Online supporting information

Divergent selection, reproductive isolation, and genomic divergence

We focused on heterogeneous genomic divergence during the process of population differentiation and speciation. This requires considering the relationship between selection and reproductive isolation. Loci under divergent selection and loci causing reproductive isolation are similar in exhibiting reduced introgression (and thus greater divergence) between populations relative to other loci (Barton 1979, 1983; Barton and Hewitt 1989; Mallet 1995, 2005, 2006; Wu 2001; Wu and Ting 2004; Nosil et al. 2005). Indeed, an allele 'a' that confers a poor fit of the phenotype to the environment can be selected against, contributing to population divergence, whether the afflicted allele resides in the parental species (e.g., homozygote 'aa') or in a hybrid individual (heterozygote 'Aa'). Yet, typically, the former scenario would be considered a case of selection and the latter an example of postmating reproductive isolation. Recognizing that selection against immigrants itself represents a form of reproductive isolation (Nosil et al. 2005), and that hybrid inviability is a manifestation of selection, helps clarify the relatedness of these concepts. Moreover, the study of gene regions differentiating under selection becomes inseparable from that of gene regions causing reproductive isolation when adaptively relevant loci pleiotropically promote reproductive isolation (Muller 1942; Funk 1998; Bradshaw and Schemske 2003; Rundle and Nosil 2005), or when selection drives the population divergence of genes causing genetic incompatibilities between populations (i.e., intrinsic postmating isolation, Presgraves et al. 2003; Orr et al. 2004; Wu and Ting 2004; Dettman et al. 2007). This may also be true when genomic regions of divergence contain loci affecting phenotypic traits under selection as well as loci affecting forms of reproductive isolation such as hybrid inviability (Noor et al. 2001; Rieseberg 2001). For simplicity, the present paper focuses on evaluating genomic differentiation in terms of divergent selection per se. Thus, selection and reproductive isolation are treated conceptually together, while recognizing that: (1) they might act at different stages in life history, with consequences for heterogeneous genomic divergence, and (2) some forms of reproductive isolation will evolve due to processes other than divergent selection, such as genetic drift.

More detailed summary of the general genetic barrier to neutral gene flow caused by selection

We draw heavily on a summary by Gavrilets (2004 p. 147-148) based on his own work and that of Barton and colleagues (e.g., Bengtsson 1985; Barton and Bengtsson 1986; Pialek and Barton 1997; Gavrilets and Cruzan 1998; Navarro and Barton 2003; Gavrilets and Vose 2005).

The basic scenario is one in which a population is subject to continuous immigration. Due to divergent local adaptation, immigrants have lower fitness than residents, yielding selection against immigrants. In this case, the spread of neutral alleles between immigrant and resident populations will be slowed, to some extent, by selection against both immigrants and subsequent immigrant-resident hybrids. In this fashion, selection against incoming locally adapted alleles will – through the death of immigrants or hybrids – act as a general, if partial, genetic barrier to the spread of neutral alleles between populations. To describe this effect, Bengtsson (1985) introduced the notion of the 'gene flow factor', η, defined as the probability

that a neutral allele carried by immigrants makes it into the local genetic background. The inverse of η is known as the 'strength of genetic barrier'; Barton and Bengtsson 1986; Pialek and Barton 1997). If the migration rate (i.e., the proportion of the local population replaced by immigrants each generation) is m, then with a genetic barrier the proportion of resident neutral alleles replaced by immigrant neutral alleles per generation is:

(1) $$m_e = \eta m$$

Thus, equation (1) defines the effective migration rate of neutral alleles. To better characterize η, now assume that immigrating adults differ from residents at two genes: a gene reducing the viability of F_1 hybrids to $1 - s$ (where the viability of residents is 1) and a neutral gene unlinked to the selected gene. Assuming random mating, the probability that the neutral allele makes it to the next generation is $(1 - s) / 2$. The probability that the allele survives to the next generation but remains associated with the deleterious allele is also $(1 - s) / 2$. After many generations, the probability of inclusion of the neutral allele into the local genetic background is (see Gavrilets 2004 for derivation):

(2) $$\eta = \frac{1-s}{1+s}$$

The effective immigration rate of neutral alleles is slowed even further if there is assortative mating (equation 4 in Gavrilets 2004, p. 148).

