Species-specific dynamics may cause deviations from general biogeographical predictions – evidence from a population genomics study of a New Guinean endemic passerine bird family (Melampittidae).

Ingo A. Müller[1,3,4*], Filip Thörn[1,3,4], Samyuktha Rajan[2], Per Ericson[3], John P. Dumbacher[5], Gibson Maiah[6], Mozes Blom[4], Knud A. Jønsson[3], Martin Irestedt[3]

[1]Department of Zoology, Division of Systematics and Evolution, Stockholm University, Stockholm, Sweden
[2]Department of Zoology, Division of Ethology, Stockholm University, Stockholm, Sweden
[3]Department of Bioinformatics and Genetics, Swedish Museum of Natural History, Stockholm, Sweden
[4]Museum für Naturkunde, Leibniz Institut für Evolutions- und Biodiversitätsforschung, Berlin, Germany
[5]Institute for Biodiversity Science and Sustainability, California Academy of Sciences, San Francisco, CA, USA
[6]New Guinea Binatang Research Centre, Madang, Papua New Guinea

[*] Corresponding author
E-mail: ingo.muller@nrm.se (IM)

Abstract

New Guinea, the largest tropical island, is topographically complex and is dominated by a large central mountain range surrounded by multiple smaller isolated mountain regions along its perimeter. The island is biologically hyper-diverse and harbours an avifauna with many species found only there. The family Melampittidae is endemic to New Guinea and consists of two monotypic genera: *Melampitta lugubris* (Lesser Melampitta) and *Megalampitta gigantea* (Greater Melampitta). Both Melampitta species have scattered and disconnected distributions across New Guinea in the central mountain range and in some of the outlying ranges. While *M. lugubris* is common and found in most montane regions of the island, *M. gigantaea* is elusive and known from only six localities in isolated pockets on New Guinea with very specific habitats of limestone and sinkholes. In this project, we apply museomics to determine the population structure and demographic history of these two species. We re-sequenced the genomes of all seven known *M. gigantaea* samples housed in museum collections as well as 24 *M. lugubris* samples from across its distribution. By comparing population structure between the two species, we investigate to what extent habitat dependence, such as in *M. gigantaea*, may affect population connectivity. Phylogenetic and population genomic analyses, as well as acoustic differentiation, revealed that *M. gigantaea* consists of a single population in contrast to *M. lugubris* that shows much stronger population structure across the island. This work sheds new light on the mechanisms that have shaped the intriguing distribution of the two species within this family and is a prime example of the importance of museum collections for genomic studies of poorly known and rare species.

Introduction

What determines the build-up of biodiversity in space and through time is a long-standing question within biology. The accumulation of phenotypic and genetic differences between populations can only be generated through reproductive isolation that impedes genetic exchange between populations [1,2]. More explicitly, gene flow between diverging populations must be sufficiently limited so that genetic exchange does not exceed the accumulation of differentiation. Barriers underlying reproductive isolation, may differ markedly. They may be postzygotic and arise from genetic incompatibilities, which produce hybrid offspring that have either reduced fitness or are infertile [3–5]. Alternatively, barriers may be prezygotic and decrease the probability of mating events between populations, due to mating preferences, or through geographical (allopatric) or habitat barriers that separate different populations [4–6].

Mountains represent a classic example of geographical barriers both as physical barriers for populations but also because they harbour highly differentiated environments at different elevations. For sedentary lowland populations, mountains may represent unsurpassable barriers, which may over time lead to isolation and differentiation of separate lowland populations. Evidence for such montane barriers restricting gene flow between lowland populations are known from various organismal groups such as amphibians, spiders and coniferous trees [7–9]. Alternatively, extensive lowland valleys can also act as barriers to geneflow between populations adapted to high elevations. Lowland environments may be unsuitable for such mountain-adapted individuals, which over time become isolated on a series of mountaintops or “sky islands” [10,11] as known from some groups of birds, lizards and plants [12–16].

Related to this is the observation that older taxa are often found at higher elevations, while young lineages that are generally widespread, good dispersers and show little

differentiation inhabit the lowlands (e.g. 16–19). Such observations (mostly from island systems) have led to the formulation of the concept of taxon cycles, in which taxa pass through phases of expansions and contractions. The concept predicts that over time, taxa move into high elevation habitats either because they are outcompeted by new young taxa in their original (lowland) habitats or because they specialise and adapt to new environments at higher elevations [17,18,20–22].

Recent work on the New Guinean avifauna has provided empirical evidence in favour of species originating in the lowlands from where they move into the highlands over time and become relictual specialists [16,22,23], although some colonisation from mountaintop to mountaintop has also been shown to occur [15]. In addition, recent Pleistocene speciation events on New Guinea are mainly the result of changes in habitat distributions due to climate fluctuations, as this has caused species with continuous distributions to become geographically fragmented [24–26]. Pliocene speciation events, on the other hand, are driven mainly by geological processes such as montane uplift, which is known to have caused barriers to gene flow [27–30].

The Melampittidae represents an example of an old passerine family with only two deeply diverged species in monotypic genera. Their taxonomic affinities have been difficult to establish, but recent genetic results have placed the family as sister to crows (Corvidae) and shrikes (Laniidae) with an estimated divergence time from these at ca. 17 Mya [31]. One of the species, *Melampitta lugubris* (Lesser Melampitta) is relatively common at high elevations (1150-3500 m asl.), in accordance with the notion that older species tend to occupy higher elevations [21,22,32]. The other species, *Megalampitta gigantea* (Greater Melampitta) is only known from six localities at mid-elevations (650 -1400 m asl.) scattered across New Guinea. Based on few field observations, it is considered to be sedentary and to have limited flight capabilities [33,34]. Within its range, *M. gigantea* is associated with very specific karstic

habitats where it has been observed to spend considerable time nesting in narrow limestone sinkholes in which the birds have to climb in and out [33]. In contrast to *M. lugubris* the distribution of *M. gigantea* does not fit the general pattern that old taxa tend to occupy higher elevation.

In this genomic study we determine the population structure within *M. lugubris* and *M. gigantea* to understand how habitat connectivity across space and through time has shaped differentiation in these two species. Based on the contemporary distributions of the two species we hypothesize that:

1) The 7 individuals of *M. gigantea* represent several distinct evolutionary entities/populations, as the species is a poor disperser and has a fragmented distribution across New Guinea where it is associated with specific karst limestone habitat with sinkholes.
2) Individuals of *M. lugubris* represent a relatively cohesive group, yet with some population structure as deep lowland valleys may prevent gene flow between the various montane populations in the Central Range, the Huon mountains in the northeast and the Arfak mountains in the northwest.

Material & Methods

DNA sampling, sequencing and read processing

In this study, we follow the taxonomy of the IOC World Bird List [35]. We sampled 24 individuals of *Melampitta lugubris* of which 22 were footpads from museum specimens and two were fresh tissue samples. Additionally, we sampled 7 individuals of *Megalampitta gigantaea*, which represent all known samples present in museum collections. One of these

113 samples was a fresh blood sample. The rest were footpads from historical samples (for a
114 detailed list of samples and the museum collections in which they are stored see S1 Table). The
115 work is mainly based on old museum specimens for which the Nagoya Protocol does not apply.
116 The few fresh tissue samples included in the study are from already preserved samples at
117 natural history museums, for which all required permits are available.
118 DNA from fresh blood/tissue samples was extracted using Qiagen's DNeasy Blood and Tissue
119 kits. For DNA extraction and sequencing library preparation of historical samples, we followed
120 a modified version of Meyer and Kircher [36] that proved suitable for avian museum samples
121 [37]. In short, we extract DNA from toepad tissue mainly following the instructions from
122 Qiagen for animal tissue with the addition of Dithiothreitol (DTT) to improve the ligation yield.
123 During library preparation, we treat our samples with USER enzyme to reduce deamination
124 patterns that are typical for fragmented DNA from historical or ancient samples [38]. For a
125 detailed protocol see [37]. Whole genome re-sequencing was performed on Illumina NovaSeq
126 6000 machines on S4 flow cells going through 200 cycles with a read length of 2 x 100 bp at
127 the National Genomics Infrastructure (NGI) in Stockholm. Up to 24 samples with four libraries
128 each were multiplexed on a single flow cell lane.
129 Sequenced reads were then polished using the reproducible *Nextflow* workflow *nf-polish*
130 (https://github.com/MozesBlom/nf-polish) [39,40] (see S2 Table for specific github commits
131 used). This pipeline performs multiple polishing steps, including deduplication, adapter- and
132 quality-based trimming, read merging and the removal of low-complexity reads. Polished reads
133 were then mapped onto a reference genome using *nf-µmap* (https://github.com/IngoMue/nf-
134 umap) applying bwa-mem2 as mapping algorithm and which also allowed us to evaluate the
135 quality control after mapping and investigate damage patterns that are typical for historical
136 DNA [41,42]. We used the hooded crow (*Corvus cornix*, Refseq GCF_000738735.5, [43]) as

137 our reference genome as it represents a reasonable closely related species with a high-quality
138 chromosome level assembly [31].

