American Genetic Association

Journal of Heredity, 2020, 21–32
doi:10.1093/jhered/esz057
Symposium Article
Advance Access publication November 14, 2019

OXFORD

Symposium Article

Adaptive Radiation Genomics of Two Ecologically Divergent Hawai'ian Honeycreepers: The 'akiapōlā'au and the Hawai'i 'amakihi

Michael G. Campana, André Corvelo, Jennifer Shelton, Taylor E. Callicrate, Karen L. Bunting, Bridget Riley-Gillis, Frank Wos, Justin DeGrazia, Erich D. Jarvis, and Robert C. Fleischer

From the Center for Conservation Genomics, Smithsonian National Zoo and Conservation Biology Institute, Washington, DC 20008 (Campana, Callicrate, and Fleischer); New York Genome Center, New York, NY 10013 (Corvelo, Shelton, Bunting, Riley-Gillis, Wos, and DeGrazia); Species Conservation Toolkit Initiative, Chicago Zoological Society, Brookfield, IL 60513 (Callicrate); The Rockefeller University, New York, NY 10065 and Howard Hughes Medical Institute, Chevy Chase, MD 20815 (Jarvis).

Address correspondence to Michael G. Campana at the address above, or e-mail: campanam@si.edu.

Received March 4, 2019; First decision May 9, 2019; Accepted September 30, 2019.

Corresponding Editor: Rosemary Gillespie

Abstract

The Hawai'ian honeycreepers (drepanids) are a classic example of adaptive radiation: they adapted to a variety of novel dietary niches, evolving a wide range of bill morphologies. Here we investigated genomic diversity, demographic history, and genes involved in bill morphology phenotypes in 2 honeycreepers: the 'akiapōlā'au (*Hemignathus wilsoni*) and the Hawai'i 'amakihi (*Chlorodrepanis virens*). The 'akiapōlā'au is an endangered island endemic, filling the "woodpecker" niche by using a unique bill morphology, while the Hawai'i 'amakihi is a dietary generalist common on the islands of Hawai'i and Maui. We de novo sequenced the 'akiapōlā'au genome and compared it to the previously sequenced 'amakihi genome. The 'akiapōlā'au is far less heterozygous and has a smaller effective population size than the 'amakihi, which matches expectations due to its smaller census population and restricted ecological niche. Our investigation revealed genomic islands of divergence, which may be involved in the honeycreeper radiation. Within these islands of divergence, we identified candidate genes (including *DLK1*, *FOXB1*, *KIF6*, *MAML3*, *PHF20*, *RBP1*, and *TIMM17A*) that may play a role in honeycreeper adaptations. The gene *DLK1*, previously shown to influence Darwin's finch bill size, may be related to honeycreeper bill morphology evolution, while the functions of the other candidates remain unknown.

Subject areas: Genomics and gene mapping, Conservation genetics and biodiversity
Keywords: bill morphology, *Chlorodrepanis virens*, demography, *Hemignathus wilsoni*, islands of divergence, natural selection

Adaptive radiation is a process that involves branching speciation with differential adaptation (Futuyma 1986). It usually involves morphological change that allows species to occupy alternative niches, and derived species may become more ecologically specialized. The genomics of this process has only recently begun to be studied (Berner and Salzburger 2015; Wolf and Ellegren 2017; Malinsky et al. 2018; Salzburger 2018), with some key examples deriving from studies of groups such as African lake cichlids (Brawand et al. 2014;

Henning and Meyer 2014; Malinsky et al. 2018; Salzburger 2018), *Heliconius* butterflies (Joron et al. 2011; Reed et al. 2011; Supple et al. 2013; Kozak et al. 2018), and *Anolis* lizards (Tollis et al. 2018). Of particular interest is discovering the genomic architectures that enable adaptive radiation to proceed, and the number and types of genes and genetic changes involved in the processes of speciation and adaptation (e.g., Salzburger 2018; Tollis et al. 2018). In birds, the latter aspect has gained insight via the study of the radiation of Darwin's finches (Geospizinae), through both gene expression analyses (Abzhanov 2004, 2006), genome scans and genome-wide association studies (GWAS) (Lamichhaney et al. 2015, 2016). These studies identified genes (*ALX1* and *BMP4*) and the calmodulin pathway as being critical to the adaptation of Darwin's finches' bill morphologies to varying environments and diets. Unfortunately, few other avian radiations have been analyzed using such methods to date.

The Hawai'ian honeycreepers (drepanids) are a spectacular example of rapid adaptive radiation in a species-depauperate island system. Currently classified in the Carduelinae subfamily of the Fringillidae (finch) family (Fleischer et al. 2001; James 2004; Lerner et al. 2011; Zuccon et al. 2012), the more than 50 honeycreeper species arose from a single colonization of the remote Hawai'ian islands by rosefinch-like (*Carpodacus* spp.) ancestors between 5.2 and 8.4 million years ago (Mya; Lerner et al. 2011). In addition to the typical "finch" granivorous niche (e.g., palila [*Loxioides bailleui*]), honeycreepers fill nectarivorous (e.g., 'apapane [*Himatione sanguinea*]), generalist (e.g., Hawai'i 'amakihi [*Chlorodrepanis virens*]), arthropodivorous (e.g., 'akikiki [*Oreomystis bairdi*]), and frugivorous (e.g., 'o'u [*Psittirostra psittacea*]) roles. To fill these ecological niches, honeycreepers evolved numerous bill morphologies, including some relatively unique forms of unknown function in some of the extinct species (James 2004), the genetics behind which are currently unknown.

One of the most extreme examples of bill morphological adaptation is the 'akiapōlā'au (*Hemignathus wilsoni*; Figure 1), the last confirmed extant species in its genus. The 'akiapōlā'au is endemic to the dry and montane moist forests of the youngest island of Hawai'i (Ralph and Fancy 1996). An ecological specialist, the 'akiapōlā'au fills the "woodpecker" niche, using its shorter, straight mandible to peck bark and its longer, curved maxilla to extract insects for consumption. The species prefers to forage on koa (*Acacia koa*) trees, but will also utilize 'ōhi'a lehua (*Metrosidros polymorpha*) and naio (*Myoporum sandwicense*) (Ralph and Fancy 1996). It has small census population sizes (currently ~1000 individuals: Gorresen et al. 2007) due to its limited niche and slow population growth rates (Ralph and Fancy 1996). Even compared to its closest known relatives, the extinct nukupu'us (*Hemignathus hanapepe*, *H. lucidus*, and *H. affinis*: Fleischer 2009; Fleischer RC, Campana MG, et al., unpublished data), the 'akiapōlā'au has a unique bill morphology—the nukupu'us had longer bills and a curved lower mandible, while the 'akiapōlā'au has a straight and stout lower mandible (Pratt 2005). The nukupu'us apparently became extinct by the late 1990s (although accuracy of sightings post-1900 is debated: Elphick et al. 2010; Roberts et al. 2009). The 'akiapōlā'au is itself endangered (International Union for the Conservation of Nature [IUCN] Red List categories: B1ab(i,ii,iii,iv,v); C2a(ii)) (IUCN 2019), necessitating conservation action to prevent its extinction (e.g., Gorresen et al. 2007).

Unfortunately, the Hawai'ian honeycreepers in general are an extremely threatened clade: of the 17 confirmed extant species (18 if Maui 'amakihi [*Chlorodrepanis virens wilsoni*] is considered a separate species rather than subspecies), the IUCN rates 6 as critically endangered, 4 as endangered, and 5 as vulnerable. Another 22 species (and the Lāna'i 'alauahio subspecies [*Paroreomyza montana montana*]) either went extinct or likely went extinct (no recent confirmed sightings) since the arrival of Europeans to the Hawai'ian islands. At least 17 additional extinct species are known only from subfossil material (James and Olson 1991). In addition to the grave conservation concern for these emblematic Hawai'ian forest birds, the endangerment and extinction of most members of the clade hinders our ability to research most of these taxa due to the limited availability of samples.

Callicrate et al. (2014) published a reference-assisted de novo assembly of the Hawai'i 'amakihi genome. The Hawai'i 'amakihi is a dietary generalist honeycreeper species endemic to the islands of Hawai'i and Maui. Each island has a unique subspecies, and the 2 subspecies are genetically differentiated (Tarr and Fleischer 1993; Fleischer et al. 1998; Lerner et al. 2011). Rated as Least Concern by the IUCN, the Hawai'i 'amakihi is common throughout its range (Pratt 2005), even occupying lower elevations on Hawaii Island despite the presence of invasive avian malaria (*Plasmodium relictum* strain GRW4; Beadell et al. 2006)—lethal to most honeycreepers—in these regions (Foster et al. 2007; Atkinson et al. 2013). It has evolved a relatively small and thin bill and feeds largely on a mixed diet of invertebrates and nectar (Pratt 2005; Figure 1).

Currently, we can only compare the 'amakihi genome with distantly related nondrepanid passerine genomes, severely limiting our ability to detect traits related to the honeycreeper adaptive radiation. Additional genomic resources (including genomes, transcriptomes, exomes, genome-wide marker panels, etc.) are critical to understand the honeycreeper radiation and evolution of bill morphologies. These resources would permit the interrogation of the complete genome (including regulatory regions, putative noncoding sequences, and genes of unknown function) for signs of selection

Figure 1. Comparison of bill morphologies between the 'akiapōlā'au (*Hemignathus wilsoni*) and the Hawai'i 'amakihi (*Chlorodrepanis virens*). The specialist 'akiapōlā'au uses its straight mandible to peck bark and its curved maxilla to extract insects. The 'amakihi has a relatively small and thin bill and feeds largely on a mixed diet of invertebrates and nectar. Photographs by Jack Jeffrey (used with permission).