Table S1. Robustness of studies listed in Table 2 of the main text. 'Multiple methods' indicates whether more than one primary type of analysis was used to detect outliers (N = No; Y = Yes, one program/simulation method was run in different ways; Y+ = Yes, more than one program or statistical procedure was used). This column refers to the actual implementation of different programs, whereas other ways of confirming outlier status (e.g., replication of outlier status across different population pairs) are deal with in subsequent columns. 'Quantiles' indicates the threshold of expected neutral differentiation used to determine whether a locus was an outlier. 'Replication across population pairs' indicates whether the replication of outlier status across multiple population pairs was evaluated (N = No; Y = Yes; Y+, direct = Yes, and the study also explicitly evaluated if outliers were associated with a specific ecological variable, for example by being outliers only in comparisons between population pairs that differ in that variable, and never outliers in population pairs similar for that variable; Y+, indirect = same as Y+, but although an association of outliers with an ecological variable was evident in the data, it was not explicitly noted). 'Type I error' indicates whether this type of error was accounted for (N = No; Y = Yes, by correcting for the number of loci within a comparison (e.g., via Bonferroni correction); Y+ = Yes, via additional consideration of the number of population pairs in which the locus was an outlier). 'Mutation rate variation' refers to whether this possible confounding factor was discussed (N = No; Y = Yes, by arguing that gene flow negates the effects of mutation rate variation, or by comparing differentiation between regions known to differ in mutation rate; Y+ = Yes, by evaluating outliers that were associated with a specific ecological variable, a pattern unlikely to arise via mutation rate variation. We note here only studies that actually discussed this issue, but any study examining parallel divergence, and particularly those that noted associations between outliers and ecological variables, indirectly argue that mutation rate variation to be an unlikely cause of outlier behavior). Past studies suggest that divergence-based methods for detecting divergent selection are robust to demographic variability (Beaumont and Balding 1996), but we further note whether the potential confounding effects of 'demography' were discussed. 'Background selection' refers to whether it was considered. When the discussion of a particular factor was particularly explicit, we note the relevant page number.

Study system from Table 2	Multiple methods	quantiles	replication across population pairs	Type I error	Mutation rate variation	Demography	Background selection
1.	Y+	0.99	Y+, indirect	Y+, p. 616	N	N	N
2.	N	0.95	Y	Y+	N	N	N
3.	Y, Y+, p. 1069	0.99	Y	Y+, p. 1074	Y+	N	N
4.	Y, Y+, p. 775	0.95, 0.99	Y+, indirect	Y+	N	Y, p. 775	N
5.	Y+	0.95, 0.99	Y+, direct	Y	N	N	N
6.	N	0.95	Y+, direct	N	Y+, p. 98	N	N
7.	Y+, p. 2396	0.99	Y+, direct	Y, p. 2401	N	N	N
8.	Y, Y+	0.95, 0.99	Y+, direct, p. 321	Y+, p. 323	Y,Y+, p. 322	N	N
9.	Y, Y+	0.95, 0.99	Y+, direct	Y+, p. 1167	Y,Y+, p. 1167	N	N

10.	Y+	0.95	Y+, direct	Y	N	N	N
11.	Y	0.95	Y+, indirect	N	N	N	N
12.	Y+	0.95, 0.99	Y+, direct	Y+, p.5163	N	Y, p.5160	
13.	N	0.95	N	N	N	N	N
14.	N	0.95	Y	N	N	N	N
15.	Y+	0.95	Y	Y, p. e285	N	Y, p. e285	Y, p. e285
16.	Y+	0.985	Y	N	Y, p. 734	N	N
17.	Y	0.99	N	Y	N	N	N
18.	Y,Y+, p.3473	0.99	Y+, direct	Y+	N	N	N
19.	N	0.95	N	N	N	N	N
20.	Y	0.05	Y	Y	N	N	N

Table S2. Summary of studies providing information on Isolation-by-Adaptation (IBA), where IBA refers to a positive association between the degree of adaptive phenotypic (or ecological) divergence and the level of genetic differentiation (here, at putatively neutral genetic markers). Provided is the study system, a description of the taxa or ecological forms examined, the type and number of markers used, the study design, and the main result with respect to IBA and IBD (where applicable). When reporting findings for IBA, the results reported are independent of geographic distance. Three main types of studies are reported: 1) population genetic studies explicitly examining IBA, generally using distance matrices of adaptive and neutral genetic divergence, 2) population genetic studies using an AMOVA framework (Excoffier et al. 1992), where we report the percent of total genetic variation observed between ecological types versus that observed among populations within ecological types, and 3) illustrative examples of a pattern consistent with IBA in a mosaic hybrid zone and in a phylogeographic study.