139 Phylogenetic analyses on mitochondrial and nuclear DNA

140 To assemble the full mitogenome from our polished reads we used nf_mito-mania with default
141 settings (https://github.com/FilipThorn/nf_mito-mania) [44]. Variant calling implemented in
142 this pipeline filters sites with a depth-of-coverage below 20 or above three times the average
143 depth-of-coverage across the whole mitogenome of each individual. The resulting consensus
144 sequences of every individual were aligned using *MAFFT* (v7.407, [45], see S2 File for specific
145 flags). An occasional artefact of Mitobim where mitochondrial assemblies become longer than
146 they are supposed to be resulted in overhanging sequences in some individuals. These
147 overhangs were then cut out of the alignment after visual inspection using *Geneious Prime*
148 *2023.0.4* so that the final alignment consisted only of overlapping reads (total length 17 112 bp
149 including gaps), which were then used as input for *RAxML-NG* (v1.1.0, [46] using the GTR+G
150 substitution model, 100 bootstrapping replicates and ten randomized parsimony starting trees
151 to generate a mitochondrial phylogenetic tree. Mitochondrial assemblies were also forced into
152 diploid variant calls to check for contamination in our samples. As mitochondria are haploid,
153 heterozygote sites are not expected and could therefore be indicative of cross-contamination.
154 For the nuclear phylogenetic tree, we used the previously mapped .bam files excluding
155 individuals with very low mean depth-of-coverage (n = 3, DoC < 4 *x*) to call variants for each
156 individual using *freebayes* (v1.3.1-dirty, [47]). Polymorphic sites were filtered based on their
157 quality score (> 20), allelic balance (≥ 0.2), and minimum and maximum depth-of-coverage (3
158 *x* / 100 *x*). We also decomposed multiple nucleotide polymorphisms (MNPs) into single
159 nucleotide polymorphisms (SNPs) and masked heterozygous positions and indels. These
160 filtered .vcf files were then used as input files for the reproducible *Nextflow* workflow *nf-phylo*
161 with default settings (https://github.com/MozesBlom/nf-phylo) [48] to generate both

162 concatenated (*IQtree2*) and summary coalescent (*ASTRAL3*) species trees based on 'gene' trees
163 of different window-sizes (2000, 5000, 10000, 20000 base pairs).

164 Population structure and differentiation

165 To quantify population substructure and to estimate levels of differentiation between samples,
166 we used a genotype likelihood approach as implemented in ANGSD (v0.938) as this is better
167 suited for low coverage data [49]. Specific commands and the filters used are explained in the
168 S2 File. The filters we used for admixture and principal component analyses (PCAs) were
169 slightly different from those used to calculate nucleotide diversity, heterozygosity and Tajima's
170 D. As we did not have an ancestral genome available, we used the reference genome as
171 ancestral sequence and folded the site frequency spectra (SFS). PCAs were performed through
172 *PCAngsd* [50] and plotted with custom R scripts through RStudio (v 2023.03.0 build 386, R
173 version 4.1.1, [51,52]). Admixture analyses to determine population structure was run through
174 *NgsAdmix* [53] running up to K = 10 with ten replicates for each K and visualised with custom
175 R scripts. Individual heterozygosity was estimated by generating a site frequency spectrum for
176 each individual and dividing the number of sites with one derived allele divided by the total
177 number of sites as performed by e.g. Hansen et al. [54]. Using SFS for each species, nucleotide
178 diversity and Tajima's D were both estimated for each chromosome as well as in 20 kb
179 windows sliding in steps of 10 kb using the *thetaStat* command. We divided the pairwise theta
180 estimator (*tP*) by the total number of Sites (*nSites*) of each chromosome/window to calculate
181 nucleotide diversity. Statistical significance of differences in heterozygosity and nucleotide
182 diversity between the two species was checked using Welch's t-test after verifying normal
183 distributions and inequal variances within the data.

184 ## Estimation of effective population sizes through time and
185 divergence times

186 To estimate effective population sizes through time, we used pairwise sequentially Markovian
187 coalescent (PSMC) [55] (for details on the method see S1 file). As an estimate of the neutral
188 genomic mutation rate per generation we used $4.6*10^{-9}$ as obtained in a study of the collared
189 flycatcher *Ficedula albicollis* [56]. We set the estimated generation time for *M. lugubris* to
190 3.90 years and for *M. gigantea* to 4.58 years [57]. The parameters for the PSMC analysis were
191 set to " -N30 -t5 -r5 -p 4 + 30*2 + 4 + 6 + 10" following Nadachowska-Brzyska et al. [58]. The
192 authors observed no significant change in curve shape when modifying the atomic vectors
193 parameter (-p) and applied the same settings to several different avian species. We only ran
194 PSMCs for the two samples of *M. gigantea* with the highest depth-of-coverage and for each of
195 the five identified clusters within *M. lugubris* (West, Central, East, Huon, Southeast). False
196 negative rates (FNRs) were adjusted based on depth-of-coverage. If depth-of-coverage was
197 higher than 15 X, FNR was kept at 0. However, if individual A had a depth-of-coverage higher
198 than 15 X and individual B had a depth-of-coverage below 15 X, then individual B would have
199 an FNR of $0.1 + 0.1 * x$, where x is the depth-of-coverage of individual A divided by depth-
200 of-coverage of individual B. If both individuals had a depth-of-coverage < 15 X, we used 0.1
201 for both individuals.

202 To estimate divergence times between the two species, but also between the different
203 subpopulations of *M. lugubris*, we first ran F_1-hybrid PSMC (hPSMC, [59]) using the same
204 parameters as for the previous PSMC analyses and implementing 100 bootstraps replicates.
205 Additionally, we estimated mitochondrial divergences within and between the two species and
206 subpopulations of *M. lugubris* using the previously generated mitochondrial alignments. We
207 applied the simple "2% rule" which assumes that in birds, the average divergence between two

species is 2% per million years [60]. We compared these divergence time estimates with those obtained from the hPSMC analyses.

Acoustic recordings and analysis

Acoustic recordings of 10 *M. gigantea* individuals and 28 *M. lugubris* individuals were obtained from an online repository of avian vocalizations (https://xeno-canto.org/), which covered different locations across New Guinea. We included all types of vocalisations- songs, calls and vocalisations of an unknown type in the analysis, unless the function of the vocalisation was specified by the recordist (eg: alarm). This is due to the high uncertainty in estimating the type of vocalisation in *M. lugubris,* and visual comparison between vocalisations classified as 'songs' versus 'calls' between individuals recorded in the same location, often showed that they were the same. The vocalisations of each individual (median = 9 vocalisations/individual) were measured by a single author (SR) using the Luscinia sound analysis program (version 2.17.11.22.01, [61]).

Each vocalisation was visualised using a Gaussian windowing function with the following spectrogram settings: 13 kHz maximum frequency, 5 ms frame length, 221 spectrograph points, 80% spectrograph overlap, 80 dB dynamic range, 30% dereverberation, and 50 ms of dereverberation range. Elements were measured as continuous sound traces and then grouped into syllables within each vocalisation (each vocalisation contained only one syllable).

