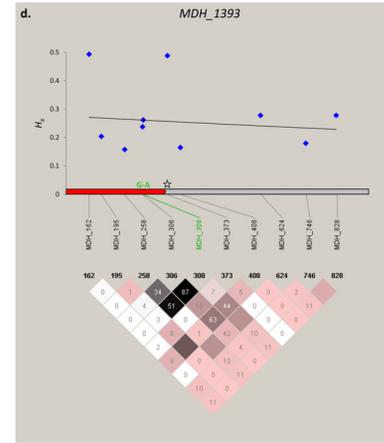
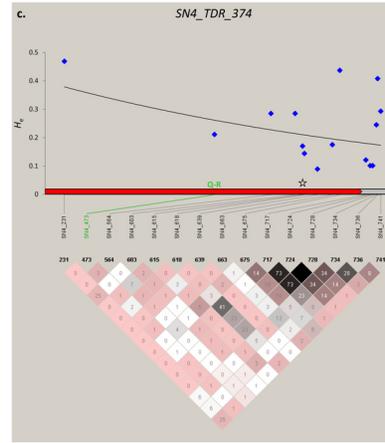
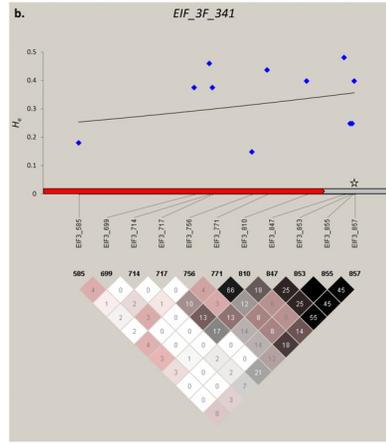
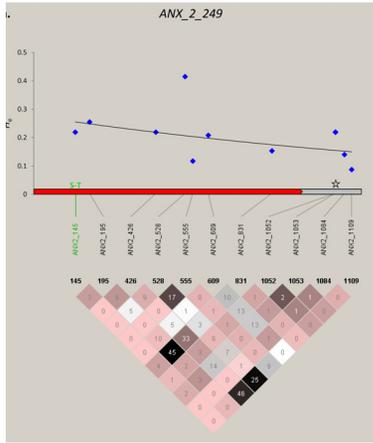


Figure S4 Expected heterozygosity and linkage disequilibrium structure along the sequenced transcribed region for each of the 13 genes associated with explanatory variables. Blue diamonds correspond to values of H_e , showed with a fitted exponential curve. Each square in the triangular matrix represents the level of LD calculated between a pair of polymorphic sites, which position in the gene has been indicated above the matrix. The color indicates the level of LD based on the D' statistic (white for no linkage, shades of pink for intermediate levels of LD and black for strong LD) and the r^2 value (given as a percentage) is displayed for each pair of SNPs. Variable sites in green represent amino-acid polymorphisms in the coding gene regions shown in red. The focal SNP showing significant association with explanatory variables is indicated by a star symbol.

