EVOLUTIONARY BIOLOGY

Rapid adaptive radiation of Darwin's finches depends on ancestral genetic modules

Carl-Johan Rubin[1,2]†, Erik D. Enbody[1]†‡, Mariya P. Dobreva[3], Arhat Abzhanov[3], Brian W. Davis[4], Sangeet Lamichhaney[5], Mats Pettersson[1], Ashley T. Sendell-Price[1,6], C. Grace Sprehn[1]§, Carlos A. Valle[7], Karla Vasco[7], Ola Wallerman[1], B. Rosemary Grant[8], Peter R. Grant[8], Leif Andersson[1,4,9]*

Recent adaptive radiations are models for investigating mechanisms contributing to the evolution of biodiversity. An unresolved question is the relative importance of new mutations, ancestral variants, and introgressive hybridization for phenotypic evolution and speciation. Here, we address this issue using Darwin's finches and investigate the genomic architecture underlying their phenotypic diversity. Admixture mapping for beak and body size in the small, medium, and large ground finches revealed 28 loci showing strong genetic differentiation. These loci represent ancestral haplotype blocks with origins predating speciation events during the Darwin's finch radiation. Genes expressed in the developing beak are overrepresented in these genomic regions. Ancestral haplotypes constitute genetic modules for selection and act as key determinants of the unusual phenotypic diversity of Darwin's finches. Such ancestral haplotype blocks can be critical for how species adapt to environmental variability and change.

INTRODUCTION

Identification of factors that promote or constrain the process of adaptive radiation—rapid morphological and ecological diversification from a common ancestor—provides opportunities for understanding the origins of biodiversity. Species that radiate rapidly are thought to share some common features (*1*), promoting their ability to evolve into diverse forms (*2*, *3*), whereas depauperate clades may lack them (*4*, *5*). One such feature, evolvability, may be determined in part by the modularity of phenotypic traits (*6*), allowing some species to exploit ecological opportunity more readily (*3*). Two factors that influence why some species exhibit greater evolvability than others are phenotypic plasticity and the genetic potential for diversification (*3*). While rapid speciation in adaptive radiations provides limited time to generate de novo genetic variation, ancestral polymorphisms can facilitate rapid accumulation of diverse combinations of alleles (*7*–*12*). Under this model, ancestral variation is sorted in unique combinations in descendent lineages (*13*) and/or is transmitted across lineages through introgression (*9*, *11*). Introgressive hybridization may lead to loss of genetic differentiation (*14*, *15*); on the other hand, it enhances the potential for selection by increasing phenotypic and genetic variation (*16*, *17*). Identification of genetic variants underlying phenotypic variation is essential for understanding the role of ancestral genetic variation in evolutionary change. This remains an outstanding challenge when comparing species because causal variants are greatly outnumbered by neutral variants. However, recent adaptive radiations, particularly those that still hybridize, are excellent groups for studying the origins of genetic variation and their effects on phenotype because gene flow has homogenized the genetic background, thus facilitating the identification of loci contributing to phenotypic differences among species (*18*, *19*).

The Darwin's finch radiation comprises 18 species, 17 present in Galápagos and one on Cocos Island. The group is highly unusual in that no species is known to have become extinct because of human activities, in contrast to some other avian radiations (*20*). The species have experienced current and historical gene flow (*21*–*24*), and diversification involved a key ecological trait, beak morphology, that mediates the efficient use of different food sources (insects, seeds of various sizes, cactus fruits, and even blood from other birds) (*25*). Previous genetic studies have revealed a few loci where ancestral alleles explain variation in beak morphology: *ALX1* affecting beak shape (*21*), a genomic region controlling beak size including *HMGA2* and three other genes (*MSRB3*, *LEMD3*, and *WIF1*) (*26*, *27*) and, in addition, a number of suggestive loci under selection (*21*, *26*–*28*). Whether the presence of ancestral alleles is a general property of the genetic architecture, governing phenotypic variation in the radiation is unknown. Here, we present a high-quality chromosome-scale reference genome and leverage a natural scaling transformation in beak size (*29*) across three species of ground finches (*Geospiza*) to identify 28 loci under selection. We show that the origin of these haplotype blocks linked to phenotypic divergence predates speciation events. These genetic modules have been reused over the past million years, were exchanged by gene flow, and contributed to the rapid phenotypic evolution and speciation among Darwin's finches.

RESULTS

High-quality assembly of the *Camarhynchus parvulus* genome

The previously reported genome assembly based on Illumina short reads of a medium ground finch (*Geospiza fortis*) is highly fragmented

[1]Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden. [2]Institute of Marine Research, Nordnesgaten 50, 5005 Bergen, Norway. [3]Department of Life Sciences, Imperial College London, Silwood Park Campus, SL5 7PY Ascot, UK. [4]Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, TX, USA. [5]Department of Biological Sciences, Kent State University, Kent, OH, USA. [6]Edward Grey Institute, Department of Zoology, University of Oxford, Oxford, UK. [7]Colegio de Ciencias Biológicas y Ambientales, Galápagos Science Center GSC, Universidad San Francisco de Quito USFQ, Quito, Ecuador. [8]Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, USA. [9]Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden.
*Corresponding author. Email: leif.andersson@imbim.uu.se
†These authors contributed equally to this work.
‡Present address: Department of Biomolecular Engineering, University of California, Santa Cruz, Santa Cruz, CA, USA.
§Present address: DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA, USA.

(contig N50 = 30.5 kb) (*30*). We therefore developed a high-quality, highly contiguous assembly for Darwin's finches by combining long-read data with chromatin contact (HiC) data (fig. S1). Because of uncertainty in safely exporting tissue samples with intact long DNA molecules from the Galápagos National Park, we carried out Oxford Nanopore Technologies (ONT) sequencing at the Galápagos Science Center. Genomic DNA prepared from a male small tree finch (*C. parvulus*) was of high molecular weight, and this individual was selected for ONT sequencing. The close evolutionary relationship among all species of Darwin's finches (pairwise interspecies d_{xy} in the range of 0.2 to 0.3% (*31*)) implies that this reference assembly can be used across the phylogeny. We generated 35x ONT long-read sequence coverage from the reference individual and assembled the haploid genome (fig. S1). Erroneous base calls were corrected using linked-read data, and chromosome-sized scaffolds were generated using HiC data. The resulting assembly is of similar quality to current state-of-the art genome assemblies in contiguity and accuracy [96% of the sequence assigned to chromosomes, scaffold N50 = 71.1 Mb, Benchmarking Universal Single-Copy Orthologs (BUSCO) = 96.1% complete, gaps = 0.01%] and shows a high degree of conserved synteny to the zebra finch genome assembly (Fig. 1A). Field-collected tissue samples were used to generate RNA sequencing (RNA-seq) data for annotation (table S1 and fig. S1). Genome annotation of these data using the Ensembl annotation pipeline (*32*) generated 17,167 gene modules that include noncoding RNA and microRNA. We further used 25 *C. parvulus* individuals to generate a linkage disequilibrium–based recombination map (fig. S2) (*33*). Consistent with other avian species (*34*, *35*), recombination rates are generally elevated at the ends of chromosomes and correlated with nucleotide diversity [coefficient of determination (R^2) = 0.19, $P < 0.001$], particularly on the Z chromosome ($R^2 = 0.27$, $P < 0.001$). This relationship is consistent with the widespread effects of background selection known in birds (*35*, *36*).

Twenty-eight trait loci explaining phenotypic differentiation

We used admixture mapping (*37*) to search for loci contributing to genetic differentiation among three closely related species that differ primarily in a scaling factor for beak and body size from small to large (*29*): the small, medium, and large ground finches (*Geospiza fuliginosa*, *G. fortis*, and *Geospiza magnirostris*, respectively) (fig. S3). This trio was selected on the basis of their notable phenotypic differentiation in beak and body traits alone and low genome-wide genetic differentiation (pairwise F_{ST} = 0.02 to 0.10) because this reduces the background noise due to genetic drift. In this study, we generated whole-genome, short-read sequence data from 28 individuals across the three species and combined them with previously published samples for a total of 75 birds on nine islands (mean coverage = 17 ± 9; table S2 and data S1) and applied phenotypic