The vocalisations were then compared using the dynamic time warping algorithm (DTW) in Luscinia, following the same settings used in Wheatcroft et al. [62] that has provided reliable grouping outputs for other songbird species. The final output of the DTW analysis was an acoustic dissimilarity matrix, from which we extracted 10 principal components using nonmetric multidimensional scaling.

231 # Results

232 Our evaluation of mapped reads against the *Corvus cornix* genome showed a median depth-of-
233 coverage (DoC) of ~ 9.252 *x* (min: 0.026 *x*, max: 30.935 *x*, SD: 7.444) and a median percentage
234 of mapped reads at 89.5 % (min 0.2 %:, max: 95.8 %, SD: 23.693). Detailed values for each
235 individual are listed in S1 Table.
236 During contamination control using mitochondrial assemblies, we observed an increased
237 amount of heterozygote sites across the libraries in 6 individuals (S3 Table). Upon manual
238 inspection using *Geneious Prime* we found that these heterozygote positions mostly appear in
239 blocks and often within the same regions. This suggests that they were in fact nuclear
240 mitochondrial sequences (NUMTs) that were wrongly mapped onto the mitochondrial genome
241 instead of being a result of contamination. We also observed that non-reference alleles often
242 appeared at a lower frequency (98.093 % of heterozygote sites had a reference allele frequency
243 > 0.5, median reference allele frequency across all heterozygote sites at 0.874) and therefore
244 disappeared during consensus calling, as the more frequent allele gets chosen during this step.
245 Nonetheless, we manually excluded two regions from all samples with blocks (in total 5 700
246 bp out of the entire alignment's 17 112 bp) of heterozygote sites shared across the majority of
247 individuals. The remaining 11 412 bp were used to generate the mitochondrial phylogenies.

248 ## Phylogenetic analyses on mitochondrial and nuclear DNA

249 We found high congruence between phylogenies built from mitochondrial and nuclear
250 genomes (Fig. 1 A and S1 Fig.). Different window sizes and summary coalescent vs
251 concatenated nuclear phylogenies also had little effect on the topology. We recovered three
252 main clusters within *M. lugubris* that correspond to the geographic location of the samples on
253 an east to west axis (Fig. 1). These clusters also align with previously described subspecies of
254 *M. lugubris* [34]. The first cluster within *M. lugubris* consists of individuals inhabiting the

255 Birds-Head of north-western New Guinea as well as an individual in the westernmost part of
256 the Central Range. The next cluster inhabits the western and central parts of the Central Range
257 of New Guinea, and the third cluster inhabits the eastern and south-eastern section of the
258 Central Range as well as the isolated outlying Huon mountains. *M. gigantaea*, on the other
259 hand, shows little differentiation between individuals compared to *M. lugubris*. Relationships
260 within *M. gigantaea* are also in accordance with the geographical locality of the samples.

261 **Fig. 1. A Distribution map and sampling sites** of Megalampitta gigantaea (orange distribution) and
262 M. lugubris (blue distribution), subclusters of M. lugubris are also coloured differently. Shapefiles for
263 administrative boundaries were obtained from geoBoundaries *[63]*, the map was created using QGIS
264 *[64]*. **B Nuclear phylogeny of Melampittidae** (based on concatenated 5 kbp window alignments)
265 highlighting the subdivisions within M. lugubris (West, Central, East, Huon, Southeast), support values
266 next to the main branches show Bootstraps/site concordance factors (sCF)/window concordance factors
267 (wCF).

268 Population structure and differentiation

269 We recover the same pattern of lower levels of differentiation in *M. gigantaea* compared to *M.*
270 *lugubris* in the PCAs (Fig. 2 A), Heterozygosity (Fig. 2 B), nucleotide diversity (π, S4 Table,
271 S2 Fig.) and admixture (Fig. 2 C). For *M. lugubris* Tajima's D was consistently negative with
272 a mean value of -0.902 (SD: 0.126, median: -0.897, S5 Table, S3 Fig.) which is indicative of
273 either population expansion or a selective sweep. In *M. gigantaea* values for Tajima's D were
274 slightly above zero in the range of 0 – 0.2 (mean 0.103, SD: 0.040, median: 0.112, S5 Table,
275 S3 Fig.). Positive values of Tajima's D could indicate a reduction in population size or
276 balancing selection acting, however as the values are so close to zero the population may just
277 evolve neutrally. In the PCA (Fig. 2 A), PC1 separates the two species, afterwards *M. gigantaea*
278 remains closely clustered up to PC4, while subgroups corresponding to geographic localities
279 make up clusters within *M. lugubris* (S4 Fig.). The two distinct clusters of *M. lugubris* on PC2
280 (Fig. 2 A) separate eastern New Guinean populations and northwestern New Guinean
281 populations as also observed in the phylogenetic tree (Fig. 1 B). Both heterozygosity and
282 nucleotide diversity were significantly lower in *M. gigantaea* than in *M. lugubris*. Although we

observed a clear trend of increasing heterozygosity with higher depth-of-coverage, the slopes for each population were similar and consistently higher in all but one population of *M. lugubris* (S5 Fig.). Admixture analysis revealed no substructure within *M. gigantaea* from K=2 to K = 7. For *M. lugubris* the observed clusters between K = 2-6 align with the clusters observed in the phylogenetic trees and in the PCAs. Further subdivisions within the main clusters of *M. lugubris* at higher values of K are also corresponding to the populations' geographical location.

Fig. 2. Genetic differentiation in Melampittidae. A) PCA showing the first two principal components for both species. Subclusters of Melampitta lugubris are also coloured differently. **B)** Admixture analysis from K = 1 to K = 6 **C)** Heterozygosity for all individuals between both species.

Estimation of effective population size in time and divergence times

PSMC curves (Fig. 3) for samples from the same populations had similar shapes, but not entirely overlapping as depth-of-coverage varied between samples. Within *M. lugubris* the shape of the curves varied, but most of this variation could be ascribed to population specific events. In *M. gigantaea*, we observe an effective population size peak at around 200 Kya followed by a steady decline in effective population size up until around 40 Kya.

The divergence time obtained from hPSMCs curves for the split between *M. gigantaea* and *M. lugubris* was estimated to about 10 mya (S6 Fig.). Splits between subgroups within *M. lugubris* were estimated more recently with the split between Western+Vogelkop and Eastern populations at around 4-5 mya (S7 Fig.) and between Western and Vogelkop populations at about 3-4 mya (S8 Fig.). The next divisions within Eastern *M. lugubris* populations (East, Huon and Southeast) happened at similar times around 1 mya (S8 and S9 Figs.).

Fig. 3. PSMC plots for two representative individuals of **A)** Megalampitta gigantaea **B)** Melampitta lugubris western population **C)** M. lugubris central population **D)** M. lugubris eastern population and Huon, colours in **D)** represent the Huon (blue) and (South)Eastern (purple) subclusters.

Mitochondrial divergences were, as expected, highest for comparisons between the *M. gigantaea* and *M. lugubris* and its subpopulations at a range of 9.8 – 13.3 %. Divergence within *M. gigantaea* was also lower (mean 0.912 %) than within *M. lugubris* (mean 4.979 %) or even in some of its subpopulations. For an extensive table with all comparisons of mitochondrial divergence see S11 and S12 Figs.. Divergence times obtained through the 2% rule were 4.9 – 6.6 mya for the split between *M. gigantaea* and *M. lugubris*, 3.7 – 5.6 mya for splits between Western+Vogelkop populations from Eastern populations of *M. lugubris* and Western populations from Vogelkop populations at 3.6 - 3.8 mya. Subdivisions within the Eastern populations were estimated at 0.2 – 2.7 mya.