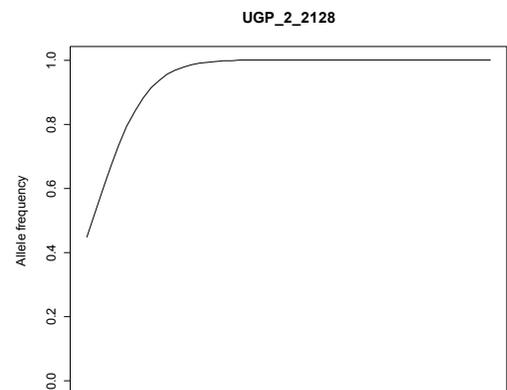
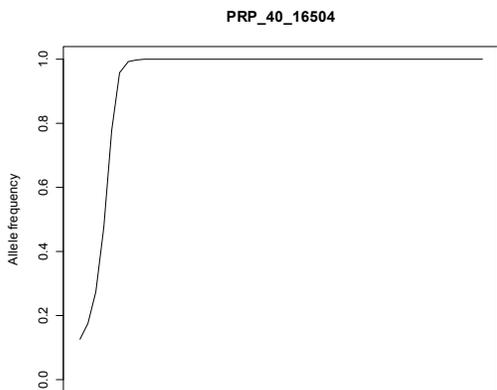
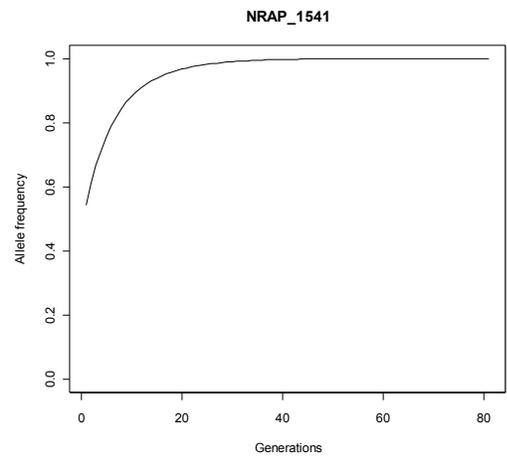
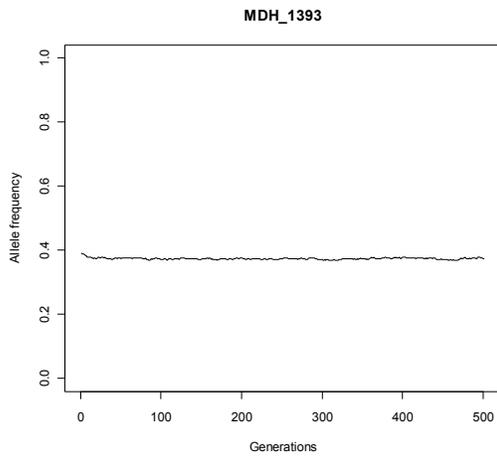
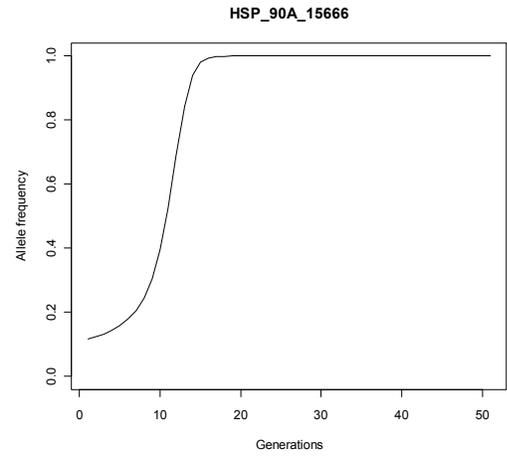
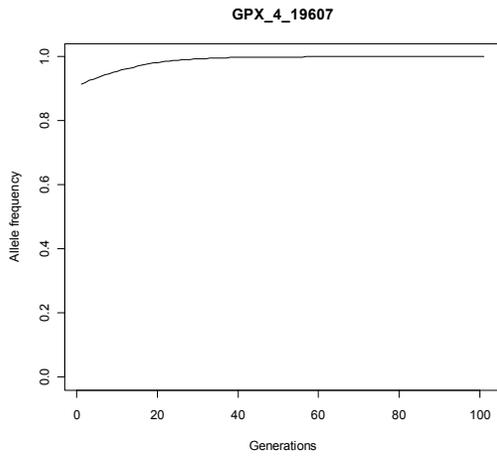
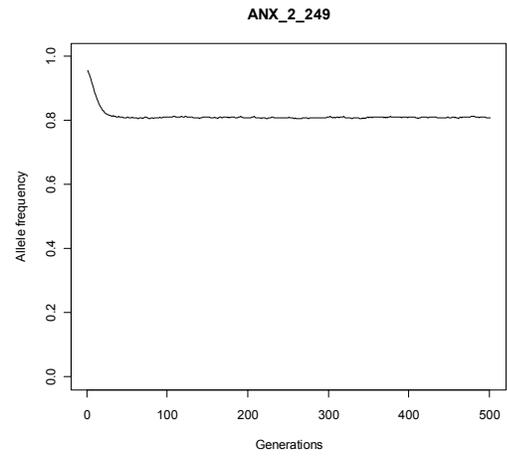
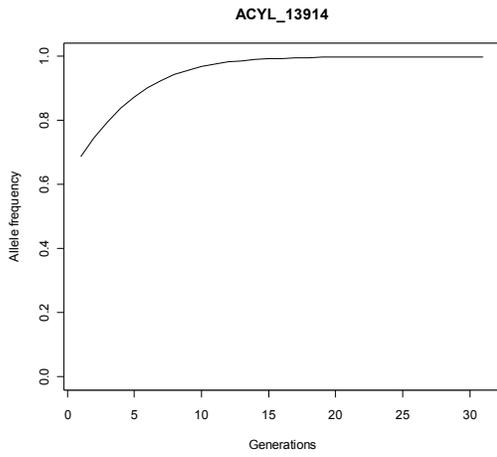


Figure S5 Simulated evolution of allelic frequencies at each of the 8 loci statistically associated with explanatory variables in the *GLASS8* category. Uniform contribution among niches and a population effective size of 10^5 are assumed.

Table S1 Locus name, gene annotation, contig length, BLASTX e-value and variation type for each of the 73 genotyped SNPs.

Locus	Gene	Contig length	Blastx e-value	SNP
40S_S18_1401	40S ribosomal protein S18	413	3.00E-73	A/G
60S_L10A_21874	60S ribosomal protein L10a	437	3.00E-48	A/G
60S_L27A_12416	60S ribosomal protein L27a	890	3.00E-52	C/T
ACTB_21752	Beta-actin	1005	1.00E-170	C/T
ACT_A3B_8646	Actinin alpha 3b	838	6.00E-150	C/T
ACP_13914	Acyl carrier protein, mitochondrial precursor	718	5.00E-57	A/G
ADH_3_533	Alcohol dehydrogenase class-3	707	3.00E-19	C/T
ADSS_L1_15447	Adenylosuccinate synthetase isozyme 1	325	2.00E-42	C/T
ALDH_2_16634	Aldehyde dehydrogenase 2	567	7.00E-91	C/T
ALD_R_2449	Aldose reductase	362	3.00E-48	A/C
ANK_R_13478	Ankyrin repeat domain-containing protein 1	1212	3.00E-96	A/T
ANN_A11_16176	Annexin A11	1097	2.00E-82	C/G
ANX_2_249	Annexin A2-A	1082	2.00E-127	A/G
ARF_4_18099	ADP-ribosylation factor 4	455	4.00E-35	C/T
ATP_BC_212	ATP-binding cassette sub-family A member 1	1145	1.00E-121	C/G
BPNT_1_18778	3'(2'),5'-bisphosphate nucleotidase 1	331	3.00E-39	A/C
CLIC_5_10148	Chloride intracellular channel 5	390	2.00E-49	C/G
COI_17591	Cytochrome oxidase subunit I, mitochondrial	601	1.00E-62	C/T
COP_9_18132	26S protease regulatory subunit 7	569	7.00E-68	A/G
CSDE_1_11069	Cold shock domain-containing protein E1	1505	0	G/T
CSDE_1_19713	Cold shock domain-containing protein E1	1623	0	C/G
CST_21113	Cystatin precursor	722	2.00E-33	A/G
CYT_BC1_9061	Cytochrome b-c1 complex subunit 2, mitochondrial precursor	852	4.00E-95	G/T
EF2_10494	Translation elongation factor 2	2190	0	C/G
EF_1G_4796	Translation elongation factor 1 gamma	484	2.00E-65	C/T
EIF_3F_341	Translation initiation factor 3 subunit F	902	4.00E-137	C/T
EIF_3J_11587	Translation initiation factor 3 subunit J	890	8.00E-41	A/T
FER_H_20955	Ferritin heavy subunit	1152	7.00E-98	G/T
GAPDH_20355	Glyceraldehyde-3-phosphate dehydrogenase	1223	0	C/T
GDE1_2508	Glycerophosphocholine phosphodiesterase	1127	3.00E-147	A/G
GOG_B1_15792	Golgin subfamily B member 1	457	2.00E-29	A/G
GPX_4_19607	Glutathione peroxidase 4	895	1.00E-90	C/T
HMG_T_9973	High mobility group-T protein	1131	2.00E-80	A/G
HSPE_1_17854	10 kDa heat shock protein, mitochondrial	410	2.00E-44	A/G
HSP_90A_15666	Heat shock protein 90 alpha	1108	1.00E-160	C/T
HSP_90B_21100	Heat shock protein 90 beta	1131	2.00E-118	C/T