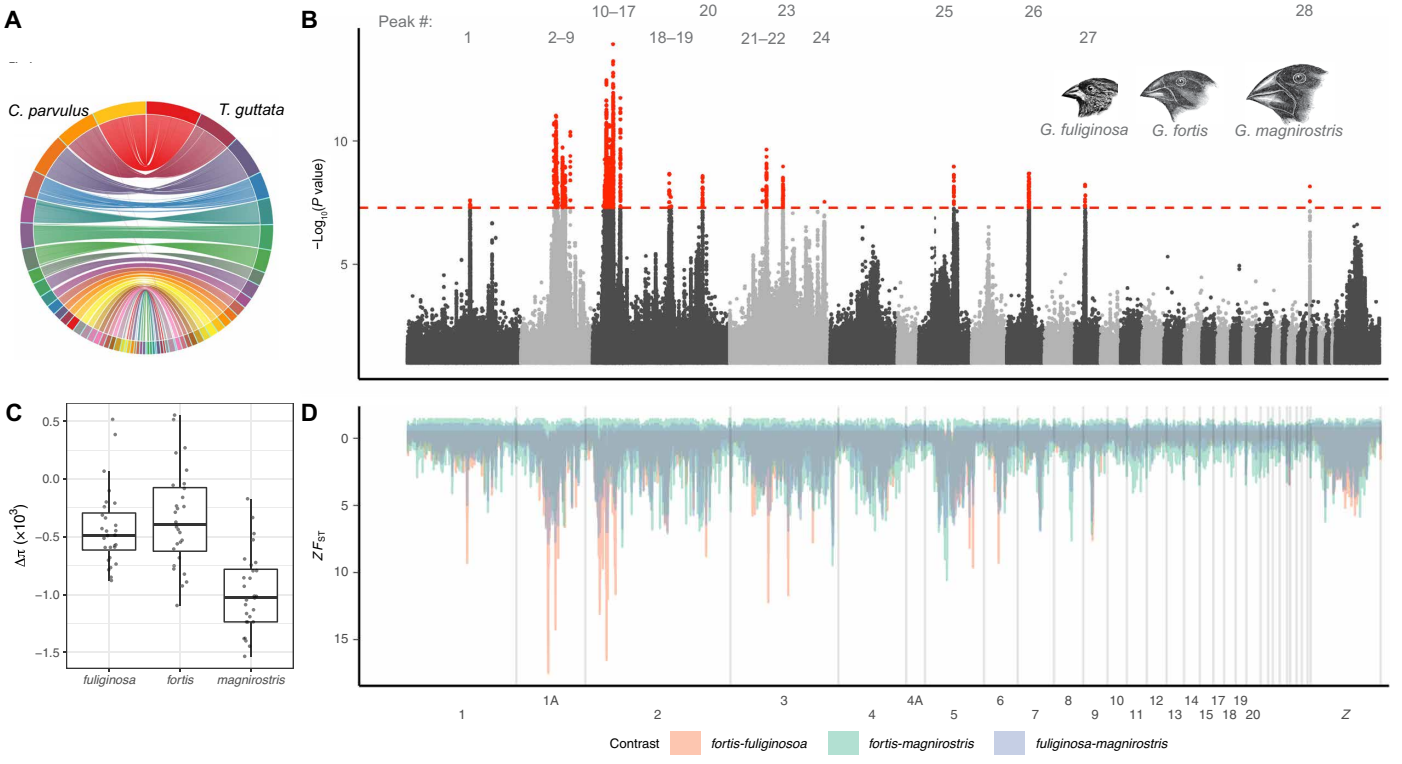

Fig. 1. Genome assembly, genetic diversity, and differentiation among three species of Darwin's finches. (**A**) Illustration of conserved synteny to zebra finch (*Taenopygia guttata*). (**B**) Genome-wide admixture mapping using three species sorted ascendingly by beak and body size: 0 = *G. fuliginosa*, 1 = *G.* fortis, and 2 = *G. magnirostris*. The red dashed line indicates the significance threshold set by permutation. Illustrations of the three species are adapted from P.R.G. and Darwin (*65*). (**C**) Boxplot showing the difference in nucleotide diversity between the regions of association marked in (B) and regions outside the main area of association. Centerline indicates the median, bounded by the 25th and 75th percentiles, with whiskers extending to 1.5× the interquartile range. (**D**) Genome-wide *Z*-transformed F_{ST} for all three possible pairwise combinations of *Geospiza* considered here. Lines are colored by the comparison of interest. Many highly divergent regions are shared between contrasts and overlap with regions of association in (B) (data S2 and S3).

scores of 0, 1, and 2 to reflect the increasing beak and body size of *G. fuliginosa* < *G. fortis* < *G. magnirostris* because phenotypic data were not available for each individual. The experimental setup is similar to an earlier study of these three species from only one island and which used reduced representation sequencing (*27*). Admixture mapping (*37*) revealed 28 loci that exceeded the significance threshold set by permutation (Fig. 1B) and represent independent loci (data S2). The size of these regions ranges from hundreds of kilobases (kb) to 2.7 Mb (Fig. 2) and contained between 0 and 35 genes (data S2). Outlier loci were clustered on macrochromosomes and included the previously described *ALX1* and *HMGA2* loci affecting the beak morphology, both located on chromosome 1A and only ~7 Mb apart (loci 4 and 9 in Fig. 1B). These loci are separated by a recombination hotspot (fig. S2), consistent with previous results showing that these loci do not show strong linkage disequilibrium (*21*, *26*).

Regions of association identified with admixture mapping largely mirrored the results of F_{ST}-based contrasts (Fig. 1D) and strongly correlated with per-window estimates in the two contrasts involving *G. fuliginosa* ($R^2_{fortis\text{-}fuliginosa}$ = 0.84, $R^2_{magnirostris\text{-}fuliginosa}$ = 0.69; data S2). We do not expect a perfect match between the results of admixture mapping and F_{ST} analysis because the former is based on a linear comparison of the trio, while the latter is derived from pairwise comparisons of species. In the 28 regions of association, *G. fuliginosa* and *G. magnirostris* were often homozygous for different haplotypes, while *G. fortis* exhibited intermediate allele frequencies (fig. S4 and data S3). This is highlighted at the *HMGA2* locus on chromosome 1A, where measures of Tajima's *D* are strongly negative for *G. fuliginosa* and *G. magnirostris* but strongly positive for *G. fortis* (fig. S4), consistent with balancing selection maintaining haplotype diversity in the phenotypically variable *G. fortis* population (*25*). Nucleotide diversity was reduced in *G. magnirostris* relative to the genomic background in 23 of the 28 regions (Fig. 1C), consistent with selective sweeps in *G. magnirostris* or an ancestor. The 28 loci also fell in genome-wide low recombination regions (mean ρ in peaks = 1.4 compared with mean ρ outside = 2.0), with the exception of one peak on chromosome 25 (ρ = 4.7), consistent with previous studies in Darwin's finches showing that regions of elevated differentiation often lie in recombination coldspots (*31*). Low recombination in these regions has likely facilitated the persistence of large haplotype blocks despite high gene flow among Darwin's finch species, as predicted from theoretical studies (*36*).

To explore the extent to which the loci detected in the ground finch contrast are also associated with phenotypic diversity among tree finches (*Camarhynchus*), we next performed a similar admixture analysis comparing small, medium, and large tree finches (*C. parvulus*, *Camarhynchus pauper*, and *Camarhynchus psittacula*, respectively), also classified as 0, 1, and 2, respectively, based on beak and body size. These samples were previously sequenced and include 46 individuals (n = 10 to 25 each) from eight islands. This replicated a signal for the *HMGA2* locus affecting beak size ($P = 4 \times 10^{-16}$) (fig. S5), as expected from previous work (*26*). No other locus showed such a notable signal of genetic differentiation, but we noted an overlap of higher genetic differentiation, approaching genome-wide significance, at several of the loci detected in the ground finch contrast (fig. S5). The identification of regions associated with phenotypic variation in size among *Camarhynchus* is hampered by the relatively small sample size (n = 46 versus n = 75 for *Geospiza*). Furthermore, we should not expect a perfect overlap because the beak proportions,

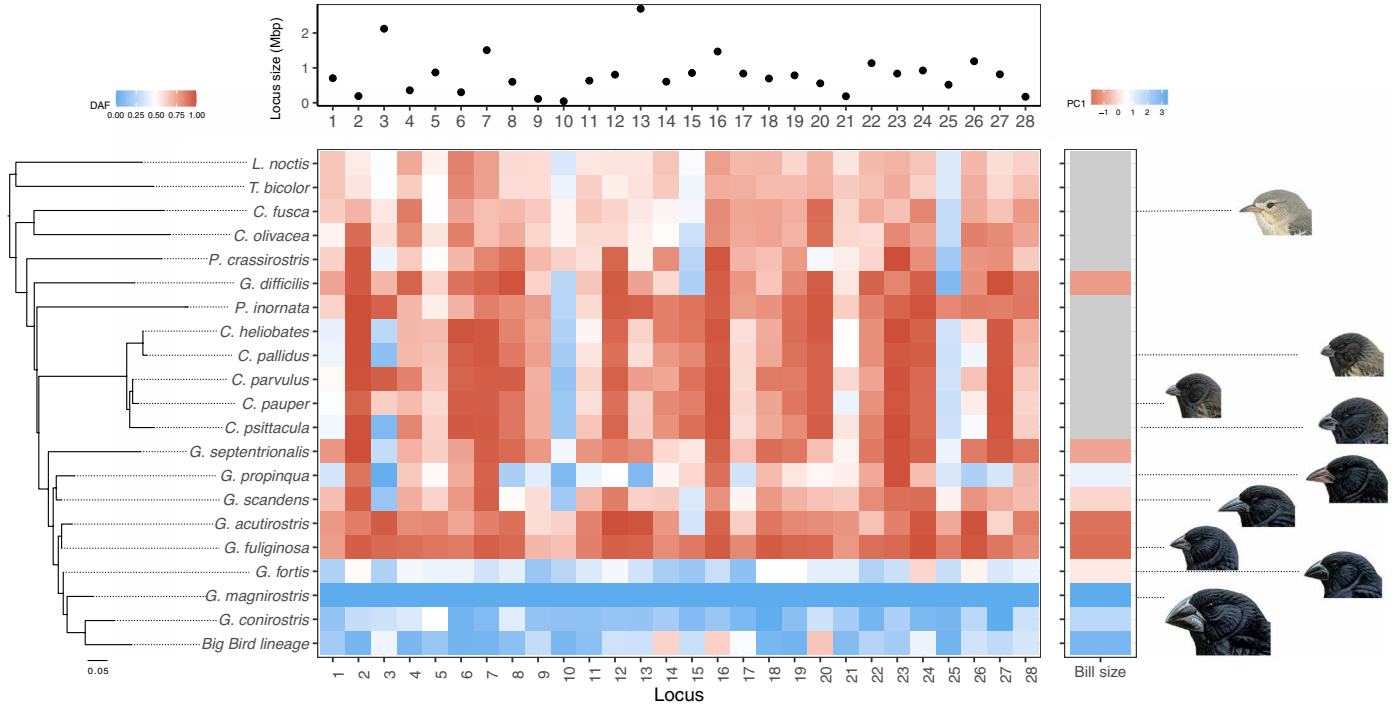

Fig. 2. Haplotype variation across the Darwin's finch phylogeny. The heatmap displays average delta allele frequency (DAF) based on 34 to 328 SNPs/locus for each species compared to *G. magnirostris*, the species with the largest beak. On the right, bill size is presented according to a principal component (PC) analysis of three beak dimensions averaged across all island populations for each species. Only *Geospiza* species are shown. Above, the size of each genomic region (in mega-base-pairs) is marked in a dot plot. Right: Finch illustrations reproduced by permission of Lynx Edicions.

skull architecture, and musculature of the three tree finches differ from those in the ground finches and thus may have a different genetic basis (*38*).