Acoustic recordings and analysis

The first ten principal components collectively explained 97% of the variation in vocalisations across the two *Melampitta* species. PC1, which explained 44.5% of the variation in all vocalisations, was more varied for *M. lugubris* (standard deviation (SD) = 0.086) compared to *M. gigantaea* (SD = 0.036). The same was true for PC2, where the standard deviation was once again higher for *M. lugubris* (SD = 0.11) compared to *M. gigantaea* (SD = 0.017). These results show that *M. lugubris* has greater acoustic diversity than *M. gigantaea* (Fig. 4).

Fig. 4. Acoustic variation across species: Principal component space (PC1-2) of vocalisations from M. gigantaea and M. lugubris. PC1 and PC2 scores are averaged within individuals and triangles represent species centroids. Ellipses contain 95% of vocalisations of each species. The Significant outlier within M. lugubris may represent an odd vocalisation that is not directly comparable with the other vocalisations included here. Note that vocalisations for M. gigantaea were only available from three localities (the Fakfak mountains in the Bird's Neck, a locality in the southern Bird's Neck and Tabubil in the central highlands) and vocalisations for M. lugubris were only available from two of the three distinct clades (samples were available from the western and central but no vocalisation data was available from the eastern and Huon populations).

335 # Discussion

336 The formation of the avifauna on New Guinea largely follows the predictions of taxon cycles
337 [15,16] whereby new species form in or colonise through the lowlands and over time move
338 upwards and become relictual at high elevations. The family Melampittidae is a species-poor
339 old endemic lineage of New Guinea [31]. The family includes two extant species of which one
340 (*Melampitta lugubris*) follows the general taxon-cycle expectation in that it is an old lineage
341 that inhabits montane forests of New Guinea. The other species, *Megalampitta gigantea,*
342 however, has a distribution associated with specific karst habitats at lower elevations and in
343 foothills [34].

344 Our divergence time estimates suggest that *M. gigantea* and *M. lugubris* diverged from
345 each other in the Miocene (at approximately 10 Mya based on hPSMC results), which is
346 slightly younger than the divergence time estimated by Jønsson et al.[65] and slightly older
347 than the divergence time estimated by McCullough et al. [66]. The three main populations of
348 *M. lugubris* (Fig. 1 B) diverged from each other in the early Pliocene (at approximately 4-5
349 Mya based on hPSMC curves). A Pliocene divergence of *M. lugubris* populations coincides
350 with major uplift of various mountain regions on New Guinea [67–69], which may have shaped
351 the present population structure of *M. lugubris*. The distributional pattern of populations of *M.*
352 *lugubris*, with one distinct Vogelkop population and a division of an eastern and a western
353 population along the central mountain range, is also a pattern similar to that found in other New
354 Guinean mountain birds with Pliocene divergences [30,70,71].

355 The PSMC curves of the three main populations of *M. lugubris* differ (Fig. 3), yet with
356 a general trend of increasing population sizes towards the present. The exception to this is the
357 population of the Huon mountains, which shows a continuous decrease in population size since
358 approximately 100 Kya. Our interpretation is that eastern and south-eastern populations of the

359 Central Range have maintained continuous gene flow, while the connectivity with the Huon
360 population was broken or at least severely reduced as this population became isolated in the
361 outlying Huon Mountain range.
362
363 Given a presumed poor dispersal capacity [33] and a patchy distribution at mid-elevations, we
364 initially hypothesised that *M. gigantea* would exhibit a clear population structure. However,
365 contrary to expectations, all our samples of *M. gigantea*, from localities scattered across New
366 Guinea, cluster tightly together genetically (Fig. 1 and Fig**. 2**). Analysis of vocalisations also
367 shows a similar pattern, as *M. gigantea* compared to *M. lugubris* exhibits less vocal
368 differentiation (Fig. 4). This is fascinating and difficult to explain. Below, we discuss three
369 scenarios that may provide possible explanations for these patterns. First, it is possible that
370 continuous migration (or high rates of juvenile dispersal) of *M. gigantea* individuals maintains
371 contact and gene flow between populations. However, an exclusively ground-dwelling lifestyle
372 and the lack of long-distance flight capabilities, suggested by its morphology and field
373 observations contradict this scenario [33,34]. Second, it is possible that their presently known
374 fragmented distribution does not properly reflect their actual distribution, which may be more
375 extensive [34,72]. Karst regions are generally species-poor in comparison to the species-rich
376 tropical forests of New Guinea and such localised karstic areas dispersed throughout New
377 Guinea may therefore have commanded less attention by ornithological surveys. Finally, it is
378 possible that *M. gigantea* once had a wider more continuous distribution and that a recent
379 decline has left scattered populations in small pockets of Karst habitat. The PSMC analyses
380 support this scenario by showing that the population size of *M. gigantea* has dropped
381 dramatically within the last 200 Ky (Fig. 3). The fact that *M. gigantea* is highly adapted to a
382 very specific habitat type (nesting in deep holes in karst limestone that they have to climb out
383 of [33] is, however, difficult to reconcile with this scenario. However, one may speculate that

M. gigantea in the past had broader habitat preferences not only restricted to the present karst limestone habitats. Perhaps during the last 200 Ky, increased competition from other species forced *M. gigantea* to retract to a particular low-diversity habitat type, leaving behind the scattered distribution that we see today. Overall, we find it most plausible, that *M. gigantea* had a larger and more continuous distribution in the past, yet we acknowledge that the present distribution may be underestimated. Additional ornithological surveys to suitable habitats may, thus, reveal further *M. gigantea* populations.

Conclusions

In this study, the rather surprising population structure of the two species of an old New Guinean avian family have been elucidated by genomic data largely obtained from historical museum collections. While the population structure of *Melampitta lugubris* is similar to those found in other mountain birds of New Guinea with similar age, the population structure of *Megalampitta gigantea* is intriguing. The study is an example of how intrinsic properties, such as those exhibited by *M. gigantea*, may cause their population dynamics to deviate from general biogeographical predictions. The study is also an example of how important museum collections are for increasing the knowledge of rare taxa that occur in remote regions. The levels of divergence between the three major populations of *M. lugubris* are well above those at which ornithologists would normally assign species rank. Consequently, we tentatively propose that these three populations should be elevated to species rank, *M. lugubris* (Schlegel, 1871) in the Vogelkop region, *Melampitta rostrata* (Ogilvie-Grant, 1913) in the western central range and *Melampitta longicauda* (Mayr & Gilliard, 1952) in the eastern central range.

Acknowledgements

Computations were enabled by resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS) and the Swedish National Infrastructure for Computing (SNIC) at UPPMAX partially funded by the Swedish Research Council through grant agreements no. 2022-06725 and no. 2018-05973. Furthermore, the authors acknowledge support from the National Genomics Infrastructure in Stockholm funded by Science for Life Laboratory, the Knut and Alice Wallenberg Foundation and the Swedish Research Council, and SNIC/Uppsala Multidisciplinary Center for Advanced Computational Science for assistance with massively parallel sequencing and access to the UPPMAX computational infrastructure. Lastly, we would like to thank Nikolay Oskolkov for assistance with bioinformatic analyses as part of the drop-in services and the Swedish Bioinformatics Advisory Program of the National Bioinformatics Infrastructure Sweden (NBIS).

References

1. Dobzhansky TG. Genetics of the evolutionary process. Vol. 139. Columbia University Press; 1970.
2. Westram AM, Stankowski S, Surendranadh P, Barton N. What is reproductive isolation? J Evol Biol. 2022;35(9):1143–64.
3. Dobzhansky TG. Studies on Hybrid Sterility. II. Localization of Sterility Factors in Drosophila Pseudoobscura Hybrids. Genetics. 1936 Mar;21(2):113–35.
4. Coyne JA. Genetics and speciation. Nature. 1992 Feb;355(6360):511–5.
5. Turelli M, Barton NH, Coyne JA. Theory and speciation. Trends Ecol Evol. 2001 Jul;16(7):330–43.
6. North HL, Caminade P, Severac D, Belkhir K, Smadja CM. The role of copy-number variation in the reinforcement of sexual isolation between the two European subspecies of the house mouse. Philos Trans R Soc B Biol Sci. 2020 Aug 31;375(1806):20190540.
7. Sánchez-Montes G, Wang J, Ariño AH, Martínez-Solano Í. Mountains as barriers to gene flow in amphibians: Quantifying the differential effect of a major mountain ridge on the genetic structure of four sympatric species with different life history traits. J Biogeogr. 2018;45(2):318–31.
8. Salgado-Roa FC, Pardo-Diaz C, Lasso E, Arias CF, Solferini VN, Salazar C. Gene flow and Andean uplift shape the diversification of Gasteracantha cancriformis (Araneae: Araneidae) in Northern South America. Ecol Evol. 2018;8(14):7131–42.

