Phylogenomics of a 'great speciator' reveals Indonesian Archipelago as the center of lineage diversity

Chyi Yin Gwee[1]

[1]Affiliation not available

September 12, 2020

Chyi Yin Gwee, Kritika M. Garg, Balaji Chattopadhyay, Keren R. Sadanandan, Dewi M. Prawiradilaga, Martin Irestedt, Fumin Lei, Luke M. Bloch, Jessica G.H. Lee, Mohammad Irham, Tri Haryoko, Malcolm C.K. Soh, Kelvin S.-H. Peh, Karen M.C. Rowe, Teuku Reza Ferasyi, Shaoyuan Wu, Guinevere O.U. Wogan, Rauri C.K. Bowie, Frank E. Rheindt

INTRODUCTION

Archipelagoes are ideal settings for unravelling complex evolutionary patterns as they constitute natural laboratories for the study of factors contributing to speciation (MacArthur and Wilson 2001; Whittaker and Fernández-Palacios 2007; Lohman et al. 2011). Among vertebrate groups that occur across archipelagoes, island radiations of birds are most well-studied (Lerner et al. 2011; Lamichhaney et al. 2015). These avian models display a great deal of variability in their diversification rates across islands, which are fundamentally linked to species' capability to disperse over water (Diamond et al. 1976). In particular, the so-called 'great speciators', first characterized by Diamond et al. (1976), stand out from all other birds based on their paradoxical ability to disperse widely and colonize entire archipelagoes while, at the same time, diversifying into multiple daughter species in spite of a continuing potential for overwater gene flow (Cai et al. 2020).

One of the few classical examples of 'great speciators' identified by Diamond et al. (1976) is the songbird genus *Zosterops* , or white-eyes, which are excellent dispersers capable of differentiating rapidly from source populations (Clegg et al. 2002; Moyle et al. 2009). The genus *Zosterops* comprises c. 100 species that have radiated across the Old World and Oceania within the past 1 to 3.5 million years, reflecting perhaps the fastest diversification rate of any vertebrate clade (Warren et al. 2006; Moyle et al. 2009; Leroy et al. 2019; Cai et al. 2020). An overwhelming proportion (more than 70%) of these species occurs exclusively in archipelagoes distributed across the Atlantic, Indian and Pacific Oceans (Fig. 1). As such, the remarkable radiation of white-eyes serves as an ideal model system with which to explore island biogeography theory (Diamond et al. 1976; Moyle et al. 2009).

In order to achieve an understanding of the underlying processes driving the white-eye radiation, its phylogeny first needs to be resolved to provide a reliable backbone for hypothesis testing. In *Zosterops* , however, traditional methods that rely on morphological tools to infer how species are related to one another have proven to be unreliable, as plumage features of ecologically distinct and geographically disjunct *Zosterops* species are often indistinguishable (Mees 1957; Mayr 1965). Although a more recent application of genetic methods has helped disentangle the white-eye radiation to some extent, most studies have concentrated on Melanesian and Indian Ocean members of the genus (Slikas et al. 2000; Warren et al. 2006; Moyle et al. 2009; Cox et al. 2014; Linck et al. 2016; Wickramasinghe et al. 2017; Manthey et al. 2020). There continues to be a dearth of knowledge on this radiation across the core of its Asian distribution due to limited sampling and lack of genetic data. In particular, it is crucial to unravel the phylogenetic affinities of white-eyes distributed across the Indonesian archipelago, which – comprising more than 17,000 islands – is the largest archipelago

1

in the world and harbors about 20 endemic *Zosterops*species (Fig. 1), including two undescribed species that were discovered in the last two decades (Eaton et al. 2016; O'Connell et al. 2019). The high density of *Zosterops* species across the Indonesian archipelago hints at the possible importance of this region in white-eye evolution.

Apart from incomplete geographic sampling, the lack of resolution of the white-eye radiation has largely been a consequence of sparse genomic sampling: most phylogenetic studies of white-eyes have been restricted to one or a few genetic markers, resulting in trees that are plagued by unresolved polytomies, hampering useful evolutionary inference (Slikas et al. 2000; Warren et al. 2006; Moyle et al. 2009; Oatley et al. 2012; Nyári and Joseph 2013; Cox et al. 2014; Husemann et al. 2016; Linck et al. 2016; Round et al. 2017; Wickramasinghe et al. 2017; Shakya et al. 2018; Cai et al. 2019; Lim et al. 2019; O'Connell et al. 2019). Disentangling relationships within rapid and recent radiations, such as white-eyes, requires overcoming the challenges of heterogenous gene trees due to biological factors such as incomplete lineage sorting (Edwards et al. 2005; Song et al. 2012). The multispecies coalescent (MSC) model offers a promising avenue to overcoming gene tree discordance by allowing the evolutionary histories of each locus to be inferred independently (Song et al. 2012; Liu et al. 2015).

An equally debilitating source of bias in reconstructing the phylogeny of rapid and recent radiations such as white-eyes is the potential for member species to engage in secondary gene flow or genetic introgression long after a speciation event has occurred (Rheindt and Edwards 2011; Edwards et al. 2016). Such introgression will be reflected in the phylogenetic signal of a varying proportion of loci, thereby leading astray efforts to search for the true species tree. Multiple analytical approaches have been devised to account for secondary gene flow, such as through tree-based analysis as implemented in Phylogeographic Inference using Approximate Likelihoods (PHRAPL) (Jackson et al. 2017a; Jackson et al. 2017b), or through SNP-based analysis like the ABBA-BABA test, which detects an excess of shared derived alleles between populations (Green et al. 2010; Patterson et al. 2012).

In this study, we acquired historical and modern DNA samples from museums and recent material from our own fieldwork at remote locations to achieve a wide geographic representation of white-eyes (Table S1). We designed RNA probes using *Z. lateralis* (Cornetti et al. 2015) as a reference genome, targeting 832 loci at high coverage to overcome the limitations of missing data expected from degraded DNA of historical samples, thereby recovering a comparable set of loci across both historical and fresh samples (Templeton et al. 2013). Our target capture methodology addresses the recalcitrant persistence of unresolved polytomies in the phylogeny of this rapid radiation by making use of a large set of loci (Cai et al. 2019). We employed three different coalescent species tree methods to assess topological incongruence across tree-building approaches (Liu et al. 2009; Liu et al. 2010; Vachaspati and Warnow 2015). Recognizing that gene flow is commonly observed in recent radiations, we conducted PHRAPL (Jackson et al. 2017b) analysis and performed ABBA-BABA tests (Patterson et al. 2012) to assess introgression between closely related species with incongruent topologies and ultimately elucidate the likely evolutionary history of this complex radiation.
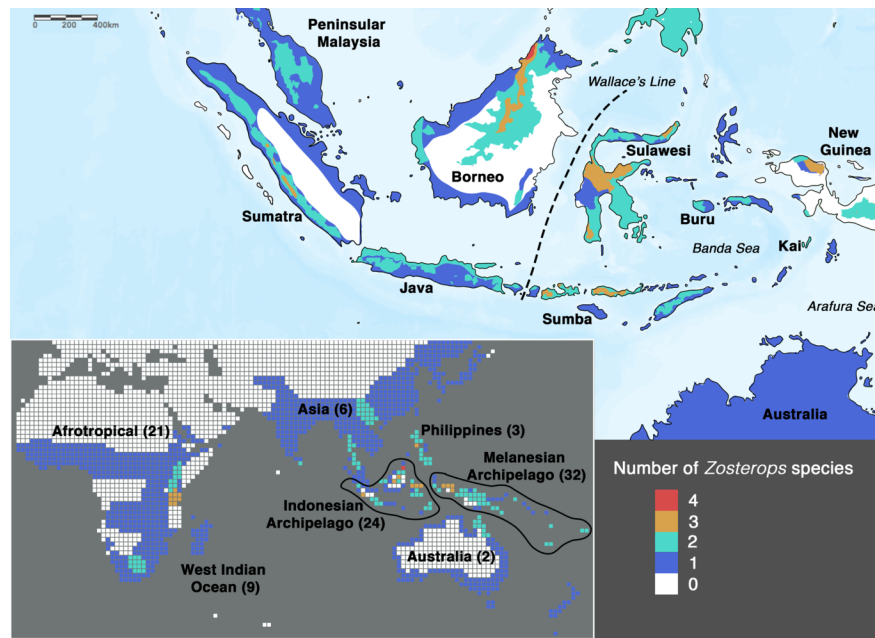
2

Figure 1. Species density map of the genus *Zosterops* across the Indonesian archipelago (main map), and across the entire distribution range (bottom left inset). Islands referred to in the text are specifically labelled on the main map. The total number of *Zosterops* species is shown in brackets beside each labelled region on the inset. The Micronesian archipelago, which constitutes several islands too small to be depicted, harbors eight *Zosterops* species.