IF_RF2_19747	Interferon regulatory factor 2	334	2.00E-51	C/T
JAM_3_13916	Junctional adhesion molecule 3b	474	7.00E-58	C/T
KRT_12_20618	Keratin 12	884	2.00E-104	A/G
KRT_A_15738	Keratin alpha-like	847	7.00E-44	A/G
LDH_B_9441	Lactate dehydrogenase B	1528	1.00E-159	A/G
MDH_1393	Malate dehydrogenase, mitochondrial precursor	918	3.00E-56	C/G
MYH_14857	Superfast myosin heavy chain	1204	1.00E-107	C/T
NADH1_10_21119	NADH dehydrogenase 1 alpha subcomplex subunit 10	894	9.00E-88	C/T
NADH_4_21742	NADH dehydrogenase subunit 4, mitochondrial	1221	1.00E-141	A/G
NADH_5_17101	NADH dehydrogenase subunit 5, mitochondrial	2823	0	G/T
NCP_2_15547	Nucleolar complex protein 2	424	3.00E-62	A/G
NEX_19953	Nexilin	802	4.00E-33	C/T
NGD_21138	Neuroguidin	365	2.00E-35	G/T
NRAP_1541	Nebulin-related anchoring protein	1624	2.00E-145	C/T
PA2G4_2600	Proliferation-associated protein 2G4	1211	0	C/G
PFN_15113	Profilin-2	687	4.00E-16	G/T
PGD_18096	6-phosphogluconate dehydrogenase, decarboxylating	486	6.00E-54	A/G
PGI_1	Phosphoglucose isomerase-1	927	9.00E-76	A/G
PGI_2_1862	Phosphoglucose isomerase-2	1953	1.00E-130	C/T
PGK_1_11454	Phosphoglycerate kinase 1	1136	1.00E-106	A/G
PRP_40_16504	Pre-mRNA-processing factor 40 homolog A	585	7.00E-29	C/G
PSA_4_21534	Proteasome subunit alpha type-4	978	2.00E-122	C/T
RFC_3_18186	Replication factor C subunit 3	359	2.00E-38	C/G
RTF_1_21288	RNA polymerase-associated protein RTF1 homolog	655	2.00E-38	G/T
SLC_25A5_19808	ADP/ATP translocase 2	991	1.00E-153	C/T
SM_22_6449	Transgelin	775	5.00E-112	C/T
SN4_TDR_374	Staphylococcal nuclease domain-containing protein 1	1248	8.00E-119	C/G
TENT_02_11046	No hit	884	-	G/T
TENT_03_12589	Collagen type XXVIII alpha 1 a	275	2.00E-26	C/T
TENT_05_19704	No hit	400	-	C/G
TENT_06_16512	Protein phosphatase regulatory subunit	410	2.00E-04	G/T
TRIM_35_8416	Tripartite motif-containing protein 35	771	1.00E-81	A/G
TTN_B_20952	Titin b	884	4.00E-127	C/T
TUB_A_19211	Tubulin alpha 2	733	5.00E-67	C/T
UBI_A52_5049	Ubiquitin A-52 residue ribosomal protein fusion product 1	715	5.00E-66	A/G
UGP_2_2128	UDP-glucose pyrophosphorylase 2	824	2.00E-129	C/G
ZETA_15177	Tyr 3-monooxygenase/Trp 5-monooxygenase activation protein	724	2.00E-54	A/G

Table S2 List of the KASPar primers used for the 73 individual SNP assays.