9. Li YS, Shih KM, Chang CT, Chung JD, Hwang SY. Testing the Effect of Mountain Ranges as a Physical Barrier to Current Gene Flow and Environmentally Dependent Adaptive Divergence in Cunninghamia konishii (Cupressaceae). Front Genet [Internet]. 2019 [cited 2023 Jan 19];10. Available from: https://www.frontiersin.org/articles/10.3389/fgene.2019.00742
10. Heald WF. Sky Island [Internet]. Van Nostrand; 1967. Available from: https://books.google.se/books?id=jPWeuQAACAAJ
11. McCormack JE, Huang H, Knowles LL, Gillespie R, Clague D. Sky islands. Encycl Isl. 2009;4:841–3.
12. Robin VV, Sinha A, Ramakrishnan U. Ancient Geographical Gaps and Paleo-Climate Shape the Phylogeography of an Endemic Bird in the Sky Islands of Southern India. PLOS ONE. 2010 Oct 13;5(10):e13321.
13. Wagner P, Greenbaum E, Malonza P, Branch W. Resolving sky island speciation in populations of East African Adolfus alleni (Sauria, Lacertidae). Salamandra. 2014;50(1):1–17.
14. Zhang D, Hao G, Guo X, Hu Q, Liu J. Genomic insight into "sky island" species diversification in a mountainous biodiversity hotspot. J Syst Evol. 2019 Nov;57(6):633–45.
15. Pujolar JM, Blom MPK, Reeve AH, Kennedy JD, Marki PZ, Korneliussen TS, et al. The formation of avian montane diversity across barriers and along elevational gradients. Nat Commun. 2022 Dec;13(1):268.
16. Kennedy JD, Marki PZ, Reeve AH, Blom MPK, Prawiradilaga DM, Haryoko T, et al. Diversification and community assembly of the world's largest tropical island. Glob Ecol Biogeogr. 2022;31(6):1078–89.
17. Wilson EO. Adaptive Shift and Dispersal in a Tropical Ant Fauna. Evolution. 1959;13(1):122–44.
18. Wilson EO. The Nature of the Taxon Cycle in the Melanesian Ant Fauna. Am Nat. 1961;95(882):169–93.
19. Ricklefs RE, Bermingham E. The concept of the taxon cycle in biogeography. Glob Ecol Biogeogr. 2002;11(5):353–61.
20. MacArthur RH, Wilson EO. The Theory of Island Biogeography. REV-Revised. Princeton University Press; 1967.
21. Ricklefs RE, Cox GW. Taxon Cycles in the West Indian Avifauna. Am Nat. 1972 Mar;106(948):195–219.
22. Pepke ML, Irestedt M, Fjeldså J, Rahbek C, Jønsson KA. Reconciling supertramps, great speciators and relict species with the taxon cycle stages of a large island radiation (Aves: Campephagidae). J Biogeogr. 2019 Jun;46(6):1214–25.
23. Reeve AH, Kennedy JD, Pujolar JM, Petersen B, Blom MPK, Alström P, et al. The formation of the Indo-Pacific montane avifauna [Internet]. bioRxiv; 2022 [cited 2023 Jan 26]. p. 2022.06.15.495085. Available from: https://www.biorxiv.org/content/10.1101/2022.06.15.495085v1
24. Williams DJ, O'Shea M, Daguerre RL, Pook CE, Wüster W, Hayden CJ, et al. Origin of the eastern brownsnake, Pseudonaja textilis (Duméril, Bibron and Duméril) (Serpentes: Elapidae: Hydrophiinae) in New Guinea: evidence of multiple dispersals from Australia, and comments on the status of Pseudonaja textilis pughi Hoser 2003. Zootaxa. 2008 Feb 13;(1703):47–61.
25. Kearns AM, Joseph L, Omland KE, Cook LG. Testing the effect of transient Plio-Pleistocene barriers in monsoonal Australo-Papua: did mangrove habitats maintain genetic connectivity in the Black Butcherbird? Mol Ecol. 2011;20(23):5042–59.

26. Ericson PGP, Irestedt M, She H, Qu Y. Genomic signatures of rapid adaptive divergence in a tropical montane species. Biol Lett. 2021 Jul;17(7):20210089.
27. Rawlings LH, Donnellan SC. Phylogeographic analysis of the green python, Morelia viridis, reveals cryptic diversity. Mol Phylogenet Evol. 2003 Apr 1;27(1):36–44.
28. Dumbacher JP, Fleischer RC. Phylogenetic evidence for colour pattern convergence in toxic pitohuis: Müllerian mimicry in birds? Proc R Soc Lond B Biol Sci. 2001 Oct 7;268(1480):1971–6.
29. Deiner K, Lemmon AR, Mack AL, Fleischer RC, Dumbacher JP. A Passerine Bird's Evolution Corroborates the Geologic History of the Island of New Guinea. PLOS ONE. 2011 May 6;6(5):e19479.
30. Irestedt M, Batalha-Filho H, Roselaar CS, Christidis L, Ericson PGP. Contrasting phylogeographic signatures in two Australo-Papuan bowerbird species complexes (Aves: Ailuroedus). Zool Scr. 2016;45(4):365–79.
31. Oliveros CH, Field DJ, Ksepka DT, Barker FK, Aleixo A, Andersen MJ, et al. Earth history and the passerine superradiation. Proc Natl Acad Sci. 2019 Apr 16;116(16):7916–25.
32. Price TD, Hooper DM, Buchanan CD, Johansson US, Tietze DT, Alström P, et al. Niche filling slows the diversification of Himalayan songbirds. Nature. 2014 May 8;509(7499):222–5.
33. Diamond JM. Melampitta gigantea: Possible Relation between Feather Structure and Underground Roosting Habits. The Condor. 1983 Feb;85(1):89–91.
34. Del Hoyo J, Elliot A, Christie D. Handbook of the Birds of the World, Picathartes to Tits and Chickadees, vol. 12. Lynx Edicions Barc. 2007;
35. Gill F, Donsker DB, Rasmussen PC. IOC World Bird List 13.1 [Internet]. World Bird Names; 2023 [cited 2023 Mar 30]. Available from: https://www.worldbirdnames.org/ioc-lists/crossref
36. Meyer M, Kircher M. Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. Cold Spring Harb Protoc. 2010 Jun 1;2010(6):pdb.prot5448-pdb.prot5448.
37. Irestedt M, Thörn F, Müller IA, Jønsson KA, Ericson PGP, Blom MPK. A guide to avian museomics: Insights gained from resequencing hundreds of avian study skins. Mol Ecol Resour. 2022;22(7):2672–84.
38. Briggs AW, Stenzel U, Meyer M, Krause J, Kircher M, Pääbo S. Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. Nucleic Acids Res. 2010 Apr;38(6):e87–e87.
39. Di Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C. Nextflow enables reproducible computational workflows. Nat Biotechnol. 2017 Apr;35(4):316–9.
40. Blom MPK. nf-polish [Internet]. 2021. Available from: https://github.com/MozesBlom/nf-polish
41. Müller IA. nf-µmap [Internet]. 2022. Available from: https://github.com/IngoMue/nf-umap
42. Vasimuddin Md, Misra S, Li H, Aluru S. Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. In: 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS). 2019. p. 314–24.
43. Poelstra JW, Vijay N, Bossu CM, Lantz H, Ryll B, Müller I, et al. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. Science. 2014 Jun 20;344(6190):1410–4.
44. Thörn F. nf_mito-mania [Internet]. 2022 [cited 2023 Mar 30]. Available from: https://github.com/FilipThorn/nf_mito-mania

45. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–80.
46. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics. 2019 Nov 1;35(21):4453–5.
47. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing [Internet]. arXiv; 2012 [cited 2023 Sep 8]. Available from: http://arxiv.org/abs/1207.3907
48. Blom MPK. nf-phylo [Internet]. 2023 [cited 2023 Mar 30]. Available from: https://github.com/MozesBlom/nf-phylo
49. Korneliussen TS, Albrechtsen A, Nielsen R. ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics. 2014 Nov 25;15(1):356.
50. Meisner J, Albrechtsen A. Inferring Population Structure and Admixture Proportions in Low-Depth NGS Data. Genetics. 2018 Oct 1;210(2):719–31.
51. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2021. Available from: https://www.R-project.org/
52. Posit team. RStudio: Integrated Development Environment for R [Internet]. Boston, MA: Posit Software, PBC; 2023. Available from: http://www.posit.co/
53. Skotte L, Korneliussen TS, Albrechtsen A. Estimating Individual Admixture Proportions from Next Generation Sequencing Data. Genetics. 2013 Sep;
54. Hansen CCR, Hvilsom C, Schmidt NM, Aastrup P, Van Coeverden De Groot PJ, Siegismund HR, et al. The Muskox Lost a Substantial Part of Its Genetic Diversity on Its Long Road to Greenland. Curr Biol. 2018 Dec;28(24):4022-4028.e5.
55. Li H, Durbin R. Inference of human population history from individual whole-genome sequences. Nature. 2011 Jul;475(7357):493–6.
56. Smeds L, Qvarnström A, Ellegren H. Direct estimate of the rate of germline mutation in a bird. Genome Res. 2016 Jan 9;26(9):1211–8.
57. Bird JP, Martin R, Akçakaya HR, Gilroy J, Burfield IJ, Garnett ST, et al. Generation lengths of the world's birds and their implications for extinction risk. Conserv Biol. 2020;34(5):1252–61.
58. Nadachowska-Brzyska K, Li C, Smeds L, Zhang G, Ellegren H. Temporal Dynamics of Avian Populations during Pleistocene Revealed by Whole-Genome Sequences. Curr Biol. 2015 May 18;25(10):1375–80.
59. Cahill JA, Soares AER, Green RE, Shapiro B. Inferring species divergence times using pairwise sequential Markovian coalescent modelling and low-coverage genomic data. Philos Trans R Soc B Biol Sci. 2016 Jul 19;371(1699):20150138.
60. Weir JT, Schluter D. Calibrating the avian molecular clock. Mol Ecol. 2008;17(10):2321–8.
61. Lachlan R. rflachlan/Luscinia [Internet]. 2014. Available from: https://github.com/rflachlan/Luscinia
62. Wheatcroft D, Bliard L, El Harouchi M, López-Idiáquez D, Kärkkäinen T, Kraft FLH, et al. Species-specific song responses emerge as a by-product of tuning to the local dialect. Curr Biol. 2022 Dec 5;32(23):5153-5158.e5.
63. Runfola D, Anderson A, Baier H, Crittenden M, Dowker E, Fuhrig S, et al. geoBoundaries: A global database of political administrative boundaries. PLOS ONE. 2020 Apr 24;15(4):e0231866.
64. QGIS Development Team. QGIS Geographic Information System [Internet]. QGIS Association; 2023. Available from: https://www.qgis.org

65. Jønsson KA, Fabre PH, Kennedy JD, Holt BG, Borregaard MK, Rahbek C, et al. A supermatrix phylogeny of corvoid passerine birds (Aves: Corvides). Mol Phylogenet Evol. 2016 Jan 1;94:87–94.
66. McCullough JM, Oliveros CH, Benz BW, Zenil-Ferguson R, Cracraft J, Moyle RG, et al. Wallacean and Melanesian Islands Promote Higher Rates of Diversification within the Global Passerine Radiation Corvides. Ho S, editor. Syst Biol. 2022 Oct 12;71(6):1423–39.
67. Pigram C, Davies H. Terranes and the accretion history of the Papua New Guinea orogeny. BMR J Aust Geol Geophys. 1987;10:193–212.
68. Hall R. Cenozoic geological and plate tectonic evolution of SE Asia and the SW Pacific: computer-based reconstructions, model and animations. J Asian Earth Sci. 2002 Apr 1;20(4):353–431.
69. Davies HL. The geology of New Guinea-the cordilleran margin of the Australian continent. Episodes J Int Geosci. 2012;35(1):87–102.
70. Irestedt M, Batalha-Filho H, Ericson PGP, Christidis L, Schodde R. Phylogeny, biogeography and taxonomic consequences in a bird-of-paradise species complex, Lophorina–Ptiloris (Aves: Paradisaeidae). Zool J Linn Soc. 2017 Oct 1;181(2):439–70.
71. Jønsson KA, Reeve AH, Blom MP, Irestedt M, Marki PZ. Unrecognised (species) diversity in New Guinean passerine birds. Emu-Austral Ornithol. 2019;119(3):233–41.
72. Diamond J, Bishop KD. What's so special about New Guinea birds? Bull Br Ornithol Club. 2023 Jun;143(2):212–36.
73. Kendall MG. Rank correlation methods. 1948;

Supporting information

S1 Figure. Mitochondrial phylogeny for all individuals

S2 Figure. Individual nucleotide diversity (π) is significantly lower in M. gigantaea (orange) than in M. lugubris (blue). The applied statistical test was a Welch's two sample t-test for unequal variances.

S3 Figure. Tajima's D is consistently negative across all chromosomes in M. lugubris (blue) and slightly positive in M. gigantaea.

S4 Figure. Principal components 1 to 4 describing divisions within subpopulations of M. lugubris (shades of blue/purple) while M. gigantaea (orange) remains a tight cluster.

S5 Figure. Heterozygosity shows a correlation with increasing depth of coverage (DoC). Slopes are similar between populations/species and M. gigantaea still shows lower Heterozygosity than most M. lugubris populations when comparing individuals with similar DoC. To fit regression lines, we applied Kendall's rank correlation coefficient as it is recommended for smaller sample sizes containing outliers [73].

S6 Figure. PSMC for M. lugubris (blue) and M. gigantaea (orange) and their hybrid PSMC curve (red) to show the divergence time between the two species.

S7 Figure. PSMC for M. lugubris NHMD616019 from Huon (red) and M. lugubris B98503 from Central New Guinea (blue) and their hybrid PSMC curve (purple) to show the divergence time between Eastern populations and Western + Central populations of M. lugubris.

S8 Figure. PSMC for M. lugubris B98503 from Central New Guinea (red) and M. lugubris AMNH293751 from Western New Guinea (blue) and their hybrid PSMC curve (purple) to show the divergence time between Central populations and Western populations of M. lugubris.

S9 Figure. PSMC for M. lugubris NHMD616019 from Huon (red) and M. lugubris AMNH590750 from the Southeast (blue) and their hybrid PSMC curve (purple) to show the divergence time between Huon populations and Southeastern populations of M. lugubris.

S10 Figure. PSMC for M. lugubris NHMD616019 from Huon (red) and M. lugubris B100613 from the East (blue) and their hybrid PSMC curve (purple) to show the divergence time between Huon populations and East populations of M. lugubris.

S11 Figure. Mitochondrial divergence matrix showing the minimum (first value) and maximum (second value) for each comparison of populations and species, M. gigantaea (Meg), M. lugubris (Mel), Eastern populations (EPops) include the subpopulations East, Southeast and Huon.

S12 Figure. Mitochondrial divergence matrix showing the mean divergence for each comparison of populations and species, M. gigantaea (Meg), M. lugubris (Mel), Eastern populations (EPops) include the subpopulations East, Southeast and Huon.

S1 Table. List of samples. Additional information such as sample locality, museum voucher, tissue type, etc. are included. The table also shows mapping statistics (e.g. mapping percentage and depth-of-coverage) for each individual.