(e.g., greatly increased or decreased genomic heterozygosity and divergence), thereby greatly increasing detection power for causative traits over candidate gene approaches. For instance, only with the power of transcriptomics, genome scans, and GWAS were genes responsible for the variation in bill morphologies in Darwin's finches identified (Abzhanov 2004, 2006; Lamichhaney et al. 2015, 2016). Similarly, the 4 candidate regions that likely explain plumage differences between golden-winged (*Vermivora chrysoptera*) and blue-winged warblers (*V. cyanoptera*) required analysis of these species' complete genomes (Toews et al. 2016).

Therefore, to explore genomic effects of niche specialization, identify candidate genes and regions to explain the rapid radiation of honeycreeper bill morphology, and inform honeycreeper conservation strategies, we de novo sequenced the genome of the ʻakiapōlāʻau. The recent divergence of the Hawaiʻi ʻamakihi from the ʻakiapōlāʻau (~3.0 Mya: Lerner et al. 2011) facilitates genomic alignment. In addition, their co-occurrence on the island of Hawaii, which provides a maximum age of about 1 million years (Lipman and Calvert 2013), and their differing degrees of morphological and ecological specialization, makes the comparison of the genomes of the 2 species potentially important for determining patterns and rates of genome evolution and adaptation. We tested the following hypotheses:

1. Due to its smaller census size and limited niche, the ʻakiapōlāʻau will have lower heterozygosity, a smaller effective population size (N_e) and a differing demographic history from the more numerous, dietary generalist ʻamakihi.
2. The fully de novo ʻakiapōlāʻau genome sequence will facilitate more accurate variant calling and genomic reconstructions within honeycreepers than analyses using divergent reference sequences.
3. The ʻakiapōlāʻau and ʻamakihi genomes will be more divergent at or near loci responsible for adaptive traits such as bill morphology than in the rest of the genome. These divergent regions may be in previously identified adaptive genes that may be convergent in other species, newly discovered genes of unknown function, putatively noncoding sequences, or regulatory regions.

Methods

Genome Sequencing and Assembly

We sequenced a wild male ʻakiapōlāʻau (SOL748, USGS band number 1581-74856) captured at Solomon's Camp, Mauna Loa, Hawaii, United States on 22 February 2002. Whole blood was obtained by brachial venipuncture. Genomic DNA was extracted using a standard phenol-chloroform isolation protocol (Tarr and Fleischer 1993). Three Illumina libraries were constructed from the genomic DNA at the New York Genome Center: a standard paired-end short-insert (550 base pair [bp] average length) library and 2 mate-pair libraries with insert lengths of 3–5 kilobase pair (kbp) and 8–10 kbp, respectively. Libraries were 2 × 125 bp paired-end sequenced on 3 lanes of an Illumina HiSeq 2500.

Polymerase chain reaction (PCR) duplicates were removed using an in-house tool at the New York Genome Center. Adapters and low-quality bases were trimmed using Cutadapt 1.8.1 (Martin 2011). For the mate-pair libraries, junction adapter identification and clipping were also performed using Cutadapt. Contaminants were removed via GEM pre-rel3 by alignment against a spiked-in PhiX sequencing control (Marco-Sola et al. 2012). Joint error correction was performed using Lighter 1.0.6 (Song et al. 2014).

The genome was assembled using ABySS 1.9.0 (Simpson et al. 2009) with a *K*-mer length of 71. Gap closure was performed using ABySS 1.9.0 Sealer under default settings except that we used a larger bloom filter size (2400 MB) and maximum number of branches in the de Bruijn graph traversal (3000). We assembled the genome using a range of *K*-mer sizes (61, 71, 75, 79, 83, 87) and chose the assembly that produced the largest scaffold N50 values.

Genome assembly statistics were calculated using the assemblathon_stats.pl script from the Assemblathon 2 competition (Bradnam et al. 2013). Genic completeness was calculated using BUSCO 3.0.2 (option --long) with the aves_odb9 gene set (totaling 4915 orthologs) and the AUGUSTUS "chicken" model as a starting data set (Stanke et al. 2006a; Waterhouse et al. 2017). For comparison, we also ran BUSCO on the ʻamakihi genome (both the Callicrate assembly and the ʻakiapōlāʻau-realigned sequence; see below).

Variant Calling and Consensus Sequence Generation

The Callicrate ʻamakihi genome sequence is not directly comparable to our de novo ʻakiapōlāʻau genome due to differing sequencing technologies and assembly approaches. The ʻamakihi contigs were generated from a combination of ~2× Roche/454 and ~19× Illumina GAII sequences (Callicrate et al. 2014). The authors generated ~60× Illumina coverage, but downsampled their data for contig assembly due to both computer memory limits and the need to maintain the Roche/454 long reads in the assembly. The contigs were then oriented to the zebra finch (*Taeniopygia guttata*) genome (Warren et al. 2010) to generate a chromosome-level assembly. This genomic orientation could impact genomic analyses and interpretations by incorrectly assuming synteny where there have been genomic rearrangements.

Therefore, in order to compare the genomes of the ʻakiapōlāʻau and the ʻamakihi genomes, we realigned the short-insert Illumina reads from both species to the ʻakiapōlāʻau genome assembly using BWA-MEM 0.7.12 (Li 2013). We omitted the ʻakiapōlāʻau mate-pair and ʻamakihi Roche/454 sequences to minimize batch effects due to differing sequence technologies. PCR duplicates were removed using Picard 1.141 MarkDuplicates (Broad Institute 2015). Sequence variants were called and consensus sequences were generated using SAMtools 1.2 mpileup (option -C50) and BCFtools 1.2 call (option -c) (Li et al. 2009). We used the older consensus variant caller since downstream PSMC (pairwise sequential Markovian coalescent) analysis (Li and Durbin 2011) is not compatible with the newer multiallelic algorithm. Variant calls were converted to FASTQ format using vcfutils.pl vcf2fq. We retained single-nucleotide variants (SNVs) with quality scores of at least 20 for subsequent analysis.

To control for ʻakiapōlāʻau reference bias (defined here as false positive and negative variants derived from inaccurate mapping to a divergent reference sequence) impacting our downstream estimates of substitution rates and demography reconstructions, we also aligned the ʻakiapōlāʻau and ʻamakihi short-insert Illumina reads to the zebra finch genome (assembly taeGut3.2.4; Warren et al. 2010) and the Callicrate ʻamakihi genome assembly using BWA-MEM 0.7.17 (Li et al. 2013). Mate-pair tags were fixed and PCR duplicates were marked using SAMtools 1.9 (Li et al. 2009) fixmate (option -m) and markdup, respectively. Indels were left aligned using the Genome Analysis Toolkit 4.1.0.0 LeftAlignIndels command (McKenna et al. 2010). Variants were called using BCFtools 1.9 mpileup (option -a DP,AD) and BCFtools call (option -c) (Li et al. 2009). We retained SNVs with quality scores of at least 20.

Genomic Diversity and Divergence

Because the ‘akiapōlā‘au was male and the ‘amakihi was female, we excluded the sex chromosomes from downstream heterozygosity, divergence, and demographic analyses. Moreover, exclusion of the Z improved these estimates since the sex chromosomes are difficult to reconstruct accurately due to the presence of repeat elements and homologous sequences between the Z and W (Smeds et al. 2015). We also excluded the nonrecombining mitochondrial genome due to its maternal inheritance. For the zebra-finch- and Callicrate-‘amakihi-aligned analyses, we excluded the previously annotated mitogenome and Z chromosome assemblies (Z and Z_random). For the de novo ‘akiapōlā‘au-aligned analyses, we identified putative mitochondrial and Z chromosome scaffolds by searching the ‘akiapōlā‘au genome for matches to the previously sequenced ‘akiapōlā‘au mitogenome (GenBank accession KM078774.2; Lerner et al. 2011) and the zebra finch Z chromosome assembly (taeGut 3.2.4; Warren et al. 2010) using megaBLAST (BLASTN 2.7.1+: Camacho et al. 2008). Scaffolds with matches of bit score 4000 or greater, sequence identity 90% or greater, and alignment length 1000 or greater were considered to derive from the mitogenome or Z chromosome. We excluded 3 mitochondrial (1127, 1842, 2102) and 59 Z chromosome (28, 47, 55, 91, 101, 105, 107, 111, 120, 122, 131, 133, 141, 144, 147, 151, 154, 159, 160, 165, 179, 182, 186, 187, 189, 200, 211, 219, 225, 228, 253, 291, 295, 310, 319, 327, 353, 356, 357, 362, 363, 397, 408, 409, 412, 422, 432, 453, 466, 468, 491, 493, 513, 541, 634, 638, 1020, 1150) scaffolds.

For each of the 3 data sets (the de novo ‘akiapōlā‘au-aligned data set, the Callicrate-‘amakihi-aligned data set, and the zebra-finch-aligned data set), we then calculated autosomal heterozygosities in 10 000 nonoverlapping windows using VCFtools 0.1.15 (Danecek et al. 2011) following Callicrate et al. (2014). To identify regions of maximal autosomal divergence between the ‘akiapōlā‘au and ‘amakihi, we calculated densities of fixed autosomal SNVs separating the 2 species in 10 000 bp nonoverlapping windows using VCFtools 0.1.15 (Danecek et al. 2011). We considered divergence windows with a z-score > 7 as candidate adaptive loci for further investigation. The stringent z-score cutoff minimized false positives due to misalignments, sequence and variant errors, and multiple testing (~95 000 windows per analysis). In the ‘akiapōlā‘au-aligned data set, we discarded contigs and scaffolds below 10 000 bp in length from the heterozygosity and divergence analyses since these had artificially reduced values. Additionally, the ‘amakihi and zebra finch reference sequences include artificial “chromosomes” (‘amakihi: Un and Un2; zebra finch: Un) consisting of randomly ordered, concatenated unplaced contigs. We also excluded these “chromosomes” from the heterozygosity and divergence analyses since they had aberrant values reflecting misalignments and genotyping errors (data not shown). Finally, for each data set, we searched for runs of homozygosity of at least 1 Mbp using BCFtools 1.9 roh (Li et al. 2009) assuming allele frequencies of 50% (option --Af-dflt 0.50).