Locus	Primer Allele X	Primer Allele Y	Common Primer	Allele X	Allele Y
40S_S18_1401	CCAGAAGTGCCGGAGACCG	CCCAGAAGTGCCGGAGACCA	AGGATCTGGAGAGGCTGAAGAAGAT	G	A
60S_L10A_21874	CCCAGGCCCAAGTTCTCC	ACTCCCAGGCCCAAGTTCTCT	GCCTCATCACAGTGCTGCTGGT	G	A
60S_L27A_12416	GGAGCTTGTGTGCTMATGGCG	GGAGCTTGTGTGCTMATGGCA	TTTTAAACATTTATTCCTGCAAAAA	C	T
ACTB_21752	CACCATCGGCAACGAGCGC	ATCACCATCGGCAACGAGCGT	GGGCCTCGGGGCAACGGAA	C	T
ACT_A3B_8646	AACACTGCGTTGAGGTGGCT	CACTGCGTTGAGGTGGCC	ACATCTTGGGGATGTCCAGGTA	T	C
ACP_13914	ACATCGCAGACAAGAAAGACGTC	GTACATCGCAGACAAGAAAGACGTT	GTGGGGAGGCCCGTTATTCAT	G	A
ADH_3	AGGCAGAAGTCCAGTTGACTC	ATAAGGCAGAAGTCCAGTTGACTT	CCCCAATGACCTGACATGTATGAA	C	T
ADSS_L1_15447	GGAGACGGCTTTGGAGTTGATAATT	GAGACGGCTTTGGAGTTGATAATC	AGTGGTTGAYGGGAAGGAATATGACTT	T	C
ALDH_2_16634	GGCTACAAGGCCTCTGGGAAT	GGCTACAAGGCCTCTGGGAAC	CGGTGTAGTTGTCCAGGCCRTATT	T	C
ALD_R	GGACGCCGGGCTGGTCA	GGACGCCGGGCTGGTCC	CGGTTGAAGTTGGAGAYCCCGAT	A	C
ANK_R_13478	TTATTCTCACAGCGTCGTGCAGA	TATTCTCACAGCGTCGTGCAGT	CGCCAAAGACATGGACGGAGACA	A	T
ANN_A11_16176	CGTGTACAGAGACTTCTGGTAGG	CGTGTACAGAGACTTCTGGTAGC	GACATCCGCCAGGAGTAYGTGAA	G	C
ANX_2_249	GTAGCGCAATGGGGAGGGG	CAGTAGCGCAATGGGGAGGGG	GCTGTCTCTGTCACCCTGTCAT	G	A
ARF_4_18099	CACAGCCTCCCTCAGCTCG	GCACAGCCTCCCTCAGCTCA	RCTGCTGAAAATGCTGCAGGAGGA	C	T
ATP_BC_259	GAAGGAAGCCGTGGTGGACC	GAAGGAAGCCGTGGTGGACG	TGGCCTTCTCGTCTGCAGGAA	G	C
BPNT_1_18778	ACATCCACGGGAAGACGTACC	GACATCCACGGGAAGACGTACA	TGTGCTTGACGTCCGCGTCGTA	C	A
CLIC_5_10148	CGTGCGCRTTCTGCAGGTAC	CGTGCGCRTTCTGCAGGTAG	ACTTCGAGATCCCGCGGAGAT	C	G
COI_17591	GCCGGAGCATCTGTTGACCTG	GCCGGAGCATCTGTTGACCTA	GAGATACCTGCAAGGTGAAGTGAGAA	C	T
COP_9_18132	GATACATGACCTACAACCTAAGCAACG	AGATACATGACCTACAACCTAAGCAACA	AAAGAAAACGTATAACCGAGCRCAATCAAAA	G	A
CSDE_1_11069	GGCAGGATGTCGATGTTGGTG	GGGCAGGATGTCGATGTTGGTT	CCACCGACCGGCGGGACAA	G	T
CSDE_1_19713	CCATGTAGAAGCTGACTTTGTCTG	CCATGTAGAAGCTGACTTTGTCC	CCTGACGTACACGTCCGAGGAT	C	G
CST_21113	GTCACCTTGTGACGAACAAAGCA	CACCTTGTGACGAACAAAGCG	CGTCCAGGTCAACCTCAAGAGCAA	A	G
CYT_BC1_9061	GGGAGATGGTGTACTCTCAAAC	GGGAGATGGTGTACTCTCAAAA	CAGTGCATCTGCATTCAATGCCAGTT	G	T
EF2_10494	CCGGACTCCTCGCTCACC	CCGGACTCCTCGCTCACG	AATCGGACCGGTGGTGTCTTA	C	G
EF_1G_4796	GATGCAGGTCACAAACCAGCGT	ATGCAGGTCACAAACCAGCGC	CATTCGTCAGCCCTTCCCAA	T	C