MATERIALS and METHODS

Taxon Sampling

A total of 48 historical toepad samples and 52 fresh samples were acquired from various museums and through fieldwork conducted across peninsular Malaysia (research permit from the Economic Planning Unit of Malaysia: UPE 40/200/19/3295 to M.C.K.S. and K.S.-H.P.) and the Indonesian archipelago (research permit from the Ministry of Research & Technology of Indonesia: 10/TKPIPA/FRP/SM/X/2013 to F.E.R., and 314/SIP/FRP/E5/Dit.KI/X/2018 to C.Y.G.) (Table S1).

Probe Design for Target Capture

Target enrichment protocols have been shown to be highly effective at capturing historical DNA for phylogenomic studies (Chattopadhyay et al. 2019). We designed loci specifically targeting both conserved exons and variable intronic regions of the *Zosterops* genome (Chattopadhyay et al. 2019). We first used EvolMarkers (Li et al. 2012) to identify conserved single copy coding sequences in the genomes of *Z. lateralis* (accession no. GCA_001281735) (Cornetti et al. 2015), *Ficedula albicollis* (accession no. GCA_000247815.1) (Ellegren et al. 2012) and *Taeniopygia guttata* (accession no. GCF_003957565.1; released by the Vertebrate Genomes Project). To identify conserved exons, EvolMarkers performs a BLAST search (Altschul et al. 1990), for which we set a minimum of 55% identity and e-value of less than 10E-15. Only exons longer than 500bp were used for further downstream analysis. Then we isolated 500 bp upstream and downstream of these conserved exons from the *Z. lateralis* genome to include variable intronic regions using bedtools 2.28.0 (Quinlan and Hall 2010). We further checked for overlapping targets and merged all overlapping loci in bedtools, removing any loci with GC content less than 40% or more than 60%. Loci comprising repeat elements were identified using RepeatMasker 4.0.7 (Smit et al. 2013-2015) and removed. After filtering, our design retained 832 loci,

3

which were used by Arbor Biosciences (USA) to design a total of 63,244 RNA baits. Each locus was targeted with 4X tiling density of overlapping baits, each bait of 100 bp, for in-solution target enrichment.

Laboratory procedures

Fresh DNA was extracted following the manufacturer's protocol using the DNeasy Blood and Tissue Kit (Qiagen, Germany). The DNA of historical toepad samples was extracted under sterile conditions inside a dedicated ancient DNA facility, and extractions were performed inside a biosafety cabinet with laminar air-flow. The ancient DNA facility room was subject to at least 12 hours of UV light and thoroughly cleaned with bleach in between each session of historical DNA extractions. We used the same kit for extraction of historical DNA with slight modifications (Chattopadhyay et al. 2019). Extraction negatives were included to ensure absence of contamination. Double-stranded DNA concentrations were ascertained using a Qubit 2.0 high sensitivity DNA Assay kit (Invitrogen, USA), and fragment sizes were assessed using an AATI Fragment Analyzer (Agilent, USA). The negatives were also quantified using a Qubit 2.0 assay and AATI to ensure absence of DNA.

Fresh DNA was sheared into a targeted size of 250 bp using a Bioruptor Pico sonication device (Diagenode, Belgium) with 13 cycles of sonication prior to library preparation. Each cycle consisted of 30s of sonication followed by 30s of rest. We used NEBNext Ultra II DNA Library Prep Kits for Illumina (New England BioLabs, USA) and NEBNext 8 bp dual indexes (New England BioLabs, USA) for both fresh and historical library preparation. The libraries using fresh tissue were size selected for an insert size of 250 bp with AMPure XP (Beckman Coulter, USA) beads, giving an expected final library size of ˜370 bp with adapters and primers included. Size selection was omitted during library preparation of historical samples to reduce DNA loss. A total of 12 cycles of polymerase chain reaction (PCR) was applied and negative controls were carried out for both fresh and historical libraries.

The library preparation of historical DNA was conducted inside a dedicated PCR cabinet with laminar air-flow. The PCR cabinet was subject to 1 hour of UV light and thoroughly cleaned with bleach in between each batch of library preparations. We added NEBNext FFPE DNA repair mix (New England BioLabs, USA) to the historical DNA prior to library preparation to reduce deamination of cytosine to uracil, repair nicks, and fill in 5' overhangs of the damaged DNA. DNA quantification and assessments of libraries' fragment sizes were conducted as above. The peak fragment sizes of each library prepared with historical samples ranged between 200 bp and 300 bp, whereas the peak fragment sizes of each library prepared with fresh samples ranged between 330 bp and 420 bp. The negatives were also quantified by Qubit 2.0 and assessed by AATI to ensure that only adapters and primer-dimer DNA (single peak at ˜55 bp and ˜150 bp respectively) were present.

Target capture was performed on all historical and fresh samples using a MYbaits kit version 3 (Arbor Biosciences, USA), with a modified protocol following Chattopadhyay et al. (Chattopadhyay et al. 2019). In brief, we diluted the volume of baits to 1.85 uL per historical DNA sample (˜3X dilution) and 1.1 uL per fresh DNA sample (˜5X dilution). Biotinylated RNA baits and target sequences were hybridized at 60°C for 40 hours for historical samples and at 65°C for 20 hours for fresh samples. Following hybridization, the samples were cleaned according to the myBaits manual, and PCR was conducted using IS5 and IS6 primers with 20 cycles for historical samples and 15 cycles for fresh samples (Fortes and Paijmans 2015). DNA quantification and assessments of libraries' fragment sizes were conducted as above (see extraction). The peak fragment sizes of each historical target capture library ranged between 250 bp and 400 bp, whereas the peak fragment sizes of each fresh target capture library ranged between 370 bp and 410 bp. The target capture libraries were sequenced using the Illumina HiSeq 2500 and HiSeq 4000 platforms with 150 bp paired-end runs for all samples. Fresh and historical samples were run on separate lanes.

We additionally sequenced the whole genomes of nine *Zosterops*individuals. Their DNA was extracted using the DNeasy Blood and Tissue Kit (Qiagen, Germany) according to the manufacturer's protocol. The samples were then prepared using a Nextera Library Prep Kit (Illumina, USA) with dual indexes. The whole genome libraries were sequenced on the Illumina X10 or NovaSeq platforms at Medgenome (Foster City, California)

with 150 bp paired-end runs.

Sequence Assembly

We removed adapter sequences using Trimmomatic 0.38 (Bolger et al. 2014) and duplicates with FastUniq 1.1 (Xu et al. 2012). Paired trimmed reads of historical samples were examined with mapDamage 2.0.9 (Jónsson et al. 2013) to assess DNA deamination in read ends. We further trimmed 5 bp from the 3' ends of both forward and reverse reads as the mapDamage results show a high amount (>0.1 probability) of G to A misincorporation in the read ends, and reran mapDamage to ensure the average probability of misincorporation across samples remained below 0.1. We ran HybPiper 1.3.1 (Johnson et al. 2016) to extract target sequences. Following the HybPiper pipeline, the trimmed reads were aligned to each target gene using BWA 0.7.17 (Li 2013). We conducted a *de novo* assembly of target sequences using SPAdes 3.13 (Bankevich et al. 2012), applying a sequencing depth cut-off of at least 16X coverage per contig. The contigs generated by SPAdes were re-aligned against the target sequences using Exonerate 2.4.0 (Slater and Birney 2005) to assemble coding sequence regions (including intronic regions) and the resulting DNA sequence of each locus was extracted for downstream analyses. The length of each locus assembled for each sample was examined using the following python scripts in the HybPiper package: get_seq_lengths.py and hybpiper_stats.py. We removed four historical samples due to high amount of missing data and kept 770 loci after quality checks across all samples, ensuring each locus is present in at least 85% of individuals and contains less than 30% missing nucleotides.

