Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among Darwin's finches

Fan Han,[1] Sangeet Lamichhaney,[1] B. Rosemary Grant,[2] Peter R. Grant,[2] Leif Andersson,[1,3,4] and Matthew T. Webster[1]

[1]*Department of Medical Biochemistry and Microbiology, Science for Life Laboratory, Uppsala University, 75123 Uppsala, Sweden;* [2]*Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey 08544-2016, USA;* [3]*Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, 75007 Uppsala, Sweden;* [4]*Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, Texas 77843-4461, USA*

Genomic comparisons of closely related species have identified "islands" of locally elevated sequence divergence. Genomic islands may contain functional variants involved in local adaptation or reproductive isolation and may therefore play an important role in the speciation process. However, genomic islands can also arise through evolutionary processes unrelated to speciation, and examination of their properties can illuminate how new species evolve. Here, we performed scans for regions of high relative divergence (F_{ST}) in 12 species pairs of Darwin's finches at different genetic distances. In each pair, we identify genomic islands that are, on average, elevated in both relative divergence (F_{ST}) and absolute divergence (d_{XY}). This signal indicates that haplotypes within these genomic regions became isolated from each other earlier than the rest of the genome. Interestingly, similar numbers of genomic islands of elevated d_{XY} are observed in sympatric and allopatric species pairs, suggesting that recent gene flow is not a major factor in their formation. We find that two of the most pronounced genomic islands contain the *ALX1* and *HMGA2* loci, which are associated with variation in beak shape and size, respectively, suggesting that they are involved in ecological adaptation. A subset of genomic island regions, including these loci, appears to represent anciently diverged haplotypes that evolved early during the radiation of Darwin's finches. Comparative genomics data indicate that these loci, and genomic islands in general, have exceptionally low recombination rates, which may play a role in their establishment.

[Supplemental material is available for this article.]

What evolutionary processes drive divergence between the genomes of newly forming species? A powerful way to make inferences about these processes is to analyze variation in levels of divergence between the genomes of recently evolved species (Wolf and Ellegren 2016). One observation that has arisen from such studies is the existence of localized peaks of elevated genetic divergence (Cruickshank and Hahn 2014; Seehausen et al. 2014). Analyzing the features of these peaks has proved informative for the process of speciation. Peaks of divergence were initially referred to as "genomic islands of speciation" (Turner et al. 2005). It was suggested that these islands represent regions that are resistant to gene flow between incipient species and that drive the initial stages of reproductive isolation. However, it is now recognized that a variety of processes could cause these patterns (Noor and Bennett 2009; Cruickshank and Hahn 2014), and this metaphor has been gradually replaced with the term "genomic islands of divergence" (Nosil et al. 2009).

Genomic islands of divergence have been observed in a broad range of closely related species of plants and animals (Turner et al. 2005; Harr 2006; Lawniczak et al. 2010; Nadeau et al. 2012; Martin et al. 2013; Renaut et al. 2013; Carneiro et al. 2014; Clarkson et al. 2014; Poelstra et al. 2014; Soria-Carrasco et al. 2014; Feulner et al. 2015; Malinsky et al. 2015; Berg et al. 2016; Marques et al. 2016; Wang et al. 2016). However, the landscape of divergence varies between pairwise comparisons of species both due to their evolutionary histories and the length of time they have been separated. In addition, the interpretation of genomic islands depends on the statistics used to define them. Relative measures of divergence, such as F_{ST}-based statistics, are widely used to quantify genetic differentiation between populations. However, such relative measures are strongly influenced by intra-specific variation, and peaks of F_{ST} can be observed not only in regions with high divergence between populations but also where either of the compared populations has low genetic diversity (Cruickshank and Hahn 2014). Conversely, an absolute measure of genetic divergence between populations, d_{XY}, is less affected by local reduction in genetic variation. As described below, genomic islands with high absolute and relative divergence are predicted to result from differential gene flow between genomic regions or from the presence of anciently diverged haplotypes, whereas several additional processes are predicted to generate peaks of relative, but not absolute, divergence.

In modes of speciation where incipient species are able to exchange genes, islands of divergence are predicted to form around genomic regions where gene flow is disadvantageous because they contain variants involved in local adaptation and/or

Corresponding author: matthew.webster@imbim.uu.se
Article published online before print. Article, supplemental material, and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.212522.116.

27:1004–1015 Published by Cold Spring Harbor Laboratory Press; ISSN 1088-9051/17; www.genome.org

reproductive isolation (Wu 2001; Turner et al. 2005; Seehausen et al. 2014). Hence, genetic divergence accumulates between species in such regions but is homogenized by gene flow in the rest of the genome. Under this speciation-with-gene-flow model, genomic islands are resistant to gene flow and can expand due to divergence hitchhiking. Both absolute and relative measures of divergence are predicted to be elevated in such regions, as depicted in Supplemental Figure S1A. An example of such genomic islands with elevated d_{XY} is seen in comparisons of cichlid ecomorphs, which interbreed but maintain their distinct morphologies (Malinsky et al. 2015; McGee et al. 2016). A similar genomic pattern also could be formed without the existence of recent gene flow. In particular, if ancient, highly diverged haplotypes were present at a locus in the ancestral population of two species, then these regions could form islands of divergence in the extant species due to lineage sorting (Supplemental Fig. S1B). If the haplotypes began to diverge before a species radiation, then genomic islands of divergence at these loci might be expected to be shared across genome comparisons of multiple extant species. Such regions could also be involved in the establishment of barriers to gene flow between derived populations (Sicard et al. 2015; Pease et al. 2016).

In strictly allopatric speciation, allelic exchange between species is prevented by a geographical barrier (Mayr 1942; Coyne and Orr 2004). In this speciation-without-gene-flow model, each of the isolated populations gradually adapts to the local environment, and reproductive incompatibilities build up by selection and/or genetic drift. As gene flow is absent in this model, genomic islands where gene flow is disadvantageous are not expected. However, in modes of speciation both with and without gene flow, genomic islands of divergence may form from additional processes. For example, selective sweeps are expected to result in regions with elevated F_{ST} in comparisons of two populations. Such genomic islands likely contain loci that contribute to adaptation. However, as selective sweeps essentially convert intra-species variation into fixed differences, we do not expect d_{XY} to be elevated in such regions unless selection also results in reduction of gene flow in these genomic islands compared to the rest of the genome (Supplemental Fig. S1C; Cruickshank and Hahn 2014; Supple et al. 2015). Hence, selection due to local ecological adaptation can result in islands of divergence, but d_{XY} is only expected to be elevated in such regions if gene flow is reduced at these loci, for example, because individuals with genomes that contain introgressed segments at these loci are selected against.

Processes related to intrinsic genomic architecture could also lead to island-like patterns along the genome (Supplemental Fig. S1D). Both ongoing background selection and recurrent selective sweeps result in locally reduced levels of genetic variation, which are more extensive in regions of low recombination. This leads to a general positive correlation between genetic variation and recombination rate (Begun and Aquadro 1992) and is expected to result in a heterogeneous landscape of genome divergence (Cruickshank and Hahn 2014; Burri et al. 2015; Irwin et al. 2016; Wang et al. 2016). These processes can result in genomic islands of divergence of elevated F_{ST} but are not expected to cause elevated levels of d_{XY}. For example, a comparison of flycatcher species identified genomic islands of elevated F_{ST}, which were initially interpreted as genomic islands of speciation where gene flow has ceased (Ellegren et al. 2012). However, the observation that these islands were not elevated in d_{XY}, coupled with the observation that genomic islands were found at the same genomic positions in independent comparisons of flycatcher species pairs at different levels of divergence, led them to be reconsidered as mainly driven by low recombination rate and linked selection (Burri et al. 2015). Such genomic islands produced by recurrent background selection or selective sweeps are actually expected to have reduced d_{XY}. This is because selection reduces genetic variation in the ancestral population, which leads to a reduced time to the most recent common ancestor (TMRCA) between individuals from the two descendent populations.

In summary, a heterogeneous genomic landscape incorporating islands of divergence can be generated by (1) variation in gene flow between loci, which can be caused by ecological adaptation or genomic incompatibilities, (2), the presence of diverged haplotypes in the ancestral population that predate speciation, (3) adaptation in the absence of differential gene flow, or (4) recurrent hitchhiking/background selection acting across the genome with effects on genetic variation modulated by recombination rate. Whereas all of these models generate genomic islands of relative divergence (F_{ST}), the main difference is that d_{XY} is above average in islands caused by models 1 and 2 but is unchanged or below average in models 3 and 4. The effect of selection in present-day populations is to reduce linked genetic variation, which affects relative measures of divergence, like F_{ST}, but not absolute ones, like d_{XY}. These models are not mutually exclusive, and multiple forces can act to produce a particular pattern in the genome (Supplemental Fig. S1).