EIF_3F_341	GAGCAGAAGCTCACTTATATGTTTCTT	AGCAGAAGCTCACTTATATGTTTCTC	ACTTCACGCTTTTATTCAGTAGAGCCATA	T	C
EIF_3J_11587	AATGCACAGATCGTTCGTA CTGAGA	AATGCACAGATCGTTCGTA CTGAGT	CACGGATTTGCATAAGTTGCTAAAGAGAA	T	A
FER_H_20955	CGCCTGCCCTGGACGGAA	CGCCTGCCCTGGACGGAC	GTGACGTCCCCGCRAAGCCAT	T	G
GAPDH_20355	CCACCCACGGACGTTCCAC	CCACCCACGGACGTTCCAT	GAGCTTTCCGCCTTCGGCCTT	C	T
GDE1_2508	GGAGCGGGTGCTTCAGAGC	GGGAGCGGGTGCTTCAGAGT	CCCACCAGCCCRGGGAGAA	G	A
GOG_B1_15792	CAAATCCCCTACCTGTTCTCA	AAATCCCCTACCTGTTCTCG	GASAACGTGTAACGCCAGCAGGTA	A	G
GPX_4_19607	CACTTGGGTCTGCAGTGCC	CACTTGGGTCTGCAGTGGT	CAGAGAGGGAAAGTTGTGAAAAGATATT	C	T
HMG_T_9973	GCAAGTTCGAGGACTTGGCGAAA	CAAGTTCGAGGACTTGGCGAAG	TGTAGTCTTCATCTCCCGCTCGTA	A	G
HSPE_1_17854	TACCCGAATATGGAGGAACTAAAGTT	ACCCGAATATGGAGGAACTAAAGTC	GGAAGTAGTCTTGTCTCCAGAAT	A	G
HSP_90A_15666	GTTCTTTCCCTCTTTTGTGCTCTCA	CTTTCCCTCTTTTGTGCTCTCG	CCCCAGATTGAGGAYGTGGGAT	T	C
HSP_90B_21100	GGAGAAGGAGAAAGAGGAGGAC	AGGAGAAGGAGAAAGAGGAGGAT	CCACGTCTCGATCTTGGGCTT	C	T
IF_RF2_19747	GTGAATGGCCAGTTTCGGAAC	GTGAATGGCCAGTTTCGGAAT	GACGTGGAGAAAGACGCACCYTT	C	T
JAM_3_13916	AAGTAGCCCCGTTTATACGCACAA	GTAGCCCCGTTTATACGCACAG	GTGGTGGTGGTTCTCTGTGTATAA	T	C
KRT_12_20618	GGAACCAGTTCTCCATTCTTC	GGAACCAGTTCTCCATTCTTT	CAGTACGAGGGAATGGCCGACAA	G	A
KRT_A_15738	CCCTTCTCCACTTTGCCAC	CCCTTCTCCACTTTGCCAT	CTCATCCCACTGGGAGACCAGAA	G	A
LDH_B_9441	ACATCCATCACATCTACACTCCC	CTACATCCATCACATCTACACTCCT	AGTGTAGCAGTAAMGGTAGAACAGAGT	G	A
MDH_1393	AATTAATGAGGTCTTAATTTGGCTGCC	AATTAATGAGGTCTTAATTTGGCTGCG	GAGGACTTTGTACAGAACATGAAATGAGAA	C	G
MYH_14857	TTGGTTCTGTGTTTTGAGCTCA	CTTTGGTTCTGTGTTTTGAGCTCG	GGTGGAGGTCTCTGAGAGACACAA	T	C
NADH1_10_21119	GCACAATCAATGAGATATAACTTATTTGTTAAT	CACAATCAATGAGATATAACTTATTTGTTAAC	GYTTTCCTTTCTCTGATCCTGTTCTA	T	C
NADH_4_21742	GATGATACTAGTCCGTGGGCAATT	ATGATACTAGTCCGTGGGCAATC	CCCATGAGGATTYACAGGAGCAATTAT	A	G
NADH_5_17101	CAACTATTCATTGGCTGAGAAGGC	TCAACTATTCATTGGCTGAGAAGGA	CGTATCATCAYCCAATTAGGAGAAATGATA	G	T
NCP_2_15547	GATCTGCCGGCACAAGCAGG	AAGATCTGCCGGCACAAGCAGA	GCTTGAGGACGGGGTTGAGGTA	G	A
NEX_19953	CCTCTCCAGTCGCCGC	CTCTCTCCAGTCGCCGT	GAGGGCTGAAGAGGAGGCCAA	C	T
NGD_21138	ATGACATTCCCAGCATGCCTCG	GATGACATTCCCAGCATGCCTCT	TCGCCGAGAGAAAAGTAAGCGGAA	G	T
NRAP_1541	CTTGGCGTGCTGGAAGTCA	CTCTTGGCGTGCTGGAAGTGC	CCAGGAGACCTGCTCYATTACAA	T	C
PA2G4_2600	AAACTCACCTCCTTCTCGTG	AAACTCACCTCCTTCTCGTGC	ACGAGCTGCTGCAGCCCTTCAA	C	G
PFN_15113	CAAACAGGCTATCGCGGTCATG	GCAAACAGGCTATCGCGGTCATT	AGAGGAGGTCCAGGTGCTGCA	G	T
PGD_18096	CCCGGTTTCGACAGAAGCTCA	CCGGGTTTCGACAGAAGCTCG	GGGATTATTTGGYGCGCACACCTA	A	G