Organism	Divergent forms	Marker	Study design	IBA results	IBD results	Reference
Population genetic studies generally using distance matrices						
1. *Isoodon obesulus* (brown bandicoot)	gradient in rainfall and swamp vs. forest habitat types	39 RAPDS	genetic distance was related to habitat divergence and to geographic distance using 36 population pairs	IBA detected for both annual rainfall and habitat type	IBD not detected	Cooper 2000
2. *Anolis roquet* (Carribean lizard)	gradient from xeric coastal woodland to transitional forest to montane rainforest	7 microsatellite loci	genetic distance was compared among pairs of adjacent localities from three different transects, one of which cut through the ecological gradient (‘habitat transect’) and two of which did not (seven to ten localities per transect)	IBA detected, strong genetic differentiation observed only in the habitat transect, at habitat boundaries; population structuring by habitat further supported by AMOVA	IBD not detected	Ogden and Thorpe 2002
3. *Poecilia reticulate* (guppies)	high vs. low predation habitats	7 microsatellite loci	genetic distance was related to ecology (predation regime)	no evidence for IBA	IBD detected (and also an effect of the biogeographical	Crispo et al. 2006

			and geography (distance, waterfalls) using from 54 to 190 population pairs		barrier of waterfalls)	
4. *Canis lupus* (grey wolf in Europe)	various ecological factors	14 microsatellite loci, mtDNA	genetic distance was related to numerous ecological variables and to distance, using from 16 to 59 populations	IBA detected, genetic differentiation among local populations was correlated with climate, habitat type, and wolf diet composition.	IBD detected, topographic barriers nor past fragmentation could explain spatial genetic structure	Pilot et al. 2006
5. *Coregonus clupeaformis* (freshwater fish ecotypes)	dwarf vs. normal lake ecotypes (but quantitative indices of morphological divergence analyzed)	6 microsatellite loci	the correlation between genetic distance and adaptive morphological differentiation between six sympatric pairs was examined (thus there is no geographic distance between any of the six pairs)	IBA detected when results are pooled across loci, 5 of 6 loci exhibit fairly strong evidence of IBA individually (r = 0.72-0.84)	N/A	Lu and Bernatchez 1999
6. *Timema cristinae* (herbivorous insect)	*Adenostoma* and *Ceanothus* host plant ecotypes (but quantitative indices of adaptive divergence analyzed)	209 AFLP loci, mtDNA	genetic distance was related to quantitative indices of host-associated adaptive divergence and to geographic distance using 15 population	IBA not significant when AFLP loci were pooled; 10% of putatively neutral (i.e., non-outlier)	weak IBD detected, stronger for mtDNA than for AFLPs	Nosil et al. 2008

			pairs	individual AFLP loci show significant IBA; mtDNA shows IBA		
7. *Dubautia arborea* and *D. ciliolate* (Hawaiian silverswords)	gradient in leaf characteristics	7 microsatellite loci	using two species, genetic distance between ten population pairs was related to indices of morphometric divergence in leaf traits and to geographic distance	IBA detected	IBD not detected	Friar et al. 2007
8. *Littorina saxatilis* (intertidal snails)	upper vs. lower shore ecotypes	275 AFLP loci	genetic distance was related to ecology (shore ecotype) and geographic distance, at two different shores	IBA detected, for a given geographic distance, stronger differentiation between ecotypes relative to between samples within ecotypes	IBD not detected	Grahame et al. 2006
9. *Geum urbanum* (forest herb)	herb-layer community similarity	6 microsatellite loci	genetic distance was related to ecology (herb community) and geographic distance using 18 populations (153 pairs)	IBA not detected	IBD not detected	Vandepitte et al. 2007
10. *Canis lupus* (grey wolf in	tundra vs. forest types	14 microsatellite loci	genetic distance was related to habitat type (tundra, taiga or	IBA detected	IBD detected	Musiani et al. 2007