Darwin's finches are a classic example of an adaptive radiation. Based on the traditional morphological taxonomy and recent genetic dating, it has been shown that 18 species of Darwin's finches evolved from a common ancestor that colonized the Galápagos ~1.5 million years ago (Petren et al. 2005; Lamichhaney et al. 2015). After colonizing different islands, they adapted to different ecological conditions and diversified in their beak morphology and feeding habits (Grant and Grant 2008). This isolated speciation system provides a perfect model to investigate genomic divergence and speciation in nature. Assortative mating occurs between species and even within morphs of the same species (Grant and Grant 1989; Huber et al. 2007). However, hybridization between species is also observed, and hybrids are both viable and fertile (Grant and Grant 2008). It is unknown whether any genetic factors that reduce fertility in hybrids exist, and species barriers are incomplete (Grant and Grant 1996). Recent analyses of genome evolution in Darwin's finches have identified genomic regions associated with phenotypic variation. In particular, two genomic regions of high differentiation containing the *ALX1* and *HMGA2* genes, respectively, are associated with differences in beak shape and size (Lamichhaney et al. 2015, 2016). However, the genomic landscape of divergence has not yet been comprehensively characterized.

In this study, we compared the genomes of 12 species pairs of Darwin's finches with different combinations of geographic distributions and beak genotypes in order to examine the genomic patterns under alternative models of genomic islands of divergence. We chose pairs of species to specifically quantify how the divergence landscape is affected by different beak morphologies and the opportunities for gene flow, based on geographical distance between the species.

Results

Number and size of genomic islands correlate with genomic divergence

We have previously reported whole-genome sequence (WGS) data for the 18 currently recognized species of Darwin's finches and

used these data to study the evolution of these species and their beaks (Lamichhaney et al. 2015, 2016). For the purpose of this study, we used data on Darwin's finches from six islands of the Galápagos archipelago and one from Cocos Island (Fig. 1A; Table 1). Additionally, we included two outgroup species, *Loxigilla noctis* and *Tiaris bicolor* from Barbados. In order to test the potential effect of gene flow on the divergence landscape, we chose (1) four sympatric pairs from two islands of the Galápagos, (2) five allopatric pairs from five islands of the Galápagos, and (3) three allopatric pairs from islands of the Galápagos and the distant Cocos Island (Fig. 1A). In addition, we made two pairwise comparisons of Darwin's finches with outgroup species. To explore the relationship between genomic islands and regions associated with phenotypic traits under selection (Lamichhaney et al. 2015, 2016), we examined patterns under different combinations of haplogroups at two beak-associated loci, *ALX1* and *HMGA2*, in each geographic group (Table 2). We analyzed WGS data from at least four individuals of each population of Darwin's finches, sequenced to ~10× coverage per individual. Approximately 7 million variant sites were identified on average in each species pair, using stringent criterion for SNP calling.

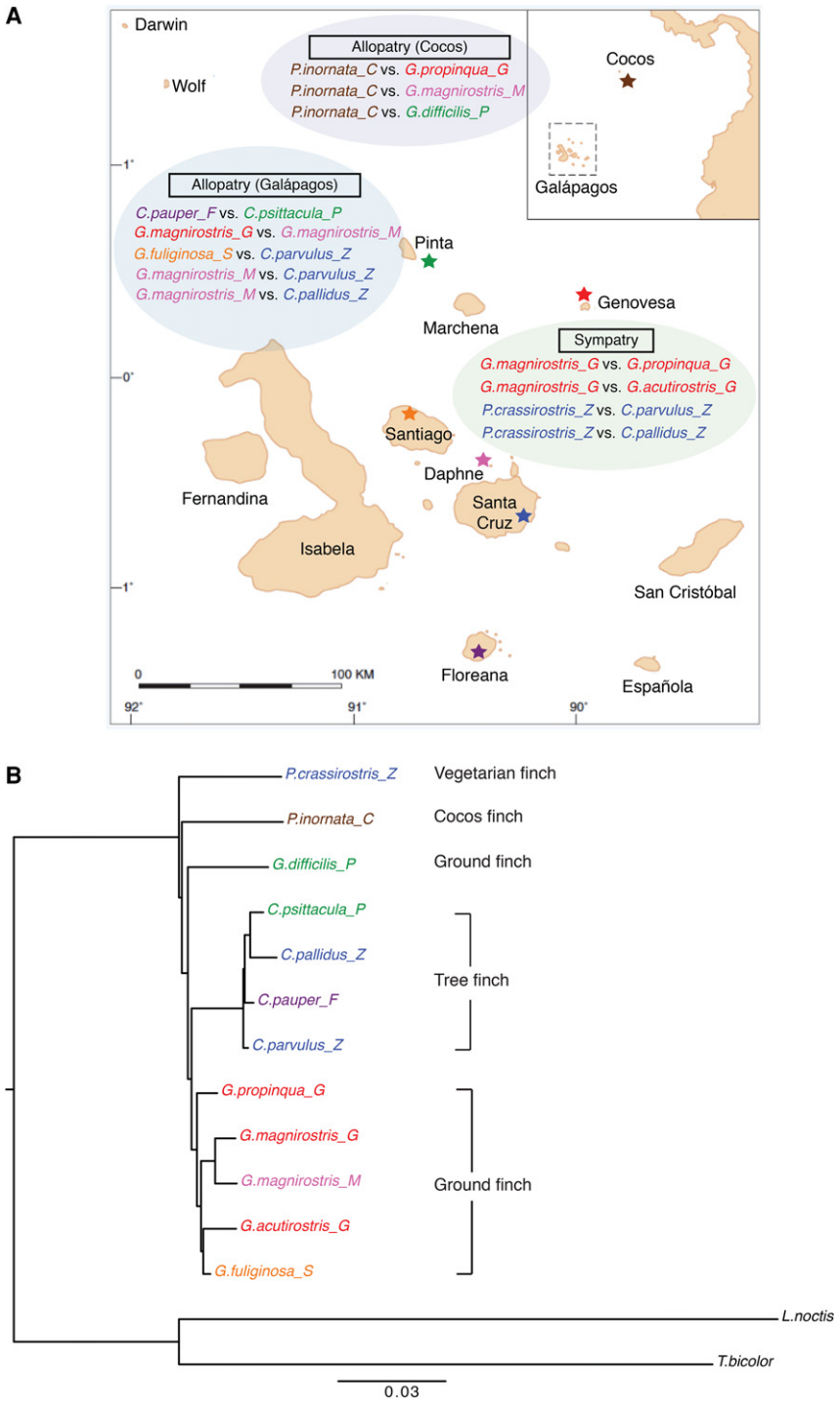

Figure 1. Species distribution and phylogeny of Darwin's finches. (*A*) Geographical distribution of 12 species pairs of Darwin's finches. Color of species represents the inhabited island. Coordinates of *x*- and *y*-axes, respectively, refer to latitude and longitude of the map. The map is modified from Grant and Grant (2014) (reprinted by permission from Princeton University Press © 2014). (*B*) Neighbor-joining tree of 12 species of Darwin's finches. Color of species represents the inhabited island.

We constructed a species tree based on all the autosomal sites from the 14 species (Fig. 1B). The topology of the tree was consistent with the phylogeny reported in our previous study (Lamichhaney et al. 2016). We first scanned the whole genome in each species pair using 50-kb nonoverlapping windows and estimated genomic parameters including nucleotide diversity (π), Tajima's D, number of fixed differences (d_f), d_{XY}, and F_{ST} in each window. Genome-wide average levels of F_{ST} of each pair varied between 0.09 and 0.41, in concordance with previous studies (Table 3; Lamichhaney et al. 2015). F_{ST} scores were Z-transformed separately in each pair, and genomic regions with $ZF_{ST} \geq 4$ were identified as outlier windows (Fig. 2A). Neighboring outlier windows were merged into single genomic islands.

Overall, the number of genomic islands in a species comparison displayed a strong negative correlation with average genomic divergence across all species pairs ($r = -0.80$, $P = 1.72 \times 10^{-3}$; Pearson's correlation coefficient test) (Fig. 2B). In *P. crassirostris_Z* vs. *C. pallidus_Z* ($F_{ST} = 0.40$) and *P. inornata*_C vs. *G. difficilis*_P ($F_{ST} = 0.41$) pairs, only five and six islands were detected, respectively (see Table 1 for origin of populations). In the pair *P. inornata* vs. *T. bicolor*, no genomic islands were detected ($F_{ST} = 0.70$) (Table 3). The mean size of genomic islands also decreased with the genomic divergence ($r = -0.47$, $P = 0.12$; Pearson's correlation

Table 1. Species names, geographic islands, and sample size of Darwin's finches and outgroup species

Island	Species	Population ID	Sample size
Santa Cruz	*Camarhynchus pallidus*	*C. pallidus_Z*	5
	Camarhynchus parvulus	*C. parvulus_Z*	12
	Platyspiza crassirostris	*P. crassirostris_Z*	5
Genovesa	*Geospiza propinqua*	*G. propinqua_G*	10
	Geospiza acutirostris	*G. acutirostris_G*	4
	Geospiza magnirostris	*G. magnirostris_G*	5
Pinta	*Camarhynchus psittacula*	*C. psittacula_P*	10
	Geospiza difficilis	*G. difficilis_P*	10
Floreana	*Camarhynchus pauper*	*C. pauper_F*	12
Santiago	*Geospiza fuliginosa*	*G. fuliginosa_S*	12
Daphne Major	*Geospiza magnirostris*	*G. magnirostris_M*	10
Cocos	*Pinaroloxias inornata*	*P. inornata_C*	8
Barbados	*Loxigilla noctis*	*L. noctis*	5
	Tiaris bicolor	*T. bicolor*	3

coefficient test) (Fig. 2C). The observation that fewer genomic islands occur in comparisons of more distantly related species suggest that they are obscured by increased divergence throughout the genome (Renaut et al. 2013).

