**Supplementary Material**

**Recovering the evolutionary history of crowned pigeons (Columbidae: *Goura*): implications for the biogeography and conservation of New Guinean lowland birds**

Jade Bruxaux *et al.*

The supplementary material includes:

**Supplementary Methods 1:** Identifying causes for the presence of minor mitochondrial DNA-like variants

**Supplementary Methods 2:** Phylogenomic analyses with filtered nuclear data

**Supplementary Table 1:** List and details of samples used in this study

**Supplementary Table 2:** Number of reads obtained per sample with Hi-Seq, results summary of mitochondrial and nuclear ribosomal DNA mapping and GenBank accessions of mitogenomes and nrDNA clusters

**Supplementary Table 3:** Description of each partition obtained for the whole mitogenome excluding the repeated sequence of the control region

**Supplementary Table 4:** Description of each partition obtained for the nuclear ribosomal DNA cluster

**Supplementary Table 5:** Description of the two partitions obtained for the mitochondrial coding genes in the dating analysis

**References**

**Supplementary Methods 1:** Identifying causes for the presence of minor mitochondrial DNA-like variants

Unlike the nuclear genome, the mitogenome is maternally transmitted in most animal species and is therefore expected to be homoplasmic (Dawid and Blackler, 1972). By using a genome skimming approach, the mitogenome assembling should then give a unique sequence, with minor variants mostly due to sequencing errors in very small proportion. However, a minor variant representing more than 10% of all the reads aligned against a reference genome was observed in 17 cases (out of 39) and this may have three putative causes: nuclear copy of part or whole mitogenome [numt; Lopez et al. (1994)], heteroplasmy (Kvist et al., 2003), or cross-contamination with samples of the same species or a close relative (Ballenghien et al., 2017).

**Methods**

We used three complementary approaches to determine the origins of minor mitochondrial DNA-like variants:

- *Phylogenetic approach*: First, we reconstructed a minority-rule sequence of the 17 individuals, choosing the minor variant instead of the major one at each site where a minor variant was present in more than 10% of the reads. When two of these sites were sufficiently close to have been sequenced on the same fragment, we checked for linkage disequilibrium. We distinguished two groups of samples with minor variants: the first one (group I) included samples for which all variable sites had a minor variant present in at least 20% of the reads with a relatively constant frequency (never exceeding 33%), and the second one (group II) contained samples with heterogeneity in the frequency of minor variants between variable sites. We then included these sequences into our mitogenome dataset to investigate their placement relative to other sequences in a phylogenetic tree. Phylogenetic analyses were performed as described in the article. Our expectations were as follows: a) when a minor variant was identical to a mitogenome isolated from another individual, it would indicate a potential cross-contamination between samples; b) heteroplasmic and numt sequences should have evolved differently and then have different placement in the phylogenetic tree. A true minor mitogenome (in a case of heteroplasmy), differing from the major one by only a few bases due to its recent origin, will be obviously related to other mitogenomes of this species. On the other hand, a nuclear copy could have been formerly integrated and have evolved less rapidly in the mitochondria (Gray et al., 1999), and consequently should be phylogenetically distant from other mitogenomes.

- *Relative nuclear-mitochondrial genome sequencing:* Second, we hypothesized that numts should be detected in samples with deepest nuclear sequencing. We thus compared for each sample the number of reads mapping against the mitogenome and the number of reads mapping against the conserved nuclear markers (McCormack et al., 2013; Prum et al., 2015). We compared the ratio nuclear / mitochondrial reads between the three groups of samples (*i.e.* no minor variant, group I, and group II) with a non-parametric test (Kruskal-Wallis test; Kruskal and Wallis, 1952) followed by a Dunn’s test (Dunn, 1964) with a Holm correction, using R v.3.3.3 (R Core Team, 2017).

