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. Webster1

1Department of Medical Biochemistry and Microbiology, Science for Life Laboratory, Uppsala University, 75123 Uppsala, Sweden; 2Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey 08544-2016, USA; 3Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, 75007 Uppsala, Sweden; 4Department 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 (FST) 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 (FST) and absolute divergence (dxy). 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 dxy 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 HMGAL2 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; Renault 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 FST-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 FST 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, dxy, 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
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-without-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_{ST} 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 (Sacard 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 descendant 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 (d_{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 $\sim 10\times$ coverage per individual. Approximately 7 million variant sites were identified on average in each species pair, using stringent criterion for SNP calling.

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 ($\pi$), 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 coefficient test).
Table 1. Species names, geographic islands, and sample size of Darwin’s finches and outgroup species

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

coefficients (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 FST between pairs in sympatry and in allopatry, which may have otherwise biased this comparison (mean FST: 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 (Δe) 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

<table>
<thead>
<tr>
<th>Geographic distribution</th>
<th>Species pairs</th>
<th>ALX1 haplogroup</th>
<th>HMGA2 haplogroup</th>
<th>Divergence time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sympathy</td>
<td>G. magnirostris_G</td>
<td>Blunt</td>
<td>Large</td>
<td>172 ± 13</td>
</tr>
<tr>
<td></td>
<td>G. propinqua_G</td>
<td>Blunt</td>
<td>Large</td>
<td>220 ± 3</td>
</tr>
<tr>
<td></td>
<td>G. acutirostris_G</td>
<td>Blunt</td>
<td>Small</td>
<td>517 ± 3</td>
</tr>
<tr>
<td></td>
<td>P. crassirostris_Z</td>
<td>Pointed</td>
<td>Large</td>
<td>521 ± 11</td>
</tr>
<tr>
<td></td>
<td>C. pauper_F</td>
<td>Pointed</td>
<td>Small</td>
<td>185 ± 6</td>
</tr>
<tr>
<td></td>
<td>C. psittacula_P</td>
<td>Pointed</td>
<td>Large</td>
<td>83 ± 8</td>
</tr>
<tr>
<td></td>
<td>G. magnirostris_G</td>
<td>Blunt</td>
<td>Large</td>
<td>353 ± 1</td>
</tr>
<tr>
<td></td>
<td>G. magnirostris_M</td>
<td>Blunt</td>
<td>Small</td>
<td>347 ± 4</td>
</tr>
<tr>
<td></td>
<td>G. fuliginosa_S</td>
<td>Pointed</td>
<td>Small</td>
<td>367 ± 11</td>
</tr>
<tr>
<td></td>
<td>C. pauper_F</td>
<td>Pointed</td>
<td>Small</td>
<td>409 ± 3</td>
</tr>
<tr>
<td></td>
<td>C. psittacula_P</td>
<td>Pointed</td>
<td>Large</td>
<td>444 ± 5</td>
</tr>
<tr>
<td>Allopatry (Galápagos)</td>
<td>G. magnirostris_M</td>
<td>Blunt</td>
<td>Large</td>
<td>520 ± 2</td>
</tr>
<tr>
<td></td>
<td>C. pauper_F</td>
<td>Pointed</td>
<td>Large</td>
<td>409 ± 3</td>
</tr>
<tr>
<td></td>
<td>C. psittacula_P</td>
<td>Pointed</td>
<td>Large</td>
<td>444 ± 5</td>
</tr>
<tr>
<td>Allopatry (Cocos)</td>
<td>P. inornata_C</td>
<td>Pointed</td>
<td>Small</td>
<td>520 ± 2</td>
</tr>
<tr>
<td></td>
<td>G. propinqua_G</td>
<td>Pointed</td>
<td>Large</td>
<td>409 ± 3</td>
</tr>
<tr>
<td></td>
<td>P. inornata_C</td>
<td>Pointed</td>
<td>Small</td>
<td>444 ± 5</td>
</tr>
<tr>
<td></td>
<td>G. magnirostris_M</td>
<td>Blunt</td>
<td>Large</td>
<td>520 ± 2</td>
</tr>
</tbody>
</table>

*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.

In thousands of years, based on data from Lamichhaney et al. (2015).
windows across comparisons is low, which indicates that the genic-
omic 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 addi-
tion, 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 hap-
lotypes 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 com-
pare 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 sympathy: \( P < 0.03 \); 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: \( P < 1 \times 10^{-4} \); randomiza-
tion test), except in one comparison (\( G. m. G \) vs. \( G. m. M \)), where \( d_{XY} \) 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 indi-
cates 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 com-
paring 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 (\( G. m. G \) vs.
\( L. n. \), \( C. p. F \) vs. \( L. n. \), \( P. i. C \) vs. \( L. n. \)) (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**