North America)			boreal coniferous forest) and geographic distance using 11 population groupings			
11. *Hordeum spontaneum* (wild barley)	basalt and terra rossa soil types	117 RAPD loci	genetic distance was related to habitat type (basalt and terra rossa soil types) and transect position, using two topographically separated transects	44% of loci exhibit IBA	36% of loci exhibit IBD (i.e., a correlation with transect)	Owuor et al. 1999
12. *Canis lupus* (grey wolf in North America)	various ecological variables including annual temperature, rainfall, vegetation, behavior and species of primary prey	15 microsatellite loci	genetic variation was related to a variety of ecological variables, to water barriers to gene flow, and to spatial position	IBA detected, at least for some ecological variables	not explicitly tested, although there were spatial components to genetic structure	Carmichael et al. 2007
13. *Alopex lagopus* (arctic foxes)	NA	13 microsatellite loci	genetic structure was analyzed in STRUCTURE, due to the detection of only a single genetic cluster, further matrix analyses were not conducted	no evidence for IBA, across a broad geographic area, only a single genetic cluster was detected	no evidence for IBD	Carmichael et al. 2007
14. *Zostera marina*	tidal creek versus tidal flat	25 EST-derived and	genetic distance was related to ecology	no evidence for IBA	IBD detected	Oetjen and Reusch

(marine flowering plant)		anonymous microsatellite markers	(tidal ecotype) and geographic distance using 15 population pairs			2007
15. *Biscutella laevigata* (terrestrial plant)	continuous population varying in habitat characteristics	102 AFLP loci	genetic distance was related to ecology (habitat type) and geographic distance in a continuous population	IBA detected	IBD detected	Parisod and Christin 2008
16. *Neochlamisus bebbinae* (leaf beetle)	willow and maple host forms	381 AFLP loci, mtDNA	genetic distance was related to quantitative indices of host-associated adaptive divergence and to geographic distance using 15 population pairs	IBA detected for pooled AFLP loci; 10% of putatively neutral (i.e., non-outlier) individual AFLP loci show significant IBA; mtDNA does not show IBA	IBD detected for mtDNA, but not for AFLPs	Funk et al., in review
Studies employing primarily an AMOVA (or similar) framework						
17. *Parus major* (great and, respectively)	deciduous vs. mixed-coniferous forests	9 microsatellite loci and 4 allozyme loci	partitioning of genetic structure within and among habitat types was examined using AMOVA	no evidence for IBA, genetic variance among habitats: 0.19%, $p > 0.05$ among populations within habitats: 1.60% $p < 0.05$	not directly tested	Blank et al. 2007
18. *Parus caeruleus* (blue tit)	deciduous vs. mixed-coniferous forests	9 microsatellite loci and 4	partitioning of genetic structure within and among	no evidence for IBA, genetic variance	not directly tested	Blank et al. 2007

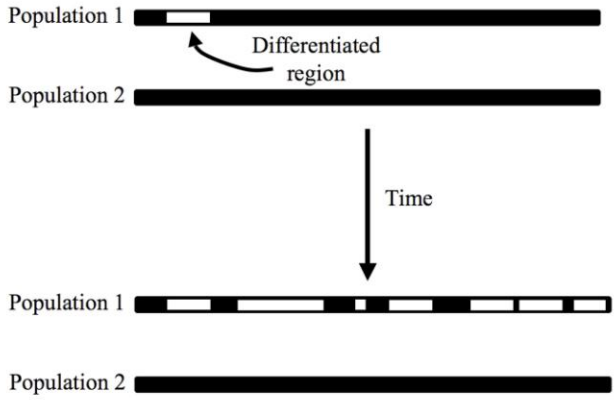
		allozyme loci	habitat types was examined using AMOVA	among habitats: <0.01%, $p > 0.05$ among populations within habitats: 7.93%, $p < 0.05$		
19. *Osmerus mordax* (rainbow smelt)	giant, normal, and dwarf morphotypes	5 microsatellite loci	partitioning of genetic structure among morphotypes versus among populations within morphotypes was examined using AMOVA	no evidence for IBA, no genetic structure among morphotypes (0%, $p > 0.05$), despite appreciable structure among populations within types (10%, $p < 0.05$)	not directly tested	Curry et al. 2004
20. *Loxia curvirostra* complex (red crossbills)	eight morphologically and vocally differentiated 'call types'	440 AFLP loci	AMOVA models used to examine genetic variation between call types and among populations within call types	IBA detected, between call-type differentiation is greater (7.0%, $p < 0.05$) than that found among different geographic locations within call types (3.5%, $p < 0.05$)	IBD not detected, despite explicit tests using distance matrices	Parchman et al. 2006
21. *Hesperotettix viridis* (grasshopper)	*Solidago mollis* vs. *Gutierrezia sarothrae* host plant forms	222 AFLPs	partitioning of genetic structure among host plant forms versus among populations within	IBA detected, strong (20%) and significant variance among host forms,	no evidence for IBD, although not directly tested, insignificant (1%) variation among	Sword et al. 2005