- *Evidence for a nuclear origin of the minor variant in GC88:* Last, we used the most recent sample (GC88, collected in 2014) and worked more specifically on a 1000-bp part with a high density of variable sites. We used the minor variant sequence of this part as a reference to map reads with Geneious v. 9.0.5 (Biomatters Ltd., Auckland, New Zealand). We used a custom sensitivity with a minimum overlap identity of 99% and reiterate the process to increase the sequence length. At each step, the consensus sequence was increased by choosing the minor variant (before meeting a repeated nuclear region; see below). We carefully checked the sequencing depth and identified the nature of the resulting sequence with BLASTn (Altschul et al., 1990).

**Results and discussion**

*Phylogenetic evidence*

We identified two types of minor variants: some sequences (group I) are similar (but never identical) to majority-rule sequences and are thus embedded in clades that support monophyly of the four *Goura* species. In contrast, sequences that belong to group II are sister (and paraphyletic) to clades supported by majority-rule sequences (species or pair of species; Fig. S1).

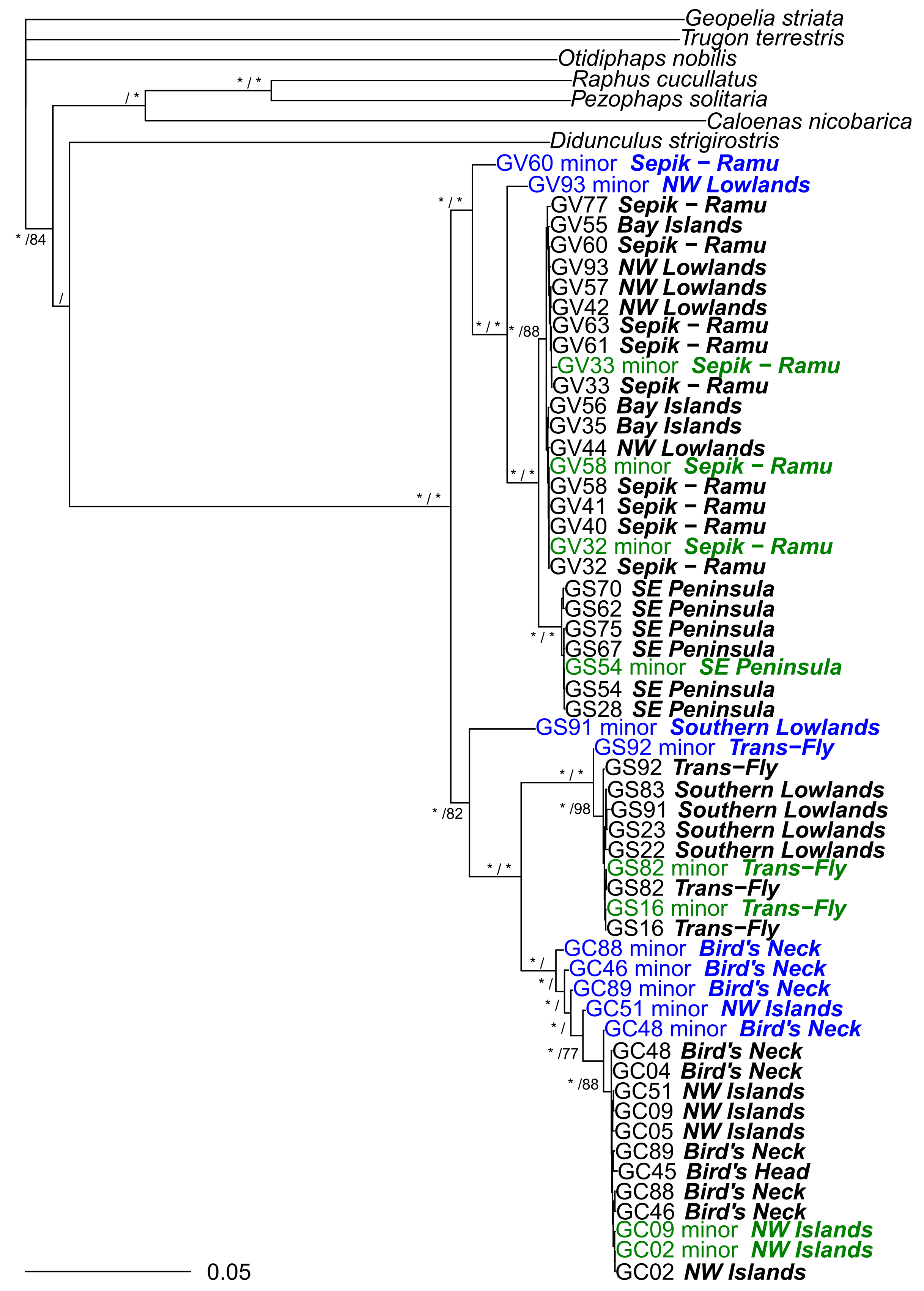
The group I (eight samples) is compatible both with true heteroplasmy or contamination. However, the minor variant is never found identical to one of the samples analyzed, which thus favors the hypothesis of a heteroplasmic origin. Among the eight samples of this group, seven show only one variable site. This could indicate a recent heteroplasmy, occurring in the individual or transmitted by its mother. On the other side, the last sample (GC09) shows 21 variable sites along the whole mitogenome. Such a pattern may be due to an event of bi-parental transmission of mitochondria (Kvist et al., 2003). In contrast, the successive sister placements of sequences of group II (nine samples) compared to clades supported by majority-rule sequences should indicate ancient nuclear incorporation of part or the whole mitogenome (numt).