<table>
<thead>
<tr>
<th>Geographic distribution</th>
<th>Species pairs a</th>
<th>Genome-wide mean ( F_{ST} )</th>
<th>Island mean ( F_{ST} )</th>
<th>No. of islands</th>
<th>Mean size b</th>
<th>Max size b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sympathy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( G. m. G )</td>
<td>0.11</td>
<td>0.44</td>
<td>58</td>
<td>71.6</td>
<td>350</td>
</tr>
<tr>
<td></td>
<td>( G. p. G )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( G. m. G )</td>
<td>0.14</td>
<td>0.64</td>
<td>84</td>
<td>107.7</td>
<td>600</td>
</tr>
<tr>
<td></td>
<td>( G. m. G )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( P. c. Z )</td>
<td>0.27</td>
<td>0.52</td>
<td>36</td>
<td>63.9</td>
<td>150</td>
</tr>
<tr>
<td></td>
<td>( C. p. P )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( G. m. G )</td>
<td>0.10</td>
<td>0.38</td>
<td>61</td>
<td>66.9</td>
<td>300</td>
</tr>
<tr>
<td></td>
<td>( G. m. M )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( L. n. )</td>
<td>0.29</td>
<td>0.73</td>
<td>42</td>
<td>67.9</td>
<td>300</td>
</tr>
<tr>
<td></td>
<td>( T. b. )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allopatry (Galápagos)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( C. p. F )</td>
<td>0.09</td>
<td>0.32</td>
<td>61</td>
<td>64.8</td>
<td>300</td>
</tr>
<tr>
<td></td>
<td>( C. p. P )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( G. m. G )</td>
<td>0.12</td>
<td>0.36</td>
<td>52</td>
<td>99.0</td>
<td>1050</td>
</tr>
<tr>
<td></td>
<td>( C. p. Z )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( G. m. M )</td>
<td>0.21</td>
<td>0.52</td>
<td>54</td>
<td>66.7</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>( C. p. Z )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( G. m. M )</td>
<td>0.29</td>
<td>0.73</td>
<td>42</td>
<td>67.9</td>
<td>300</td>
</tr>
<tr>
<td>Allopatry (Cocos Island)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( P. i. C )</td>
<td>0.26</td>
<td>0.50</td>
<td>45</td>
<td>66.7</td>
<td>250</td>
</tr>
<tr>
<td></td>
<td>( G. p. G )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( P. i. C )</td>
<td>0.31</td>
<td>0.68</td>
<td>61</td>
<td>75.4</td>
<td>550</td>
</tr>
<tr>
<td></td>
<td>( G. m. M )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( P. d. P )</td>
<td>0.41</td>
<td>0.67</td>
<td>6</td>
<td>58.3</td>
<td>100</td>
</tr>
<tr>
<td>Outgroup</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( G. f. S )</td>
<td>0.49</td>
<td>0.78</td>
<td>20</td>
<td>82.5</td>
<td>400</td>
</tr>
<tr>
<td></td>
<td>( L. n. )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( P. i. C )</td>
<td>0.70</td>
<td>–</td>
<td>0</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>( T. b. )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

aLetter 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.

bNumbers 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, (AG) allopatry within Galápagos; (AC) 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 LRRQ1, 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, MSRBB3, LEMD3, and WIFI 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, $F_{ST}$ 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, $P$. cassinistris 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 ($G$. magnisnistris $G$ vs. $G$. magnisnistris $M$ 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 $G$. magnisnistris $G$ vs. $G$. magnisnistris $M$, 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 introgression 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 $G$. magnisnistris 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 recombinational 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 ($p$) 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...
genomic islands are elevated in both of many other species (Cruickshank and Hahn 2014), is that many
In model 4, genomic islands are formed by ongoing background selections and/or recurrent selective sweeps and reflect variation in recombination rate. A striking feature of the genomic islands 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 2010; 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 \(\sim 900,000\) years ago, and \(HMGA2\) formed \(\sim 1\) 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. magisteri. 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. parvalus 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 allopatic 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. Introggression 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 sympathy 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. nootis and T. bicolor sampled from Barbados. In each sympatric/allographic 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), 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 (see Supplemental Material). Between-population sequence divergence d_{XY} was estimated according to the formula described by Nei and Li (1979) as

\[ d_{XY} = \sum_{i} x_i y_i d_{ij}, \]

where \( x_i \) is the \( i \)th sequence from population \( X \), \( y_j \) is the \( j \)th sequence from population \( Y \), and \( d_{ij} \) is the number of nucleotide differences between \( i \)th and \( j \)th sequences. Fixed difference \( d_i \) 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

\[ Z_{F_{ST}} = \frac{\text{per window } F_{ST} - \text{mean } F_{ST} \text{ of the pairwise comparison}}{\text{standard deviation of } F_{ST} \text{ of the pairwise comparison}}, \]

and determined windows with \( Z_{F_{ST}} \) of at least four (four standard deviations departing from the median \( F_{ST} \)) as outliers based on the \( Z_{F_{ST}} \) 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. parufer \( F \), 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 \( 1 \times 10^{5} \) steps and \( 1 \times 10^{5} \) steps in the MCMC chain. Only the 18 largest chromosomes were included in the recombination map. The estimated recombination rate (\( \rho \)) 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**


Genomic divergence among Darwin’s finches


Received July 8, 2016; accepted in revised form February 14, 2017.
Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among Darwin’s finches

Fan Han, Sangeet Lamichhaney, B. Rosemary Grant, et al.

Genome Res. 2017 27: 1004-1015 originally published online April 25, 2017
Access the most recent version at doi:10.1101/gr.212522.116

Supplemental Material
http://genome.cshlp.org/content/suppl/2017/04/25/gr.212522.116.DC1

References
This article cites 50 articles, 15 of which can be accessed free at:
http://genome.cshlp.org/content/27/6/1004.full.html#ref-list-1

Creative Commons License
This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see http://genome.cshlp.org/site/misc/terms.xhtml). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.

Email Alerting Service
Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.