To determine the evolutionary origin of haplotypes at the 28 loci detected in the ground finch contrast, we compared allele frequencies at the most differentiated single-nucleotide polymorphisms (SNPs) across all 18 species of Darwin's finches and two outgroups *Loxigilla noctis* and *Tiaris bicolor*. We also included the Big Bird hybrid lineage, which was formed by the mating of a *Geospiza conirostris* male and two *G. fortis* females (*24*). The sample included genome sequences from previously published data and 62 new individuals (n = 321 in total; data S1). If genetic differentiation among *Geospiza* is caused primarily by de novo mutations that occurred after the split from *Camarhynchus*, then we would expect to find little shared haplotype structure in non-*Geospiza* species as a consequence of random accumulation of variants. In sharp contrast to this expectation, the allele frequency comparison revealed a nonrandom pattern (Fig. 2) and broad haplotype structures across the most differentiated SNPs at the 28 loci (data S4). Notably, a few *G. magnirostris* haplotypes are present at a relatively high frequency across the radiation (e.g., 10 and 25), while most are consistently highly differentiated from haplotypes in other species, except those with relatively large beaks. Furthermore, a large portion of *G. magnirostris* major alleles, 67% (1273 of 1914), were derived relative to outgroups *L. noctis* and *T. bicolor*, consistent with selective sweeps in the *G. magnirostris* lineage and after Galápagos colonization (data S5). Together, these results imply that the "large" and "small" haplotype blocks identified by admixture mapping in *Geospiza* predate the separation of *Geospiza* and *Camarhynchus*.

The radiation of finches in the Galápagos proceeded rapidly, and phylogenetic reconstructions are characterized by short branch lengths, gene tree discordances, and incomplete lineage sorting (ILS). We therefore evaluated the hypothesis that these haplotype blocks are ancestral by examining the haplotype structure at each locus. Neighbor-joining trees based on the most differentiated SNPs at the 28 loci are not consistent with the genome-wide phylogeny of Darwin's finch (fig. S6), which is driven by haplotype structuring at each locus (locus 24 in Fig. 3A and all 28 loci in data S4). We also find support for phylogenetic discordance at each locus using topology weighting, which demonstrates that the placement of *G. magnirostris* at each locus is inconsistent with the species tree (Fig. 4 and fig. S7). Next, we estimated the divergence time between haplotypes among all pairwise combinations of *G. magnirostris*, *G. fuliginosa*, and *Certhidea olivacea* (green warbler finch) based on pairwise nucleotide divergence (d_{xy}). These data indicate that the divergence times between the *G. magnirostris* and *G. fuliginosa* haplotypes at the 28 loci tend to be older than the coalescent time for the two species (~260,000 years before present; Fig. 3A) and in the range of 380,000 to 800,000 years before present (Fig. 3B, fig. S8, and data S2), which means that the great majority predate the split between *Geospiza* and *Camarhynchus* about 400,000 years ago (Fig. 3A). The estimated divergence times involving *C. olivacea* haplotypes consistently exceeded the divergence time between haplotypes from the two *Geospiza* species, suggesting that they evolved after the split between warbler finches and other finches in the phylogeny. However, these data do not exclude the possibility of a divergence before this early split for some loci, because genetic exchange between haplotypes is possible in species that segregate for both alleles like many do in *G. fortis* (data S4). To explore this possibility, we estimated the time to most recent common ancestor (TMRCA) among all haplotypes of each strongly associated SNP using a probabilistic modeling approach that incorporates both mutation and recombination rate. We found that strongly associated SNPs have a broad distribution of TMRCA (fig. S9), but, on average, the 28 loci have an upper allele age of 937,000 years ago (SD = 206,000 years).

Hybridization and the mixing of ancestral haplotypes

The presence of distinct combinations of haplotypes across the phylogeny (Fig. 2 and data S4) indicates ILS or hybridization and introgression. The Genovesa cactus finch *Geospiza propinqua* has not only the fourth largest beak of all *Geospiza* but also the pointed beak characteristic of other cactus finches and carries a mixture of large and small haplotypes (Fig. 2). Gene flow from *G. magnirostris* to *G. propinqua* has been implicated from field observations (*16*). We estimated the fraction of introgression from *G. magnirostris* to *G. propinqua* on Genovesa using d_f, which incorporates d_{xy} into an extension of ABBA/BABA D statistics (*39*), and found that regions of high *G. magnirostris* similarity share an excess of derived alleles (fig. 3C) and often reduced genetic divergence d_{xy} (data S6), consistent with introgression and not ILS. In general, it remains an outstanding challenge to distinguish between ILS and introgression using D statistics (*40*), but the genomic evidence presented here for introgression in *G. propingua* demonstrates that gene flow can transfer these haplotypes among species. The role of hybridization in generating distinct combinations of haplotypes is demonstrated in the Big Bird lineage (Fig. 2). The Big Bird lineage is characterized by a proportionally large beak for its body size (*24*). The combination of *G. conirostris*, a sister species to *G. magnirostris*, and *G. fortis* alleles resulted in a unique phenotype and genotype with *G. magnirostris* haplotypes predominating at 22 of the 28 loci (Fig. 2).

Trait loci are enriched for developmental genes

The loci identified here segregate with phenotypic variation among the species and are expected to contain genes important for beak and body size variation. However, because this result is based on between-species contrasts, we cannot exclude the possibility that some loci reflect unmeasured phenotypic differences. We therefore explored previously described genotype-phenotype relationships in other taxa for the genes within these regions and their expression patterns. We conducted an enrichment analysis using mouse orthologs for the genes in the vicinity of the 28 loci using the software GREAT (genomic regions enrichment of annotations tool) (*41*) and found that deleterious mutations at these loci were significantly associated with abnormal development of cartilage and bones (Fig. 5A and fig. S10). Furthermore, these mouse orthologs were significantly enriched for genes expressed during craniofacial and limb development, consistent with our expectation that genetic changes at many of these 28 loci affect beak development and body size variation. The enrichment includes genes spread across the 28 loci (data S2).

Because of the difficulties in interpreting gene-by-term enrichment with data from other species (*42*), we performed two types of analysis to validate the association of beak gene expression with the differentiated loci. Using RNA-seq data from Darwin's finches, we compared genes expressed at embryonic day 7 (E7) in the developing upper beak (nine embryos representing six species) with five other tissues (brain, gut, heart, left forelimb, and trunk from six embryos representing three species) (table S1). This analysis revealed that many of the genes within the 28 loci are expressed in the developing