We next tested the difference in number and size of genomic islands between allopatric and sympatric groups. No significant difference was observed in the number of genomic islands (mean number: sympatry = 45, allopatry Galápagos = 54, allopatry Cocos = 37). There was no significant difference in F_{ST} between pairs in sympatry and in allopatry, which may have otherwise biased this comparison (mean F_{ST}: sympatry = 0.23, allopatry = 0.22). This is because species that are currently found in sympatry are not always the most closely related and probably evolved initially in allopatry. However, in our study the sizes of genomic islands in sympatric and allopatric groups did not differ on average (mean size: sympatry = 75.8 kb, allopatry Galápagos = 73.1 kb, allopatry Cocos = 66.8 kb). Hence, genomic islands of divergence are widespread in all comparisons of Darwin's finches, and the number and size of the genomic islands are not evidently different between sympatric and allopatric pairs (number of islands: $P = 0.80$; size of islands: $P = 1.00$; Wilcoxon rank-sum test). There is also no clear association between number of genomic islands and differences in beak morphology. We also compared the other genomic parameters of islands to the background (Supplemental Fig. S2). We found that nucleotide diversity (π) was reduced at genomic islands compared to the genomic background in both species per comparison, and Tajima's D was generally reduced in at least one of the species (with the exception of one pair). Fixed differences (d_f) were overall elevated in the genomic islands except in two species pairs (Supplemental Fig. S2).

We next examined how the genomic landscape of divergence is shared between pairwise comparisons. We identified shared islands by counting the number of 50-kb windows that were shared in the following comparisons: (A) between pairs of ground vs. ground and tree vs. tree finches; (B) between two pairs of ground vs. tree finches; and (C) between more distantly related pairs (Supplemental Table S1). It should be noted that only category (A) allows identification of islands that are present in independent pairwise comparisons, whereas the categories (B) and (C) involve shared lineages, which means that islands can be shared simply due to shared ancestry. In general, however, the number of shared

Table 2. Geographic distribution and beak-related haplogroups of 11 species pairs

Geographic distribution	Species pairs[a]	*ALX1* haplogroup	*HMGA2* haplogroup	Divergence time[b]
Sympatry	*G. magnirostris_G* *G. propinqua_G*	Blunt Pointed	Large Large	172 ± 13
	G. magnirostris_G *G. acutirostris_G*	Blunt Pointed	Large Small	220 ± 3
	P. crassirostris_Z *C. parvulus_Z*	Pointed Pointed	Large Small	517 ± 3
	P. crassirostris_Z *C. pallidus_Z*	Pointed Pointed	Large Large	521 ± 11
Allopatry (Galápagos)	*C. pauper_F* *C. psittacula_P*	Pointed Pointed	Small Large	185 ± 6
	G. magnirostris_G *G. magnirostris_M*	Blunt Blunt	Large Large	83 ± 8
	G. fuliginosa_S *C. parvulus_Z*	Pointed Pointed	Small Small	353 ± 1
	G. magnirostris_M *C. parvulus_Z*	Blunt Pointed	Large Small	347 ± 4
	G. magnirostris_M *C. pallidus_Z*	Blunt Pointed	Large Large	367 ± 11
Allopatry (Cocos)	*P. inornata_C* *G. propinqua_G*	Pointed Pointed	Small Large	409 ± 3
	P. inornata_C *G. magnirostris_M*	Pointed Blunt	Small Large	444 ± 5
	P. inornata_C *G. difficilis_P*	Pointed Pointed	Small Small	520 ± 2

[a]Letter after species name refers to the island of occurrence: (C) Cocos; (F) Floreana; (G) Genovesa; (M) Daphne Major; (P) Pinta; (S) Santiago; (Z) Santa Cruz.
[b]In thousands of years, based on data from Lamichhaney et al. (2015).

windows across comparisons is low, which indicates that the genomic islands identified in this study are unlikely to be caused solely by an intrinsic genomic property such as low recombination rate, which is conserved between independent comparisons. In addition, we noticed that if both compared pairs had different beak haplogroups, then a proportion of shared windows were present at *ALX1* and *HMGA2* loci. The presence of these loci in multiple comparisons is consistent with the high divergence between haplotypes at these loci and indicates they were probably established as polymorphisms during the early stages of the evolution of Darwin's finches.

Absolute differentiation is elevated in genomic islands

Based on the genomic islands defined by F_{ST}, we used d_{XY} to compare the level of absolute divergence in genomic islands to that of genomic background in each species pair. If recent gene flow were common, then we would expect to observe elevated d_{XY} in genomic islands in sympatric pairs of species, as they have more opportunities to exchange genes than species in allopatry. We found that d_{XY} was significantly elevated in genomic islands of sympatric species (four pairs in sympatry: $P < 0.03$; randomization test) (Fig. 3A). However, we also found that d_{XY} was significantly elevated in genomic islands compared with genomic background in allopatric comparisons (four pairs in allopatry Galápagos: $P < 2 \times 10^{-4}$; randomization test), including those involving Cocos Island (three pairs in allopatry Cocos: $P < 1 \times 10^{-4}$; randomization test), except in one comparison (*G. magnirostris_G* vs. *G. magnirostris_M*), where d_{XY} was decreased in island regions ($P < 1 \times 10^{-10}$; randomization test). The elevated d_{XY} in genomic islands is consistent with a model where haplotypes at genomic islands became genetically isolated before the rest of the genomes of the species pairs under comparison. Further, the observation that similar patterns are observed in allopatry and sympatry indicates that recent gene flow between incipient species has not been a major factor in generating these genomic islands. Divergent haplotypes in these regions may therefore have been present in the ancestral populations before the species we are comparing began to diverge from each other.

We next analyzed divergence between species pairs of Darwin's finches and a more distantly related outgroup, *L. noctis*, which shares very few genetic variants with Darwin's finches (<1% of all variable sites are shared) (Lamichhaney et al. 2015). We compared d_{XY} in genomic islands defined in the original comparisons of each geographic group with the background. No elevated d_{XY} was evident in genomic islands in any of the comparisons (*G. magnirostris_G* vs. *L. noctis*, *C. pauper_F* vs. *L. noctis*, *P. inornata_C* vs. *L. noctis*) (Supplemental Fig. S3). This supports the suggestion that elevated d_{XY} in genomic islands in Darwin's finches from the Galápagos is not caused by elevated substitution rates.

Table 3. Statistics of heterogeneous genomic regions among Darwin's finches

Geographic distribution	Species pairs[a]	Genome-wide mean F_{ST}	Island mean F_{ST}	No. of islands	Mean size[b]	Max size[b]
Sympatry	*G. magnirostris_G* *G. propinqua_G*	0.11	0.44	58	71.6	350
	G. magnirostris_G *G. acutirostris_G*	0.14	0.64	84	107.7	600
	P. crassirostris_Z *C. parvulus_Z*	0.27	0.52	36	63.9	150
	P. crassirostris_Z *C. pallidus_Z*	0.40	0.64	5	60.0	100
Allopatry (Galápagos)	*C. pauper_F* *C. psittacula_P*	0.09	0.32	61	64.8	300
	G. magnirostris_G *G. magnirostris_M*	0.10	0.38	61	66.9	300
	G. fuliginosa_S *C. parvulus_Z*	0.12	0.36	52	99.0	1050
	G. magnirostris_M *C. parvulus_Z*	0.21	0.52	54	66.7	200
	G. magnirostris_M *C. pallidus_Z*	0.29	0.73	42	67.9	300
Allopatry (Cocos Island)	*P. inornata_C* *G. propinqua_G*	0.26	0.50	45	66.7	250
	P. inornata_C *G. magnirostris_M*	0.31	0.68	61	75.4	550
	P. inornata_C *G. difficilis_P*	0.41	0.67	6	58.3	100
Outgroup	*G. fuliginosa_S* *L. noctis*	0.49	0.78	20	82.5	400
	P. inornata_C *T. bicolor*	0.70	–	0	–	–

[a]Letter after species name refers to the island of occurrence: (C) Cocos; (F) Floreana; (G) Genovesa; (M) Daphne Major; (P) Pinta; (S) Santiago; (Z) Santa Cruz.
[b]Numbers are in kilobases.