Estimation of Genomic Substitution Rates

Accurate estimation of genomic substitution rates is challenging due to a wide range of factors, including (but not limited to) accuracy and precision of selected calibration points (e.g., Fleischer et al. 1998), genomic variant sequencing depth and quality, and time dependency of calculated rates (Ho et al. 2005, 2007). Therefore, we calculated autosomal substitution rates using 3 procedures to observe the range of estimates. To account for differing levels of calibration precision between estimators, we rounded all values to 2 significant figures and used an autosomal genome length of 950 Mbp for all analyses.

As a first estimate of autosomal substitution rates, we assumed complete lineage sorting within the honeycreeper tree and that all observed within-species polymorphisms arose after colonization of Hawai‘i island. Although this assumption could be incorrect (incomplete lineage sorting is common in birds: Jarvis et al. 2014), the patterns of differentiation (Tarr and Fleischer 1993; Eggert et al. 2009) and extremely low degree of hybridization among honeycreeper taxa (Knowlton et al. 2014) suggest that gene flow from Maui or other older islands was not likely for either species and probably ceased once they became differentiated from their sister taxa. Since each mutation could occur on either parental chromosome copy (Malinsky et al. 2018), we estimated the within-species substitution rates using the following formula:

$$\mu_C = N_H/L/T_C/2$$

Here μ_C is the genomic single-nucleotide substitution rate, N_H is the number of heterozygous single-nucleotide polymorphisms within the branch, L is the sequence length, and T_C is the time of colonization. We calculated μ_C using 1.7 million and 3.9 million heterozygous sites for the ‘akiapōlā‘au and ‘amakihi, respectively. We used T_C values of 0.43 Mya (Carson and Clague 1995; Fleischer et al. 1998) and 1.0 Mya (Lipman and Calvert 2013).

The first model may overestimate the long-term genomic substitution rate due to the effects of time dependency on molecular rate estimates at recent time scales (e.g., lineage extinction and loss of nonfixed alleles: Ho et al. 2005, 2007) and by ignoring incomplete lineage sorting. To estimate long-term genomic substitution rates, we calculated the rate of divergence between the 2 species using the coalescent formula:

$$\mu_D = N/L/T_{MRCA}/2$$

Under this model, μ_D is the mean substitution rate based on the divergence between the 2 species, N is the number of single-nucleotide substitutions, and T_{MRCA} is the time of most recent common ancestor. We calculated μ_D using the total number (9.9 million) of observed substitutions among lineages. The ‘akiapōlā‘au and Hawai‘i ‘amakihi divergence time was 3.0 Mya (Lerner et al. 2011). This model calculates the average substitution rate and does not account for rate heterogeneity between the honeycreeper branches.

Finally, we repeated the divergence-based analysis using the alignment against the zebra finch. We used 75 million observed substitutions from the zebra finch sequence and an estimated genome divergence time from zebra finch of 15 Mya (based on the estimated separation of Fringillidae [represented by *Fringilla*] and Estrildidae [represented by *Estrilda*]: Moyle et al. 2016). As this model calculates the average rate between the Fringillidae and Estrildidae, the estimated honeycreeper rate will only be valid if evolutionary rates are similar between these 2 deeply diverged lineages.

Genomic Population History Analysis

We reconstructed genomic population histories for the 2 species using PSMC (Li and Durbin 2011). PSMC estimates ancestral changes in N_e from a single genome using the coalescent (Li and Durbin 2011). The program divides the genome into windows and scores each window as heterozygous or homozygous. Loci are defined as contiguous segments of windows with a similar heterozygosity rate.

PSMC uses a Hidden Markov Model to infer recombination points between loci where the heterozygosity rate changes. The number of heterozygous windows within each locus is proportional to the T_{MRCA} of that window. The number of loci at each T_{MRCA} is inversely proportional to N_e at that time. The resulting population history is then scaled by the mutation rate and the generation time. The interpretation of PSMC is thus sensitive to inaccurate mutation rates and generation times (Li and Durbin 2011). PSMC does not recover recent events accurately due to the limited number of heterozygous SNVs. It also tends to interpret rapid population size changes (e.g., a strong bottleneck effect) as gradual clines and can be confused by population structure (Li and Durbin 2011).

To control for the impact of reference bias on demographic reconstruction, we analyzed the de novo ʻakiapōlāʻau genome aligned sequences, the Callicrate ʻamakihi assembly aligned sequences, and the zebra-finch-aligned genomes independently. We explored the impact of the concatenated, unplaced contig "chromosomes" (Un for zebra finch; Un and Un2 for ʻamakihi) on demographic reconstructions by repeating the analyses with these sequences excluded. As recommended by the PSMC authors, we generated depth-filtered consensus sequences for each assembly setting the minimum and maximum depths to one-third and twice the average short-insert sequencing depths (38×–228× for ʻakiapōlāʻau and 20×–120× for ʻamakihi) using BCFtools 1.9 mpileup and BCFtools call (option -c) (Li et al. 2009). Variant calls were converted to FASTQ format using vcfutils.pl vcf2fq. We retained variants with qualities of at least 20 for the analysis. Bootstrapped PSMC (100 replicates) was performed using PSMC 0.6.5 (options -N 25, -t 15, -r 5, -p "64*1"). We scaled the PSMC reconstructions using the range of estimated genomic substitution rates (see above) and generation times of 2 years for the ʻakiapōlāʻau and 1 year for the ʻamakihi (Pratt 2005, p. 154). We also scaled the PSMC reconstruction using the collared flycatcher (*Ficedula albicollis*) germ-line mutation rate (4.6×10^{-9} substitutions/site/generation; Smeds et al. 2016) since no drepanid rate is currently available.

Genome Annotation and Gene Prediction

We annotated the de novo ʻakiapōlāʻau assembly, the realigned ʻamakihi consensus sequence, and the original Callicrate ʻamakihi genome sequence for comparison. We trained AUGUSTUS gene prediction using the BUSCO 3.0.2 retraining results (Waterhouse et al. 2017). We annotated repeats for the 3 assemblies using RepeatMasker 4.0.7 (Smit et al. 2013–2015) and Repbase Update 20170127 (all organisms; Bao et al 2015). We obtained the zebra finch (assembly taeGut 3.2.4; Warren et al. 2010) proteins and cDNAs from the Ensembl database. We mapped the zebra finch proteins and cDNAs to the genome assemblies using Exonerate 2.2 protein2genome (Slater and Birney 2005) and BLAT 36x1 (minimum sequence identity 92%; Kent 2002), respectively. We converted the annotated repeats, proteins, and cDNAs to hints for AUGUSTUS 3.3 and predicted genes using the BUSCO-trained model (Stanke et al. 2006a, 2006b, 2008). Hint weights used AUGUSTUS defaults except that the RepeatMasker hints were weighted as 1.15 for "nonexonpart" and 1 for all other states. We identified orthologous predicted genes between the 3 assemblies using Proteinortho 5.16b (Lechner et al. 2011). We classified predicted proteins' domains, functions, and pathways using InterProScan 5.33-72.0 (options -goterms -pa; Jones et al. 2014) against all eukaryotic databases. We used the InterProScan classifications to investigate predicted proteins that overlapped highly divergent regions between the ʻakiapōlāʻau and ʻamakihi identified in the ʻakiapōlāʻau-aligned divergence analysis. For comparison, we investigated Ensembl-annotated zebra finch genes (annotation taeGut3.2.4) that overlapped divergent regions in the ʻamakihi-aligned and zebra-finch-aligned data sets.

Bill Morphology Gene Analysis

We compiled a list of genes previously demonstrated to be associated with bill morphology in birds (*ALX1*, *BMP4*, *CALM1*, *CALM2*, *CALML3*, *DLK1*, *HMGA2*: Abzhanov et al. 2004, 2006; Lamichhaney et al. 2015, 2016; Chaves et al. 2016). We aligned these candidate genes' protein sequences annotated in Ensembl for zebra finch (assembly taeGut 3.2.4; Warren et al. 2010) against the ʻakiapōlāʻau and ʻamakihi genome sequences using TBLASTN 2.7.0+ (maximum expect value: 0.1; Altschul et al. 1997). We retained the best hits for each exon. We then manually inspected retained alignments to identify sequence variants for further analysis. To control for reference bias in the ʻamakihi genome, we assessed both the Callicrate assembly and the ʻakiapōlāʻau-realigned sequence.

TBLASTN did not identify exon 1 of *HMGA2* in both the ʻakiapōlāʻau and ʻamakihi genomes (see below). To confirm this result, we searched for the corresponding exon 1 DNA sequence using megaBLAST (BLASTN 2.7.1+: Camacho et al. 2008).

Since we identified 2 *DLK1* amino acids that differed between the ʻakiapōlāʻau and ʻamakihi (see below), we investigated this gene phylogenetically to determine the uniqueness of these variants. We obtained *DLK1* protein sequences predicted using the National Center for Biotechnology Information annotation pipeline for 66 avian species in GenBank. We aligned them using the Geneious and ClustalW 2.1 (Larkin et al. 2007) aligners in Geneious Prime 2019.0.4 (Biomatters Ltd, Auckland, New Zealand). We trimmed the alignment to the Ensembl-annotated zebra finch sequence. We manually inspected the trimmed alignment for erroneous stop codons and indel misalignments. We then constructed a rapid bootstrapping (100 replicates), maximum likelihood phylogenetic tree using the RAxML 8.2.11 under the GAMMA BLOSUM62 protein substitution model with a random starting tree (Stamatakis 2014).

Results

Genome Assembly

We sequenced the ʻakiapōlāʻau genome to a nominal coverage of ~134× (Table 1). We generated 915 million short-insert paired reads (~114×), 107 million 3–5 kbp mate pairs (~13×), and 58 million 8–10 kbp (~7×). The genome was assembled into 40 124 contigs (contig N50: 75 423 bp) and 9908 scaffolds (scaffold N50: 3 340 285 bp). A total of 32 612 (98.7%) contigs were assembled into scaffolds. The total contig assembly length was 1 014 901 717 bp and the total scaffolded length was 1 032 508 528 bp, similar to previously sequenced bird genomes (Jarvis et al. 2014).