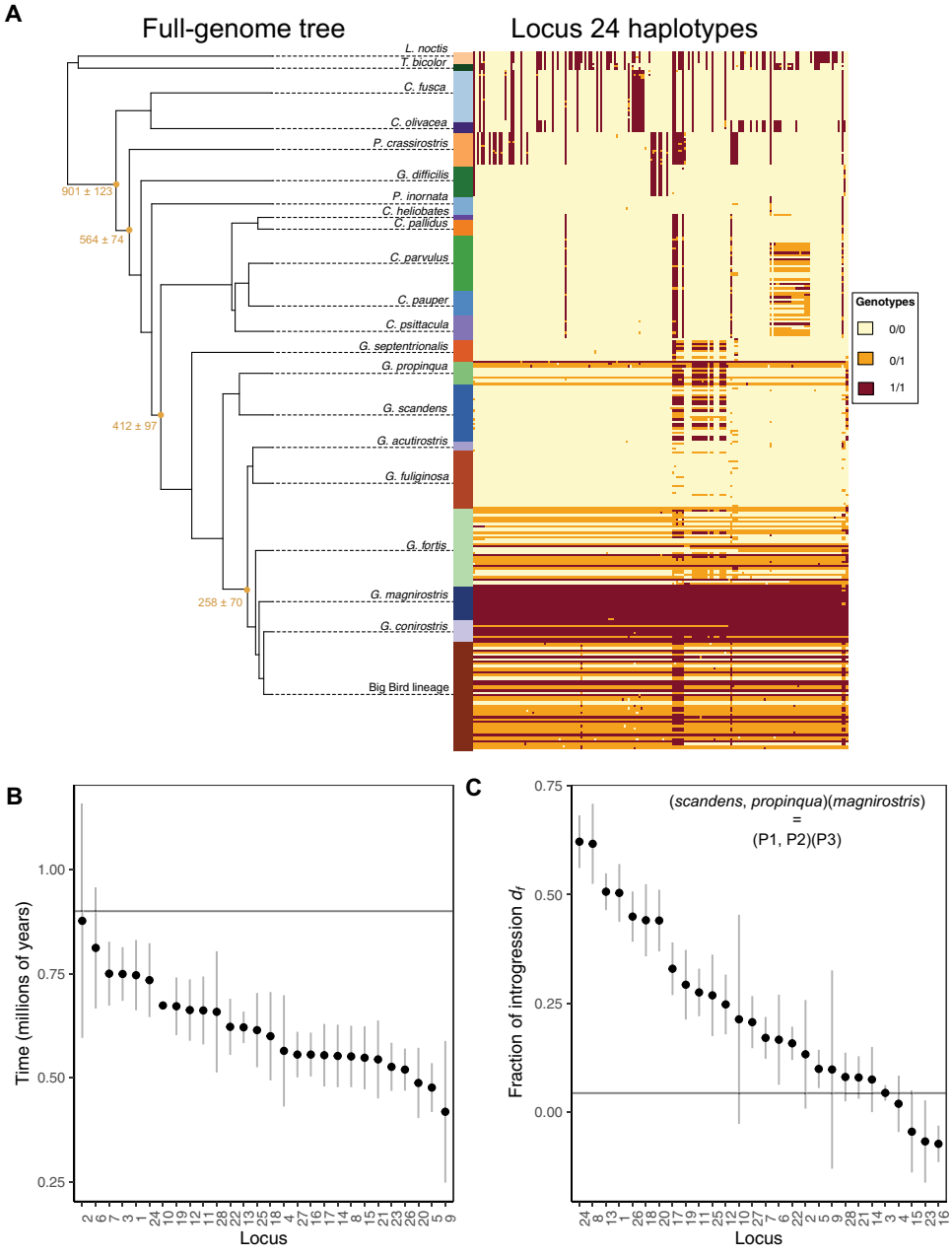

Fig. 3. Characterization of 28 adaptive loci. (**A**) Left: A neighbor-joining tree for all species of Darwin's finches based on 4.9 million SNPs. The tree was converted to a chronogram using ape (*66*), and branching times are from Lamichhaney *et al.* (*21*). Right: A haplotype plot for each of the top associated variants on locus 24 with 0/0 for homozygous reference (yellow), 1/1 homozygous alternate allele (red), and 0/1 for heterozygous positions. Species names are colored by a-priori clade assignments, and a co-phylo diagram (*67*) highlights the changes in topology between the trees. (B) Time to most recent common ancestor (TMRCA) between *G. fuliginosa* and *G. magnirostris* haplotypes for all 28 loci. Time was estimated by the conversion of genetic divergence (d_{xy}) to time using $T = d_{xy}/(2\ \mu)$ and a mutation rate (μ) of 2.04×10^{-9} (*21*). The horizontal line marks 900,000 years, the approximate time of divergence between warbler finches and all other finches in the radiation (*21*). (**C**) Fraction of introgression from *G. magnirostris* to *G. propinqua* on Genovesa Island for each of the 28 loci, as measured by d_f. Loci are arranged by the strength of the introgression, and a line is drawn at the median of genome-wide $d_f = 0.04$. The trio arrangement is written on the graph and ABBA/BABA statistics listed under Methods. In (B) and (C), error bars represent 95% confidence intervals.

beak, and these loci were 14-fold enriched for genes with higher expression levels (M≥5 = fold change ≥32) in the developing beak compared with other tissues (χ^2 test, $P = 7.4 \times 10^{-24}$, d.f. = 1; Fig. 5B). We carried out in situ hybridization (ISH) of two candidates for craniofacial development (*ALX1*, locus 7; *RUNX2*, locus 24). ISH data on a total of seven zebra finch (*Taeniopygia gutatta*) and 27 Darwin's finch embryos (of nine species) revealed that *ALX1* expression was strongly biased to the beak region over the developmental period when beak size and shape are established (E6 to E7) (Fig. 5C and figs. S11 and S12A) (*43*). In addition, we confirmed a similar expression pattern in zebra finch embryos for *RUNX2* (alias *OSF2*)—an essential gene for ossification of the mesenchyme (fig. S12B) (*44*)

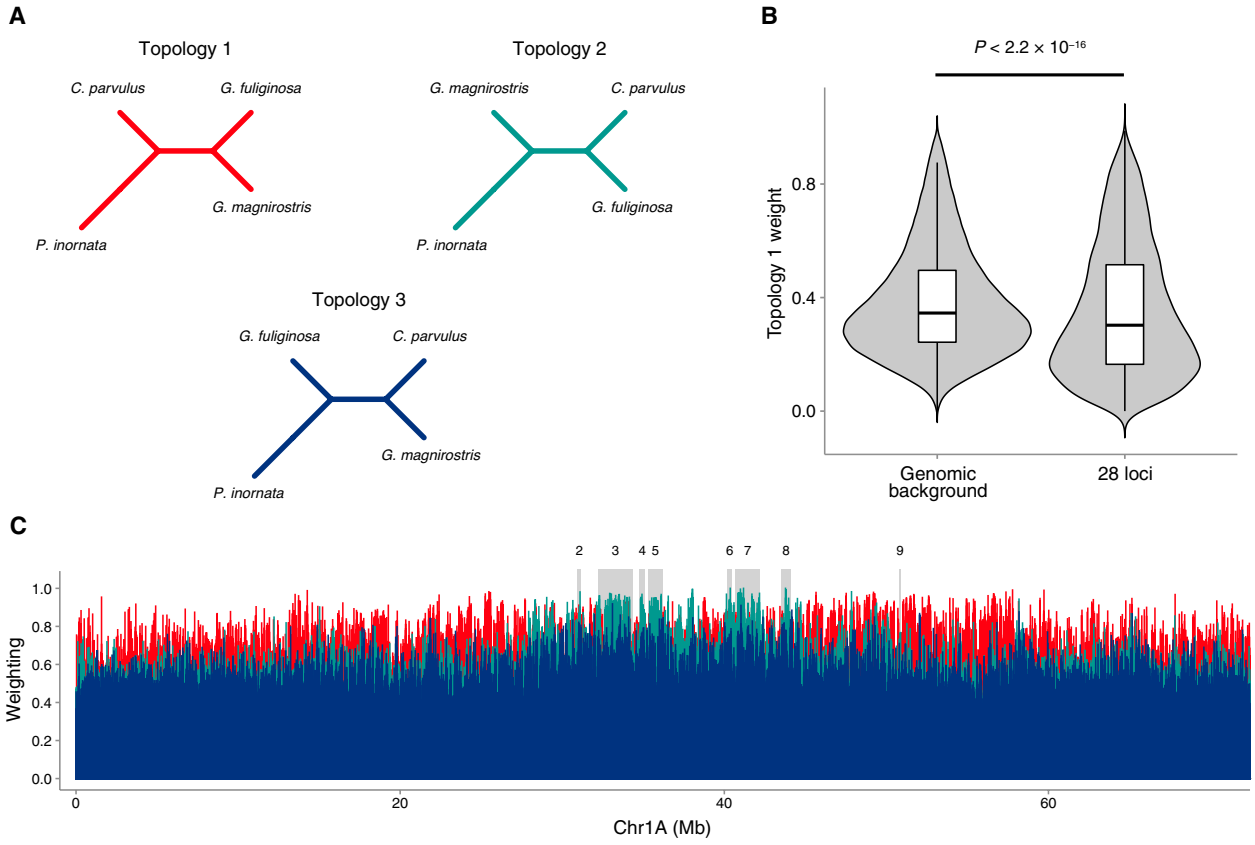

Fig. 4. Phylogenetic discordance at the eight loci on chromosome 1A. (**A**) The three possible topologies representing relationships among four taxa: *P. inornata*, *C. parvulus*, *G. magnirostris*, and *G. fuliginosa*. Topology 1 is consistent with the genome-wide phylogeny (Fig. 3A), whereas topologies 2 and 3 are discordant, separating the monophyletic *Geospiza* species. (**B**) Violin/box plots showing weightings for topology 1 (the species tree) across the eight chromosomes containing the 28 loci regions. (**C**) Topology weightings calculated in nonoverlapping 100 SNP windows across chromosome 1A (chr1A). Locations of the 28 loci regions within this chromosome are indicated by gray shading.

located within a strong signal of differentiation among *Geospiza* (fig. S4B).

DISCUSSION

We have identified ancestral haplotypes at 28 loci that contribute to the unique phenotypic diversity of the Darwin's finch radiation. Our functional characterization of these loci contributes to a growing body of literature, suggesting that genetic differences between species of Darwin's finches are enriched for genes involved in the key pathways for growth and development of beaks (*21*, *26*–*28*, *43*). In this study, we show that the distribution of these genetic modules across the phylogeny reflects natural selection and most likely both ILS and introgression (*21*, *22*). It is not only the presence/absence of these haplotype blocks that affects the phenotype but also their frequency within species, which is illustrated by intermediate haplotype frequencies at many of these loci in *G. fortis* (Fig. 2 and data S4 and S5). Intermediate frequencies, indicative of balanced polymorphism, provide the underlying variation for selection to sort adaptive haplotypes during speciation. In *G. fortis*, variation may have been crucial for survival during strong selection events (*26*). Our findings support previous suggestions that ancestral variants contribute to phenotypic diversity, as indicated for pigmentation phenotypes among other songbird species (*8*, *13*), color morphs in the common wall lizard (*45*), color patterns in *Heliconius* butterflies (*11*), various phenotypic traits in cichlids (*7*, *9*), craniofacial morphology in pupfish (*46*), winter coat in snowshoe hares (*47*), and adaptation to high altitude in humans (*48*). This type of ancestral variation is also retained within large populations preceding speciation, as illustrated in Atlantic herring and stickleback where ecotypes show differences in the frequency of haplotype blocks at hundreds of loci, all underlying ecological adaption (*49*, *50*).

Characteristic features of these 28 loci are the large size of the haplotype blocks, often spanning hundreds of kilobases, and their ancient origins (Fig. 3B). The haplotype structure is in contrast to another ancestral polymorphism in Darwin's finches, at the *BCO2* locus controlling nestling beak color, where a single base change constitutes the likely causal mutation (*51*). The identification of causal variants within the haplotype blocks described here is challenging because of strong linkage disequilibrium among many sequence variants within each region. Such large haplotypes could include structural variants, which have been proposed as a key determinant of adaptive evolution and speciation (*52*, *53*), but these 28 loci do not represent large inversions (>5 Mb) and tend to be relatively small (0.1 to 2.7 Mb) compared to previously described inversions in vertebrates (*49*, *54*–*57*). Furthermore, none of the loci exhibit the sharp borders in our association analysis that are characteristic of inversions maintained as balanced polymorphisms (*49*, *54*).