*Relative nuclear-mitochondrial genome sequencing*

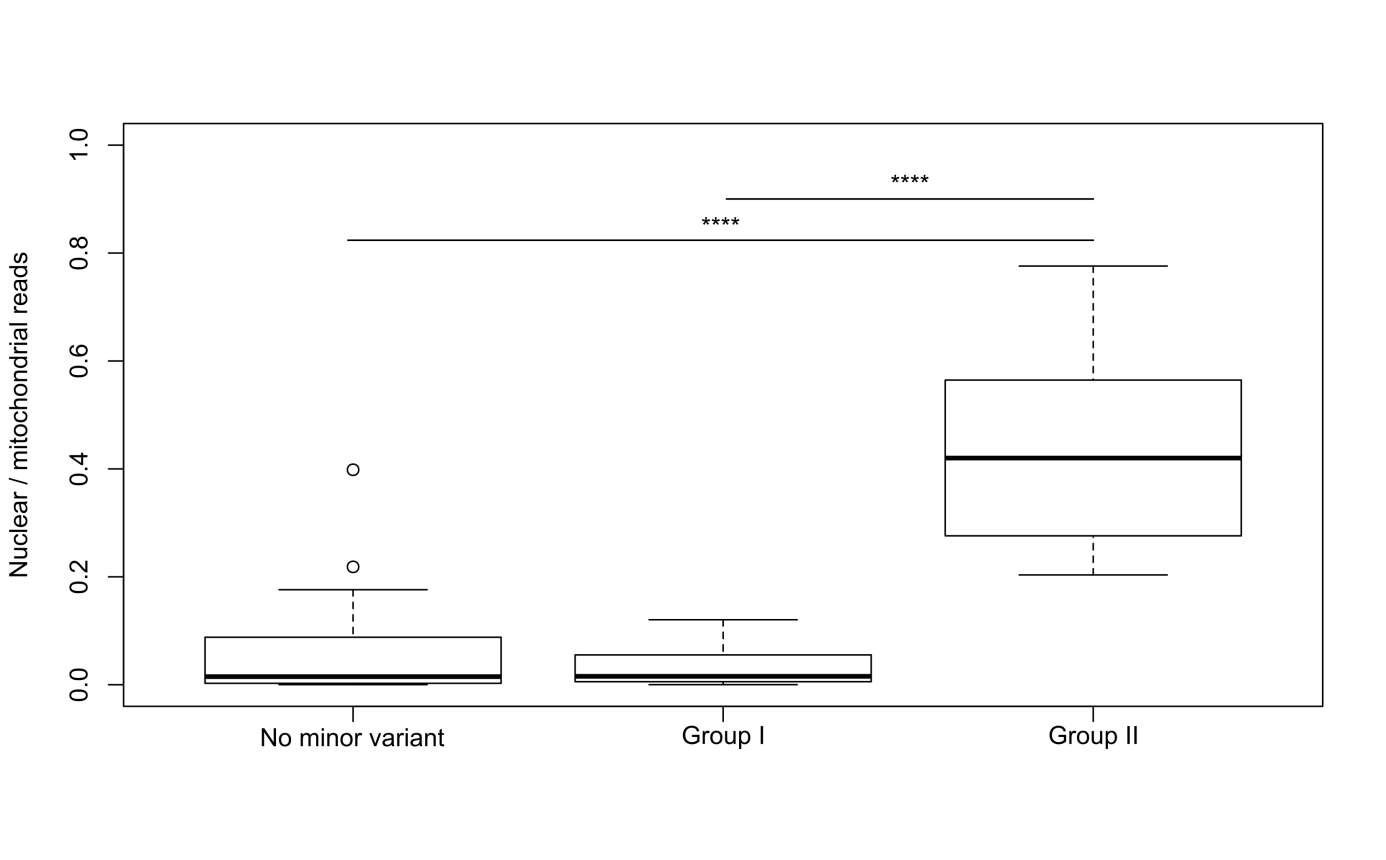
Samples belonging to group II show a significantly higher ratio of nuclear / mitochondrial reads (χ²= 19.996, df = 2, *p*-value < 0.001; Fig. S2) and these minor variant sequences are likely numts. In contrast, samples belonging to group I or those showing no minor variants were not significantly different based on the ratio of nuclear / mitochondrial reads.