Genic content was very complete for all assemblies. BUSCO found ~94% of the 4915 BUSCO groups in all assemblies, with low rates of duplication (~1%) and missingness (2% in the ʻakiapōlāʻau, 2.5% in the ʻamakihi) (Table 2). We observed slightly better BUSCO scores for the original Callicrate ʻamakihi assembly than the realigned sequence, reflecting variance between BUSCO runs or possible reference bias.

Genomic Diversity and Divergence

In the ʻakiapōlāʻau-aligned analyses, we observed 1 665 867 heterozygous autosomal SNVs in the ʻakiapōlāʻau or one every 570 bases

Table 1. ʻAkiapōlāʻau genome assembly statistics

Total scaffold length:	1 032 508 528 bp	Total contig length:	1 014 901 717 bp
Number scaffolds:	9908	Number contigs:	40 124
Scaffold N50:	3 340 285 bp	Contig N50:	75 423 bp
Scaffold L50:	85	Contig L50:	3815
Longest scaffold:	15 969 922 bp	Longest contig:	629 073 bp
Mean scaffold length:	104 210 bp	Mean contig length:	25 294 bp
Median scaffold length::	1321 bp	Median contig length:	7069 bp
Contig %A:	29.03%	Contig %C:	20.95%
Contig %G:	20.97%	Contig %T:	29.04%
Mean break length between scaffolded contigs	582 bp	Nominal sequencing coverage:	134×

Table 2. BUSCO results

BUSCO category	ʻAkiapōlāʻau	ʻAmakihi (Callicrate)	ʻAmakihi (realigned)
Complete	4635 (94.3%)	4609 (93.8%)	4604 (93.7%)
Complete, single-copy	4575 (93.1%)	4560 (92.8%)	4549 (92.6%)
Complete, duplicated	60 (1.2%)	49 (1.0%)	55 (1.1%)
Fragmented	181 (3.7%)	181 (3.7%)	189 (3.8%)
Missing	99 (2.0%)	125 (2.5%)	122 (2.5%)
Total BUSCOs searched	4915	4915	4915

Total numbers of identified BUSCOs of each category are given, with the percentage of the total found in parentheses.

across the ~950 Mbp autosomal genome. The realigned ʻamakihi had 3 938 436 or one every 241 bases. The 2 genomes were separated by 4 174 597 fixed SNVs. The ʻakiapōlāʻau's local heterozygosities (mean ± standard deviation: 0.00172 ± 0.00111 heterozygous sites/bp) were less than half (0.416×; unpaired *t*-test $P << 0.0001$) the ʻamakihi's (0.00413 ± 0.00192 heterozygous sites/bp) (Figure 2). We observed one 590 kbp region of the ʻamakihi genome (corresponding to ʻakiapōlāʻau scaffold 261) with exceptionally high heterozygosity (0.0127 heterozygous sites/bp, *z*-score = 4.46), suggestive of a collapsed repetitive region unique to the ʻamakihi genome (Treangen and Salzberg 2012). We also observed elevated heterozygosities in the scaffolds between 10 and 100 kbp (especially in the ʻakiapōlāʻau), which likely derive from assembly or alignment errors of repetitive sequences (Treangen and Salzberg 2012).

There were 1 705 302 ʻakiapōlāʻau SNVs and 3 856 230 ʻamakihi heterozygous autosomal SNVs in the ʻamakihi-aligned data set and 1 931 190 ʻakiapōlāʻau SNVs and 3 063 769 ʻamakihi heterozygous autosomal SNVs in the zebra-finch-aligned data set. In the ʻamakihi-aligned data set, local heterozygosities were similar to the ʻakiapōlāʻau-aligned data set in the ʻakiapōlāʻau (0.00174 ± 0.00159 heterozygous sites/bp), but slightly decreased in the ʻamakihi (0.00393 ± 0.00212 heterozygous sites/bp). The zebra-finch-aligned data set inferred higher ʻakiapōlāʻau local heterozygosities (0.00197 ± 0.00194 heterozygous sites/bp) and substantially lower ʻamakihi values (0.00312 ± 0.00198 heterozygous sites/bp). As a result the ratio of ʻakiapōlāʻau to ʻamakihi heterozygosities increased in these data sets (ʻamakihi-aligned: 0.443×; zebra-finch-aligned: 0.631×; unpaired *t*-test $P << 0.0001$ in both cases). Variance in local heterozygosities was greater in both the ʻamakihi-aligned and zebra-finch-aligned data sets than in the ʻakiapōlāʻau-aligned one. In the ʻamakihi-aligned and zebra-finch-aligned data sets, both species had numerous peaks of apparent local heterozygosity, especially at the telomeres of the chromosomes, suggestive of misaligned, collapsed repetitive elements (Supplementary Figures 1 and 2; Treangen and Salzberg 2012). This indicates that these 2 data sets include less accurately assembled regions excluded from the de novo ʻakiapōlāʻau genome assembly, decreasing overall analytical quality.

While local genomic divergences were relatively constant across the genome within all 3 data sets (ʻakiapōlāʻau-aligned: 0.00439 ± 0.00143 fixed SNVs/bp; ʻamakihi-aligned: 0.00429 ± 0.00170 fixed SNVs/bp; zebra-finch-aligned: 0.00321 ± 0.00118 fixed SNVs/bp), there was decreased mean divergence with reference distance from ʻakiapōlāʻau. In the ʻakiapōlāʻau-aligned data set, divergence exceeded 0.015 fixed SNVs/bp (*z*-scores > 7.4) in 7 distinct regions on ʻakiapōlāʻau scaffolds 39, 62, 119, 349, 548, 788, and 810 (Figure 2). The elevated diversity regions on scaffolds 39, 62, and 119 were 50–70 kbp long, while the remaining were 10–40 kbp. We observed 20 divergent regions in the ʻamakihi-aligned data set across chromosomes 1 ($n = 1$), 2 ($n = 5$), 3 ($n = 1$), 4 ($n = 2$), 4A ($n = 1$), 8_random ($n = 1$), 9 ($n = 1$), 10 ($n = 1$), 15 ($n = 2$), 18 ($n = 1$), 20 ($n = 1$), 26 ($n = 1$), 22_random ($n = 1$), and 27 ($n = 1$). Most (14) of these were ~10 kbp long. Six longer (30–60 kbp) divergent regions were observed on chromosomes 4, 4A, 8_random, 9, 10, and 15. We observed only 4 divergent regions in the zebra-finch-aligned data set: 3 of these matched the longer divergent regions on chromosomes 4A, 8_random, and 10 in the ʻamakihi-aligned data set and the remaining was a short (~10 kbp) region on chromosome 2 close (~55 kbp away) to one of the divergent regions in the ʻamakihi-aligned data set.

We detected no long (1 Mbp) runs of homozygosity in either genome in the ʻakiapōlāʻau- or zebra-finch-aligned sequences. In agreement with Callicrate et al. (2014), we observed 8 long runs of homozygosity on chromosomes 1 ($n = 2$) and 6 ($n = 6$) in the ʻamakihi in the ʻamakihi-aligned data set. No runs of homozygosity were detected in the ʻakiapōlāʻau in this data set. The absence of these runs of homozygosity in the other data sets are likely due to a combination of inaccurate allele calls in the divergent reference sequences and the absence of chromosome-level contiguity information in the scaffold-level ʻakiapōlāʻau genome.

Overall, we found strong evidence of reference bias impacting estimates of heterozygosity and divergence and confounding identification of runs of homozygosity (Supplementary Figures 1 and 2). As expected, alignment against the ʻamakihi genome produced heterozygosity and divergence estimates that were more similar to those generated using the ʻakiapōlāʻau as reference than either were to those using the more divergent zebra finch.

Genomic Substitution Rates

Under the first model (μ_C), the ʻakiapōlāʻau rate was 0.9–2.1 × 10^{-9} substitutions/site/year, and the ʻamakihi μ_C was 2.1–4.8 × 10^{-9} substitutions/