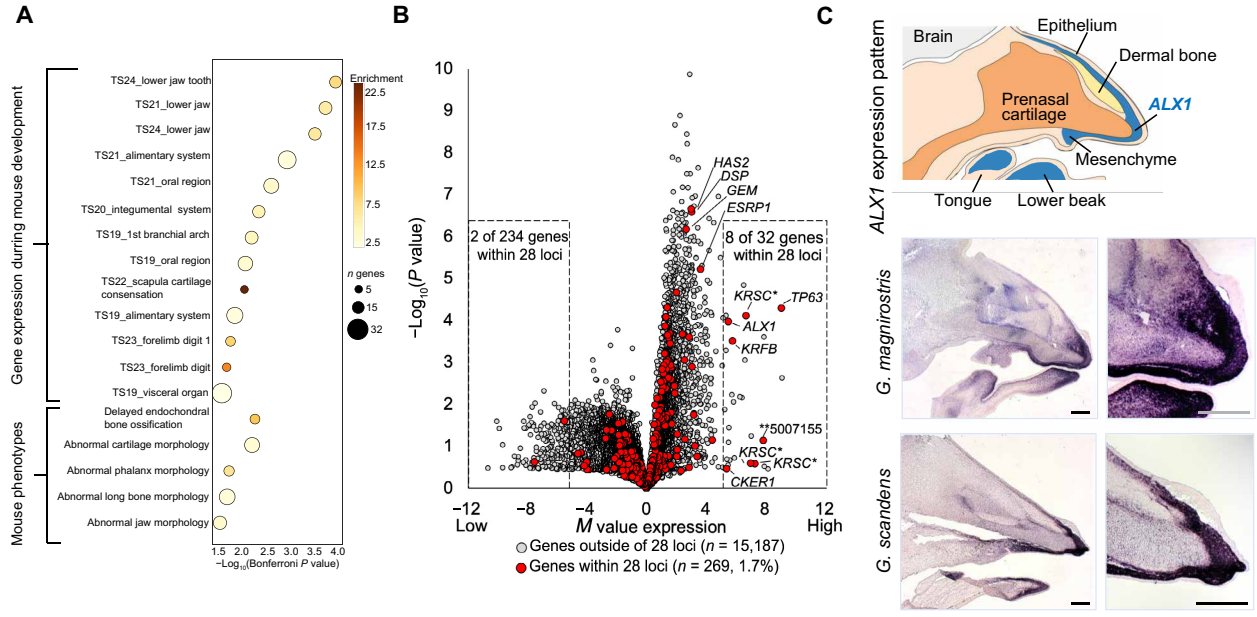

Fig. 5. Enrichment analysis and gene expression. (**A**) Annotation term enrichment analysis. GREAT (*41*) was used to screen for enrichment of gene annotation terms associated with the 28 differentiated regions. $-\text{Log}_{10}$ Bonferroni-corrected *P* values of significantly enriched terms are shown on the *x* axis. Fold enrichment is indicated using dot colors. Dot sizes indicate numbers of genes belonging to each annotation term. (**B**) Gene expression in Darwin's finch upper beak versus other tissues. RNA-seq gene expression levels in upper beak ($n = 9$) were compared with expression levels in noncraniofacial tissues ($n = 7$, representing four different tissues. $-\text{Log}_{10}$ (*P* values) for differential expression are shown on the *y* axis, and *M* values [$\log_2$(fold change beak samples versus other tissues)] are shown on the *x* axis. Dashed boxes show genes with *M* values < −5 and > +5, representing genes with lower (left) and higher (right) expression in beak samples (right) compared with other tissues. The asymmetry in $-\log_{10}$(*P* values) is a result of comparing one tissue (beak) to several other tissues. Gene names are shown for selected genes. * indicates separate genes belonging to a gene cluster of scale keratins. ** is a gene of uncertain function (see data S7). (**C**) In situ hybridization (ISH). Top: Schematic representation of the expression pattern of *ALX1* (blue) in E7 embryos of zebra finch ($n = 5$) and Darwin's finches (nine species, $n = 21$). Bottom: mRNA expression of *ALX1* (dark purple) in mid-face longitudinal sections through the heads of Darwin's finch embryos. The developing beak region is shown. Magnified area of the left images is shown on the right. Scale bars, 250 μm.

The block structures at the 28 loci most likely reflect large-effect haplotypes composed of clusters of multiple causal variants that have accumulated during the evolution of Darwin's finches (*58*), similar to the evolution of alleles in domestic animals by the sequential accumulation of causal mutations during the past 10,000 years (*59*). The occurrence of these haplotype blocks in low recombination regions likely facilitated their evolution (data S2). The reuse of ancestral genetic modules is a much faster route to adaptive change than the slow accumulation of adaptive de novo mutations (*7*). Our study is comprehensive in surveying genomic variation across all 18 extant species of a single adaptive radiation, and yet, the principle finding, of repeated reassembly of ancient haplotype blocks in the formation of species, is likely to be a general feature of rapid radiations (*8*, *46*) and of general importance for how species adapt to environmental variability and change (*25*).

METHODS

Sample collection

Blood samples were collected from various Galápagos islands as part of sampling described elsewhere (*21*, *24*) and stored on EDTA-soaked filter paper in DRIERITE to preserve red blood cells for DNA extraction later. This included 101 individuals of eight different species (data S1). An additional 18 samples of six species were captured using mist nests on San Cristóbal in 2018 for this study and resequenced using short reads (data S1). In total (combined with sequences from previous studies), our sampling includes all 18 species of Darwin's finches, the hybrid Big Bird lineage, and two outgroup species that sum to a combined sampling of 321 individuals (data S1). Of the individuals captured on San Cristóbal in 2018, one small tree finch was sampled for targeted long-read sequencing (see below). Sampling was conducted in accordance with the protocols of Princeton University's Animal Welfare Committee. Embryos for RNA-seq for annotation ($n = 11$, three species) and for differential expression ($n = 9$, six species) analyses and for ISH ($n = 27$, nine species) were collected on Santa Cruz, Genovesa, and Pinta as described (table S1) (*60*). Embryos were stored in methanol or RNAlater (Thermo Fisher Scientific, CA) until further use.

ONT sequencing

Pilot experiments before the expedition indicated that avian DNA (chicken and Darwin's finches) was not sequenced efficiently. We decided to develop a protocol to optimize yield from avian DNA sequencing with the MinION and hypothesized that the issue was partially due to avian DNA being enriched in molecules carrying certain motifs or creating certain types of secondary structures, which promoted blocking of nanopores and thus resulted in premature loss of actively sequencing nanopores and ultimately poor sequencing yields. We further hypothesized that T7 endonuclease I treatment of DNA before library generation would increase yields because T7 endonuclease I has been used to increase sequencing yield in an ONT protocol for sequencing DNA amplified using bacteriophage Φ29 polymerase and has a broad substrate specificity involving four-way junctions, various branched structures, and single-base

mismatched heteroduplexes (*61*). We first used g-TUBEs (Covaris) to achieve a controlled fragmentation of the isolated DNA to approximately 6 to 10 kb. Electrophoresis of the original g-TUBE–fragmented DNA side by side with the T7 endonuclease I–treated sample revealed a low-molecular weight smear unique to the treated sample. To efficiently remove DNA cut by T7 endonuclease I, we used a custom SPRI bead mixture (SeraMag SPRI beads, Thermo Fisher Scientific) capable of retaining DNA above the size of approximately 4 to 5 kb, thus selecting against the DNA cut by T7 endonuclease I. We used the T7 endonuclease I cleavage protocol to prepare DNA for library generation for libraries aimed to generate reads in the 6 to 10 kb size range. For libraries where we aimed at sequencing longer molecules, we did not conduct any T7 cleavage. In the case of T7 cleavage of DNA before library generation, we ran agarose gel electrophoresis before advancing to library isolation to ascertain successful removal of the low–molecular weight DNA smear by the SeraMag speed bead mixture.

DNA was isolated by two main protocols, depending on whether we aimed to construct DNA libraries in the 6- to 10-, 20- to 30-, or 50+-kb size ranges. For all samples, we started out with small amounts of blood (5 to 20 μl) sampled from wing veins of birds using glass capillaries. Immediately following sampling, the blood was deposited in a tube containing 1 ml of cold phosphate-buffered saline (10 mM EDTA). The sampled blood was kept cold in a Styrofoam box containing ice packs until back at the laboratory facility. Nuclei were isolated by adapting the protocol for cell culture in the QIAGEN Genomic DNA Handbook using ice-cold buffer C1 (QIAGEN). Following isolation of nuclei, we used either blood and tissue spin columns (QIAGEN) or NaCl/ethanol precipitation to isolate <30 kb or high–molecular weight DNA, respectively. Regardless of means of DNA isolation, we included a DNA cleanup step using SPRI beads before library generation. SeraMag SPRI beads were prepared by mixing 10 ml of 5 M NaCl, 500 μl of 1 M tris, 100 μl of 0.5 M EDTA, and 10 ml of H_2O in a 50-ml Falcon tube. SeraMag beads (1 ml) were washed five times in 1 ml of tris-EDTA buffer (10 mM tris, 1 mM EDTA) using a magnet stand, and the final 1-ml volume was added to the 20.6 ml of solution. A 50% (w/v) solution of PEG8000 in water was prepared, and 18 ml of this mix was added to the 21.6 ml of bead solution. The content was rigorously vortexed. Tween 20 (27.5 μl) was added to the mixture, and the volume was adjusted to 50 ml with H_2O. The final mixture was vortexed and then aliquoted to 1.5-ml Eppendorf tubes. The performance of the bead mix at different bead/sample volume ratios was evaluated using a mixture of bacteriophage lambda DNA and 1 kb of DNA ladder.

From one small tree finch (*C. parvulus*) individual (STF5), we produced four LSK-108 libraries from g-TUBE–fragmented DNA and five LSK-108 libraries made from high–molecular weight DNA using the protocol Genome sequencing by ligation, selecting for long reads (ONT). The libraries were sequenced on a GridION instrument (ONT) according to the manufacturer's instructions. Together, from individual STF5, we generated 32.3 Gb of sequence data, corresponding to approximately 30× coverage.

HiC analysis

We were not able to perform HiC analysis at the Galápagos Science Center but were able to export fresh blood from a *G. fortis* individual; therefore, the HiC analysis was based on a *G. fortis* individual, although the long-read data were from a *C. parvulus* individual. The blood was kept refrigerated until departure from Galápagos and was kept chilled (2° to 10°C) until arrival at Uppsala University, where a nuclei isolation protocol (buffer C from the QIAGEN Genomics Buffer) was applied to isolate erythrocyte nuclei. Isolated nuclei were immediately snap frozen in liquid nitrogen and kept at −80°C until further use. Nuclei were used for HiC library generation using the Arima Genomics HiC Kit v1, and the resulting HiC library was sequenced on a NextSeq 2000 instrument (Illumina).