An additional 12 *Zosterops* samples were included by extracting target sequences from their whole genomes using blastn in BLAST+ 2.6.0 (Camacho et al. 2009). Three of the whole genomes were obtained from NCBI (Cornetti et al. 2015; Leroy et al. 2019), and the other nine whole genomes were generated for this study (Appendix I). Target sequences were also extracted from the *Mixornis gularis* whole genome (Tan et al. 2018) to be used as an outgroup for some downstream analyses. For the nine whole genomes that were generated for the present study, we first cleaned the raw reads using a modified Perl script to remove exact PCR duplicates and low complexity reads (Bi et al. 2012; Singhal 2013), and used Trimmomatic 0.30 (Bolger et al. 2014) for adapter removal. Contaminants were removed by aligning the raw reads to the reference genomes of potential contaminant sources (such as ribosomal RNA, human, and bacterial DNA) with Bowtie 2.0.1 (Langmead and Salzberg 2012), followed by another cleanup of reads using Cutadapt 1.16 (Martin 2011). We merged overlapping paired reads using FLASH 1.2.11 (Magoč and Salzberg 2011), and aligned the cleaned paired reads against the *Zosterops lateralis* genome using BWA-MEM in BWA 0.7.8 (Li 2013). The data were then converted into bam file format and sorted using SAMtools 1.5 (Li et al. 2009). We used mpileup in BCFtools (Li et al. 2009) to calculate the genotype likelihoods of each site of the reads, and generated a consensus sequence in fasta format using BCFtools. For all 12 whole genome resequenced samples, we used blastn in BLAST+ 2.6.0 (Camacho et al. 2009) to extract the same set of loci as that used in target capture.

Phylogenetic Analyses

MAFFT 7.0 (Katoh and Standley 2013) was run to ensure each locus direction was consistent throughout all samples. All 770 loci were concatenated, giving a final alignment of 1,635,155 bp with 7.60% gaps. RAxML 8.2.12 (Stamatakis 2014) was run to construct a Maximum Likelihood (ML) tree using a GTR + GAMMA + Invariant Sites model with 100 rapid bootstraps to search for the best-scoring ML tree, and the tree was rooted with *M. gularis* .

Best-scoring ML gene trees were inferred for each locus with RAxML 8.2.12 (Stamatakis 2014) with 20 independent searches from a random starting tree and using a GTR + GAMMA + Invariant Sites substitution model. Each gene tree was run with 100 bootstraps for node support. A total of 768 gene trees were rooted with *M. gularis* , while no outgroup sequence was present for two gene trees which had to be discarded for species tree inference.

We adopted the multispecies coalescent model using three different algorithm methods: MP-EST 2.0 (Liu et al. 2010), STAR (Liu et al. 2009), and ASTRID 1.4 (Vachaspati and Warnow 2015). All three species tree methods were run with the 768 best-scoring ML gene trees to infer the species tree topology, and with 100

different sets of input gene trees to infer bootstrap support. Nodes with less than 68% bootstrap support were collapsed.

We assessed the level of congruence in the phylogenetic placement of *Zosterops* species between the present and previously published datasets following the species-level classification by del Hoyo et al. (del Hoyo et al. 2016) and the more recent taxonomic revisions within the Asiatic white-eye complex by Lim et al. (Lim et al. 2019). Specifically, we assigned species to one of three main clades that emerged within the genus in our analyses. We expanded our clade assignment to 30 *Zosterops* taxa not sampled in our study but found to be embedded within one of the three main clades with high bootstrap support (>90%) by at least one previous study (Table S2). We also included two undescribed *Zosterops* species in this expanded clade assignment (Eaton et al. 2016; O'Connell et al. 2019).

Mitochondrial Tree

We observed non-specific binding during hybridization of probes and sample DNA, allowing us to assemble mitochondrial DNA from the raw reads of each individual prepared by target capture. The raw reads were first mapped to the reference mitogenome of *Z. lateralis* (accession no. NC029146) using bwa 0.7.17 (Li 2013), and converted to bam files using SAMtools 1.9 (Li et al. 2009). The bam files were then imported into CLC Genomics Workbench 7.0.4, remapped to the same reference mitogenome, and locally re-aligned. A consensus mitogenome of each individual was extracted with a minimum coverage of five, otherwise an ambiguous base 'N' was inserted. Finally, we extracted 1041 bp of ND2 sequence from each individual by aligning each assembled mitogenome to the ND2 sequence of *Z. lateralis* . Some samples were removed due to extensive missing nucleotides, and the ND2 sequences of 68 individuals were retained. As a means to assess the presence of artifacts from DNA damage, especially in toepad DNA, we compared the ND2 sequences generated in this study with the ND2 sequences of the same taxa deposited on GenBank by previously published studies. We also added the Genbank sequences of 16 *Zosterops* species not represented in our sampling regime. A maximum likelihood tree was generated using RAxML with 10,000 bootstrap replicates under the GTR + GAMMA model.

Testing Gene Flow with ABBA-BABA Statistics

We computed SNP-based ABBA-BABA statistics to test for gene flow among a sub-set of taxa with topological incongruence among trees, as well as populations which overlap in distribution. The bam files generated during locus assembly were used as input for SNP calling for the fresh samples, while the bam files of the historical samples were first subjected to mapDamage 2.0.9 (Jónsson et al. 2013) to rescale the quality scores of possibly deaminated sites. We used ANGSD 0.923 (Korneliussen et al. 2014) to call SNPs with the following filters applied: minimum depth of 20, block size of 50,000 bp, remove transitions, minimum mapping quality of 30, and minimum base quality of 20. We computed D statistics and used the jackknife.R script in ANGSD 0.923 (Korneliussen et al. 2014) to compute critical values (Z) and test for significance. A positive critical value $Z > 3$, corresponding to a p-value below 0.0013, suggests a significant excess of ABBA-like alleles as compared to BABA-like alleles (Green et al. 2010; Patterson et al. 2012). This critical value is widely applied as a threshold for detecting introgression and reduces the likelihood of false positives (Zheng and Janke, 2018).

Demographic Analysis

We additionally assessed the presence of secondary gene flow using PHRAPL 0.62 (Jackson et al. 2017b) by simulating the probability of observing a set of gene trees across various demographic models. We conducted simulations on a smaller subset of populations exhibiting topological incongruence: *Z. auriventer* (A), *Z. atricapilla* (B), *Z. simplex* (C), *Z. emiliae* (D) and *Z. melanurus* (E). A total of four combinations (ACD, ACE, BCD and BDE) were tested with 770 gene trees, each dataset consisting of three populations and an outgroup, *Z. senegalensis* . Each population was randomly subsampled to at most four individuals with ten replicates per locus: four out of five *Z. auriventer* individuals, four out of ten *Z. simplex* individuals, four out of seven *Z. melanurus* individuals, one out of two *Z. atricapilla* samples, and one *Z. emiliae* individual. We generated 48 possible demographic models under the following settings: an overall maximum of three

free parameters (K=3), a maximum of two coalescent events (K=2), either complete isolation or migration event(s) with a single rate (K=1), no variation in population size and growth (K=1), only fully resolved trees were assumed, and only symmetrical migration between populations was assumed. Each dataset was simulated under these 48 different models with 10,000 trees using grid search. The probability of observing the set of gene trees under each model was assessed using Akaike Information Criterion (AIC) scores and the top two models with the lowest AIC values for each dataset were plotted in R 3.6.1 (R Core Team 2019). We assessed the second-best model for each combination because a previous study has shown that the second-best model may emerge as the correct demographic model with further testing (Morales et al. 2017) . Additionally, we computed the genealogical divergence index (*gdi*) of each combination to assess the overall level of divergence between the sister taxa inferred from each model with the combined effects of genetic drift and gene flow (Jackson et al. 2017a). A *gdi* index closed to 0 suggests panmixia, while an index closed to 1 suggests strong divergence.

RESULTS

Congruent Phylogenetic Trees Revealing Three Distinct Lineages

To shed light on the phylogenetic relationships of *Zosterops* species, we employed both concatenation methods, in which sequence data from individual loci are combined into one larger sequence, as well as MSC approaches (MP-EST (Liu et al. 2010), STAR (Liu et al. 2009), and ASTRID (Vachaspati and Warnow 2015)), which account for individual gene tree stochasticity in a coalescent framework (Edwards et al. 2007). All four tree inference methods produced a similar phylogeny with a congruent tree topology for highly supported nodes, except for the placement of a Sundaic group consisting of *Z. atricapilla* and *Z. auriventer* (Fig. 2). These two Sundaic taxa are embedded within the Australasian clade in the concatenated tree, but emerged within the Asiatic clade in the species trees constructed with MP-EST (henceforth our baseline species tree) and ASTRID (Fig. 2). The other species tree method, STAR, shows a weak bootstrap support for an unresolved placement of these Sundaic taxa.