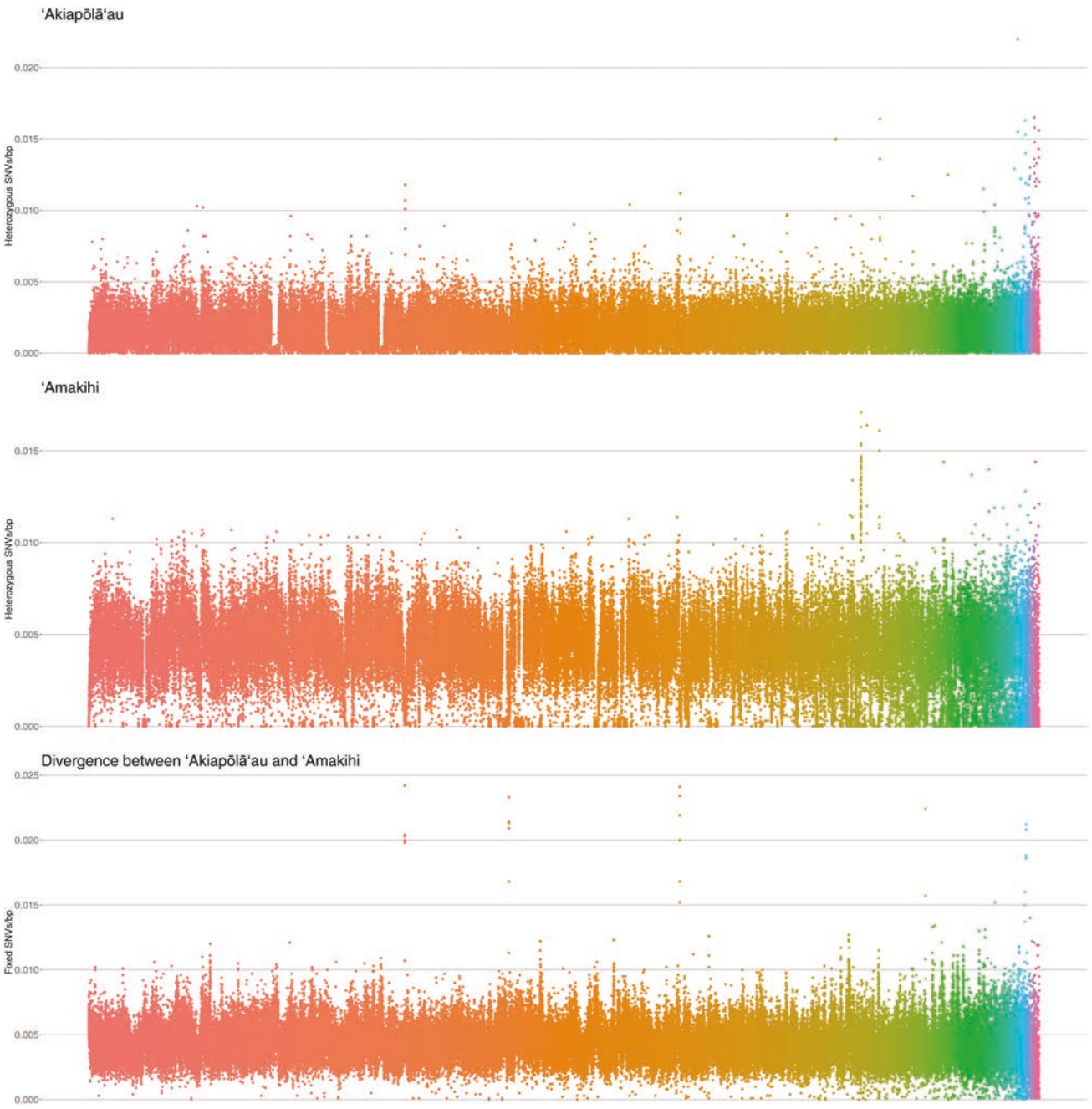

Figure 2. Genomic heterozygosities and SNV densities of the 'akiapōlā'au and 'amakihi calculated in 100 kbp windows. Colors correspond to individual scaffolds. Scaffolds are sorted by decreasing length with longer ones to the left. The de novo 'akiapōlā'au genome was used as the reference sequence.

site/year. The increased rate in the 'amakihi likely reflects its higher effective population size and/or decreased rates of purifying selection due to being an ecological generalist. We estimated the mean divergence substitution rate (μ_D) as 1.7×10^{-9} substitutions/site/year (divergence between 'akiapōlā'au and 'amakihi) to 2.6×10^{-9} substitutions/site/year (divergence from zebra finch). The increased substitution rate from the zebra finch-based estimate compared to the within-honeycreeper estimates could represent a faster evolutionary rate along the Estrildid lineage artificially inflating the estimated honeycreeper rate. An increased evolutionary rate along the Estrildid lineage is likely since the zebra finch population is large (IUCN 2019), permitting the evolution and maintenance of an increased number of mutations.

Genomic Population History Analysis

The PSMC reconstructions showed markedly divergent population histories for the 2 species (Figure 3). The 'amakihi shows an initial decrease in N_e ending 2–5 Mya in the 'akiapōlā'au-aligned and 1–3 Mya in the zebra-finch-aligned data sets, respectively (Figure 3; Supplementary Figures 3 and 4). We did not detect this decline in the 'amakihi aligned data set. After a period of stability, the 'amakihi

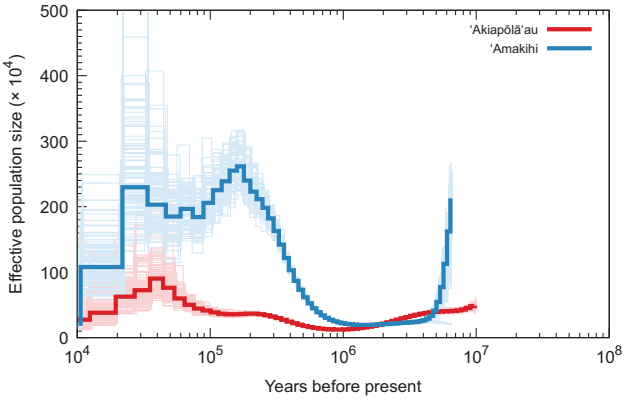

Figure 3. Pairwise sequential Markovian coalescent reconstruction of ʻakiapōlāʻau and ʻamakihi demographic histories. Demographic reconstructions used substitution rates assuming colonization of Hawai'i island 1.0 Mya. The ʻakiapōlāʻau and ʻamakihi generation times were 2 years and 1 year, respectively. Initial results are plotted using darker lines, with bootstrap replicates in lighter hues.

lineage underwent a rapid growth phase starting 0.3–1 Mya in the ʻakiapōlāʻau-aligned and ʻamakihi-aligned data sets and 0.2–0.8 Mya in the zebra-finch-aligned data set. In the ʻakiapōlāʻau-aligned and ʻamakihi-aligned data sets, this was followed by a secondary population decline starting 70–200 thousand years ago (kya) and ending between 40–90 kya (ʻakiapōlāʻau-aligned data set) or 20–70 kya (ʻamakihi-aligned data set). In both reconstructions, this was followed by a period of population stability or slight expansion until 10–30 kya. Conversely, in the zebra-finch-aligned reconstructions, instead of a decline after the expansion, there is a ~50 thousand-year period of population stability. The stable period is followed by a secondary recent expansion (Supplementary Figure 3). Depending on the substitution rate, we detected a secondary rapid recent decline in the ʻakiapōlāʻau-aligned and the zebra-finch-aligned ʻamakihi data sets. As the confidence interval around this secondary decline is quite large, this is likely an artifact due to there being insufficient heterozygous SNVs to infer more recent demographic history accurately.

The ʻakiapōlāʻau N_e remained relatively stable, but smaller than the ʻamakihi's, throughout the population history since after the ʻamakihi population expansion. The ʻakiapōlāʻau underwent slow population growth after ~1 Mya, with a recent population decline starting 15–50 kya. Additionally, using the substitution rate derived from the emergence of Hawai'i at 1 Mya and the zebra finch reference, the ʻakiapōlāʻau undergoes an apparent dramatic expansion within the last ~20 000 years (Supplementary Figure 3). This is likely an artifact due to insufficient variable sites at recent time scales. The confidence interval includes population stasis to expansion. The appearance of this expansion is dependent on the substitution rate and does not appear with higher rates (data not shown).

While the reconstructions using the ʻakiapōlāʻau and ʻamakihi references were similar to each other, we observed significant impact on inferred demographic history using the zebra finch reference. ʻAmakihi N_e estimates and degrees of population size changes were markedly smaller in the zebra-finch-aligned reconstructions (Supplementary Figure 3). The decreased ʻamakihi population sizes and apparent population stasis period are likely derived from incorrect variant calls in the zebra-finch-aligned reconstructions since they are not concordant with the ʻakiapōlāʻau-aligned or ʻamakihi-aligned results. Moreover, N_e estimates for both species and timing of the population size changes were determined by the substitution rate scaling (Supplementary Figure 4). Exclusion of the artificial "chromosomes" Un and Un2 had negligible impact on demographic reconstructions using the ʻamakihi reference (Supplementary Figure 5). In the zebra-finch-aligned data set, exclusion of "chromosome" Un had minimal impact on the ʻakiapōlāʻau reconstruction, but reduced inferred N_e in the ʻamakihi (Supplementary Figure 6). The ʻamakihi's demographic history topology however remained similar.

Genome Annotation and Candidate Gene Identification

AUGUSTUS predicted 34 450 genes in the ʻakiapōlāʻau genome, 33 083 genes in the Callicrate ʻamakihi assembly, and 34 620 genes in the realigned ʻamakihi sequence. Proteinortho identified 27 774 orthologous clusters between the 3 assemblies. Thirteen predicted ʻakiapōlāʻau and 10 predicted ʻamakihi genes overlapped the highly divergent genomic regions. One gene classified as "hydrocephalus-inducing-like" with immunoglobulin-like folds was found in both genome sequences on ʻakiapōlāʻau scaffold 788. A cadherin-like gene (scaffold 119) and an *ACE1-Sec16*-like gene (scaffold 62) were identified in the ʻakiapōlāʻau. We found 5 genes deriving from retroviruses: 3 in the ʻamakihi on scaffold 119 (an avian retrovirus envelope protein, an integrase, and an RNase H), another RNase H in the ʻamakihi on scaffold 548, and a retroviral aspartyl protease in the ʻakiapōlāʻau on scaffold 548. The remaining predicted genes had no identified functions.

The divergent regions from the ʻamakihi-aligned data set overlapped 10 protein coding genes (*ENSTGUG00000000646*, *ENSTGUG00000001862*, *ENSTGUG00000002283*, *ENSTGUG00000009266*, *FOXB1*, *KIF6*, *MAML3*, *PHF20*, *RBP1*, and *TIMM17A*) and an RNA gene (*ENSTGUG00000018063*). Only *ENSTGUG00000009266*, *FOXB1*, and *ENSTGUG00000018063* were found using the zebra-finch-aligned data set.

Bill Morphology Genes

All candidate bill morphology genes were discovered on single ʻakiapōlāʻau scaffolds, supporting the accuracy of the ʻakiapōlāʻau genome assembly. These genes were also identified in the original Callicrate ʻamakihi assembly to be located on the expected chromosomes based on assumed conserved synteny with the zebra finch. This indicates that we recovered the correct orthologs for these proteins.