RNA sequencing

We prepared RNA-seq libraries for genome annotation using a variety of library preparation protocols, tissues, and finch species. We used a ribosomal RNA depletion kit from NEBNext for six samples of lower beak, jaw muscles, brain, gut and heart, left forelimb, and brain. An additional four samples were prepared using poly(A) enrichment protocols from NEBNext for brain and jaw muscle tissues. Last, a single *G. fortis* trunk sample was enriched using a Lexogon SENSE total poly(A)-enriched RNA-seq kit and sequenced at three times higher depth than the remaining samples. The specific kits and sample accessions are listed in table S1. All 11 libraries were multiplexed and sequenced on a flow cell of an Illumina (San Diego, CA) SP chip. For differential expression analysis, we extracted RNA from the upper beak primordia of nine embryos from six species of Darwin's finches (table S1), prepared cDNA libraries with the NEBNext Ultra RNA Library Prep Kit for Illumina (New England Biolabs, MA) with poly(A) selection, and sequenced them on HiSeq 4000 (Illumina, CA).

Contrasting gene expression in beak and other tissues

RNA-seq data were aligned to the genome assembly using STAR v2.7.2b (*62*), guided by the GTF file from genome annotation. For each sample, uniquely mapping reads for each annotated gene were extracted using the STAR option "--quantMode GeneCounts." Per-sample gene counts were normalized to transcripts per kilobase million (TPM) values, by first dividing observed counts with the longest isoform length in kilobases and then by dividing those values with the total numbers (in millions) of uniquely mapping reads. Obtained TPM values were used to contrast upper beak development samples ($n = 9$) with samples corresponding to other tissues ($n = 7$) (table S1). *P* values for differential expression from two-sided *t* tests were calculated for each gene. *M* values were calculated as $-\log_2$(average TPM beaks / average TPM other tissues).

Short-read sequencing

One hundred one individuals were extracted using a custom salt preparation method [described by Enbody *et al.* (*51*)] and sequenced using the TruSeq Kit (Illumina, CA). Sixteen additional whole-genome libraries were prepared using a custom Tn5 transposon–based tagmentation protocol derived from Picelli *et al.* (*63*) and detailed by Enbody *et al.* (*51*). Briefly, we assembled the Tn5 transposon construct using the stock Tn5 (prepared by Karolinska Institutet Protein Science Facility) and the following primers (*63*): Tn5MErev: (5′-[phos]CTG TCTCTTATACACATCT-3′), Tn5ME-A (Illumina FC-121-1030) (5′-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3′), and Tn5ME-B (Illumina FC-121-1031) (5′-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-3′).

In situ hybridization

Tissue sectioning and fixation, probe hybridization, and signal detection were performed according to the previously published protocols

(*64*) with probe concentration of 0.5 ng/μl. Digoxigenin (DIG)–labeled antisense riboprobes were generated by polymerase chain reaction followed by RNA synthesis according to the standard procedures using T3/T7 promoter primers combined with the following gene-specific primers: *ALX1* [581 base pairs (bp)] (forward: 5′-CAGGACAGCAACGTCAACTA-3′; reverse: 5′-AAGCCTGTGTAGCCAGAATC-3′), *COL2A1* (683 bp) (forward: 5′-GCAAGGCCAAGGAGAAGAA-3′; reverse: 5′-TGATTCTGGTGTTGGGATGAG-3′), *COL9A1* (608 bp) (forward: 5′-CTGGCCCAAAGGGTAATAGAG-3′; reverse: 5′-ACCAAATTCTGGCCTCCTAAG-3′), and *RUNX2* (619 bp) (forward: 5′-GAACCAGGTGGCCAGATTTA-3′; reverse: 5′-GACTGGCGGTGTATAGGTAAAG-3′).

Genome assembly

The methods used for genome assembly, gene and repeat annotation, and construction of linkage map are provided in Supplementary Methods.

Population genomics analysis

The Supplementary Methods contains a full description of the methods used for short-read variant analysis, genotype phasing, admixture mapping, genetic diversity and divergence analyses, allele age estimation, analysis of introgression, gene enrichment analysis, phylogenetic reconstruction, analyses of allele frequencies, and topology weighting.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.abm5982

REFERENCES AND NOTES

1. R. G. Gillespie, G. M. Bennett, L. De Meester, J. L. Feder, R. C. Fleischer, L. J. Harmon, A. P. Hendry, M. L. Knope, J. Mallet, C. Martin, C. E. Parent, A. H. Patton, K. S. Pfennig, D. Rubinoff, D. Schluter, O. Seehausen, K. L. Shaw, E. Stacy, M. Stervander, J. T. Stroud, C. Wagner, G. O. U. Wogan, Comparing adaptive radiations across space, time, and taxa. *J. Hered.* **111**, 1–20 (2020).
2. D. Schluter, *The Ecology of Adaptive Radiation* (Oxford Univ. Press, 2000).
3. J. T. Stroud, J. B. Losos, Ecological opportunity and adaptive radiation. *Annu. Rev. Ecol. Evol. Syst.* **47**, 507–532 (2016).
4. I. J. Lovette, E. Bermingham, R. E. Ricklefs, Clade-specific morphological diversification and adaptive radiation in Hawaiian songbirds. *Proc. R. Soc. Lond. B* **269**, 37–42 (2002).
5. P. R. Grant, B. R. Grant, *How and Why Species Multiply: The Radiation of Darwin's Finches* (Princeton Univ. Press, 2008).
6. D. Melo, A. Porto, J. M. Cheverud, G. Marroig, Modularity: Genes, development and evolution. *Annu. Rev. Ecol. Evol. Syst.* **47**, 463–486 (2016).
7. D. A. Marques, J. I. Meier, O. Seehausen, A combinatorial view on speciation and adaptive radiation. *Trends Ecol. Evol.* **34**, 531–544 (2019).
8. K. F. Stryjewski, M. D. Sorenson, Mosaic genome evolution in a recent and rapid avian radiation. *Nat. Ecol. Evol.* **1**, 1912–1922 (2017).
9. J. I. Meier, D. A. Marques, S. Mwaiko, C. E. Wagner, L. Excoffier, O. Seehausen, Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nat. Commun.* **8**, 14363 (2017).
10. R. F. Guerrero, M. W. Hahn, Speciation as a sieve for ancestral polymorphism. *Mol. Ecol.* **26**, 5362–5368 (2017).
11. N. B. Edelman, P. B. Frandsen, M. Miyagi, B. Clavijo, J. Davey, R. B. Dikow, G. García-Accinelli, S. M. Van Belleghem, N. Patterson, D. E. Neafsey, R. Challis, S. Kumar, G. R. P. Moreira, C. Salazar, M. Chouteau, B. A. Counterman, R. Papa, M. Blaxter, R. D. Reed, K. K. Dasmahapatra, M. Kronforst, M. Joron, C. D. Jiggins, W. O. McMillan, F. D. Palma, A. J. Blumberg, J. Wakeley, D. Jaffe, J. Mallet, Genomic architecture and introgression shape a butterfly radiation. *Science* **366**, 594–599 (2019).
12. R. D. H. Barrett, D. Schluter, Adaptation from standing genetic variation. *Trends Ecol. Evol.* **23**, 38–44 (2008).
13. S. P. Turbek, M. Browne, A. S. Di Giacomo, C. Kopuchian, W. M. Hochachka, C. Estalles, D. A. Lijtmaer, P. L. Tubaro, L. F. Silveira, I. J. Lovette, R. J. Safran, S. A. Taylor, L. Campagna, Rapid speciation via the evolution of pre-mating isolation in the Iberá Seedeater. *Science* **371**, eabc0256 (2021).
14. M. Todesco, M. A. Pascual, G. L. Owens, K. L. Ostevik, B. T. Moyers, S. Hübner, S. M. Heredia, M. A. Hahn, C. Caseys, D. G. Bock, L. H. Rieseberg, Hybridization and extinction. *Evol. Appl.* **9**, 892–908 (2016).
15. A. M. Kearns, M. Restani, I. Szabo, A. Schrøder-Nielsen, J. A. Kim, H. M. Richardson, J. M. Marzluff, R. C. Fleischer, A. Johnsen, K. E. Omland, Genomic evidence of speciation reversal in ravens. *Nat. Commun.* **9**, 906 (2018).
16. P. R. Grant, B. R. Grant, Hybridization increases population variation during adaptive radiation. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 23216–23224 (2019).
17. P. R. Grant, B. R. Grant, Phenotypic and genetic effects of hybridization in Darwin's finches. *Evolution* **48**, 297–316 (1994).
18. N. H. Barton, G. M. Hewitt, Adaptation, speciation and hybrid zones. *Nature* **341**, 497–503 (1989).
19. B. A. Payseur, Using differential introgression in hybrid zones to identify genomic regions involved in speciation. *Mol. Ecol. Resour.* **10**, 806–820 (2010).
20. C. Van Riper III, J. M. Scott, S. Conant, Evolution, ecology, and management of Hawaiian birds: A vanishing avifauna. *Stud. Avian Biol.* **22**, 1–428 (2001).
21. S. Lamichhaney, J. Berglund, M. S. Almén, K. Maqbool, M. Grabherr, A. Martinez-Barrio, M. Promerová, C.-J. Rubin, C. Wang, N. Zamani, B. R. Grant, P. R. Grant, M. T. Webster, L. Andersson, Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* **518**, 371–375 (2015).
22. S. Lamichhaney, F. Han, M. T. Webster, B. R. Grant, P. R. Grant, L. Andersson, Female-biased gene flow between two species of Darwin's finches. *Nat. Ecol. Evol.* **4**, 979–986 (2020).
23. H. L. Farrington, L. P. Lawson, C. M. Clark, K. Petren, The evolutionary history of Darwin's finches: Speciation, gene flow, and introgression in a fragmented landscape. *Evolution* **68**, 2932–2944 (2014).
24. S. Lamichhaney, F. Han, M. T. Webster, L. Andersson, B. R. Grant, P. R. Grant, Rapid hybrid speciation in Darwin's finches. *Science* **359**, 224–228 (2018).
25. P. R. Grant, B. R. Grant, *40 Years of Evolution. Darwin's Finches on Daphne Major Island* (Princeton Univ. Press, 2014).
26. S. Lamichhaney, F. Han, J. Berglund, C. Wang, M. S. Almén, M. T. Webster, B. R. Grant, P. R. Grant, L. Andersson, A beak size locus in Darwin's finches facilitated character displacement during a drought. *Science* **352**, 470–474 (2016).
27. J. A. Chaves, E. A. Cooper, A. P. Hendry, J. Podos, L. F. de León, J. A. M. Raeymaekers, W. O. MacMillan, J. A. C. Uy, Genomic variation at the tips of the adaptive radiation of Darwin's finches. *Mol. Ecol.* **25**, 5282–5295 (2016).
28. L. P. Lawson, K. Petren, The adaptive genomic landscape of beak morphology in Darwin's finches. *Mol. Ecol.* **26**, 4978–4989 (2017).
29. O. Campàs, R. Mallarino, A. Herrel, A. Abzhanov, M. P. Brenner, Scaling and shear transformations capture beak shape variation in Darwin's finches. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 3356–3360 (2010).
30. G. Zhang, P. Parker, B. Li, H. Li, J. Wang, The genome of Darwin's Finch (*Geospiza fortis*). *GigaScience* (2012); http://dx.doi.org/10.5524/100040.
31. F. Han, S. Lamichhaney, B. R. Grant, P. R. Grant, L. Andersson, M. T. Webster, Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among Darwin's finches. *Genome Res.* **27**, 1004–1015 (2017).
32. B. L. Aken, S. Ayling, D. Barrell, L. Clarke, V. Curwen, S. Fairley, J. F. Banet, K. Billis, C. G. Girón, T. Hourlier, K. Howe, A. Kähäri, F. Kokocinski, F. J. Martin, D. N. Murphy, R. Nag, M. Ruffier, M. Schuster, Y. A. Tang, J.-H. Vogel, S. White, A. Zadissa, P. Flicek, S. M. J. Searle, The Ensembl gene annotation system. *Database* **2016**, baw093 (2016).
33. A. H. Chan, P. A. Jenkins, Y. S. Song, Genome-wide fine-scale recombination rate variation in *Drosophila melanogaster*. *PLOS Genet.* **8**, e1003090 (2012).
34. M. A. Groenen, H. H. Cheng, N. Bumstead, B. F. Benkel, W. E. Briles, T. Burke, D. W. Burt, L. B. Crittenden, J. Dodgson, J. Hillel, S. Lamont, A. P. de Leon, M. Soller, H. Takahashi, A. Vignal, A consensus linkage map of the chicken genome. *Genome Res.* **10**, 137–147 (2000).
35. S. Singhal, E. M. Leffler, K. Sannareddy, I. Turner, O. Venn, D. M. Hooper, A. I. Strand, Q. Li, B. Raney, C. N. Balakrishnan, S. C. Griffith, G. M. Vean, M. Przeworski, Stable recombination hotspots in birds. *Science* **350**, 928–932 (2015).
36. B. Charlesworth, M. T. Morgan, D. Charlesworth, The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**, 1289–1303 (1993).
37. X. Zhou, M. Stephens, Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **44**, 821–824 (2012).
38. R. I. Bowman, Morphological differentiation and adaptation in the galápagos finches. *Univ. Calif. Publ. Zool.* **58**, 1–302 (1961).
39. B. Pfeifer, D. D. Kapan, Estimates of introgression as a function of pairwise distances. *BMC Bioinformatics* **20**, 207 (2019).
40. S. H. Martin, J. W. Davey, C. D. Jiggins, Evaluating the use of ABBA–BABA statistics to locate introgressed loci. *Mol. Biol. Evol.* **32**, 244–257 (2014).