*Evidence for a nuclear origin of the minor variant in GC88*

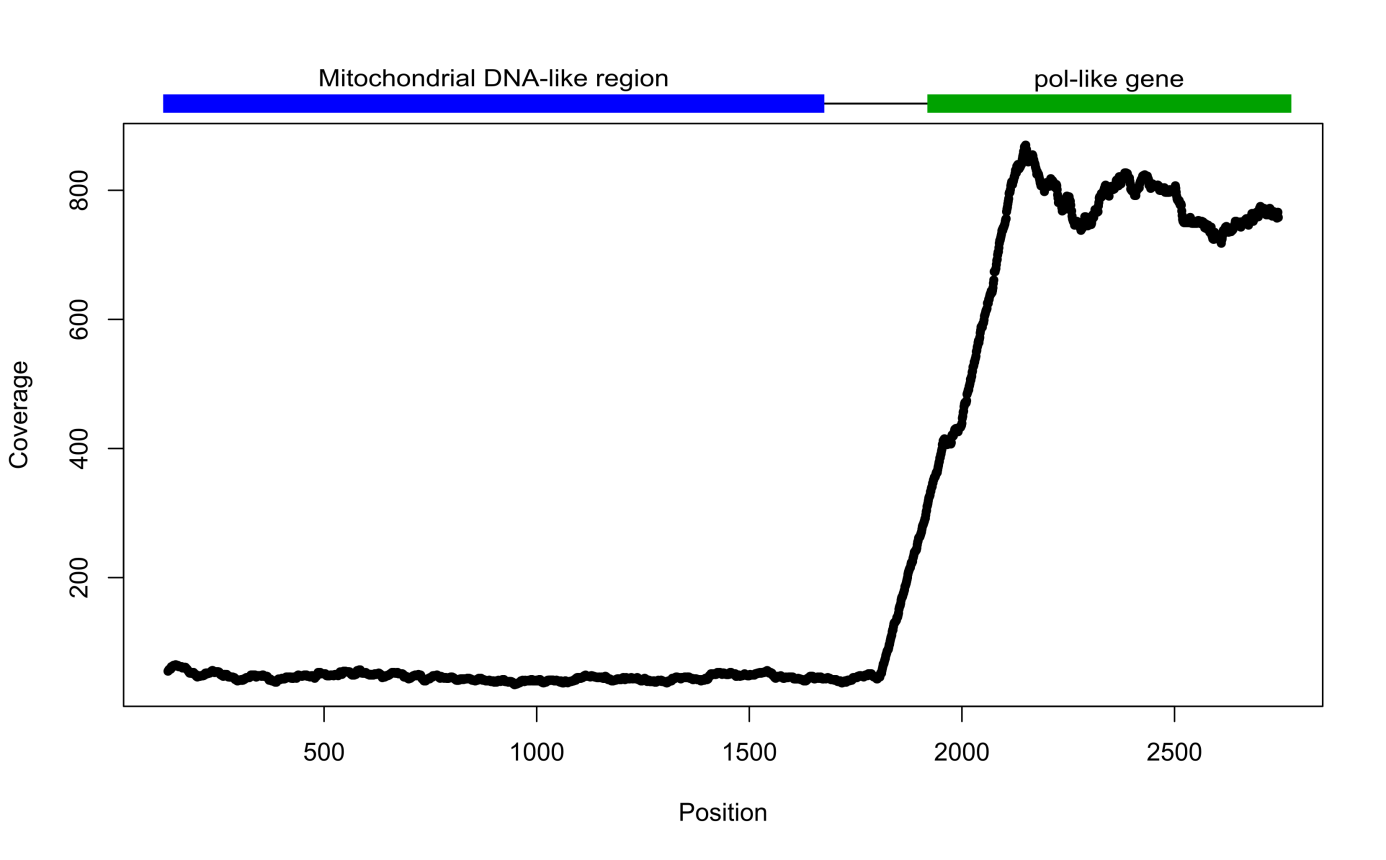
From the 1,000-bp segment used as reference, the minor variant (belonging to group II) was extended to approximately 1,600 bp (with a sequencing depth of approximately 45×), where a sudden increase of sequencing depth (ca. 750×) occurred around position 1,700 (Fig. S3). The first 1,611 bp shows a high homology with the *Goura* mitogenome. In contrast, after position 1,858, the best matches correspond to a gene encoding a nuclear pol-like protein, similar to those present in LTR retrotransposons (Wessler, 2006), suggesting that this minor variant may be a copy of a region of mitochondria formerly inserted in the nuclear genome (numt hypothesis).