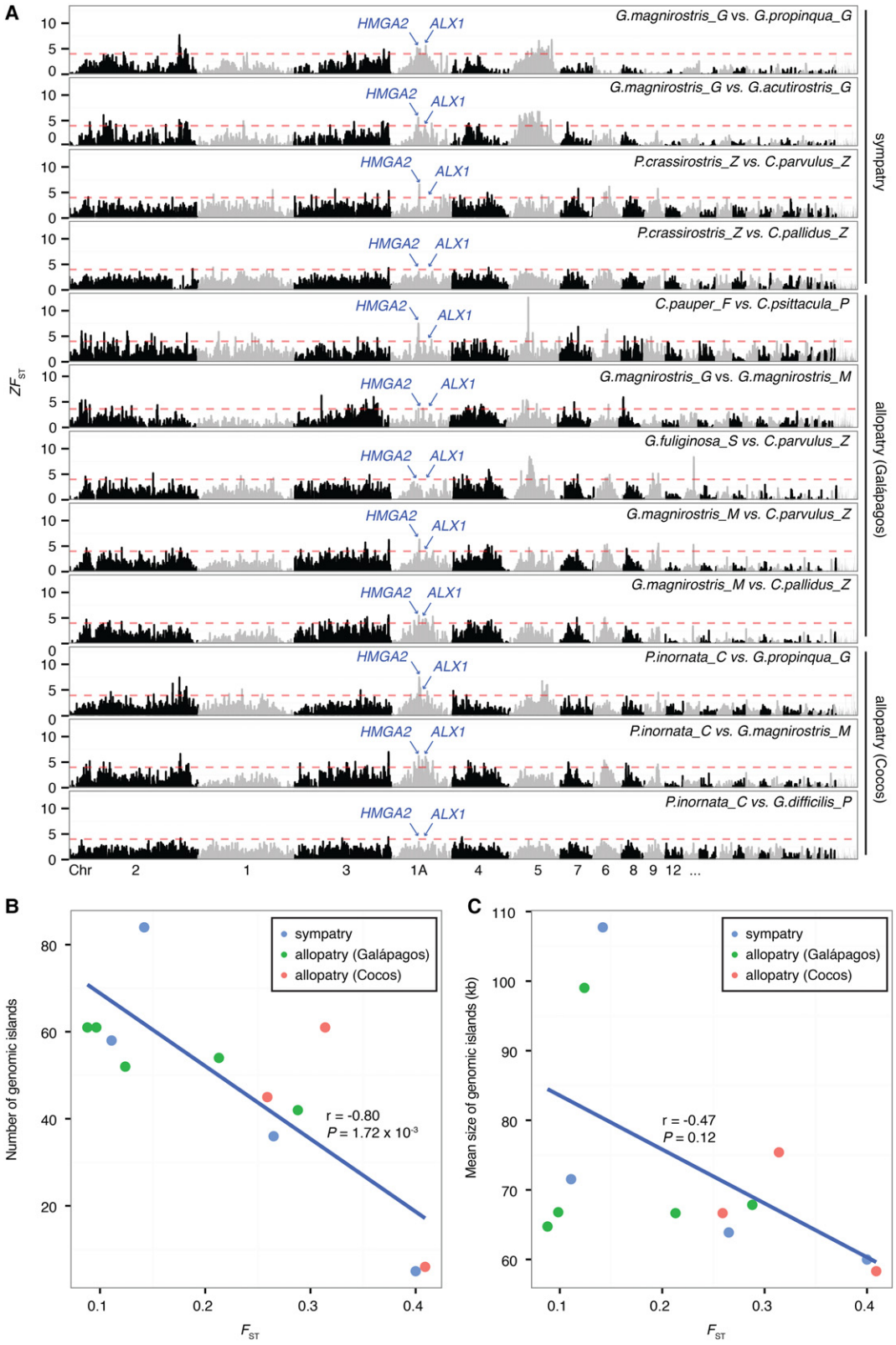

Figure 2. Landscape of divergence across the genome and correlation between genomic islands and genome-wide divergence. (*A*) Genome-wide ZF_{ST} screen in 50-kb windows across 12 species pairs of Darwin's finches. Scaffolds were ordered along the zebra finch genome, and chromosomes are numbered according to the zebra finch nomenclature. Red dashed lines indicate a ZF_{ST} of four as the threshold to define islands of divergence. Genomic locations of the *ALX1* and *HMGA2* related to beak morphology are indicated. (*B*) Number of genomic islands against genomic divergence. (*C*) Mean size of genomic islands against genomic divergence. Each dot represents a species pair. Correlations are tested with Pearson's correlation coefficient test.

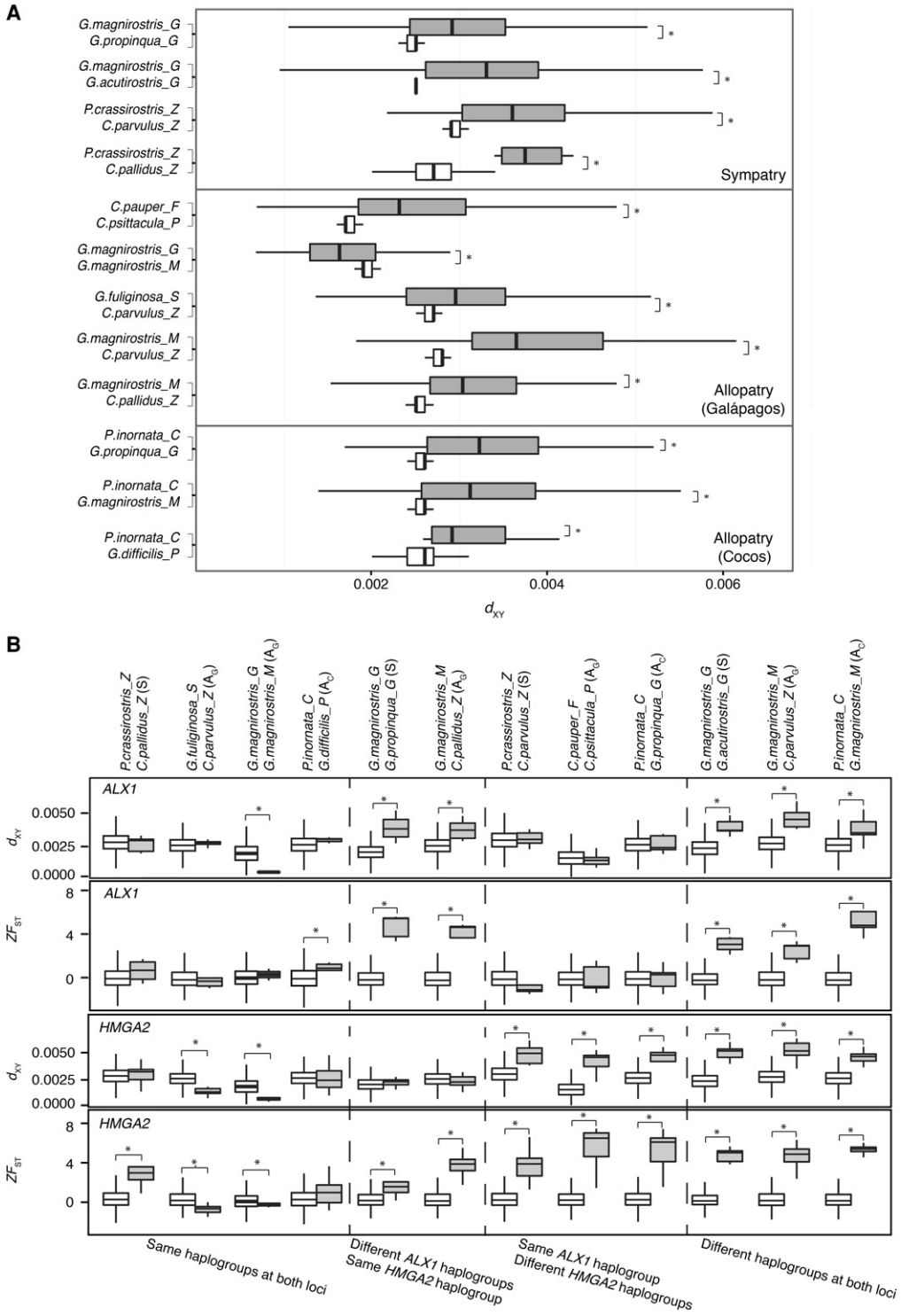

Figure 3. Measures of divergence at genomic islands and beak-related loci among Darwin's finches. (*A*) Absolute measures of genomic divergence (d_{XY}). Gray boxes represent values in genomic islands and white boxes represent genomic background. Vertical lines from *left* to *right* of each box refer to first quartile, median, and third quartile, respectively. The three panels, from *top* to *bottom*, are sympatric pairs, allopatric pairs within Galápagos, and allopatric pairs from Cocos Island and Galápagos. Species pairs are sorted by divergence time in each panel. Significance is tested on the basis of a randomization test. (*) $P < 0.05$. Outliers are not shown in the plot. (*B*) Relative measures (ZF_{ST}) and absolute measures of divergence (d_{XY}) at *ALX1* and *HMGA2*. The four panels, from *top* to *bottom*, are d_{XY} at *ALX1*, ZF_{ST} at *ALX1*, d_{XY} at *HMGA2*, and ZF_{ST} at *HMGA2*, respectively. Gray boxes represent values in each target locus and white boxes represent genomic background. Horizontal lines from *top* to *bottom* of each box refer to first quartile, median, and third quartile, respectively. Abbreviations in parentheses after each species pair refer to geographic distribution: (S) sympatry, (A_G) allopatry within Galápagos; (A_C) allopatry from Cocos Island and Galápagos. *Bottom* categories illustrate the haplotype differentiations. Boxes with asterisks indicate that d_{XY} or ZF_{ST} is significantly higher than the background, tested with the Wilcoxon rank-sum test. (*) $P < 0.05$. Outliers are not shown in the plot.