We recovered the coding sequence of *ALX1* (4 exons), *BMP4* (2 exons), and *CALML3* (1 exon) in their entirety. The 3 genes were located on ʻakiapōlāʻau scaffolds 83, 436, and 751 (Callicrate chromosomes 1A, 5, and 1A), respectively. The ʻakiapōlāʻau and ʻamakihi amino acid sequences were identical for all 3 genes. The honeycreeper *ALX1* differed from the zebra finch peptide by 3 amino acid replacements (p.L103F, p.I221V, and p.T266S). The honeycreeper *BMP4* differed from the zebra finch's by 2 amino acid replacements (p.H23Y and p.T213S). *CALML3* differed by one replacement (p.K31R).

CALM1 and *CALM2* (both calmodulins) produce identical amino acid sequences. We identified both copies of this sequence on ʻakiapōlāʻau scaffolds 54 and 192. The scaffold 54 sequence corresponded to Callicrate ʻamakihi chromosome 3, while the scaffold 192 sequence matched Callicrate ʻamakihi chromosome 5. This indicates that the scaffold 54 copy is *CALM2* and the scaffold 192 sequence is *CALM1* assuming that the original ʻamakihi assembly was correctly oriented to the zebra finch in this region. Nevertheless, both copies covered amino acids 11 to 139 of the calmodulin sequences (from an

expected length of 148 amino acids across 6 exons). No amino acid replacements were observed in either the ʻakiapōlāʻau or ʻamakihi.

While the initial TBLASTN search only identified exon 2 of *HMGA2* (from a total of 2 exons) on ʻakiapōlāʻau scaffold 24 (Callicrate ʻamakihi chromosome 1A), the subsequent megaBLAST search found exon 1 on the same scaffold as exon 2. The honeycreeper peptide sequences were identical and shared a single replacement (p.A33T) from the zebra finch peptide.

For *DLK1*, we recovered all 4 exons in their entirety for the ʻakiapōlāʻau and the realigned ʻamakihi (ʻakiapōlāʻau scaffold 27/Callicrate ʻamakihi chromosome 5). Exon 1 was missing from the original ʻamakihi genome assembly. Phylogenetic analysis clustered the honeycreeper *DLK1* sequences with canary (*Serinus canaria*: GenBank accession XP_009099991.2; Supplementary Figure 7; del Hoyo et al. 2019). The ʻakiapōlāʻau, ʻamakihi and canary shared 5 amino acid replacements from the zebra finch (p.N3S, p.Q106K, p.Y176H, p.T188P, and p.K223R). The ʻamakihi had an asparagine at amino acid 76, while the ʻakiapōlāʻau, canary, and zebra finch shared a threonine residue at this position. The ʻamakihi asparagine residue was unique across all investigated avian species. At amino acid 220, the ʻakiapōlāʻau and canary had a valine residue, while the zebra finch had a methionine. Amino acid 220 was not resolved in the realigned ʻamakihi, but was an alanine in the original assembly. Inspection of the ʻamakihi realignment revealed a C/T SNV at amino acid 220, indicating that both alanine and valine residues were present at this site.

Discussion

Genomic Diversity and Demography

Our heterozygosity estimates and demographic reconstructions supported the expected genomic patterns for the 2 Hawaiʻian honeycreeper species. The ʻakiapōlāʻau is less heterozygous and has a smaller effective population size than the ʻamakihi, consistent with an ecological specialist restricted to a small endemic range with low census population numbers. The ʻamakihi is more heterozygous and has a much greater N_e, consistent with its comparatively larger census population and wider ecological niche. As previously reported by Callicrate et al. (2014), the ʻamakihi's runs of homozygosity on chromosomes 1 and 6 are suggestive of a selective sweep on the same chromosomes, possibly due to the recent evolution of tolerance for avian malaria in lowland ʻamakihi. Encouragingly for its long-term survival, the ʻakiapōlāʻau's individual heterozygosity (~0.0017 heterozygous sites/bp) is higher than the range previously observed for threatened bird species (0.0001–0.00091 heterozygous sites/bp: Li et al. 2014; Cortes-Rodriguez et al. 2019) such as New Zealand hihi (*Notiomystis cincta*; 0.00069 heterozygous sites/bp: de Villemereuil et al. 2019), Hawaiʻian ʻalala (*Corvus hawaiiensis*; 0.0004 sites/bp: Sutton et al. 2018), and Marianas åga (*Corvus kubaryi*; 0.0001 sites/bp: Cortes-Rodriguez et al. 2019). Furthermore, the absence of large runs of homozygosity in the ʻakiapōlāʻau suggests that it is not recently inbred (e.g., Dobrynin et al. 2016). Additional genome sequences are needed to determine whether inbreeding is occurring in other currently unsampled individuals.

Despite the current range of substitution rate estimates, our PSMC reconstructions permit us to tentatively identify genomic events likely related to the honeycreeper radiation. While we are unable to link the initial population decline in the ʻamakihi to a known historical event due to its imprecise dating (ending 1–5 Mya), it may represent a bottleneck caused by a colonization or speciation event before its ancestors' colonization of Hawaiʻi island. We posit that that the ʻamakihi rapid population growth period corresponds to the colonization of Hawaiʻi and subsequent expansion into new habitat. Scaling the reconstruction using most estimated rates dates the onset of the expansion to 0.5–1 Mya (Figure 3; Supplementary Figure 3), correlating well with more recent estimates of the first subaerial emergence of the island (Lipman and Calvert 2013). The latest estimated dates for the start of the growth phase is ~200–300 kya (depending on reference sequence). However, these dates derive from the substitution rates estimated with the emergence of the youngest island of Hawaiʻi at 0.43 Mya, which is probably an underestimate of its age, and the flycatcher germ-line mutation rate, which may be inappropriate for honeycreepers. Nevertheless, in all cases, the ʻamakihi expansion followed shortly after the date used to anchor the emergence of Hawaiʻi, strongly supporting the argument that this expansion represents a colonization event. The slow population growth in the ʻakiapōlāʻau after ~1 Mya may also represent its colonization of Hawaiʻ and/or population expansion due to bill morphological adaptation permitting better exploitation of its dietary niche compared to a presumed nukupuʻu-like ancestor.

The recent population declines in both species are more difficult to explain. One possibility is that it may represent the recent 19th and 20th century bottleneck in the honeycreeper populations due to the introduction of *Avipoxvirus* and avian malaria to the Hawaiʻian islands (Warner 1968; van Riper et al. 1986; Samuel et al. 2011, 2015). The sequenced ʻamakihi was a low elevation bird (Callicrate et al. 2014; Cassin-Sackett et al. 2019), so the bottleneck due to the introduction of avian malaria and the strong selection for tolerance to the parasite is probably reflected in its genome (Foster et al. 2007; Atkinson et al. 2013). PSMC often incorrectly infers gradual declines instead of bottlenecks and struggles to accurately infer recent demographic events due to limited variation and recombination events (Li and Durbin 2011). Alternatively, these may represent true population declines in the past. For instance, population declines could derive from restriction of suitable habitat caused by climate change (e.g., increased global temperatures since the Last Glacial Maximum). Further research using specimens that predate the recent bottleneck will clarify the timing of this population decline. Similarly, analysis of Hawaiʻi ʻamakihi individuals from populations presumably less affected by malaria and the bottleneck (e.g., from high elevations) could help determine the population specificity and timing of the observed decline.

Honeycreeper Adaptive Radiation Genomics

Our genomic scan identified small strongly divergent regions that distinguished the ʻakiapōlāʻau and ʻamakihi genomes, suggesting that the honeycreeper radiation is driven by diversifying selection on a small number of genes and pathways rather than uniform selection across the genome (Chaves et al. 2016; Wolf and Ellegren 2017). This pattern has been observed in birds previously: for instance, only 6 genomic regions differed between golden-winged and blue-winged warblers, of which 4 were linked to plumage genes (Toews et al. 2016). While the honeycreeper genomic islands of differentiation may contain adaptation or speciation genes, further analysis is required to determine whether the identified genes in these regions are truly involved in honeycreeper speciation or are the result of other processes (e.g., linked selection or population divergence; Wolf and Ellegren 2017). Nevertheless, the identified genes in these regions are candidates to explain honeycreeper evolution. Intriguingly, *FOXB1* expression regulates numerous developmental processes, notably

neurological development and visual learning (UniProt Consortium 2019). Evolution in *FOXB1* expression could contribute to behavior differences between the 2 species. Furthermore, in humans, genic regulation of brain growth and development is involved with the development of congenital hydrocephalus (a disorder characterized by abnormal accumulation of cerebral spinal fluid within the cerebral ventricles; Kahle et al. 2016). Therefore, similar to *FOXB1*, the identified "hydrocephalus-inducing-like" gene suggests genic regulation of neurological development in honeycreeper evolution. The predicted ʻakiapōlāʻau cadherin-like gene suggests a role in the calcium-binding pathway in honeycreeper evolution. The calmodulin pathway (which also involves calcium binding) was previously shown to influence bill length in Darwin's finches (Abzhanov et al. 2006). It is therefore plausible that this gene is involved in the ʻakiapōlāʻau's unique bill morphology. While the identified avian retroviruses may also play a role in honeycreeper evolution, endogenous retrovirus expression and regulation remains poorly understood (Hu et al. 2016).

The remaining named identified genes are involved in a variety of cellular processes: *KIF6* is involved in ATP binding and microtubule movement, *MAML3* is part of the Notch signaling pathway and regulates transcription, *PHF20* contributes to histone acetylation, DNA binding, and transcription regulation, *RBP1* binds retinol, and *TIMM17A* regulates transmembrane protein transport (UniProt Consortium 2019). The ʻakiapōlāʻau *ACE1-Sec16*-like gene is too poorly identified to reliably infer function, but is likely involved in protein transport or binding (UniProt Consortium 2019). Due to the low-level nature of these genes, selection on them likely contributes to a range of phenotypes, which cannot be disentangled without further controlled experimental study. The zebra-finch-predicted (*ENSTGUG00000000646*, *ENSTGUG00000001862*, *ENSTGUG00000018063*, *ENSTGUG00000002283*, *ENSTGUG00000009266*) and remaining ʻakiapōlāʻau/ʻamakihi-predicted genes' functions are currently unknown, but could be critical to honeycreeper evolution.