PGI_1	CCTCCACCAACGGGCTCATCA	CTCCACCAACGGGCTCATCG	TGGAGCTCARGCGTAGTTTTCTTGAT	A	G
PGI_2_1862	ACATGCACCGCTTTGCTGCC	GTACATGCACCGCTTTGCTGCT	GATTCATGTCTCCCTGCTGGAAAT	C	T
PGK_1_11454	GGGCGAAGTTGTCCCACTCA	GGGCGAAGTTGTCCCACTCG	ACGGGCCGGTCGGCGTGTT	A	G
PRP_40_16504	AGAGTGACGTCTCAAAGGCC	AGAGTGACGTCTCAAAGGCC	TGCGGGAGAGATTCTGAAGGAAT	G	C
PSA_4_21534	GGAGGAGGGACACCAGAGGA	GAGGAGGGACACCAGAGGG	GGACAAGAAGGAGAAGGAGCAGAAA	T	C
RFC_3_18186	AAGAAAACGCGGATCATGTGTCTG	AAGAAAACGCGGATCATGTGTCTC	CGGGTCCGTACAGCTCGCGTA	G	C
RTF_1_21288	GATGGGGGACGAGGAGGACA	ATGGGGGACGAGGAGGACC	GCTCCCTCTCCTTCTCCGTCAT	T	G
SLC_25A5_19808	ACAGTGTCGAAGGGGTAGGAC	GACAGTGTCGAAGGGGTAGGAT	CTCACATTGTGGTCAGCTGGATGAT	C	T
SM_22_6449	AGAGGGGCGTCCAAGCG	CAGAGGGGCGTCCAAGCA	GGGGTCGRCCATATCCTGTCAAT	C	T
SN4_TDR_374	GAACTCGTCGGCGTCGTCC	GAACTCGTCGGCGTCGTCCG	GAACATCTGGCGTTACGGRGACTT	C	G
TENT_02_11046	CGAGAAAATCACATGACCTGTCC	GTCGAGAAAATCACATGACCTGTCA	ACACAAACATGTTCTCAGCAAACATKCT	G	T
TENT_03_12589	GTCTGGCCACTCGTCCG	GTCTGTCTGGCCACTCGTCCA	AAGGGARGCGCACTTCTGTGAAA	C	T
TENT_05_19704	CCTTGTGAATCGCTCGACGCC	CCTTGTGAATCGCTCGACGCG	GTTAGCTCAGTGTGTGAGAGCAGAA	G	C
TENT_06_16512	TCTGTTTTGGTCATTTCCCTGATTTT	CTTCTGTTTTGGTCATTTCCCTGATTTA	CTGTCAATCAACGTGATTATGCACCTTAT	G	T
TRIM_35_8416	GKTTGAGATCCCCGCACTACTA	GAGATCCCCGCACTACTG	GGCACTGGCACTGCCARAGTA	A	G
TTN_B_20952	ACATAATAATCCAAACCTTCATGAACAAC	GTACATAATAATCCAAACCTTCATGAACAAT	CCTTTAGCATTTGGACCTCAGATTGAA	C	T
TUB_A_19211	GGGATCACTTTGTCTGTCTTTCAA	GGATCACTTTGTCTGTCTTTCAAG	CCCTTGAAGCAGGATAGTATAAGACATTT	T	C
UBI_A52_5049	GGGAGGGCTCAATGATCCCG	AGGGAGGGCTCAATGATCCCA	AGAGTCCACCTCCACCTGGTA	G	A
UGP_2_2128	CCCGACACAATCTTGTCTCCAA	CCCGACACAATCTTGTCTCCAG	TCGACATTCCGGCAGGAGCCAT	T	C
ZETA_15177	AGGAATGGAGAAGGTAAAAAGGTTTC	AATAGGAATGGAGAAGGTAAAAAGGTTTT	GCACACCGACTCYTCTGCCAAA	G	A

Table S3 Results of within-category tests for Hardy-Weinberg Equilibrium, for each of the 70 diploid loci. Significant *p*-values

after correcting for False Discovery Rate are shown in bold.

Locus Name	p-value	p-value	p-value	p-value
	<i>SAR7</i>	<i>GLASS8</i>	<i>OY09</i>	<i>GLASS9</i>
40S_S18_1401	1.00000	0.12154	0.22574	0.18342
60S_L10A_21874	1.00000	1.00000	1.00000	1.00000
60S_L27A_12416	1.00000	0.33219	0.73350	0.28553
ACTB_21752	1.00000	0.91921	0.01548	0.43163
ACT_A3B_8646	1.00000	0.21700	0.36195	1.00000
ACP_13914	1.00000	0.13962	0.33264	0.09231
ADH_3	1.00000	0.00088	0.00003	0.00700
ADSS_L1_15447	1.00000	0.01291	0.69547	1.00000
ALDH_2_16634	0.76938	0.37673	0.43921	0.42038
ALD_R	1.00000	0.12263	0.52380	0.44679
ANK_R_13478	1.00000	0.10961	0.27621	0.78760
ANN_A11_16176	1.00000	1.00000	1.00000	1.00000
ANX_2_249	0.06563	0.00000	0.00052	0.00000
ARF_4_18099	0.40586	0.18323	0.47020	1.00000
ATP_BC_259	0.22725	0.77055	0.31773	0.11295
BPNT_1_18778	0.59720	0.22555	0.13523	0.55784
CLIC_5_10148	0.18338	0.10316	0.76834	1.00000
COP_9_18132	1.00000	0.00743	0.32456	0.01337
CSDE_1_11069	1.00000	1.00000	0.76649	0.75550
CSDE_1_19713	0.14806	0.13634	0.02794	0.85559
CST_21113	0.77052	0.30015	0.72836	0.28627
CYT_BC1_9061	0.02557	0.27349	0.59605	0.47821
EF2_10494	1.00000	0.00251	0.59933	0.36706
EF_1G_4796	0.04617	0.42028	0.80844	1.00000
EIF_3F_341	0.42879	0.00000	0.00000	0.00000
EIF_3J_11587	1.00000	0.00000	0.00107	0.06416
FER_H_20955	0.77235	0.73993	0.30569	0.22289
GAPDH_20355	1.00000	0.05372	0.78629	0.80586
GDE1_2508	1.00000	0.72005	1.00000	0.23356
GOG_B1_15792	1.00000	0.14666	1.00000	0.64949
GPX_4_19607	1.00000	0.07780	1.00000	0.61542
HMG_T_9973	1.00000	0.13886	0.15097	1.00000
HSPE_1_17854	1.00000	0.22323	0.00423	0.26826
HSP_90A_15666	1.00000	0.02031	0.13498	0.36947