Elevated d_{XY} at beak-associated loci in species with different beak morphology reflects the role of natural selection in the ancestral population of Darwin's finches

Beak morphology is one of the most diverse features of Darwin's finches. It directly mirrors the local adaptations of finches to ecological conditions. From previous work, two major genomic loci that contain the *ALX1* and *HMGA2* genes were identified to be associated with beak morphology. The *ALX1* locus, containing part of *LRRIQ1*, the entire coding sequence of the *ALX1* gene and flanking sequences, is a 240-kb region controlling beak shape (Lamichhaney et al. 2015). Two divergent haplotype groups (haplogroups), *B* and *P*, associated with blunt and pointed beaks, respectively, occur at this locus. The other major locus, *HMGA2,* is a 525-kb region that is associated with beak size in finches (Lamichhaney et al. 2016). It contains the *HMGA2*, *MSRB3*, *LEMD3*, and *WIF1* genes. Similar to the *ALX1* locus, the *HMGA2* locus harbors two divergent haplogroups, *S* and *L*, associated with small and large beaks, respectively. To assess the contribution of these loci to genomic islands of divergence, we calculated the relative and absolute measures of divergence at these two loci relative to the rest of the genome (background) (Fig. 3B). At the *ALX1* locus, ZF_{ST} was greatly elevated relative to background in the species pairs that had different *ALX1* haplogroups, but not in the pairs with the same haplogroup. This pattern is also shown at the *HMGA2* locus when different haplogroups are compared and in some comparisons of the same haplogroup. It is important to note that there is considerable sequence diversity within each of the haplogroups; for example, *P. crassirostris* and *C. pallidus* cluster on the same side of the *HMGA2* tree (Lamichhaney et al. 2016) but still have a considerable divergence between them. In addition, there is large variation in beak morphologies, and it is possible that additional variation exists within each of the *ALX1* and *HMGA2* haplogroups that subdivides the haplotypes into more than two functional alleles.

At both loci, we also found significantly elevated d_{XY} associated with segregation of haplotypes in all sympatric and allopatric pairs; i.e. when two populations contained different haplogroups at either the *ALX1* or *HMGA2* locus, they tended to have significantly elevated d_{XY} in these regions. In contrast, no distinct differentiation was seen in most of the pairs that shared the same haplogroup. This reflects the high divergence between haplogroups at these loci, which is consistent with their relatively ancient origin, prior to the radiation of ground and tree finches (Lamichhaney et al. 2015, 2016). Hence, when two species carry different haplotypes at these anciently diverged beak loci, the genomic regions appear as genomic islands with elevated absolute divergence.

Notably, however, two pairwise comparisons (*G. magnirostris_G* vs. *G. magnirostris_M* and *G. fuliginosa_S* vs. *C. parvulus_Z*) showed a different pattern at these loci (Fig. 3B). Both of these pairs carry concordant haplogroups at both loci, and d_{XY} in these regions is significantly reduced compared to the genomic background. In the comparison of *G. magnirostris_G* vs. *G. magnirostris_M*, both the *ALX1* and *HMGA2* regions have reduced d_{XY}, and in the comparison of *G. fuliginosa_S* vs. *C. parvulus_Z*, d_{XY} is reduced at *HMGA2* and is similar to the genomic background at *ALX1*. One explanation for this pattern could be that adaptive introgression has occurred at these loci. This could be due to a haplotype introgressing between populations after they split due to an increase in fitness related to the effects of the haplotype on beak morphology. In the case of the comparison involving two *G. magnirostris* populations, it could also reflect recent selective sweeps at these loci that affected both populations and reduced haplotype diversity at these loci.

Recombination rate is reduced in genomic islands

To further explore the mechanisms involved in shaping genomic islands of divergence, we investigated whether genomic islands tend to be located in regions with reduced recombination rates. Such an association is expected if genomic islands are generated by recurrent selective sweeps or background selection, which tend to reduce genetic variation and hence elevate F_{ST} in regions of low recombination. Recent data indicate that the recombination landscape is well conserved among birds (Singhal et al. 2015). We therefore utilized the fine-scale recombination map that was available for zebra finch. We sorted the 50 kb-windows of the medium ground finch reference genome along the zebra finch genome based on pairwise alignment between the two assemblies and mapped the recombination estimates of zebra finch (Singhal et al. 2015) to each window. On average, 91% of the windows had a successful hit to the genome of zebra finch. Average relative recombination rates (ρ) in the genomic islands were significantly lower than the background in all species pairs ($P < 5.1 \times 10^{-3}$; randomization test) (Fig. 4).

We investigated whether we could distinguish a subset of genomic islands involved in speciation or ancient adaptation (Supplemental Fig. S1A,B) from a subset of genomic islands formed due to ongoing background selection (Supplemental Fig. S1D). Genomic islands involved in speciation or ancient adaptation are expected to have elevated d_{XY} but are not expected to be preferentially found in regions with low recombination rates. Genomic islands formed by recurrent background selection are not expected to have elevated d_{XY} but are more likely to occur in regions with lower recombination rates. We therefore examined the distribution of recombination rates and d_{XY} in the genomic island regions. We found that nearly all the islands were below the mean recombination rate, and no significant correlation was observed between recombination rate and d_{XY} (Supplemental Fig. S4). We also estimated recombination rates at the *ALX1* and *HMGA2* loci. We found that these loci also had much lower recombination rates than the genomic background (percentile: *ALX1* = 1.6%, *HMGA2* = 4.3%) (Fig. 4). It therefore seems that these regions, which have been proven to be involved in ecological adaptation, also tend to be associated with lower recombination rates.

Discussion

We identified genomic islands of divergence in several pairwise comparisons of Darwin's finch species using the F_{ST} statistic. These genomic islands are distinguished by the following features: (1) they are elevated in d_{XY} in addition to F_{ST} in most species pairs; (2) they are present in similar numbers and sizes in both allopatric and sympatric comparisons; (3) they include two major regions known to be associated with variation in beak shape and size (*ALX1* and *HMGA2*); (4) they do not form independently in the same location (i.e., islands that have evolved in ground finches and tree finches do not overlap); and (5) they have significantly lower recombination rates than the genome average.

A major aim of this study was to distinguish among four main evolutionary models that can lead to genomic islands of divergence (Supplemental Fig. S1). In model 1, genomic islands

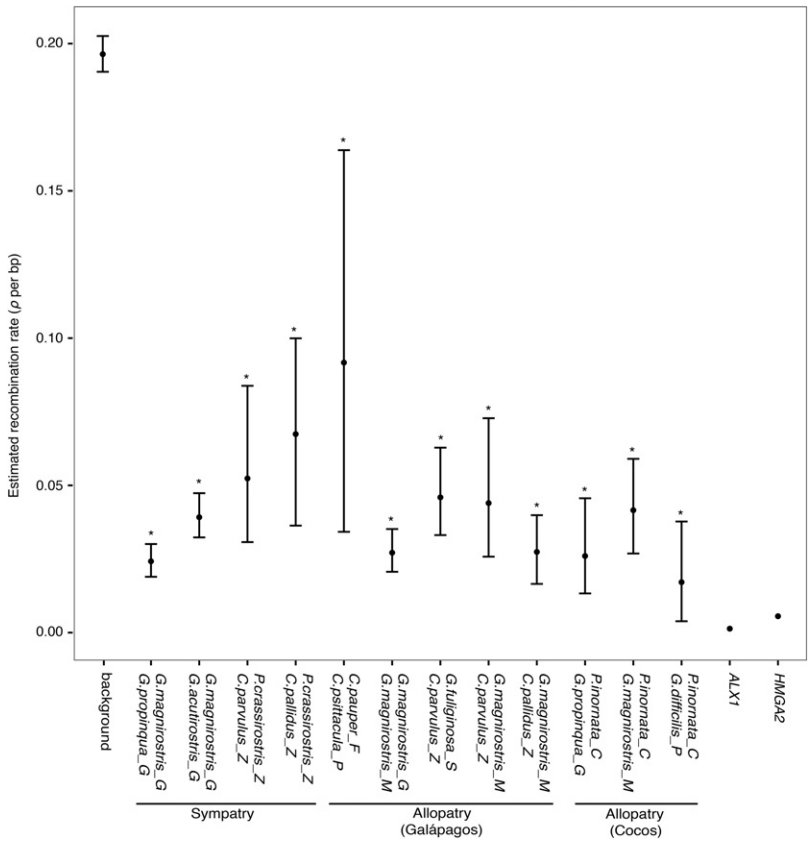

Figure 4. Estimated recombination rate (ρ) in island regions of each pair and at beak-related loci in comparison to the genomic background. Points refer to the average recombination rate in genomic islands from each species pair or beak-related locus. *Upper* and *lower* bounds of each point are the corresponding 95% confidence interval from bootstrapped resamples of each species pair. Significance is tested with a randomization test. (*) $P < 0.05$.