With the exception of *DLK1*, we found no evidence of between-honeycreeper differences in the candidate bill morphology genes derived from the literature, suggesting adaptation in different genes or in regulatory regions (Wittkopp and Kalay 2012). Gene regulation, rather than protein sequence evolution, can be a primary source of phenotypic variation leading to speciation (Mack and Nachman 2017). Gene regulation evolution could explain the concurrent rapid phenotypic and limited genotypic evolution within the honeycreeper clade (Lerner et al. 2011). Since we observed variation in *DLK1* coding sequence between honeycreeper species, the gene is a candidate for further analysis to explain honeycreeper bill morphology. *DLK1* plays roles in calcium ion binding, cell differentiation, and negative regulation of the Notch signaling pathway (UniProt Consortium 2019). *DLK1* has previously been associated with variation in bill size in Darwin's finches (Lamichhaney et al. 2015; Chaves et al. 2016). If honeycreeper bill morphologies evolved using the same gene, this would suggest that bill morphology evolution is constrained to a limited set of genic pathways within passerines. Interestingly, we observed the unique *DLK1* mutation in the generalist ʻamakihi, rather than the specialist ʻakiapōlāʻau. This is contrary to the expectation that the more derived morphology would correspond to the derived gene sequence. However, while not as morphologically unique, the ʻamakihi has also evolved a different bill phenotype from that of its presumed granivorous finch-like ancestor (Lerner et al. 2011). While this also indicates that the ʻakiapōlāʻau's unique bill morphology does not derive from protein sequence evolution within *DLK1*, it does not rule out variation in *DLK1* expression as playing a role.

An obvious next step to better understand the evolution of the ʻakiapōlāʻau's unique bill morphology would be to search the nukupuʻus' genomes for signatures of selection that could explain this phenomenon. Due to the nukupuʻus' apparent extinction, such an analysis will require the use of museum specimens and ancient DNA techniques. In addition, developmental transcriptomic analyses of Hawaiian honeycreepers that are not endangered with differing bill morphologies (as has been done for Darwin's finches: Abzhanov et al. 2004, 2006) may enable discovery of other candidate genes involved in the differentiation of bill morphology in this extensive adaptive radiation.

Conclusions

Our de novo ʻakiapōlāʻau genome assembly provides a critical resource for genomic investigation of the endangered honeycreeper radiation. Comparison of the ʻakiapōlāʻau and ʻamakihi genomes revealed the expected pattern of reduced heterozygosity and effective population size in the ʻakiapōlāʻau. The age of the island on which both species occur provides, with assumptions, a means to assess comparative substitution rates within each species. Speciation within the drepanids appears to be driven by selection on a small number of islands of divergence, rather than uniform diversification across the genome. Most notably, the gene *FOXB1* was discovered within one of these genomic islands of divergence, suggesting a role in neurological or behavioral adaptation between the 2 species. The amino acid changes in only one bill morphology associate gene, *DLK1*, indicates that it could be a strong candidate for controlling bill morphology differences in honeycreepers.

Supplementary Material

Supplementary data are available at *Journal of Heredity* online.

Data Availability

The ʻakiapōlāʻau genome sequence and raw data are available in GenBank (accession SGIP00000000) and as NCBI BioSample SAMN10867508 under BioProject PRJNA520783.

Funding

This work was supported by the Smithsonian Institution; and the National Science Foundation (grant numbers 1547168, 1717498).

Acknowledgments

We thank Jack Jeffrey for providing the honeycreeper photographs. This research used the Smithsonian Institution High Performance Cluster (SI/HPC).

References

Abzhanov A, Kuo WP, Hartmann C, Grant BR, Grant PR, Tabin CJ. 2006. The calmodulin pathway and evolution of elongated beak morphology in Darwin's finches. *Nature*. 442:563–567.

Abzhanov A, Protas M, Grant BR, Grant PR, Tabin CJ. 2004. *Bmp4* and morphological variation of beaks in Darwin's finches. *Science*. 305:1462–1465.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 25:3389–3402.

Atkinson CT, Saili KS, Utzurrum RB, Jarvi SI. 2013. Experimental evidence for evolved tolerance to avian malaria in a wild population of low elevation Hawai'i 'amakihi (*Hemignathus virens*). *Ecohealth*. 10:366–375.

Bao W, Kojima KK, Kohany O. 2015. Repbase update, a database of repetitive elements in eukaryotic genomes. *Mob DNA*. 6:11.

Beadell JS, Ishtiaq F, Covas R, Melo M, Warren BH, Atkinson CT, Bensch S, Graves GR, Jhala YV, Peirce MA, *et al.* 2006. Global phylogeographic limits of Hawaii's avian malaria. *P R Soc B*. 273: 2935–2944.

Berner D, Salzburger W. 2015. The genomics of organismal diversification illuminated by adaptive radiations. *Trends Genet*. 31:491–499.

Bradnam KR, Fass JN, Alexandrov A, Baranay P, Bechner M, Birol I, Boisvert S, Chapman JA, Chapuis G, Chikhi R, *et al.* 2013. Assemblathon 2: evaluating *de novo* methods of genome assembly in three vertebrate species. *Gigascience*. 2:10.

Brawand D, Wagner CE, Li YI, Malinsky M, Keller I, Fan S, Simakov O, Ng AY, Lim ZW, Bezault E, *et al.* 2014. The genomic substrate for adaptive radiation in African cichlid fish. *Nature*. 513:375–381.

Broad Institute. 2015. Picard version 1.141. [accessed 2015 December 16]. Available from: https://broadinstitute.github.io/picard

Callicrate T, Dikow R, Thomas JW, Mullikin JC, Jarvis ED, Fleischer RC; NISC Comparative Sequencing Program. 2014. Genomic resources for the endangered Hawaiian honeycreepers. *BMC Genomics*. 15:1098.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2008. BLAST+: architecture and applications. *BMC Bioinformatics*. 10:421.

Carson HL, Clague DA. 1995. Geology and biogeography of the Hawaiian Islands. In: Wagner W, Funk V, editors. *Hotspot archipelago*. Washington (DC): Smithsonian Institution Press. p. 14–29.

Cassin-Sackett L, Callicrate TE, Fleischer RC. 2019. Parallel evolution of gene classes, but not genes: evidence from Hawai'ian honeycreeper populations exposed to avian malaria. *Mol Ecol*. 28:568–583.

Chaves JA, Cooper EA, Hendry AP, Podos J, De León LF, Raeymaekers JA, MacMillan WO, Uy JA. 2016. Genomic variation at the tips of the adaptive radiation of Darwin's finches. *Mol Ecol*. 25:5282–5295.

Cortes-Rodriguez N, Campana MG, Berry L, Faegre S, Derrickson SR, Ha RR, Dikow RB, Rutz C, Fleischer RC. 2019. Population genomics and structure of the critically endangered Mariana Crow (*Corvus kubaryi*). *Genes*. 10:187.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, *et al.*; 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. *Bioinformatics*. 27:2156–2158.

del Hoyo J, Elliott A, Sargatal J, Christie DA, Kirwan G, editors. 2019. *Handbook of the birds of the world alive*. Barcelona (Spain): Lynx Editions. [accessed 2019 July 11]. Available from: http://www.hbw.com

de Villemereuil P, Rutschmann A, Lee KD, Ewen JG, Brekke P, Santure AW. 2019. Little adaptive potential in a threatened passerine bird. *Curr Biol*. 29:889–894.e3.

Dobrynin P, Liu S, Tamazian G, Xiong Z, Yurchenko AA, Krasheninnikova K, Kliver S, Schmidt-Küntzel A, Koepfli KP, Johnson W, *et al.* 2016. Genomic legacy of the African cheetah, *Acinonyx jubatus*. *Genome Biol*. 16:277.

Eggert LS, Beadell JS, McClung A, McIntosh CE, Fleischer RC. 2009. Evolution of microsatellite loci in the adaptive radiation of Hawaiian honeycreepers. *J Hered*. 100:137–147.

Elphick CS, Roberts DL, Reed JM. 2010. Estimated dates of recent extinctions for North American and Hawaiian birds. *Biol Conserv*. 143:617–624.

Fleischer RC. 2009. Honeycreepers, Hawaiian. In: Gillespie R, Clague DA, editors. *Encylopedia of islands*. Berkeley (CA): University of California Press. p. 410-414

Fleischer RC, McIntosh CE, Tarr CL. 1998. Evolution on a volcanic conveyor belt: using phylogeographic reconstructions and K-Ar-based ages of the Hawaiian Islands to estimate molecular evolutionary rates. *Mol Ecol*. 7:533–545.

Fleischer RC, Tarr CL, James HF, Slikas B, McIntosh CE. 2001. Phylogenetic placement of the po'ouli *Melamprosops phaeosoma* based on mitochondrial DNA sequence and osteological characters. *Stud Avian Biol*. 22:98–103.

Foster JT, Woodworth BL, Eggert LE, Hart PJ, Palmer D, Duffy DC, Fleischer RC. 2007. Genetic structure and evolved malaria resistance in Hawaiian honeycreepers. *Mol Ecol*. 16:4738–4746.

Futuyma DJ. 1986. *Evolutionary biology*. Sunderland (MA): Sinauer.

Gorresen PM, Camp RJ, Pratt TK. 2007. *Forest bird distribution, density and trends in the Ka'ū region of Hawai'i Island. U.S. Geological Survey Open-File report 2007-1076*. U.S. Geological Survey. Available from: https://pubs.usgs.gov/of/2007/1076

Henning F, Meyer A. 2014. The evolutionary genomics of cichlid fishes: explosive speciation and adaptation in the postgenomic era. *Annu Rev Genomics Hum Genet*. 15:417–441.

Ho SY, Phillips MJ, Cooper A, Drummond AJ. 2005. Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Mol Biol Evol*. 22:1561–1568.