			host plant forms was examined using AMOVA	insignificant (1%) variation among localities within forms	different geographic localities within host forms	
22. *Salamandra salamandra* (fire salamander)	stream vs. pond form	11 microsatellite loci and mtDNA sequences	partitioning of genetic structure among 33 sites was examined, in relation to geographic distance and pond type (ecology)	IBA detected, two ecologically (pond vs. stream types) differentiated groups within a relatively small forest showed signs of genetic differentiation (i.e., two main genetic clusters were correlated to larval habitat type)	IBD not detected, analysis of a large forest area (neighboring the smaller one exhibiting IBA) where all salamanders use streams showed no genetic differentiation, gene flow between ecologically similar types occurs over large distances	Steinfartz et al. 2007
Phylogeographic and hybrid zone studies						
23. *Halichoeres spp.* (tropical reef fish, wrasses)	habitat types, such as warm versus cold water habitats	mtDNA sequences	phylogeographic patterns within and among five species were attributed to effects of habitat, distance, and biogeographic barriers	IBA detected, concordance of phylogenetic partitions with habitat types	Little evidence for IBD, high genetic connectivity between similar habitats separated by thousands of kilometers	Rocha et al. 2005
24. *Bombina spp.* (toads)	pond-adapted *B. bombina* vs. puddle-adapted *B. variegata*	5 unlinked, diagnostic allozyme loci	correlates of genetic structure examined within a hybrid zone	IBA detected, genetic structure strongly associated with habitat type	N/A	MacCullum et al. 1998; see also Vines et al. 2003

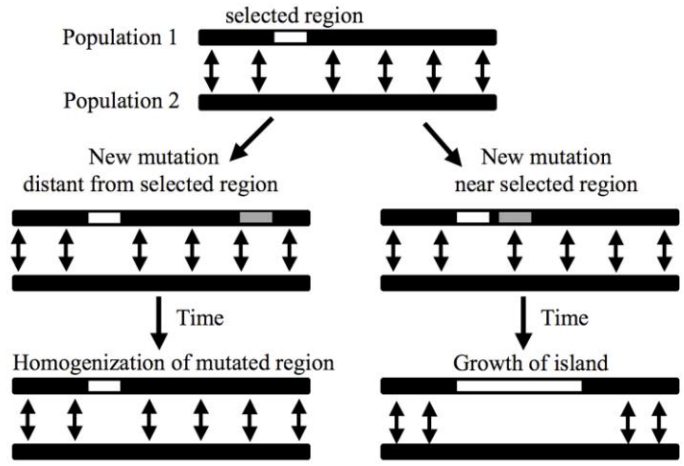
				rather than distance		

Figure S1. Conceptual models for the growth of genomic islands of divergence. In all cases, bars represent chromosomes, white boxes within them represent differentiated regions of the genome, and filled, black areas represent undifferentiated regions. Two-headed arrows represent regions of the genome where genetic exchange between populations is high. A) Allopatric model. Divergence proceeds unimpeded by gene flow, with the proportion of the genome differentiated between two populations being positively related to time since divergence. B) Ecological model. A new mutation (grey box within chromosome) arising near genomic regions under selection, and thus undergoing reduced introgression, has a higher likelihood of differentiating between populations than a new mutation arising in a region distant from those under selection. C) Structural model. A new mutation (grey box within chromosome) arising near an inversion has a higher likelihood of differentiating between populations than a new mutation arising in a region distant from the inversion.

A) Model I: Allopatric model



B) Model II: Ecological model (divergent selection) with gene flow



C) Model III: Structural model with gene flow