represent barriers to gene flow. In model 2, genomic islands represent ancient balanced polymorphisms. In model 3, genomic islands are formed by recent selection at ecologically relevant loci. In model 4, genomic islands are formed by ongoing background selection and/or recurrent selective sweeps and reflect variation in recombination rate. A striking feature of the genomic islands we observe in Darwin's finches, which is absent from comparisons of many other species (Cruickshank and Hahn 2014), is that many genomic islands are elevated in both d_{XY} and F_{ST}. This result fits a model of speciation where genomic islands harbor haplotypes that became genetically isolated from each other prior to the rest of the genome of the species pairs under comparison. Such a scenario is found in both models 1 and 2.

Another important observation is that genomic islands occur in similar numbers and are elevated in d_{XY} to the same extent in both allopatric and sympatric comparisons. This is not expected with model 1, which predicts genomic islands to be more pronounced when there are more opportunities for gene flow. However, the strict allopatric model of speciation-without-gene-flow is not applicable to Darwin's finches in the Galápagos archipelago (Grant and Grant 2010; Farrington et al. 2014). Alternatively, it may reflect differential gene flow in the past, when ancestral populations of two species were geographically closer or even sympatric. It is known that fluctuating sea level connected and disconnected some islands periodically during the radiation (Ali and Aitchison 2014; Gillespie and Roderick 2014). However, genomic islands are even seen in comparisons between the Cocos finch and species from the Galápagos Islands. Cocos Island lies ~780 km north of Galápagos, and the Cocos finch (*P. inornata*) is the only species of Darwin's finches there. It is highly unlikely that recent gene flow involving this species has occurred. It therefore appears that recent gene flow between species is unlikely to be a major factor in generating genomic islands among Darwin's finches.

We find striking genomic islands in the two loci containing the *ALX1* and *HMGA2* genes that are associated with variation in beak shape and size, respectively (Lamichhaney et al. 2015, 2016). The regions appear among the most distinct genomic islands in comparisons of species that bear different haplogroups at these loci. This overlap suggests that ecological selection at these loci has generated large divergence between haplogroups. Beak morphology of Darwin's finches is adaptive due to its role in feeding habits but is also involved in species discrimination and mate choice (Grant and Grant 1987, 1996, 2008). It is therefore probable that these loci are involved in reproductive isolation.

According to our previous estimates, the haplogroups of *ALX1* began to form ~900,000 years ago, and *HMGA2* formed ~1 million years ago, when finches had just started to radiate (Lamichhaney et al. 2015, 2016). The origin of haplogroups at these loci, and possibly other loci that are associated with morphological variation, is older than the species splits that we have studied here. The deep divergence and patterns of segregation of the haplotypes at these two loci suggest that these genomic regions were involved in the genetic adaptation early in the Darwin's finch radiation, most likely through their association with beak morphology. It is possible that a form of balancing selection maintained haplotypes at these beak-associated loci in the ancestral Darwin's finch population, prior to the species radiation. For example, frequency-dependent selection may have provided a selective advantage to rare beak morphologies, hence preserving haplotype diversity. Furthermore, the nature of selection at these loci may have fluctuated over time and space at these loci, for example, due to changing environmental conditions.

The evolutionary trees of beak-associated loci are discordant with the species tree. This could simply reflect lineage sorting of ancestral haplotypes but could also be influenced by haplotypes periodically crossing species boundaries. Such gene flow could represent adaptive introgression, which has been demonstrated in *G. fortis* on Daphne Major island (Grant and Grant 2014; Lamichhaney et al. 2015, 2016). We find signals indicative of adaptive introgression in our data set. In particular, we find reduced d_{XY} at beak-associated loci compared to the rest of the genome in two pairwise species comparisons where the same haplogroups at the beak loci are present in both species. Such a pattern indicates that gene flow has occurred at beak-associated loci more recently

than in the rest of the genome. We infer that this process of differential gene flow has affected both the *ALX1* and *HMGA2* loci in the Genovesa and Daphne Major populations of *G. magnirostris*. Individuals from these populations fly between islands and have opportunities to interbreed (Grant and Grant 2010). In addition, there is evidence for gene flow at the *HMGA2* locus between *G. fuliginosa* on Santiago and *C. parvulus* on Santa Cruz. This inference is surprising, as there is no evidence of hybridization between these species (Grant 1999). The extreme divergence at these beak-associated loci compared to the rest of the genome coupled with their pattern of segregation among Darwin's finches are consistent with a model where gene flow has been generally disadvantageous at these loci but has periodically occurred, associated with shifts in beak morphology (Lamichhaney et al. 2015, 2016).

We did not find evidence that genomic islands accumulate independently in the same genomic locations along different branches, as demonstrated in pied and collared flycatchers (Burri et al. 2015). In pairwise comparisons within tree and ground finches, few of the genomic islands overlapped (Supplemental Table S1). Hence, different species do not diverge at similar genomic locations in Darwin's finches, which is predicted in the case of model 4, where genomic islands result from recurrent selection and reflect the underlying recombination rate. However, we identify shared islands in multiple comparisons of species that have consistent differences in their haplogroups at beak-associated loci. This reflects the observation that, when the genomes of two species with different haplogroups at beak-associated loci are compared, these loci are detected as genomic islands.

Interestingly, analysis based on the estimation of recombination rate using a map from other bird genomes (Singhal et al. 2015) shows that recombination rate is significantly reduced in genomic islands. This feature, coupled with generally reduced levels of genetic variation and Tajima's *D* observed in genomic islands (Supplemental Fig. S2), suggests that a subset of them are formed as a by-product of linked selection. Although genomic islands formed by background selection (model 4) are predicted to occur in regions of low recombination rate, several features of the genomic islands presented above indicate their involvement in ecological adaptation. First, they do not arise in the same locations in independent comparisons; second, they display elevated d_{XY}; and third, the *ALX1* and *HMGA2* loci, known to be involved in ecological adaptation, have particularly low recombination rates.

Low recombination rates seem to be a consistent feature of the genomic islands we observe, and we do not find a tendency for genomic islands with particularly elevated d_{XY} to occur in genomic regions where recombination is not reduced. These observations indicate a link between low recombination rates and ecological selection, although the reason for this link is unclear. One possibility is that regions with low recombination rates could be more readily involved in "divergence hitchhiking" (Via 2012), whereby divergent selection reduces gene flow at a locus and surrounding linked regions. It is therefore possible that low recombination rates facilitate the emergence of genomic islands due to genetic linkage between multiple functional variants that are inherited together. Genomic islands might also be more prone to occur in clusters of co-expressed genes within regions of low recombination (Hurst et al. 2004). It may also be that high levels of linkage facilitate the formation of supergenes of co-adapted genetic variants maintained in favorable combinations (Thompson and Jiggins 2014). It should, however, be noted that it is possible that genomic islands with low recombination rates tend to be larger and more readily detected in our window-based analysis, leading to an overall bias toward low recombination rates in the genomic islands we identified. The results presented here, where numerous genomic islands are found in both allopatric and sympatric comparisons and are also associated with lower recombination rates, are similar to those reported between closely related species of wild sunflowers in North America (Renaut et al. 2013).

The causes of genomic islands of divergence are debated, and different processes are likely to dominate the landscape of divergence in different speciation events. A key feature of genomic islands that are resistant to gene flow, or that are derived from anciently diverged haplotypes, is that they become elevated in d_{XY}. A re-analysis of several data sets of insects, birds, and mammals found islands of divergence that were elevated in F_{ST} but not d_{XY} (Cruickshank and Hahn 2014). Such patterns can be produced by a number of evolutionary processes unrelated to speciation. Notably, in *Ficedula* flycatchers, the landscape of divergence evolves mainly as a result of heterogeneity in the underlying recombination rate (Burri et al. 2015). In contrast, however, at least a subset of genomic islands in Darwin's finches is involved in ecological adaptation. These highly diverged regions may eventually lead to genomic incompatibilities, although this stage is unknown to have been reached yet in Darwin's finches (Grant and Grant 2008).