All methods reveal a white-eye radiation divided into three main clades consisting of an Indo-African, Asiatic and Australasian group, respectively (Fig. 2). We expanded clade assignment to an additional 30 *Zosterops* taxa shown to have high bootstrap support for placement within any one of the three main clades based on previously published papers (Warren et al. 2006; Moyle et al. 2009; Cox et al. 2014; Cornetti et al. 2015; Shakya et al. 2018; Cai et al. 2019; O'Connell et al. 2019) to examine the distribution patterns of early-diverging white-eye lineages across their global range (Fig. 2, Table S2). For example, Cai et al. (2019) provide high bootstrap support (>90%) for the position of *Z. mouroniensis* as a descendant of the most recent common ancestor of two unequivocal members of the Indo-African clade (i.e., *Z. palpebrosus* and *Z. borbonicus* ; Fig. 2); thus the species is allocated accordingly, and the breeding distribution of *Z. mouroniensis* (Mt Karthala on Grande Comore Island) is shaded yellow on the global range map (Fig. 2a).

A majority of the taxa thus added are distributed across Africa, where all *Zosterops* species fall within the Indo-African clade, while five of these newly added taxa are distributed across Melanesia, where our species coverage allowed us to detect the presence of only the Australasian clade. Therefore, our mapping suggests that the Afrotropical and the Australo-Papuan regions are depauperate in deeper *Zosterops* lineage diversity (Fig. 2a; Table S2). Similarly, our results reveal that most areas in continental Asia generally harbor only one of the three main *Zosterops* clades, except East Asia where two clades co-occur in a narrow zone of overlap between the Indo-African *Z. palpebrosus* and the Asiatic *Z. simplex* (orange in Fig. 2a). In contrast, the Indonesian archipelago emerged as a center of deeper lineage diversity for *Zosterops* , with all three main clades represented on Java and Borneo, and two main clades on many other islands (Fig. 2a).
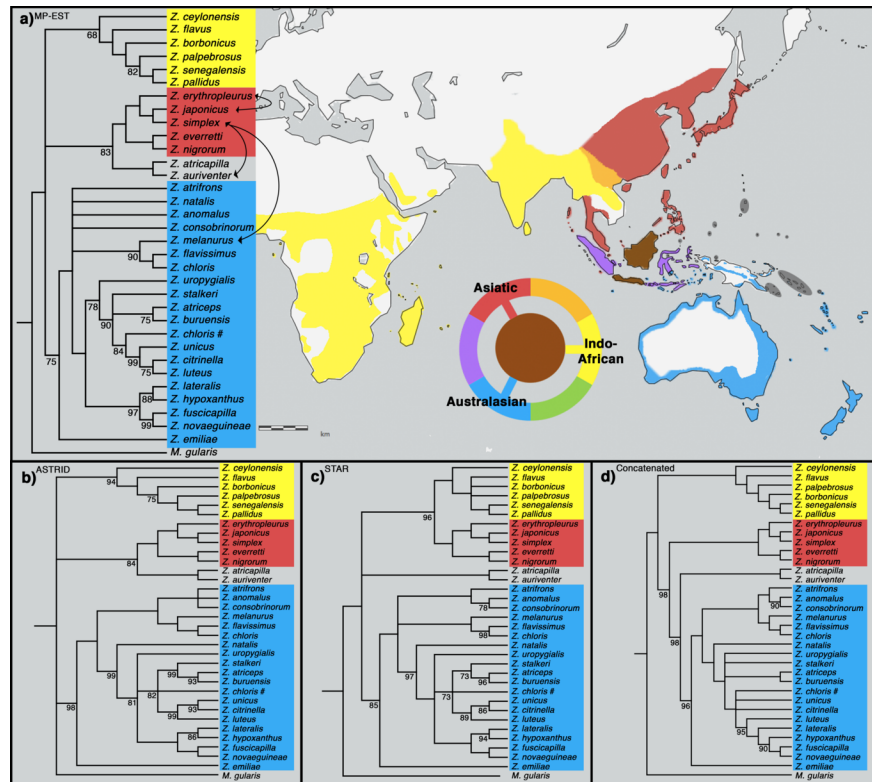
Figure 2. Phylogenetic trees of *Zosterops* generated by (a) MP-EST, (b) ASTRID, (c) STAR and (d) the concatenation method. All nodes are supported by a bootstrap value of 100 unless otherwise stated. Nodes with less than 68% bootstrap support were collapsed. The three main clades are color-coded blue (Australasian), red (Asiatic) and yellow (Indo-African). (a) The distribution of each main clade is color-coded on the distribution map. The map includes an additional 30 taxa covered by previous studies but not included in the present study. Their distributions are color-coded so long as their affinity with any of the three main clades was established by strong bootstrap. Regions shaded dark grey represent the distribution of *Zosterops* taxa of unknown origin due to insufficient phylogenetic data. Borneo and Java (shaded brown) each harbor taxa from all three main clades, while multiple islands across the rest of Indonesia (shaded purple) each harbor taxa from two main clades. Secondary gene flow detected in multiple species pairs is marked with black arrows on the tree. (d) Refer to Fig. S1 for a full concatenated tree with every individual shown.

Presence of Secondary Gene Flow

The genetic signal of recent and rapid radiations is often convoluted by the presence of secondary gene flow, leading to heterogeneous gene trees which deviate from the true phylogeny. We assessed the presence of secondary gene flow specifically between members of a Sundaic species pair (*Z. auriventer* and *Z. atricapilla*) characterized by a shifting and incongruent placement across trees with representatives from the Asiatic (i.e. *Z. simplex*) and Australasian (i.e.*Z. emiliae* and *Z. melanurus*) clades. Most of the top demographic models computed by PHRAPL produced topologies in which the Sundaic species of controversial placement emerge as sister to the Asiatic clade (Fig. 3). An overall high degree of secondary gene flow was inferred for most pairwise combinations of present-day lineages and possible ancestral lineages, suggesting that the simulated models may be oversimplified and not able to fully discern the complex relationships within this rapid radiation (Fig. 3). Convergence onto a single demographic model would require further PHRAPL simulations allowing for asymmetrical rates of gene flow, which are challenging at present computational
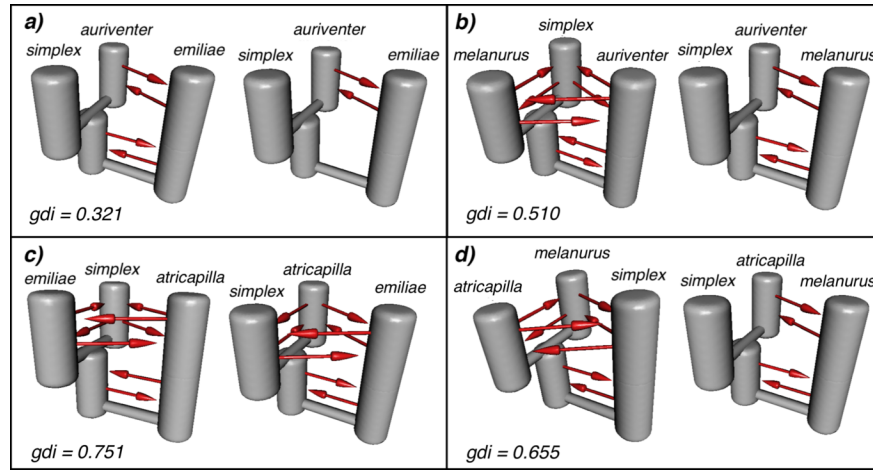
8

capabilities (Morales et al. 2017).

Figure 3. Each panel shows the top two demographic models estimated by PHRAPL to test the affinities of the Sundaic species pair consisting of *Z. auriventer* (panel a–b) and *Z. atricapilla* (panel c–d) with representatives of the Asiatic clade (*Z. simplex*) and the Australasian clade (*Z. emiliae* [panel a, c] and *Z. melanurus* [panel b, d]; see Fig. 2). The genealogical divergence index (*gdi*) value of each combination is shown in the respective panel. (a) *Z. auriventer* is more closely related to the Asiatic clade (i.e., *Z. simplex*), albeit with secondary gene flow with the Australasian clade (i.e., *Z. emiliae*). (b-c) The best models (left in each panel) show a closer relationship between the Asiatic (i.e., *Z. simplex*) and Australasian (i.e., *Z. emiliae / melanurus*) clades, but the second best models show a closer relationship between the two Sundaic species (*Z. atricapilla* or *Z. auriventer*) and the Asiatic clade (i.e., *Z. simplex*). Gene flow is detected in most directions. (d) The best model inferred *Z. atricapilla* is more closely related to the Australasian clade (i.e., *Z. melanurus*), but the second best model estimated *Z. atricapilla* is more closely related to the Asiatic clade (i.e., *Z. simplex*).