HSP_90B_21100	1.00000	0.00000	0.03145	0.00377
IF_RF2_19747	1.00000	0.37954	0.32440	1.00000
JAM_3_13916	1.00000	0.26091	0.45007	1.00000
KRT_13_20618	0.55548	0.55728	0.73019	0.87970
KRT_A_15738	1.00000	0.34404	0.26751	0.21900
LDH_B_9441	0.17316	0.34342	0.28692	0.68452
MDH_1393	0.01533	0.00000	0.02062	0.00004
MYH_14857	1.00000	0.59054	0.13159	0.33182
NADH1_10_21119	0.55818	0.32853	0.85564	0.87103
NCP_2_15547	0.25037	0.43750	0.25974	0.29391
NEX_19953	1.00000	0.00001	0.50013	0.05036
NGD_21138	0.51549	0.93132	0.48016	0.09074
NRAP_1541	0.56025	0.86245	0.47515	1.00000
PA2G4_2600	1.00000	0.03592	0.82994	0.05108
PFN_15113	0.59641	0.06795	1.00000	0.20738
PGD_18096	1.00000	0.27249	1.00000	1.00000
PGI_1	1.00000	0.29650	1.00000	0.46539
PGI_2_1862	1.00000	0.60288	0.60697	0.28459
PGK_1_11454	0.72955	0.50605	0.66184	1.00000
PRP_40_16504	1.00000	0.00000	0.76832	0.00297
PSA_4_21534	0.35604	0.09737	1.00000	0.63642
RFC_3_18186	1.00000	0.00000	0.00014	0.00000
RTF_1_21288	1.00000	0.50433	0.86476	0.35549
SLC_25A5_19808	1.00000	1.00000	0.61536	0.12976
SM_22_6449	0.24624	0.21508	0.61504	0.64701
SN4_TDR_374	0.25699	0.74257	1.00000	0.88096
TENT_02_11046	1.00000	0.00009	0.28272	0.62585
TENT_03_12589	0.24050	0.10899	1.00000	0.07961
TENT_05_19704	0.00022	0.00000	0.00146	0.00012
TENT_06_16512	1.00000	0.11943	0.70400	1.00000
TRIM_35_8416	0.14543	0.00034	0.22324	0.00645
TTN_B_20952	1.00000	0.81825	1.00000	0.39884
TUB_A_19211	0.76482	0.67939	0.85889	1.00000
UBI_A52_5049	0.21575	0.30265	0.53210	0.13803
UGP_2_2128	0.14211	0.00000	0.06281	0.00036
ZETA_15177	0.09160	0.59273	1.00000	1.00000

Table S4 Simulated evolution of allelic frequencies at the 8 loci statistically associated with explanatory variables in the *GLASS8* category under different assumptions. Models of relative contribution among niches are uniform contribution (C_UNI), normal distribution centered on the 8th and 9th niches with a variance of 10 (C_N_10), 6 (C_N_6), 4 (C_N_4), 3 (C_N_3), 2 (C_N_2) and 1 (C_N_1). The three population effective sizes considered were 10^4 , 10^5 and 10^6 . Evolution of allelic frequencies was simulated using relative fitness values either predicted from modeled or observed allelic frequencies.

Locus	Allele frequencies to estimate relative fitness	Population size	Models of relative niche contributions					
			C_UNI	C_N_10	C_N_6	C_N_4	C_N_3	C_N_2
ACP_13914	Modeled	10000	Fix. < 30 G	Fix. < 30 G	Fix. < 30 G	Fix. < 30 G	Fix. < 30 G	Fix. < 40 G
	Modeled	100000	Fix. < 20 G	Fix. < 20 G	Fix. < 30 G			
	Modeled	1000000	Fix. < 20 G	Fix. < 20 G	Fix. < 30 G			
	Observed	10000	Fix. < 30 G	Fix. < 30 G	Fix. < 30 G	Fix. < 30 G	Fix. < 30 G	Fix. < 30 G
	Observed	100000	Fix. < 20 G	Fix. < 30 G	Fix. < 30 G	Fix. < 30 G	Fix. < 30 G	Fix. < 30 G
	Observed	1000000	Fix. < 20 G	Fix. < 20 G	Fix. < 30 G			
ANX_2_249				Equ.				
	Modeled	10000	Equ. 0.81	0.805	Equ. 0.80	Equ. 0.79	Equ. 0.78	Equ. 0.77
				Equ.				
	Modeled	100000	Equ. 0.81	0.805	Equ. 0.80	Equ. 0.79	Equ. 0.78	Equ. 0.77
				Equ.				
	Modeled	1000000	Equ. 0.81	0.805	Equ. 0.80	Equ. 0.79	Equ. 0.78	Equ. 0.77
	Observed	10000	Equ. 0.68	Equ. 0.68	Equ. 0.68	Equ. 0.68	Equ. 0.68	Equ. 0.69
Observed	100000	Equ. 0.68	Equ. 0.68	Equ. 0.68	Equ. 0.68	Equ. 0.68	Equ. 0.69	
Observed	1000000	Equ. 0.68	Equ. 0.68	Equ. 0.68	Equ. 0.68	Equ. 0.68	Equ. 0.69	
GPX_4_19607	Modeled	10000	Fix. < 90 G	Fix. < 80 G	Fix. < 60 G	Fix. < 50 G	Fix. < 40 G	Fix. < 30 G
	Modeled	100000	Fix. < 80 G	Fix. < 70 G	Fix. < 50 G	Fix. < 40 G	Fix. < 30 G	Fix. < 30 G
	Modeled	1000000	Fix. < 70 G	Fix. < 60 G	Fix. < 50 G	Fix. < 40 G	Fix. < 30 G	Fix. < 30 G
	Observed	10000	Fix. < 90 G	Fix. < 80 G	Fix. < 50 G	Fix. < 50 G	Fix. < 40 G	Fix. < 40 G
	Observed	100000	Fix. < 70 G	Fix. < 70 G	Fix. < 50 G	Fix. < 40 G	Fix. < 40 G	Fix. < 30 G
	Observed	1000000	Fix. < 70 G	Fix. < 60 G	Fix. < 50 G	Fix. < 40 G	Fix. < 30 G	Fix. < 30 G
HSP_90A_15666	Modeled	10000	Fix. < 30 G	Fix. < 20 G	Fix. < 20 G	Fix. < 20 G	Fix. < 20 G	Fix. < 20 G