Our data are consistent with an evolutionary scenario where genomic islands of divergence have evolved throughout the radiation of the Darwin's finches, most often in regions with low recombination and often due to genetic adaptation, as exemplified by the beak loci. Introgression may occur through inter-island movements and interbreeding. As divergence proceeds further, introgression gradually diminishes as the islands are now "protected" by selective exclusion of foreign gene regions corresponding to the islands, and d_{XY} rises above background level. Later still, interbreeding ceases, or all but ceases, and the number of islands decreases as divergence of the background increases and the islands no longer stand out: metaphorically, the sea rises and the islands disappear.

Methods

Selection of species pairs and phylogeny reconstruction

All the whole-genome sequence data used in our study are from previous publications (Lamichhaney et al. 2015, 2016) under accession numbers PRJNA263122 and PRJNA301892 at the NCBI Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra). The total sequence length of the reference assembly (geoFor1) is 1,065,292,181 bp, and it contains 27,239 scaffolds with an N50 scaffold size of 5,255,844 bp. We aligned reads from each individual against the reference assembly using BWA v0.6.2 (Li and Durbin 2009) under BWA mem algorithm with default parameters. The alignments were strictly checked, filtered, and realigned with Picard toolkit v1.134. We performed variant calling using GATK v3.3 (McKenna et al. 2010). Stringent filtering of raw variants was applied to exclude SNPs with low quality or bad mapping based on the empirical distribution. We used BEAGLE v3.3.2 (Browning and Browning 2007) to infer the missing genotype for each population. Only biallelic variants were included in this study. We selected nine species from six islands from the Galápagos archipelago and one from Cocos Island and designed 12 species pairs based on the ranges of sympatry and allopatry and their beak haplogroups (Tables 1, 2). About 7 million SNPs were retained in each species pair. To test the influence of gene flow on the islands of divergence, allopatric comparisons within central Galápagos and the ones between populations on Cocos

Island and on Galápagos were separated. Two additional outgroup pairs were introduced involving *L. noctis* and *T. bicolor* sampled from Barbados. In each sympatric/allopatric group, we chose pairs to represent all possible different beak haplogroup combinations. We calculated allele frequency for each species and utilized PHYLIP (http://evolution.genetics.washington.edu/phylip.html) to construct a species tree from genetic distance using all the polymorphic autosomal sites from 14 species.

Genome-wide screen of divergence

For each species pair, a whole-genome scan with 50-kb nonoverlapping windows was performed for measuring both within- and between-species parameters. The former measures included nucleotide diversity (π) and sequence neutrality (Tajima's D), and the latter measures included density of fixed differences (d_f), between-population divergence (d_{XY}), and between-population difference of allele frequency (F_{ST}). We used VCFTools (v0.1.13) (Danecek et al. 2011) to estimate π, Tajima's D, and the Weir and Cockerham estimator of F_{ST}, and custom Perl scripts to calculate d_{XY} and d_f (see Supplemental Material). Between-population sequence divergence d_{XY} was estimated according to the formula described by Nei and Li (1979) as

$$d_{XY} = \sum_{ij} x_i y_j d_{ij},$$

where x_i is the ith sequence from population X, y_j is the jth sequence from population Y, and d_{ij} is the number of nucleotide differences between ith and jth sequences. Fixed difference d_f was defined as the number of variants that showed one allele in all samples of one population and meanwhile showed the other alternate allele in the compared species. All the parameters were calculated on a per-base level and averaged according to the effective window size. Since the reference genome assembly is not assembled to chromosome level, the screen was done on a per-scaffold level. This could cause problems on small scaffolds due to low mapping quality. To eliminate the false positive signals caused by low-quality mapping and variant calling, we excluded the very end window of each scaffold and the windows with less than 50 SNPs. Scaffolds smaller than the window size were not taken into account.

Analysis of the putative islands

To compare the genomic landscapes in pairs with different divergence time, we standardized per window F_{ST} in each pair to a Z score based on the formula

$$ZF_{ST} = \frac{\text{per window } F_{ST} - \text{mean } F_{ST} \text{ of the pairwise comparison}}{\text{standard deviation of } F_{ST} \text{ of the pairwise comparison}},$$

and determined windows with ZF_{ST} of at least four (four standard deviations departing from the median F_{ST}) as outliers based on the ZF_{ST} distributions of pairs. This level corresponds to the 99.2% percentile of windows from all the comparisons. The average F_{ST} of the outlier windows from all pairs is about 0.57. Adjacent outlier windows were considered in the same island and were merged into larger divergent regions for evaluating the effect of different hypotheses on the size of islands. To compare d_{XY} in the genomic islands to the background, we estimated the divergence level of background from 10,000 random resampling replicates from individual pair. Three comparisons between Darwin's finches (*G. magnirostris_G*, *C. pauper_F*, *P. inornata_C*) and the outgroup were made in order to examine the variation of substitution rate in the genomic islands. We extracted the location of the genomic islands from pairwise comparisons of Darwin's finch species and measured d_{XY} in same location of the outgroup comparisons. Background level of divergence was estimated from 10,000 random resampling replicates of individual outgroup comparisons.

Estimate of recombination rate

To make it comparable with other avian genomes, we utilized the UCSC liftOver tool (Hinrichs et al. 2006) and chain file to convert synteny blocks of the medium ground finch to genome coordinates of the zebra finch (TaeGut1) (https://genome.ucsc.edu/index.html). In order to keep the intrinsic integrity of the variants within each block, we did the conversion for the midpoint of each window and sorted all the blocks along the assembly of the zebra finch. About 91% of the windows were successfully lifted over. The average relative recombination rate of each window was calculated based on a fine-scale recombination map of zebra finch from Singhal et al. (2015), which was generated in a linkage-disequilibrium approach from haplotype estimation by the program LDhelmet v1.6 (Chan et al. 2012). The penalty block was set at 5 after a series of simulations, with a burn-in of 1×10^5 steps and 1×10^6 steps in the MCMC chain. Only the 18 largest chromosomes were included in the recombination map. The estimated recombination rate (ρ) from LDhelmet is a population-scaled recombination rate coefficient, which could be converted to cM/Mb by $\rho = 4N_e r$, where N_e is the effective population size and r is the genetic distance over an interval. Significance was tested by randomly sampling 10,000 replicates from background level of recombination rate in individual comparisons. Plots of the measures along the genomic position and statistical significance were generated from means per window with custom R scripts (R Core Team 2013).

Data access

All custom scripts from this study have been submitted to https://github.com/Fan-Han/Genomic_island and are available as Supplemental Material. Genotypes are available as VCF files from the Dryad Digital Repository (http://datadryad.org/) with the doi: 10.5061/dryad.7155d.

Acknowledgments

The US National Science Foundation (NSF) funded the collection of material under permits from the Galápagos and Costa Rica National Parks Services and the Charles Darwin Research Station and in accordance with protocols of Princeton University's Animal Welfare Committee. The project was supported by the Knut and Alice Wallenberg Foundation. Computer resources were supplied by UPPMAX. We thank Miguel Carneiro and Jonas Berglund for helpful comments and suggestions on the manuscript.

Author contributions: M.T.W., L.A., P.R.G., and B.R.G. conceived the study. P.R.G. and B.R.G. collected the material. M.T. W. and L.A. led the bioinformatic analysis of data. F.H. and S.L. performed the bioinformatic analysis. All authors were involved in writing the manuscript and approved it before submission.

References

Ali JR, Aitchison JC. 2014. Exploring the combined role of eustasy and oceanic island thermal subsidence in shaping biodiversity on the Galápagos. *J Biogeogr* **41:** 1227–1241.

Begun DJ, Aquadro CF. 1992. Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature* **356:** 519–520.

Berg PR, Star B, Pampoulie C, Sodeland M, Barth JM, Knutsen H, Jakobsen KS, Jentoft S. 2016. Three chromosomal rearrangements promote genomic divergence between migratory and stationary ecotypes of Atlantic cod. *Sci Rep* **6:** 23246.