41. C. Y. McLean, D. Bristor, M. Hiller, S. L. Clarke, B. T. Schaar, C. B. Lowe, A. M. Wenger, G. Bejerano, GREAT improves functional interpretation of *cis*-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).
42. P. Pavlidis, J. D. Jensen, W. Stephan, A. Stamatakis, A critical assessment of storytelling: Gene ontology categories and the importance of validating genomic scans. *Mol. Biol. Evol.* **29**, 3237–3248 (2012).
43. R. Mallarino, P. R. Grant, B. R. Grant, A. Herrel, W. P. Kuo, A. Abzhanov, Two developmental modules establish 3D beak-shape variation in Darwin's finches. *Proc. Natl. Acad. Sci. U.S.A.* **198**, 4057–4062 (2011).
44. P. Ducy, R. Zhang, V. Geoffroy, A. L. Ridall, G. Karsenty, Osf2/Cbfa1: A transcriptional activator of osteoblast differentiation. *Cell* **89**, 747–754 (1997).
45. P. Andrade, C. Pinho, G. P. I. de Lanuza, S. Afonso, J. Brejcha, C.-J. Rubin, O. Wallerman, P. Pereira, S. J. Sabatino, A. Bellati, D. Pellitteri-Rosa, Z. Bosakova, I. Bunikis, M. A. Carretero, N. Feiner, P. Marsik, F. Paupério, D. Salvi, L. Soler, G. M. While, T. Uller, E. Font, L. Andersson, M. Carneiro, Regulatory changes in pterin and carotenoid genes underlie balanced color polymorphisms in the wall lizard. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 5633–5642 (2019).
46. E. J. Richards, J. A. McGirr, J. R. Wang, M. E. S. John, J. W. Poelstra, M. J. Solano, D. C. O'Connell, B. J. Turner, C. H. Martin, A vertebrate adaptive radiation is assembled from an ancient and disjunct spatiotemporal landscape. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2011811118 (2021).
47. M. R. Jones, L. S. Mills, P. C. Alves, C. M. Callahan, J. M. Alves, D. J. R. Lafferty, F. M. Jiggins, J. D. Jensen, J. Melo-Ferreira, J. M. Good, Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe hares. *Science* **360**, 1355–1358 (2018).
48. X. Zhang, K. E. Witt, M. M. Bañuelos, A. Ko, K. Yuan, S. Xu, R. Nielsen, E. Huerta-Sanchez, The history and evolution of the Denisovan-*EPAS1* haplotype in Tibetans. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2020803118 (2021).
49. F. Han, M. Jamsandekar, M. E. Pettersson, L. Su, A. P. Fuentes-Pardo, B. W. Davis, D. Bekkevold, F. Berg, M. Casini, G. Dahle, E. D. Farrell, A. Folkvord, L. Andersson, Ecological adaptation in Atlantic herring is associated with large shifts in allele frequencies at hundreds of loci. *eLife* **9**, e61076 (2020).
50. G. A. R. Kingman, D. N. Vyas, F. C. Jones, S. D. Brady, H. I. Chen, K. Reid, M. Milhaven, T. S. Bertino, W. E. Aguirre, D. C. Heins, F. A. von Hippel, P. J. Park, M. Kirch, D. M. Absher, R. M. Myers, F. D. Palma, M. A. Bell, D. M. Kingsley, K. R. Veeramah, Predicting future from past: The genomic basis of recurrent and rapid stickleback evolution. *Sci. Adv.* **7**, eabg5285 (2021).
51. E. D. Enbody, C. G. Sprehn, A. Abzhanov, H. Bi, M. P. Dobreva, O. G. Osborne, C. J. Rubin, P. R. Grant, B. R. Grant, L. Andersson, A multispecies BCO2 beak color polymorphism in the Darwin's finch radiation. *Curr. Biol.* **31**, 5597–5604.e7 (2021).
52. M. Todesco, G. L. Owens, N. Bercovich, J. S. Légaré, S. Soudi, D. O. Burge, K. Huang, K. L. Ostevik, E. B. M. Drummond, I. Imerovski, K. Lande, M. A. Pascual-Robles, M. Nanavati, M. Jahani, W. Cheung, S. E. Staton, S. Muños, R. Nielsen, L. A. Donovan, J. M. Burke, S. Yeaman, L. H. Rieseberg, Massive haplotypes underlie ecotypic differentiation in sunflowers. *Nature* **584**, 602–607 (2020).
53. D. M. Hooper, T. D. Price, Chromosomal inversion differences correlate with range overlap in passerine birds. *Nat. Ecol. Evol.* **1**, 1526–1534 (2017).
54. S. Lamichhaney, G. Fan, F. Widemo, U. Gunnarsson, D. Schwochow Thalmann, M. P. Hoeppner, S. Kerje, U. Gustafson, C. Shi, H. Zhang, W. Chen, X. Liang, L. Huang, J. Wang, E. Liang, Q. Wu, S. M.-Y. Lee, X. Xu, J. Höglund, X. Liu, L. Andersson, Structural genomic changes underlie alternative reproductive strategies in the ruff (*Philomachus pugnax*). *Nat. Genet.* **48**, 84–88 (2016).
55. M. Matschiner, J. M. I. Barth, O. K. Tørresen, B. Star, H. T. Baalsrud, M. S. O. Brieuc, C. Pampoulie, I. Bradbury, K. S. Jakobsen, S. Jentoft, Supergene origin and maintenance in Atlantic cod. *Nat. Ecol. Evol.* **6**, 469–481 (2022).
56. E. R. Funk, N. A. Mason, S. Pálsson, T. Albrecht, J. A. Johnson, S. A. Taylor, A supergene underlies linked variation in color and morphology in a Holarctic songbird. *Nat. Commun.* **12**, 6833 (2021).
57. E. M. Tuttle, A. O. Bergland, M. L. Korody, M. S. Brewer, D. J. Newhouse, P. Minx, M. Stager, A. Betuel, Z. A. Cheviron, W. C. Warren, R. A. Gonser, C. N. Balakrishnan, Divergence and functional degradation of a sex chromosome-like supergene. *Curr. Biol.* **26**, 344–350 (2016).
58. M. S. Almén, S. Lamichhaney, J. Berglund, B. R. Grant, P. R. Grant, M. T. Webster, L. Andersson, Adaptive radiation of Darwin's finches revisited using whole genome sequencing. *Bioessays* **38**, 14–20 (2016).
59. L. Andersson, Molecular consequences of animal breeding. *Curr. Opin. Genet. Dev.* **23**, 295–301 (2013).
60. A. Abzhanov, Collection of embryos from Darwin's finches (Thraupidae, Passeriformes). *Cold Spring Harb. Protoc.* **2009**, pdb.prot5174 (2009).
61. R. D. Mashal, J. Koontz, J. Sklar, Detection of mutations by cleavage of DNA heteroduplexes with bacteriophage resolvases. *Nat. Genet.* **9**, 177–183 (1995).
62. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
63. S. Picelli, A. K. Björklund, B. Reinius, S. Sagasser, G. Winberg, R. Sandberg, Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* **24**, 2033–2040 (2014).
64. A. Abzhanov, In situ hybridization analysis of embryonic beak tissue from Darwin's finches. *Cold Spring Harb. Protoc.* **2009**, pdb.prot5175 (2009).
65. C. Darwin, *Journal of Researches into the Geology and Natural History of the Various Countries visited by H.M.S. Beagle, under the Command of Captain FitzRoy, R.N. from 1832 to 1936* (Henry Colbum, London, 1839).
66. E. Paradis, K. Schliep, ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).
67. L. J. Revell, phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012).
68. J. V. Peñalba, J. B. W. Wolf, From molecules to populations: Appreciating and estimating recombination rate variation. *Nat. Rev. Genet.* **21**, 476–492 (2020).
69. P. R. Grant, I. Abbott, D. Schluter, R. L. Curry, L. K. Abbott, Variation in the size and shape of Darwin's finches. *Biol. J. Linn. Soc.* **25**, 1–39 (1985).
70. P. K. Albers, G. McVean, Dating genomic variants and shared ancestry in population-scale sequencing data. *PLOS Biol.* **18**, e3000586 (2020).
71. A. Abzhanov, S. J. Rodda, A. P. McMahon, C. J. Tabin, Regulation of skeletogenic differentiation in cranial dermal bone. *Development* **134**, 3133–3144 (2007).
72. J. Ruan, H. Li, Fast and accurate long-read assembly with wtdbg2. *Nat. Methods* **17**, 155–158 (2020).
73. B. J. Walker, T. Abeel, T. Shea, M. Priest, A. Abouelliel, S. Sakthikumar, C. A. Cuomo, Q. Zeng, J. Wortman, S. K. Young, A. M. Earl, Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLOS ONE* **9**, e112963 (2014).
74. J. T. Robinson, D. Turner, N. C. Durand, H. Thorvaldsdóttir, J. P. Mesirov, E. L. Aiden, Juicebox.js provides a cloud-based visualization system for Hi-C data. *Cell Syst.* **6**, 256–258 (2018).
75. M. Seppey, M. Manni, E. M. Zdobnov, BUSCO: Assessing genome assembly and annotation completeness. *Methods Mol. Biol.* **1962**, 227–245 (2019).
76. Uniprot Consortium, UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2020).
77. A. F. A. Smit, R. Hubley, P. Green. RepeatMasker Open-4.0 (2015); https://www.repeatmasker.org/.
78. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v2 (2013).
79. D. Freed, R. Aldana, J. A. Weber, J. S. Edwards, The Sentieon Genomics Tools - A fast and accurate solution to variant calling from next-generation sequence data. *bioRxiv* , 115717 (2017).
80. R. Poplin, V. Ruano-Rubio, M. A. DePristo, T. J. Fennell, M. O. Carneiro, G. A. Van der Auwera, D. E. Kling, L. D. Gauthier, A. Levy-Moonshine, D. Roazen, K. Shakir, J. Thibault, S. Chandran, C. Whelan, M. Lek, S. Gabriel, M. J. Daly, B. Neale, D. G. MacArthur, E. Banks, Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv* , 201178 (2018).
81. P. Danecek, A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. M. Vean, R. Durbin; 1000 Genomes Project Analysis Group, The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
82. M. Martin, M. Patterson, S. Garg, S. O. Fischer, N. Pisanti, G. W. Klau, A. Schöenhuth, T. Marschall, WhatsHap: Fast and accurate read-based phasing. *bioRxiv* , 085050 (2016).
83. O. Delaneau, J.-F. Zagury, M. R. Robinson, J. L. Marchini, E. T. Dermitzakis, Accurate, scalable and integrative haplotype estimation. *Nat. Commun.* **10**, 5436 (2019).
84. R Core Team, R: A language and environment for statistical computing (R Foundation for Statistical Computing, 2020).
85. K. L. Korunes, K. Samuk, PIXY: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Mol. Ecol. Resour.* **21**, 1359–1368 (2021).
86. M. Malinsky, M. Matschiner, H. Svardal, Dsuite - Fast D-statistics and related admixture evidence from VCF files. *Mol. Ecol. Resour.* **21**, 584–595 (2021).
87. J. Armstrong, G. Hickey, M. Diekhans, I. T. Fiddes, A. M. Novak, A. Deran, Q. Fang, D. Xie, S. Feng, J. Stiller, D. Genereux, J. Johnson, V. D. Marinescu, J. Alföldi, R. S. Harris, K. Lindblad-Toh, D. Haussler, E. Karlsson, E. D. Jarvis, G. Zhang, B. Paten, Progressive Cactus is a multiple-genome aligner for the thousand-genome era. *Nature* **587**, 246–251 (2020).
88. G. Hickey, B. Paten, D. Earl, D. Zerbino, D. Haussler, HAL: A hierarchical format for storing and analyzing multiple genome alignments. *Bioinformatics* **29**, 1341–1342 (2013).
89. N. Takezaki, M. Nei, K. Tamura, POPTREEW: Web version of POPTREE for constructing population trees from allele frequency data and computing some other quantities. *Mol. Biol. Evol.* **31**, 1622–1624 (2014).
90. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
91. E. M. Ortiz, vcf2phylip v2.0: Convert a VCF matrix into several matrix formats for phylogenetic analysis (version v2.0) (Zenodo, 2019); http://doi.org/10.5281/zenodo.2540861.