	Modeled	100000	Fix. < 20 G					
	Modeled	1000000	Fix. < 20 G					
			Equ.	Equ.			Equ.	
	Observed	10000	0.755	0.755	Equ. 0.76	Equ. 0.78	0.805	Equ. 0.85
			Equ.	Equ.			Equ.	
	Observed	100000	0.755	0.755	Equ. 0.76	Equ. 0.78	0.805	Equ. 0.85
			Equ.	Equ.			Equ.	
	Observed	1000000	0.755	0.755	Equ. 0.76	Equ. 0.78	0.805	Equ. 0.85
			Equ.	Equ.			Equ.	
MDH_1393	Modeled	10000	0.375	0.365	Equ. 0.35	Equ. 0.33	0.305	Equ. 0.28
			Equ.	Equ.			Equ.	
	Modeled	100000	0.375	0.365	Equ. 0.35	Equ. 0.33	0.305	Equ. 0.28
			Equ.	Equ.			Equ.	
	Modeled	1000000	0.375	0.365	Equ. 0.35	Equ. 0.33	0.305	Equ. 0.28
					Equ.			Equ.
	Observed	10000	Equ. 0.37	Equ. 0.36	0.345	Equ. 0.33	Equ. 0.31	0.305
					Equ.			Equ.
	Observed	100000	Equ. 0.37	Equ. 0.36	0.345	Equ. 0.33	Equ. 0.31	0.305
					Equ.			Equ.
	Observed	1000000	Equ. 0.37	Equ. 0.36	0.345	Equ. 0.33	Equ. 0.31	0.305
NRAP_1541							Fix. < 100	Fix. < 120
	Modeled	10000	Fix. < 50 G	Fix. < 60 G	Fix. < 60 G	Fix. < 70 G	G	G
								Fix. < 110
	Modeled	100000	Fix. < 50 G	Fix. < 50 G	Fix. < 60 G	Fix. < 70 G	Fix. < 90 G	G
								Fix. < 110
	Modeled	1000000	Fix. < 50 G	Fix. < 50 G	Fix. < 60 G	Fix. < 70 G	Fix. < 80 G	G
						Fix. < 100		
	Observed	10000	Fix. < 60 G	Fix. < 70 G	Fix. < 80 G	G	Fix. < 90 G	Fix. < 60 G

PRP_40_16504	Observed	100000	Fix. < 60 G	Fix. < 60 G	Fix. < 70 G	Fix. < 90 G	Fix. < 80 G	Fix. < 60 G
	Observed	1000000	Fix. < 60 G	Fix. < 60 G	Fix. < 70 G	Fix. < 80 G	Fix. < 80 G	Fix. < 60 G
	Modeled	10000	Fix. < 20 G	Fix. < 20 G	Fix. < 20 G	Fix. < 20 G	Fix. < 20 G	Fix. < 20 G
	Modeled	100000	Fix. < 20 G	Fix. < 20 G	Fix. < 20 G	Fix. < 20 G	Fix. < 20 G	Fix. < 20 G
	Modeled	1000000	Fix. < 20 G	Fix. < 20 G	Fix. < 20 G	Fix. < 20 G	Fix. < 20 G	Fix. < 20 G
UGP_2_2128					Equ.	Equ.	Equ.	Equ.
	Observed	10000	Equ. 0.92	Equ. 0.93	0.935	0.945	0.955	0.975
					Equ.	Equ.	Equ.	Equ.
	Observed	100000	Equ. 0.92	Equ. 0.93	0.935	0.945	0.955	0.975
					Equ.	Equ.	Equ.	Equ.
	Observed	1000000	Equ. 0.92	Equ. 0.93	0.935	0.945	0.955	0.975
	Modeled	10000	Fix. < 30 G	Fix. < 30 G	Fix. < 30 G	Fix. < 30 G	Fix. < 30 G	Fix. < 30 G
	Modeled	100000	Fix. < 30 G	Fix. < 30 G	Fix. < 30 G	Fix. < 30 G	Fix. < 30 G	Fix. < 30 G
	Modeled	1000000	Fix. < 20 G	Fix. < 20 G	Fix. < 30 G	Fix. < 30 G	Fix. < 30 G	Fix. < 30 G
	Observed	10000	Fix. < 30 G	Fix. < 30 G	Fix. < 30 G	Fix. < 30 G	Fix. < 30 G	Fix. < 20 G
Observed	100000	Fix. < 30 G	Fix. < 30 G	Fix. < 30 G	Fix. < 30 G	Fix. < 20 G	Fix. < 20 G	
Observed	1000000	Fix. < 30 G	Fix. < 30 G	Fix. < 30 G	Fix. < 30 G	Fix. < 20 G	Fix. < 20 G	