In contrast to the difficulties of PHRAPL's demographic simulations to converge on possible avenues of secondary gene flow, the ABBA-BABA approach provided compelling evidence. Testing for an excess of shared derived alleles using strategic combinations of taxa, we were able to infer secondary gene flow that conclusively accounts for the conflicting placement of the Sundaic species pair consisting of *Z. atricapilla* and *Z. auriventer* (Figs. 2, 4). In both the species and concatenated trees, *Z. emiliae* emerges as basal to all other members of the Australasian clade (Figs. 2, 4a). Therefore, in the absence of introgression, the Sundaic *Z. atricapilla* and *Z. auriventer* should be equally closely related to both *Z. emiliae* and other members of the Australasian clade (tree topology in Fig. 4b). The ABBA-BABA statistics revealed a significant excess of derived allele sharing between the Sundaic taxa and Australasian *Z. melanurus* and *Z. citrinella*, thus supporting ancestral introgression between the two controversial Sundaic taxa and the Australasian lineage after *Z. emiliae* had diverged (Fig. 4; set 1 in Table 1), facilitating a conflicting topology produced by concatenation methods (Fig. 2d).
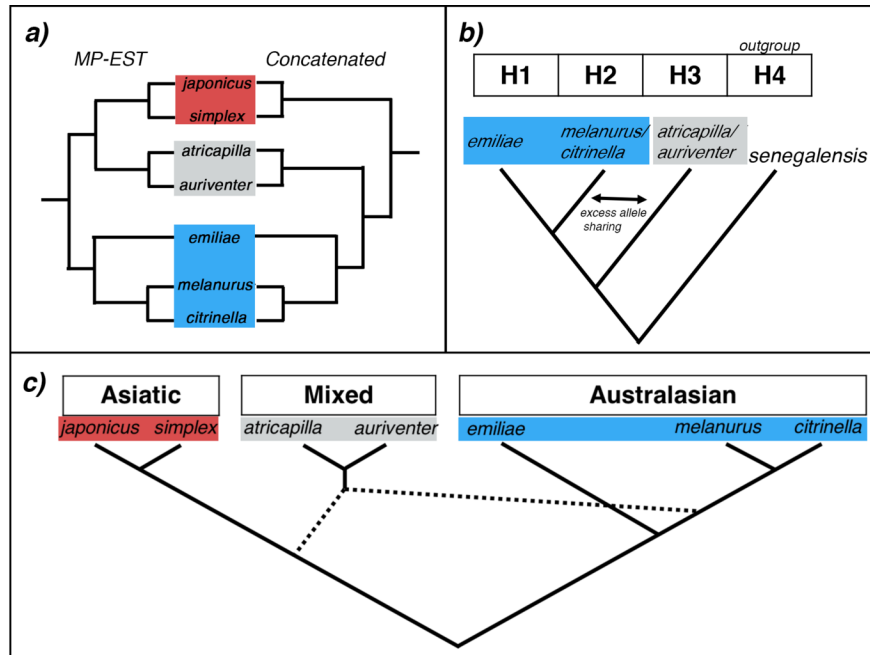
9

Figure 4. Test of secondary gene flow with ABBA-BABA statistics. (a) Placement of Sundaic pair comprising *Z. atricapilla* and *Z. auriventer* conflicts between MP-EST species tree and concatenated tree (see Fig. 2). (b) ABBA-BABA test regime for secondary gene flow, testing for an excess of derived allele sharing between H2 and H3. The regime testing for species in set 1 (Table 1) is shown; see Table 1 for additional species combinations tested. (c) Topology inferred from ABBA-BABA tests, showing the two Sundaic lineages (*atricapilla* and *auriventer*) as carriers of genomic admixture between both main clades.

In addition, secondary gene flow was also detected in several pairs of species which overlap in present-day distribution (sets 2 to 4 in Table 1). For instance, we found that *Z. melanurus* shares significantly more alleles with *Z. simplex* than with *Z. erythropleurus*, suggesting occasional hybridization between *Z. melanurus buxtoni* and *Z. simplex erwini* on Sumatra, where both occur. In a similar vein, *Z. erythropleurus* and *Z. japonicus* displayed excess allele sharing, indicating potential introgression in areas of overlap in Korea, and *Z. auriventer* and *Z. simplex* exhibited excess allele sharing that hints at occasional hybridization in parts of peninsular Malaysia and Borneo where they overlap.

Table 1. D-statistics of a selection of species combinations in which two species (H2, H3) exhibit an excess of derived allele sharing. The ABBA-BABA test was restricted to species combinations with conflicting tree topologies observed in this study (set 1), conflicting tree topologies between this study and Cai et al. (2019) (set 5), and a selection of species with a present-day geographic overlap and opportunities for secondary gene flow (sets 2 to 4). *Z. senegalensis* (H4) was used as an outgroup for all comparisons. A critical value (Z) above 3 suggests an excess of derived allele sharing between populations H2 and H3. See Fig. 4b for explanation of the testing regime involving H1, H2 and H3.

Set	H1	H2	H3	D-stat	Z	No. of ABBA	No. of BABA
1	*emiliae*	*melanurus*	*atricapilla*	0.129	3.807	192.84	148.63
	emiliae	*melanurus*	*auriventer*	0.163	5.168	190.17	136.97
	emiliae	*citrinella*	*atricapilla*	0.110	3.086	178.55	143.09
	emiliae	*citrinella*	*auriventer*	0.125	3.788	172.21	133.99
2	*citrinella*	*melanurus*	*simplex*	0.093	3.074	142.55	118.22
3	*simplex*	*japonicus*	*erythropleurus*	0.215	5.272	171.87	111.15

Set	H1	H2	H3	D-stat	Z	No. of ABBA	No. of BABA
4	*erythropleurus*	*simplex*	*auriventer*	0.108	3.169	166.20	133.74
5	*melanurus*	*emiliae*	*simplex*	-0.053	-1.456	152.33	169.34
	melanurus	*emiliae*	*japonicus*	-0.066	-1.729	149.05	170.02
	citrinella	*emiliae*	*simplex*	0.027	0.706	154.77	146.74
	citrinella	*emiliae*	*japonicus*	-0.027	-0.686	146.72	154.82

DISCUSSION

Phylogeny of Zosterops *and Presence of Secondary Gene Flow*

Given white-eyes' model status in speciation research, the evolutionary history of *Zosterops* has been addressed by a fair amount of research, but most studies have utilized single mitochondrial or few nuclear loci, resulting in trees that are plagued by unresolved polytomies (e.g., Fig. S2; (Slikas et al. 2000; Warren et al. 2006; Moyle et al. 2009; Oatley et al. 2012; Nyári and Joseph 2013; Cox et al. 2014; Husemann et al. 2016; Linck et al. 2016; Round et al. 2017; Wickramasinghe et al. 2017; Shakya et al. 2018; Cai et al. 2019; Lim et al. 2019; O'Connell et al. 2019). Using more than 700 genome-wide loci with a dense species sampling, our study produced a greatly improved phylogeny of *Zosterops* and reveals the existence of three discrete main clades characterized by an Indo-African, Asiatic and Australasian core of distribution, respectively (Fig. 2). Both concatenated and species tree analyses resulted in congruent topologies for well-supported nodes, except for the placement of a Sundaic species pair comprising *Z. atricapilla* and *Z. auriventer* (Fig. 2).

Uncertainties in the affinity of these two Sundaic species appear to be due to secondary gene flow between the incipient Asiatic and Australasian clades (Figs. 3, 4). Such genetic introgression destabilizes the robustness of the MSC model, which does not account for secondary gene flow, thereby confounding species tree estimation (Fig. 4, Table 1). Traditional phylogenetic approaches such as concatenation do not account for secondary gene flow either, but are additionally unable to cope with incomplete lineage sorting. SNP-based analysis using ABBA-BABA statistics conclusively identified introgression as an underlying cause of the conflicting placement of *Z. atricapilla* and *Z. auriventer*. Specifically, excess allele sharing between these two species and *Z. melanurus*, but not with *Z. emiliae*, suggests introgression between the incipient stages of the Asiatic and Australasian clades – after *Z. emiliae* had split off from other Australasian species (Fig. 4c). Such ancient introgression generates patterns of allele sharing that would lead to the two controversial Sundaic species partly being reflected as members of the one clade or of the other (Fig. 2, 4), depending on tree inference methods.