Browning SR, Browning BL. 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet* **81:** 1084–1097.
Burri R, Nater A, Kawakami T, Mugal CF, Olason PI, Smeds L, Suh A, Dutoit L, Bures S, Garamszegi LZ, et al. 2015. Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. *Genome Res* **25:** 1656–1665.
Carneiro M, Albert FW, Afonso S, Pereira RJ, Burbano H, Campos R, Melo-Ferreira J, Blanco-Aguiar JA, Villafuerte R, Nachman MW, et al. 2014. The genomic architecture of population divergence between subspecies of the European rabbit. *PLoS Genet* **10:** e1003519.
Chan AH, Jenkins PA, Song YS. 2012. Genome-wide fine-scale recombination rate variation in *Drosophila melanogaster*. *PLoS Genet* **8:** e1003090.
Clarkson CS, Weetman D, Essandoh J, Yawson AE, Maslen G, Manske M, Field SG, Webster M, Antao T, MacInnis B, et al. 2014. Adaptive introgression between *Anopheles* sibling species eliminates a major genomic island but not reproductive isolation. *Nat Commun* **5:** 4248.
Coyne JA, Orr HA. 2004. *Speciation*. Sinauer Associates, Sunderland, MA.
Cruickshank TE, Hahn MW. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol Ecol* **23:** 3133–3157.
Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. *Bioinformatics* **27:** 2156–2158.
Ellegren H, Smeds L, Burri R, Olason PI, Backstrom N, Kawakami T, Kunstner A, Makinen H, Nadachowska-Brzyska K, Qvarnstrom A, et al. 2012. The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature* **491:** 756–760.
Farrington HL, Lawson LP, Clark CM, Petren K. 2014. The evolutionary history of Darwin's finches: speciation, gene flow, and introgression in a fragmented landscape. *Evolution* **68:** 2932–2944.
Feulner PGD, Chain FJJ, Panchal M, Huang Y, Eizaguirre C, Kalbe M, Lenz TL, Samonte IE, Stoll M, Bornberg-Bauer E, et al. 2015. Genomics of divergence along a continuum of parapatric population differentiation. *PLoS Genet* **11:** e1004966.
Gillespie RG, Roderick GK. 2014. Evolution: geology and climate drive diversification. *Nature* **509:** 297–298.
Grant PR. 1999. *Ecology and evolution of Darwin's finches*. Princeton University Press, Princeton, NJ.
Grant BR, Grant PR. 1987. Mate choice in Darwin's finches. *Biol J Linn Soc* **32:** 247–270.
Grant BR, Grant PR. 1989. *Evolutionary dynamics of a natural population: the large cactus finch of the Galápagos*. University of Chicago Press, Chicago.
Grant BR, Grant PR. 1996. High survival of Darwin's finch hybrids: effects of beak morphology and diets. *Ecology* **77:** 500–509.
Grant PR, Grant BR. 2008. *How and why species multiply: the radiation of Darwin's finches*. Princeton University Press, Princeton, NJ.
Grant PR, Grant BR. 2010. Conspecific versus heterospecific gene exchange between populations of Darwin's finches. *Philos Trans R Soc Lond B Biol Sci* **365:** 1065.
Grant PR, Grant BR. 2014. *Forty years of evolution: Darwin's finches on Daphne Major Island*. Princeton University Press, Princeton, NJ.
Harr B. 2006. Genomic islands of differentiation between house mouse subspecies. *Genome Res* **16:** 730–737.
Hinrichs A, Karolchik D, Baertsch R, Barber G, Bejerano G, Clawson H. 2006. The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res* **34:** D590–D598.
Huber SK, De León LF, Hendry AP, Bermingham E, Podos J. 2007. Reproductive isolation of sympatric morphs in a population of Darwin's finches. *Proc Biol Sci* **274:** 1709–1714.
Hurst LD, Pál C, Lercher MJ. 2004. The evolutionary dynamics of eukaryotic gene order. *Nat Rev Genet* **5:** 299–310.
Irwin DE, Alcaide M, Delmore KE, Irwin JH, Owens GL. 2016. Recurrent selection explains parallel evolution of genomic regions of high relative but low absolute differentiation in a ring species. *Mol Ecol* **25:** 4488–4507.
Lamichhaney S, Berglund J, Almen MS, Maqbool K, Grabherr M, Martinez-Barrio A, Promerova M, Rubin CJ, Wang C, Zamani N, et al. 2015. Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* **518:** 371–375.
Lamichhaney S, Han F, Berglund J, Wang C, Almen MS, Webster MT, Grant BR, Grant PR, Andersson L. 2016. A beak size locus in Darwin's finches facilitated character displacement during a drought. *Science* **352:** 470–474.
Lawniczak MK, Emrich SJ, Holloway AK, Regier AP, Olson M, White B, Redmond S, Fulton L, Appelbaum E, Godfrey J, et al. 2010. Widespread divergence between incipient *Anopheles gambiae* species revealed by whole genome sequences. *Science* **330:** 512–514.
Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25:** 1754–1760.
Malinsky M, Challis RJ, Tyers AM, Schiffels S, Terai Y, Ngatunga BP, Miska EA, Durbin R, Genner MJ, Turner GF. 2015. Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science* **350:** 1493–1498.
Marques DA, Lucek K, Meier JI, Mwaiko S, Wagner CE, Excoffier L, Seehausen O. 2016. Genomics of rapid incipient speciation in sympatric threespine stickleback. *PLoS Genet* **12:** e1005887.
Martin SH, Dasmahapatra KK, Nadeau NJ, Salazar C, Walters JR, Simpson F, Blaxter M, Manica A, Mallet J, Jiggins CD. 2013. Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Res* **23:** 1817–1828.
Mayr E. 1942. *Systematics and the origin of species: from the viewpoint of a zoologist*. Harvard University Press, Cambridge, MA.
McGee MD, Neches RY, Seehausen O. 2016. Evaluating genomic divergence and parallelism in replicate ecomorphs from young and old cichlid adaptive radiations. *Mol Ecol* **25:** 260–268.
McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20:** 1297–1303.
Nadeau NJ, Whibley A, Jones RT, Davey JW, Dasmahapatra KK, Baxter SW, Quail MA, Joron M, ffrench-Constant RH, Blaxter ML, et al. 2012. Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Philos Trans R Soc Lond B Biol Sci* **367:** 343–353.
Nei M, Li WH. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci* **76:** 5269–5273.
Noor MAF, Bennett SM. 2009. Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity (Edinb)* **103:** 439–444.
Nosil P, Harmon LJ, Seehausen O. 2009. Ecological explanations for (incomplete) speciation. *Trends Ecol Evol* **24:** 145–156.
Pease JB, Haak DC, Hahn MW, Moyle LC. 2016. Phylogenomics reveals three sources of adaptive variation during a rapid radiation. *PLoS Biol* **14:** e1002379.
Petren K, Grant PR, Grant BR, Keller LF. 2005. Comparative landscape genetics and the adaptive radiation of Darwin's finches: the role of peripheral isolation. *Mol Ecol* **14:** 2943–2957.
Poelstra JW, Vijay N, Bossu CM, Lantz H, Ryll B, Muller I, Baglione V, Unneberg P, Wikelski M, Grabherr MG, et al. 2014. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science* **344:** 1410–1414.
R Core Team. 2013. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.
Renaut S, Grassa CJ, Yeaman S, Moyers BT, Lai Z, Kane NC, Bowers JE, Burke JM, Rieseberg LH. 2013. Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nat Commun* **4:** 1827.
Seehausen O, Butlin RK, Keller I, Wagner CE, Boughman JW, Hohenlohe PA, Peichel CL, Saetre GP, Bank C, Brannstrom A, et al. 2014. Genomics and the origin of species. *Nat Rev Genet* **15:** 176–192.
Sicard A, Kappel C, Josephs EB, Lee YW, Marona C, Stinchcombe JR, Wright SI, Lenhard M. 2015. Divergent sorting of a balanced ancestral polymorphism underlies the establishment of gene-flow barriers in *Capsella*. *Nat Commun* **6:** 7960.
Singhal S, Leffler EM, Sannareddy K, Turner I, Venn O, Hooper DM, Strand AI, Li QY, Raney B, Balakrishnan CN, et al. 2015. Stable recombination hotspots in birds. *Science* **350:** 928–932.
Soria-Carrasco V, Gompert Z, Comeault AA, Farkas TE, Parchman TL, Johnston JS, Buerkle CA, Feder JL, Bast J, Schwander T, et al. 2014. Stick insect genomes reveal natural selection's role in parallel speciation. *Science* **344:** 738–742.
Supple MA, Papa R, Hines HM, McMillan WO, Counterman BA. 2015. Divergence with gene flow across a speciation continuum of *Heliconius* butterflies. *BMC Evol Biol* **15:** 204.
Thompson MJ, Jiggins CD. 2014. Supergenes and their role in evolution. *Heredity (Edinb)* **113:** 1–8.
Turner TL, Hahn MW, Nuzhdin SV. 2005. Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biol* **3:** e285.
Via S. 2012. Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philos Trans R Soc Lond B Biol Sci* **367:** 451–460.
Wang J, Street NR, Scofield DG, Ingvarsson PK. 2016. Variation in linked selection and recombination drive genomic divergence during allopatric speciation of European and American aspens. *Mol Biol Evol* **33:** 1754–1767.
Wolf JB, Ellegren H. 2016. Making sense of genomic islands of differentiation in light of speciation. *Nat Rev Genet* **18:** 87–100.
Wu CI. 2001. The genic view of the process of speciation. *J Evol Biol* **14:** 851–865.

Received July 8, 2016; accepted in revised form February 14, 2017.