S2 Table. Used github commits when running nextflow workflows

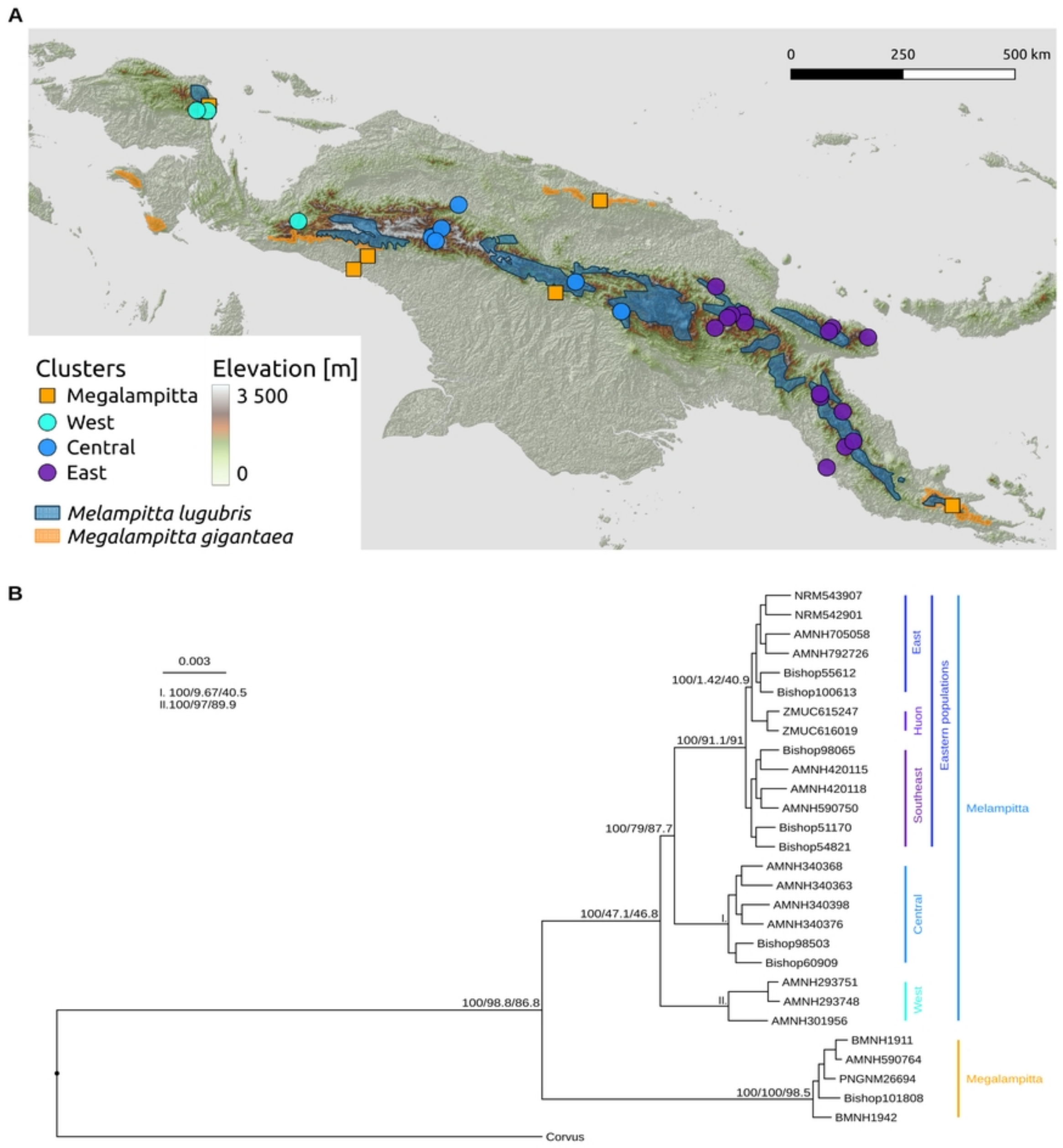
S3 Table. Filtered individuals with heterozygous blocks in mtDNA

S4 Table. Individual nucleotide diversity (π). Sheet 1 (Individual) contains statistics calculated for each species using i) all chromosomes and ii) only autosomes (aut.). Sheet 2 (Species-wide) contains estimates averaged across i) all chromosomes and ii) only autosomes (aut.)

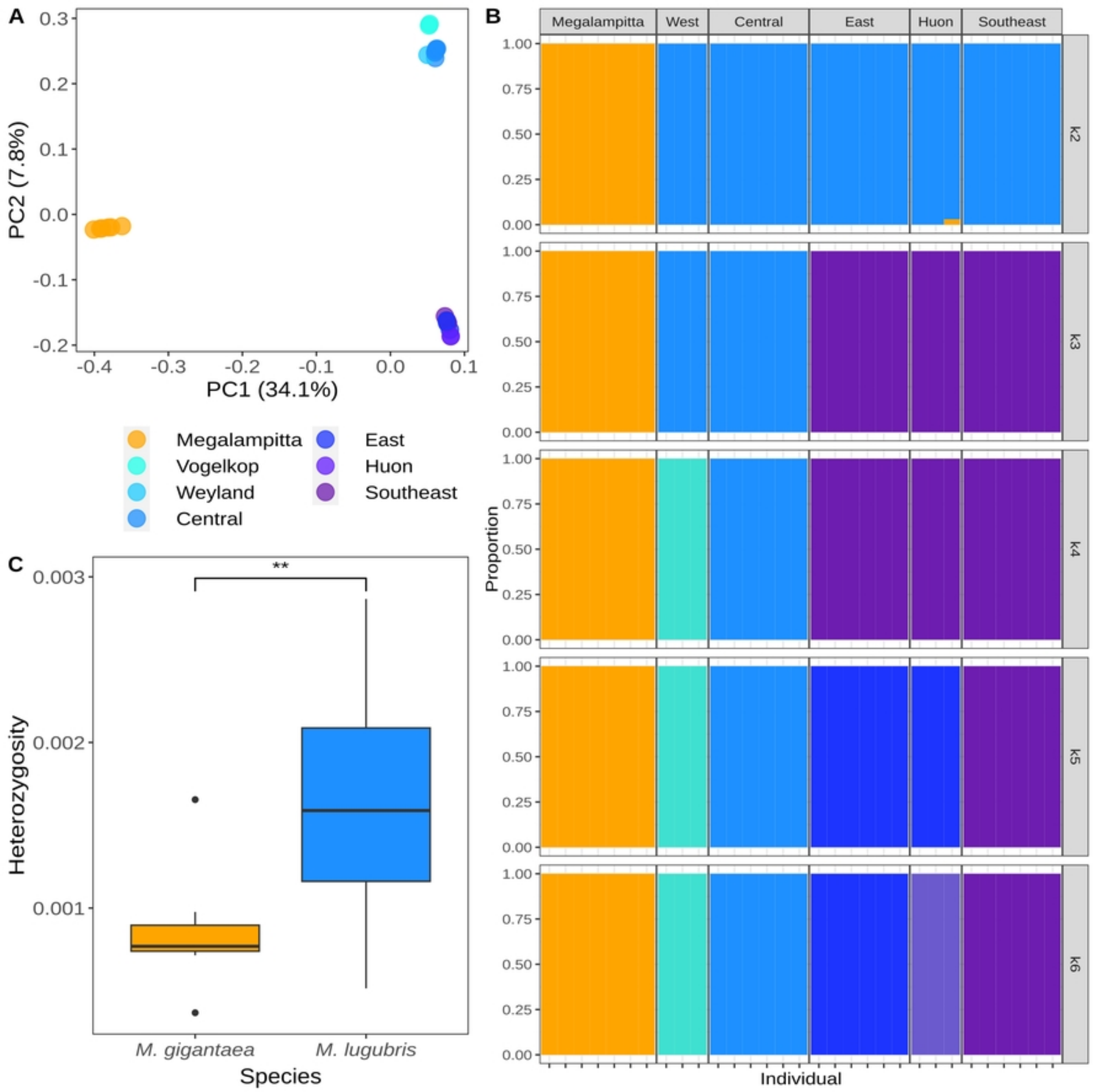
S5 Table. Species-wide Tajima's D. Values averaged across i) all chromosomes and ii) only autosomes (aut.)