More generally, the ABBA-BABA test detected rampant secondary gene flow between species that geographically overlap, regardless of their phylogenetic proximity, underscoring the pervasive nature of genetic introgression in phylogenetics (Fig. 2a; sets 2-4 in Table 1). For example, *Z. auriventer* and *Z. simplex* display excess allele sharing indicating recent gene flow in areas of Sundaland where they overlap, even though they belong to different main clades of *Zosterops* (Fig. 2, Table 1). The same is true for *Z. melanurus* and *Z. simplex*, which co-occur on Sumatra, and for *Z. japonicus* and *Z. erythropleurus*, which overlap in Korea (Table 1). Such introgression between species is likely recent, limited, and of the kind that the MSC model remains robust to (Liu et al. 2009). The detection of rampant secondary gene flow in multiple pairs of sympatric white-eye species is in agreement with the recent discovery of introgression between various non-sister white-eye species across the Solomon Islands (Manthey et al. 2020).

Indonesian Archipelago Harbors All Three Main Clades

While the highest rates of *Zosterops* diversification are known to have occurred on archipelagoes in general (Diamond et al. 1976; Moyle et al. 2009), the geographic distribution of deeper-level lineage diversity in this genus remains unexplored. Our phylogenetic results demonstrate that Africa, most parts of continental Asia and probably also all of the Australo-Papuan region each harbor representatives from only one of the three

main *Zosterops* clades, respectively (Fig. 2a), regardless of *Zosterops* species diversity. In contrast, virtually all parts of the Indonesian archipelago harbor white-eye species from two to three of the main *Zosterops* clades (Fig. 2a). Positioned between the Sunda and Sahul shelf, the Indonesian islands are a center of syntopy of lineages from either side of Wallace's Line (Moss and Wilson 1998; de Bruyn et al. 2014). This is consistent with the rapid rate of tectonic change reconstructed for the Indonesian archipelago over the last 30 million years (Hall 2002, 2012; de Bruyn et al. 2014; Nugraha and Hall 2018), which has led to a narrowing of open sea between Asia and Australia and thereby facilitated overwater dispersal of Sundaic and/or Australo-Papuan lineages for many organismic groups (e.g., (Heads 2001; Irestedt et al. 2013; Gwee et al. 2017; Ng et al. 2017; Garg et al. 2018; Reilly et al. 2019), including white-eyes). To the west of Wallace's Line, the Sundaic islands of Borneo and Java, which constitute a large part of the western Indonesian archipelago, additionally harbor the coastal endemic species *Z. flavus,* which has a surprising affinity with the Indo-African clade rather than with the geographically more proximate Asiatic and Australasian clades, demonstrating an impressive potential for dispersal capability that may in part underlie the rapid diversification rate of the genus.

Low-lying Indonesian islands to the east of Wallace's Line, such as Sumba and Kai, generally harbor only *Zosterops* species from the Australasian clade, whereas *Z. japonicus* of the Asiatic clade additionally occurs on islands that reach (sub-) montane elevations of over 1200 m. For example*, Z. japonicus* is present on mountainous Buru (c. 9500 km^2 in size) but absent on Sumba (c. 11,000 km^2 in size), which largely comprises savannah with a small hilly region not exceeding 1200m in elevation (Fig. 1), despite Sumba being a larger island. Although the elevation of an island contributes substantially to the number of main *Zosterops* clades present, it seems to have less influence on the total number of *Zosterops* species. For instance, the Kai islands, with a combined area of only approximately 1400 km^2 and an elevation of less than 700 m, harbor three distinct *Zosterops* species (including two island endemics), though all belong to the Australasian clade (Fig. 1).

Our phylogenetic results reveal that the widespread Lemon-bellied White-eye *Z. chloris* is non-monophyletic (Figs. 2, S1) and several small, low-lying islands situated between the Banda Sea and Arafura Sea, including Kai and Aru, may harbor a cryptic species morphologically identical to the Lemon-bellied White-eye. White-eyes are renowned for their conservative morphology, which contributes to the confusion that has surrounded their taxonomic treatment (Mees 1957; Mayr 1965; Lim et al. 2019; Manthey et al. 2020). Further research is required to ascertain the evolutionary status of these and other overlooked island populations.

Borneo is a Hotspot for Evolution and Harbors Deep Phylogenetic Isolates

Borneo has been identified as a major source of diversification across the Southeast Asian region for a variety of organismic groups, including birds, mammals, amphibians, and plants (de Bruyn et al. 2014). Our study reveals Borneo's unique status as the only place on Earth where members of all three main *Zosterops* clades occur, and where as many as four *Zosterops* species co-exist within a few square kilometers of one another, rendering it the center of faunal mixing for white-eyes (Figs. 1, 2). Lying in the heart of Southeast Asia and comprising distinct ecoregions (Myers 2016), Borneo provides a variety of niches accommodating mangrove species (*Z. flavus*), submontane species (*Z. atricapilla* and *Z. auriventer*), montane endemics (*Z. emiliae* and an undescribed white-eye from the Meratus range (Eaton et al. 2016)), and coastal woodland species (*Z. simplex*). Borneo constitutes the largest landmass within the Sundaic region, and its elevated count in main *Zosterops* clades strongly suggests that it has played a key role in the diversification of this important radiation.

While Borneo's exceptional biodiversity has been appreciated as early as during Alfred R. Wallace's times (Wallace 1962), most of this diversity has traditionally been interpreted as being of a Sundaic element largely shared with Sumatra and peninsular Malaysia, and is only slowly being recognized as having attained species-level depths of differentiation (Cros et al. 2020)(Cros et al. 2020). On the other hand, Borneo is known for hosting a number of deep phylogenetic isolates, such as the enigmatic Bristlehead *Pityriasis gymnocephala* (Oliveros et al. 2019). In the context of *Zosterops* diversification, we add the Mountain Black-eye *Z. emiliae* as an overlooked phylogenetic isolate (Fig. 2). Because of its unusual facial morphology, it continues to be treated as a monospecific genus (*Chlorocharis*) by some modern sources (del Hoyo et al. 2016), though it has

12

been shown to be embedded within *Zosterops* for over a decade (Moyle et al. 2009; Cai et al. 2019). Using our much-improved taxon sampling, *Z. emiliae* emerged as a basal sister to the Australasian clade with moderate support under various analytical regimes (Fig. 2). In contrast, Cai et al. (2019) placed *Z. emiliae* with members of the Asiatic clade. Our use of the ABBA-BABA test to verify whether this conflicting position may be due to ancient introgression did not identify an excess of allele sharing between *Z. emiliae* and the Asiatic members (set 5 in Table 1), suggesting that Cai et al.'s (2019) incongruent placement of *Z. emiliae* is unlikely to be a result of secondary gene flow. Instead, incomplete lineage sorting may have generated such phylogenetic conflict as the divergence of *Z. emiliae* likely fell within a time of rapid diversification within the genus, leading to its recalcitrance to phylogenetic resolution when only few loci are applied.

Conclusions

Our study presents the application of species tree methods on a large set of genome-wide markers across a comprehensive sampling of members of a rapid radiation of a classic 'great speciator'. Using approaches to account for secondary gene flow, our study demonstrates the pervasive presence of genetic introgression across this explosive radiation. The resultant phylogeny of *Zosterops* white-eyes reveals that the Indonesian archipelago, and Borneo in particular, are an evolutionary hotspot for the diversification of the genus. This archipelagic region harbors members of clades centred in the neighboring Asian and Australo-Papuan landmasses, and even from a geographically distant Indo-African clade. The identification of areas in western Indonesia as a major centre of modern phylogenetic diversity not only contributes to their recognition as an arena of important evolutionary processes, but also elevates their status as a region of global conservation relevance.

COMPETING INTERESTS

The authors have no competing interests.

REFERENCES

Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J. 1990. Basic local alignment search tool. Journal of Molecular Biology. 215:403-410.

Bankevich A., Nurk S., Antipov D., Gurevich A.A., Dvorkin M., Kulikov A.S., Lesin V.M., Nikolenko S.I., Pham S., Prjibelski A.D. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of Computational Biology. 19:455-477.

Bi K., Vanderpool D., Singhal S., Linderoth T., Moritz C., Good J.M. 2012. Transcriptome-based exon capture enables highly cost-effective comparative genomic data collection at moderate evolutionary scales. BMC Genomics. 13:403.