92. M. N. Price, P. S. Dehal, A. P. Arkin, FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLOS ONE* **5**, e9490 (2010).
93. C. C. Chang, C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, J. J. Lee, Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience* **4**, 7 (2015).
94. S. Guindon, J.-F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, O. Gascuel, New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
95. S. H. Martin, S. M. Van Belleghem, Exploring evolutionary relationships across the genome using topology weighting. *Genetics* **206**, 429–438 (2017).

Acknowledgments: We thank F. Han for assisting with bioinformatic analysis. The collection of blood samples, funded by National Science Foundation (NSF), was conducted with annual permits from the Galápagos National Parks Directorate, with approval from the Princeton University's Animal Care Committee and in accordance with its protocols and supported logistically by the Charles Darwin Research Station in Galápagos. We thank the ONT for lending us sequencing equipment and for technical assistance related to library preparation and sequencing. D. Ogeh and F. Martin at EMBL-EBI performed gene annotation. Blood samples were collected on San Cristóbal Island under permit number MAE-DNB-CM-2016-0041 to C.A.V., for which we thank Ministerio del Ambiente de Ecuador. Logistical support for researchers to enter the Galápagos and perform laboratory work was provided by GSC–Universidad San Francisco Quito. Tissue samples for expression analysis were collected with permission nos. PC-08-13 and PC-34-14 from Galápagos National Park and MAE-DNB-CM-2016-0043 from Ministerio del Ambiente de Ecuador. **Funding:** The project was financially supported by Vetenskapsrådet (2017-02907) and Knut and Alice Wallenberg Foundation (KAW 2016.0361). M.P.D. was supported by the European Union's Horizon 2020 program under the Marie Sklodowska-Curie grant agreement no. 702707. The National Genomics Infrastructure (NGI)/Uppsala Genome Center provided service in massive parallel sequencing, and the computational infrastructure was provided by the Swedish National Infrastructure for Computing (SNIC) at UPPMAX, partially funded by the Swedish Research Council through grant agreement no. 2018-05973. **Author contributions:** P.R.G. and B.R.G. collected the blood samples. L.A., C.-J.R., P.R.G., B.R.G., and E.D.E. conceived the study. C.-J.R., S.L., and K.V. performed ONT sequencing on Galápagos. C.-J.R. and E.D.E. were responsible for the bioinformatic analysis. A.A. and M.P.D. collected embryonic material, prepared RNA samples, and performed ISH. B.W.D., M.P., and A.T.S.-P. contributed to the bioinformatic analysis. C.G.S. and O.W. contributed to experimental work. C.A.V. contributed to sample collection. L.A., E.D.E., B.R.G., P.R.G, and C.-J.R. wrote the paper with input from other authors. All authors approved the manuscript before submission. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The ONT reads and the Illumina reads have been submitted to the Sequence Read Archive (www.ncbi.nlm.nih.gov/sra) under BioProject PRJNA743742. The Camarhynchus_parvulus_V1.1 reference is available under GenBank assembly accession GCA_902806625.1. Code availability: The analyses of data have been carried out with publicly available software, and all are cited in Methods. Code associated with bioinformatic analyses are available at Zenodo DOI: https://doi.org/10.5281/zenodo.6639328.

Submitted 28 September 2021
Accepted 25 May 2022
Published 8 July 2022
10.1126/sciadv.abm5982

ScienceAdvances

Rapid adaptive radiation of Darwin's finches depends on ancestral genetic modules

Carl-Johan RubinErik D. EnbodyMariya P. DobrevaArhat AbzhanovBrian W. DavisSangeet LamichhaneyMats PetterssonAshley T. Sendell-PriceC. Grace SprehnCarlos A. ValleKarla VascoOla WallermanB. Rosemary GrantPeter R. GrantLeif Andersson

Sci. Adv., 8 (27), eabm5982. • DOI: 10.1126/sciadv.abm5982

Science Advances (ISSN) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science Advances* is a registered trademark of AAAS.