S1 File. Details on PSMC methodology

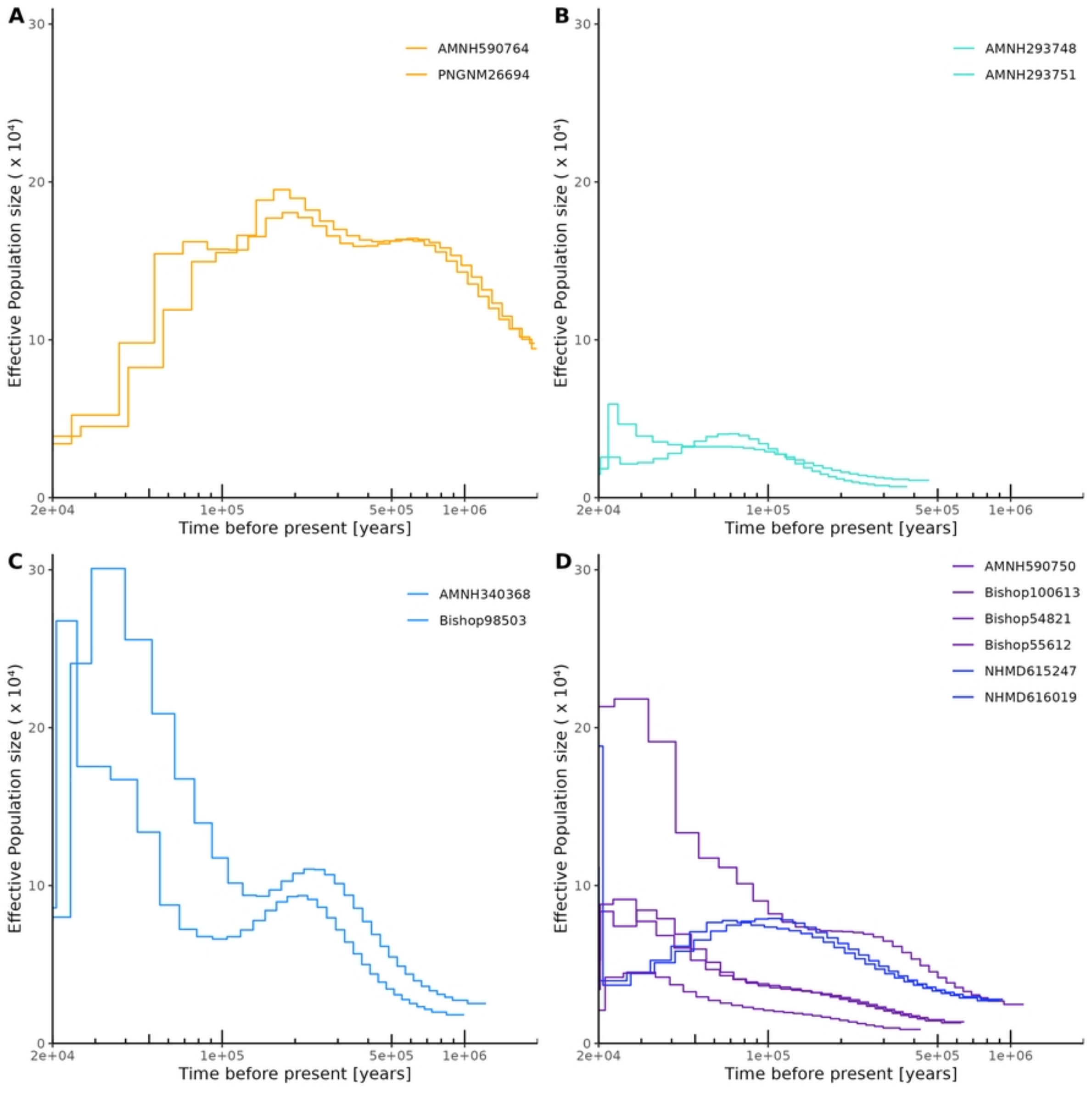
S2 File. Codes and parameters settings



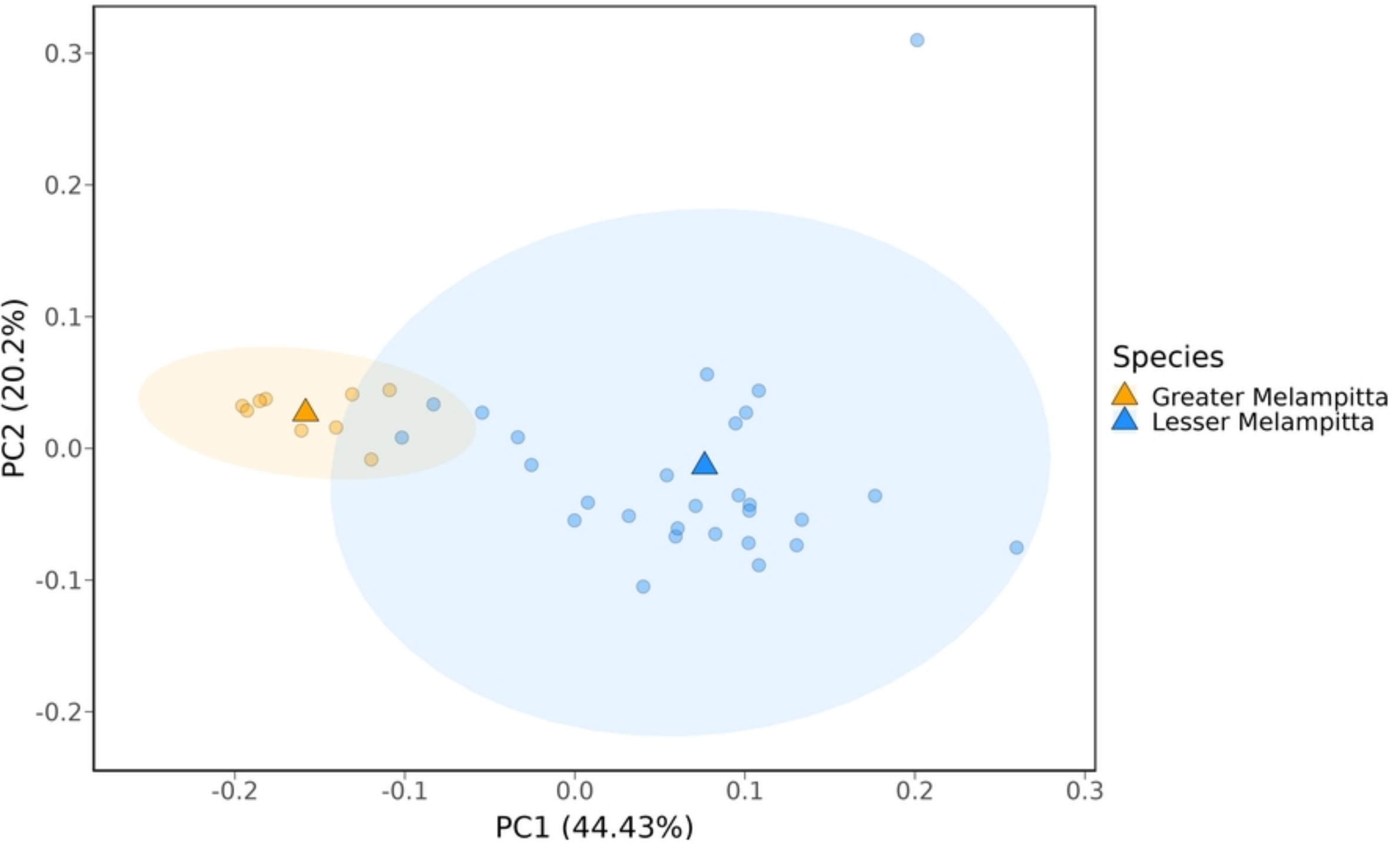
Main figure 1



Main figure 2



Main figure 3



Main figure 4