Bolger A.M., Lohse M., Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30:2114-2120.

Cai T., Cibois A., Alström P., Moyle R.G., Kennedy J.D., Shao S., Zhang R., Irestedt M., Ericson P.G., Gelang M. 2019. Near-complete phylogeny and taxonomic revision of the world's babblers (Aves: Passeriformes). Molecular Phylogenetics and Evolution. 130:346-356.

Cai T., Shao S., Kennedy J.D., Alstrom P., Moyle R.G., Qu Y., Lei F., Fjeldsa J. 2020. The role of evolutionary time, diversification rates and dispersal in determining the global diversity of a large radiation of passerine birds. Journal of Biogeography. In Press.

Camacho C., Coulouris G., Avagyan V., Ma N., Papadopoulos J., Bealer K., Madden T.L. 2009. BLAST+: architecture and applications. BMC Bioinformatics. 10:421.

Chattopadhyay B., Garg K.M., Mendenhall I.H., Rheindt F.E. 2019. Historic DNA reveals Anthropocene threat to a tropical urban fruit bat. Current Biology. 29:R1299-R1300.

Clegg S.M., Degnan S.M., Moritz C., Estoup A., Kikkawa J., Owens I.P.F. 2002. Microevolution in island forms: the roles of drift and directional selection in morphological divergence of a passerine bird. Evolution. 56:2090-2099.

Cornetti L., Valente L.M., Dunning L.T., Quan X., Black R.A., Hébert O., Savolainen V. 2015. The genome of the "great speciator" provides insights into bird diversification. Genome Biology and Evolution. 7:2680-2691.

Cox S.C., Prys-Jones R.P., Habel J.C., Amakobe B.A., Day J.J. 2014. Niche divergence promotes rapid diversification of East African sky island white-eyes (Aves: Zosteropidae). Molecular Ecology. 23:4103-4118.

Cros E., Chattopadhyay B., Garg K.M., Ng N.S.R., Tomassi S., Benedick S., Edwards D.P., Rheindt F.E. 2020. Quaternary land bridges have not been universal conduits of gene flow. Molecular Ecology. 29:2692-2706.

de Bruyn M., Stelbrink B., Morley R.J., Hall R., Carvalho G.R., Cannon C.H., van den Bergh G., Meijaard E., Metcalfe I., Boitani L. 2014. Borneo and Indochina are major evolutionary hotspots for Southeast Asian biodiversity. Systematic Biology. 63:879-901.

del Hoyo J., Collar N.J., Christie D.A., Elliott A., Fishpool L.D.C., Boesman P., Kirwan G.M. 2016. Illustrated Checklist of the Birds of the World, Volume 2: Passerines. Lynx Edicions, Barcelona & BirdLife International, Cambridge.

Diamond J.M., Gilpin M.E., Mayr E. 1976. Species-distance relation for birds of the Solomon Archipelago, and the paradox of the great speciators. Proceedings of the National Academy of Sciences. 73:2160-2164.

Eaton J.A., Mitchell S.L., Bocos C.N.G., Rheindt F.E. 2016. A short survey of the Meratus Mountains, South Kalimantan province, Indonesia: two undescribed avian species discovered. BirdingASIA. 26:107-113.

Edwards S.V., Kingan S.B., Calkins J.D., Balakrishnan C.N., Jennings W.B., Swanson W.J., Sorenson M.D. 2005. Speciation in birds: genes, geography, and sexual selection. Proceedings of the National Academy of Sciences. 102:6550-6557.

Edwards S.V., Liu L., Pearl D.K. 2007. High-resolution species trees without concatenation. Proceedings of the National Academy of Sciences. 104:5936-5941.

Edwards S.V., Potter S., Schmitt C.J., Bragg J.G., Moritz C. 2016. Reticulation, divergence, and the phylogeography–phylogenetics continuum. Proceedings of the National Academy of Sciences. 113:8025-8032.

Ellegren H., Smeds L., Burri R., Olason P.I., Backstrom N., Kawakami T., Kunstner A., Makinen H., Nadachowska-Brzyska K., Qvarnstrom A. 2012. The genomic landscape of species divergence in Ficedula flycatchers. Nature. 491:756-760.

Fortes G.G., Paijmans J.L.A. 2015. Analysis of whole mitogenomes from ancient samples. In: Kroneis T. editor. Whole Genome Amplification. USA, Springer, p. 179-195.

Garg K.M., Chattopadhyay B., Wilton P.R., Prawiradilaga D.M., Rheindt F.E. 2018. Pleistocene land bridges act as semipermeable agents of avian gene flow in Wallacea. Molecular phylogenetics and evolution. 125:196-203.

Green R.E., Krause J., Briggs A.W., Maricic T., Stenzel U., Kircher M., Patterson N., Li H., Zhai W., Fritz M.H.-Y. 2010. A draft sequence of the Neandertal genome. science. 328:710-722.

Gwee C.Y., Christidis L., Eaton J.A., Norman J.A., Trainor C.R., Verbelen P., Rheindt F.E. 2017. Bioacoustic and multi-locus DNA data of*Ninox* owls support high incidence of extinction and recolonisation on small, low-lying islands across Wallacea. Molecular Phylogenetics and Evolution. 109:246-258.

Hall R. 2002. Cenozoic geological and plate tectonic evolution of SE Asia and the SW Pacific: computer-based reconstructions, model and animations. Journal of Asian Earth Sciences. 20:353-431.

Hall R. 2012. Late Jurassic–Cenozoic reconstructions of the Indonesian region and the Indian Ocean. Tectonophysics. 570-571:1-41.

Heads M. 2001. Birds of paradise, biogeography and ecology in New Guinea: a review. Journal of Biogeography. 28:893-925.

Husemann M., Sturm S., Curto M., Meimberg H., Habel J.C. 2016. Four new mitochondrial genomes of the genus zosterops (aves: passeriformes: zosteropidae) from East Africa with a phylogenetic evaluation of the group. Mitochondrial DNA Part B: Resources. 1:544-548.

Irestedt M., Fabre P.-H., Batalha-Filho H., Jonsson K.A., Roselaar C.S., Sangster G., Ericson P.G.P. 2013. The spatio-temporal colonization and diversification across the Indo-Pacific by a 'great speciator'(Aves,*Erythropitta erythrogaster*). Proceedings of the Royal Society B: Biological Sciences. 280:20130309.

Jackson N.D., Carstens B.C., Morales A.E., O'Meara B.C. 2017a. Species delimitation with gene flow. Systematic biology. 66:799-812.

Jackson N.D., Morales A.E., Carstens B.C., O'Meara B.C. 2017b. PHRAPL: Phylogeographic Inference Using Approximate Likelihoods. Systematic Biology. 66:1045-1053.

Johnson M.G., Gardner E.M., Liu Y., Medina R., Goffinet B., Shaw A.J., Zerega N.J.C., Wickett N.J. 2016. HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. Applications in Plant Sciences. 4.

Jonsson H., Ginolhac A., Schubert M., Johnson P.L.F., Orlando L. 2013. mapDamage2. 0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics. 29:1682-1684.

Katoh K., Standley D.M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular Biology and Evolution. 30:772-780.

Korneliussen T.S., Albrechtsen A., Nielsen R. 2014. ANGSD: analysis of next generation sequencing data. BMC bioinformatics. 15:356.

Lamichhaney S., Berglund J., Almen M.S., Maqbool K., Grabherr M., Martinez-Barrio A., Promerova M., Rubin C.-J., Wang C., Zamani N. 2015. Evolution of Darwin's finches and their beaks revealed by genome sequencing. Nature. 518:371-375.

Langmead B., Salzberg S.L. 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods. 9:357.

Lerner H.R.L., Meyer M., James H.F., Hofreiter M., Fleischer R.C. 2011. Multilocus resolution of phylogeny and timescale in the extant adaptive radiation of Hawaiian honeycreepers. Current Biology. 21:1838-1844.

Leroy T., Anselmetti Y., Tilak M.-K., Berard S., Csukonyi L., Gabrielli M., Scornavacca C., Mila B., Thebaud C., Nabholz B. 2019. A bird's white-eye view on neosex chromosome evolution. bioRxiv.

Li C., Riethoven J.J.M., Naylor G.J.P. 2012. EvolMarkers: A database for mining exon and intron markers for evolution, ecology and conservation studies. Molecular Ecology Resources. 12:967-971.

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

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. 2009. The sequence alignment/map format and SAMtools. Bioinformatics. 25:2078-2079.