Ho SY, Shapiro B, Phillips MJ, Cooper A, Drummond AJ. 2007. Evidence for time dependency of molecular rate estimates. *Syst Biol*. 56:515–522.

Hu X, Zhu W, Chen S, Liu Y, Sun Z, Geng T, Wang X, Gao B, Song C, Qin A, *et al.* 2016. Expression of the *env* gene from the avian endogenous retrovirus ALVE and regulation by miR-155. *Arch Virol*. 161:1623–1632.

IUCN. 2019. The IUCN Red List of Threatened Species version 2019-1. [accessed 2019 July 10]. Available from: https://www.iucnredlist.org

James HF. 2004. The osteology and phylogeny of the Hawaiian finch radiation (Fringillidae: Drepanidini), including extinct taxa. *Zool J Linn Soc*. 141:207–255.

James HF, Olson SL. 1991. *Description of thirty-two new species of birds from the Hawaiian Islands: part II. Passeriformes*. Ornithological Monographs 46. Washington (DC): American Ornithologists' Union.

Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, Ho SY, Faircloth BC, Nabholz B, Howard JT, *et al.* 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*. 346:1320–1331.

Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, *et al.* 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics*. 30:1236–1240.

Joron M, Frezal L, Jones RT, Chamberlain NL, Lee SF, Haag CR, Whibley A, Becuwe M, Baxter SW, Ferguson L, *et al.* 2011. Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature*. 477:203–206.

Kahle KT, Kulkarni AV, Limbrick DD Jr, Warf BC. 2016. Hydrocephalus in children. *Lancet*. 387:788–799.

Kent WJ. 2002. BLAT–the BLAST-like alignment tool. *Genome Res*. 12:656–664.

Knowlton JL, Flaspohler DJ, Rotzel Mcinerney NC, Fleischer RC. 2014. First record of hybridization in the Hawaiian honeycreepers: 'i'iwi (*Vestiaria coccinea*) × 'apapane (*Himatione sanguinea*). *Wilson J Ornithol*. 126:562–568.

Kozak K, McMillan WO, Joron M, Jiggins CD. 2018. Genome-wide admixture is common across the *Heliconius* radiation. *bioRxiv*. doi:10.1101/414201

Lamichhaney S, Berglund J, Almén MS, Maqbool K, Grabherr M, Martinez-Barrio A, Promerová M, Rubin CJ, Wang C, Zamani N, *et al.* 2015. Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*. 518:371–375.

Lamichhaney S, Han F, Berglund J, Wang C, Almén MS, Webster MT, Grant BR, Grant PR, Andersson L. 2016. A beak size locus in Darwin's finches facilitated character displacement during a drought. *Science*. 352:470–474.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, *et al.* 2007. Clustal W and Clustal X version 2.0. *Bioinformatics*. 23:2947–2948.

Lechner M, Findeiss S, Steiner L, Marz M, Stadler PF, Prohaska SJ. 2011. Proteinortho: detection of (co-)orthologs in large-scale analysis. *BMC Bioinformatics*. 12:124.

Lerner HR, Meyer M, James HF, Hofreiter M, Fleischer RC. 2011. Multilocus resolution of phylogeny and timescale in the extant adaptive radiation of Hawaiian honeycreepers. *Curr Biol*. 21:1838–1844.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*. 1303.3997.

Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. *Nature*. 475:493–496.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics*. 25:2078–2079.

Li S, Li B, Cheng C, Xiong Z, Liu Q, Lai J, Carey HV, Zhang Q, Zheng H, Wei S, *et al.* 2014. Genomic signatures of near-extinction and rebirth of the crested ibis and other endangered bird species. *Genome Biol*. 15:557.

Lipman PW, Calvert AT. 2013. Modeling volcano growth on the Island of Hawaii: deep-water perspectives. *Geosphere* 9:1348–1383.

Mack KL, Nachman MW. 2017. Gene regulation and speciation. *Trends Genet*. 33:68–80.

Malinsky M, Svardal H, Tyers AM, Miska EA, Genner MJ, Turner GF, Durbin R. 2018. Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nat Ecol Evol*. 2:1940–1955.

Marco-Sola S, Sammeth M, Guigó R, Ribeca P. 2012. The GEM mapper: fast, accurate and versatile alignment by filtration. *Nat Methods*. 9: 1185–1188.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J*. 17:10-12.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, *et al.* 2010. The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 20:1297–1303.

Moyle RG, Oliveros CH, Andersen MJ, Hosner PA, Benz BW, Manthey JD, Travers SL, Brown RM, Faircloth BC. 2016. Tectonic collision and uplift of Wallacea triggered the global songbird radiation. *Nat Commun*. 7:12709.

Pratt HD. 2005. *The Hawaiian honeycreepers: Drepanidinae*. Oxford (UK): Oxford University Press.

Ralph CJ, Fancy SG. 1996. Aspects of the life history and foraging ecology of the endangered akiapolaau. *Condor*. 98:312–321.

Reed RD, Papa R, Martin A, Hines HM, Counterman BA, Pardo-Diaz C, Jiggins CD, Chamberlain NL, Kronforst MR, Chen R, *et al.* 2011. *optix* drives the repeated convergent evolution of butterfly wing pattern mimicry. *Science*. 333:1137–1141.

Roberts DL, Elphick CS, Reed JM. 2009. Identifying anomalous reports of putatively extinct species and why it matters. *Conserv Biol*. 24:189–196.

Salzburger W. 2018. Understanding explosive diversification through cichlid fish genomics. *Nat Rev Genet*. 19:705–717.

Samuel MD, Hobbelen PHF, DeCastro F, Ahumada JA, LaPointe DA, Atkinson CT, Woodworth BL, Hart PJ, Duffy DC. 2011. The dynamics, transmission, and population impacts of avian malaria in native Hawaiian birds: a modeling approach. *Ecol Appl*. 21:2960–2973.

Samuel MD, Woodworth BL, Atkinson CT, Hart PJ, LaPointe DA. 2015. Avian malaria in Hawaiian forest birds: infection and population impacts across species and elevations. *Ecosphere*. 6:1–21.

Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I. 2009. ABySS: a parallel assembler for short read sequence data. *Genome Res*. 19:1117–1123.

Slater GS, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*. 6:31.

Smeds L, Qvarnström A, Ellegren H. 2016. Direct estimate of the rate of germline mutation in a bird. *Genome Res*. 26:1211–1218.

Smeds L, Warmuth V, Bolivar P, Uebbing S, Burri R, Suh A, Nater A, Bureš S, Garamszegi LZ, Hogner S, *et al.* 2015. Evolutionary analysis of the female-specific avian W chromosome. *Nat Commun*. 6:7330.

Smit AFA, Hubley R, Green P. 2013–2015. RepeatMasker open-4.0. [accessed 2017 October 13]. Available from: http://www.repeatmasker.org

Song L, Florea L, Langmead B. 2014. Lighter: fast and memory-efficient sequencing error correction without counting. *Genome Biol*. 15:509.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 30:1312–1313.

Stanke M, Diekhans M, Baertsch R, Haussler D. 2008. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics*. 24:637–644.

Stanke M, Schöffmann O, Morgenstern B, Waack S. 2006a. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics*. 7:62.

Stanke M, Tzvetkova A, Morgenstern B. 2006b. AUGUSTUS at EGASP: using EST, protein and genomic alignments for improved gene prediction in the human genome. *Genome Biol*. 7(Suppl 1):S11.1–S11.8.

Supple MA, Hines HM, Dasmahapatra KK, Lewis JJ, Nielsen DM, Lavoie C, Ray DA, Salazar C, McMillan WO, Counterman BA. 2013. Genomic architecture of adaptive color pattern divergence and convergence in *Heliconius* butterflies. *Genome Res*. 23:1248–1257.

Sutton JT, Helmkampf M, Steiner CC, Bellinger MR, Korlach J, Hall R, Baybayan P, Muehling J, GU J, Kingan S, *et al.* 2018. A high-quality, long-read *de novo* genome assembly to aid conservation of Hawaii's last remaining crow species. *Genes*. 9:393.

Tarr CL, Fleischer RC. 1993. Mitochondrial-DNA variation and evolutionary relationships in the amakihi complex. *Auk*. 100:825–831.

Toews DP, Taylor SA, Vallender R, Brelsford A, Butcher BG, Messer PW, Lovette IJ. 2016. Plumage genes and little else distinguish the genomes of hybridizing warblers. *Curr Biol*. 26:2313–2318.

Tollis M, Hutchins ED, Stapley J, Rupp SM, Eckalbar WL, Maayan I, Lasku E, Infante CR, Dennis SR, Robertson JA, *et al.* 2018. Comparative genomics reveals accelerated evolution in conserved pathways during the diversification of anole lizards. *Genome Biol Evol*. 10:489–506.

Treangen TJ, Salzberg SL. 2012. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat Rev Genet*. 13:36–46.

UniProt Consortium. 2019. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res*. 47:D506–D515.

van Riper C III, van Riper SG, Goff ML, Laird, M. 1986. The epizootiology and ecological significance of malaria in Hawaiian (USA) land birds. *Ecol Monogr*. 56:327–344.

Warner RE. 1968. The role of introduced diseases in the extinction of the endemic Hawaiian avifauna. *Condor*. 7: 101–120.

Warren WC, Clayton DF, Ellegren H, Arnold AP, Hillier LW, Künstner A, Searle S, White S, Vilella AJ, Fairley S, *et al.* 2010. The genome of a songbird. *Nature*. 464:757–762.

Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM. 2017. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol Biol Evol*. 35:543–548.

Wittkopp PJ, Kalay G. 2012. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat Rev Genet*. 13:59–69.

Wolf JB, Ellegren H. 2017. Making sense of genomic islands of differentiation in light of speciation. *Nat Rev Genet*. 18:87–100.

Zuccon D, Prŷs-Jones R, Rasmussen PC, Ericson PG. 2012. The phylogenetic relationships and generic limits of finches (Fringillidae). *Mol Phylogenet Evol*. 62:581–596.