Lim B.T.M., Sadanandan K.R., Dingle C., Leung Y.Y., Prawiradilaga D.M., Irham M., Ashari H., Lee J.G.H., Rheindt F.E. 2019. Molecular evidence suggests radical revision of species limits in the great speciator white-eye genus *Zosterops* . Journal of Ornithology. 160:1-16.

Linck E., Schaack S., Dumbacher J.P. 2016. Genetic differentiation within a widespread "supertramp" taxon: Molecular phylogenetics of the Louisiade White-eye (*Zosterops griseotinctus*). Molecular Phylogenetics and Evolution. 94:113-121.

Liu L., Xi Z., Davis C.C. 2015. Coalescent methods are robust to the simultaneous effects of long branches and incomplete lineage sorting. Molecular Biology and Evolution. 32:791-805.

Liu L., Yu L., Edwards S.V. 2010. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. BMC Evolutionary Biology. 10:302.

Liu L., Yu L., Pearl D.K., Edwards S.V. 2009. Estimating species phylogenies using coalescence times among sequences. Systematic Biology. 58:468-477.

Lohman D.J., Bruyn M.d., Page T., Rintelen K.v., Hall R., Ng P.K.L., Shih H.-T., Carvalho G.R., Rintelen T.v. 2011. Biogeography of the Indo-Australian Archipelago. Annual Review of Ecology Evolution and Systematics. 42:205-226.

MacArthur R.H., Wilson E.O. 2001. The Theory of Island Biogeography. Princeton, New Jersey, Princeton University Press.

Magoč T., Salzberg S.L. 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics. 27:2957-2963.

Manthey J.D., Oliveros C.H., Andersen M.J., Filardi C.E., Moyle R.G. 2020. Gene flow and rapid differentiation characterize a rapid insular radiation in the southwest Pacific (Aves: *Zosterops*). Evolution. 74:1788-1803.

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

Mayr E. 1965. Relationships among Indo-Australian Zosteropidae (Aves). In: Wright N. editor. Breviora. Cambridge, Massachusetts, USA, Museum of Comparative Zoology, Harvard University, p. 1-6.

Mees G.F. 1957. A systematic review of the Indo-Australian Zosteropidae (Part I). Zoologische Verhandelingen. 35:1-204.

Morales A.E., Jackson N.D., Dewey T.A., O'Meara B.C., Carstens B.C. 2017. Speciation with gene flow in North American *Myotis* bats. Systematic Biology. 66:440-452.

Moss S.J., Wilson M.E.J. 1998. Biogeographic implications of the Tertiary palaeogeographic evolution of Sulawesi and Borneo. Biogeography and Geological Evolution of SE Asia. 133:163.

Moyle R.G., Filardi C.E., Smith C.E., Diamond J. 2009. Explosive Pleistocene diversification and hemispheric expansion of a "great speciator". Proceedings of the National Academy of Sciences. 106:1863-1868.

Myers S. 2016. Birds of Borneo. London, Bloomsbury Publishing.

Ng N.S., Wilton P.R., Prawiradilaga D.M., Tay Y.C., Indrawan M., Garg K.M., Rheindt F.E. 2017. The effects of Pleistocene climate change on biotic differentiation in a montane songbird clade from Wallacea. Molecular phylogenetics and evolution. 114:353-366.

Nugraha A.M.S., Hall R. 2018. Late Cenozoic palaeogeography of Sulawesi, Indonesia. Palaeogeography, Palaeoclimatology, Palaeoecology. 490:191-209.

Nyári Á.S., Joseph L. 2013. Comparative phylogeography of Australo-Papuan mangrove-restricted and mangrove-associated avifaunas. Biological Journal of the Linnean Society. 109:574-598.

O'Connell D.P., Kelly D.J., Lawless N., O'Brien K., Marcaigh F.Ó., Karya A., Analuddin K., Marples N.M. 2019. A sympatric pair of undescribed white-eye species (Aves: Zosteropidae: *Zosterops*) with different origins. Zoological Journal of the Linnean Society. 186:701-724.

Oatley G., Voelker G., Crowe T.M., Bowie R.C.K. 2012. A multi-locus phylogeny reveals a complex pattern of diversification related to climate and habitat heterogeneity in southern African white-eyes. Molecular Phylogenetics and Evolution. 64:633-644.

Oliveros C.H., Field D.J., Ksepka D.T., Barker F.K., Aleixo A., Andersen M.J., Alström P., Benz B.W., Braun E.L., Braun M.J. 2019. Earth history and the passerine superradiation. Proceedings of the National Academy of Sciences. 116:7916-7925.

Patterson N., Moorjani P., Luo Y., Mallick S., Rohland N., Zhan Y., Genschoreck T., Webster T., Reich D. 2012. Ancient admixture in human history. Genetics. 192:1065-1093.

Quinlan A.R., Hall I.M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 26:841-842.

R Core Team. 2019. R: a language and environment for statistical computing computer program.

Reilly S.B., Stubbs A.L., Karin B.R., Bi K., Arida E., Iskandar D.T., McGuire J.A. 2019. Leap-frog dispersal and mitochondrial introgression: Phylogenomics and biogeography of *Limnonectes* fanged frogs in the Lesser Sundas Archipelago of Wallacea. Journal of Biogeography. 46:757-769.

Rheindt F.E., Edwards S.V. 2011. Genetic introgression: an integral but neglected component of speciation in birds. The Auk. 128:620-632.

Round P.D., Manawattana S., Khudamrongsawat J., Thunhikorn S., Safoowong M., Bhummakasikara T. 2017. Disentangling avian diversity: South-East Asian mainland Oriental white-eye *Zosterops palpebrosus* constitutes two distinct lineages. Forktail. 33:103-115.

Shakya S.B., Haryoko T., Burner R.C., Prawiradilaga D.M., Sheldon F.H. 2018. Preliminary assessment of community composition and phylogeographic relationships of the birds of the Meratus Mountains, south-east Borneo, Indonesia. Bulletin of the British Ornithologists' Club. 138:45-66.

Singhal S. 2013. De novo transcriptomic analyses for non-model organisms: An evaluation of methods across a multi-species data set. Molecular Ecology Resources. 13:403-416.

Slater G.S.C., Birney E. 2005. Automated generation of heuristics for biological sequence comparison. BMC bioinformatics. 6:31.

Slikas B., Jones I.B., Derrickson S.R., Fleischer R.C. 2000. Phylogenetic relationships of Micronesian white-eyes based on mitochondrial sequence data. The Auk. 117:355-365.

Smit A., Hubley R., Green P. 2013-2015. RepeatMasker 4.0. Seattle, WA, Institute for Systems Biology.

Song S., Liu L., Edwards S.V., Wu S. 2012. Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. Proceedings of the National Academy of Sciences. 109:14942-14947.

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

Tan D.J.X., Chattopadhyay B., Garg K.M., Cros E., Ericson P.G.P., Irestedt M., Rheindt F.E. 2018. Novel genome and genome-wide SNPs reveal early fragmentation effects in an edge-tolerant songbird population across an urbanized tropical metropolis. Scientific Reports. 8:1-12.

Templeton J.E.L., Brotherton P.M., Llamas B., Soubrier J., Haak W., Cooper A., Austin J.J. 2013. DNA capture and next-generation sequencing can recover whole mitochondrial genomes from highly degraded samples for human identification. Investigative Genetics. 4:26.

Vachaspati P., Warnow T. 2015. ASTRID: accurate species trees from internode distances. BMC Genomics. 16:S3.

Wallace A.R. 1962. The Malay Archipelago: the land of the orang-utan and the bird of paradise; a narrative of travel, with studies of man and nature. New York, Courier Corporation.

Warren B.H., Bermingham E., Prys-Jones R.P., Thebaud C. 2006. Immigration, species radiation and extinction in a highly diverse songbird lineage: white-eyes on Indian Ocean islands. Molecular Ecology. 15:3769-3786.

Whittaker R.J., Fernandez-Palacios J.M. 2007. Island Biogeography: Ecology, Evolution, and Conservation. Oxford, Oxford University Press.

Wickramasinghe N., Robin V.V., Ramakrishnan U., Reddy S., Seneviratne S.S. 2017. Non-sister Sri Lankan white-eyes (genus *Zosterops*) are a result of independent colonizations. PLOS ONE. 12:e0181441.

Xu H., Luo X., Qian J., Pang X., Song J., Qian G., Chen J., Chen S. 2012. FastUniq: a fast de novo duplicates removal tool for paired short reads. PLOS ONE. 7:e52249.