**Figure S1:** Phylogenetic tree including all the mitogenome and minor variants for the 17 individuals involved. On major nodes are mentioned Bayesian posterior probabilities on the left and bootstrap value on the right. Minor variants of group I are written in green, those of group II are in blue.



**Figure S2:** Comparison of the ratio nuclear / mitochondrial reads between three groups of samples: no minor variant, Group I, and Group II. The difference is statistically significant between the first and last groups (Z=-4.273, *p*-value = 2.89e-05) and between the last two groups (Z=-3.389, *p*-value = 7.01e-04).



**Figure S3:** Sequencing depth along the GC88 minor mitochondrial DNA-like variant extended. The first part (1611 bp) shows a high homology with the *Goura* mitogenome, whereas the second part shows a high homology with a nuclear gene encoding a pol-like protein of retrotransposon (LTR).

**Conclusions**

From these three analyses, we can first conclude that the minor mitochondrial DNA-like variant present in some samples is not the result of a cross-contamination between our samples or with a close species because it was never found identical to one of the majority-rule consensus sequences reconstructed in our study. Second, we can clearly distinguish two groups of samples with a minor mitochondrial DNA-like variant:

- The first one (Group I) includes samples with only a few variable sites. These sequences cluster with sequences of the same species in the phylogenetic tree. They may correspond to true heteroplasmy, with different variants of mitogenomes within the same individual resulting either from mutations within the individual or paternal leakage during sexual reproduction.

- For the second group (Group II), we have strong evidence for the presence of numts, with samples showing more variable sites along a part or the totality of the mitogenome. These numts are detectable among samples with the highest nuclear sequencing depth (Fig. S1).The phylogenetic tree including minor sequences indicates that such numts are not embedded within the *Goura* speciesclades supported by mitogenome sequences (*i.e.* majority-rule consensus sequences). In contrast, each numt sequence is sister to clades supported by majority-rule sequences (*i.e.* species or pair of species), suggesting recurrent integrations in the nucleus after the first event of diversification in *Goura*. In addition, numts show slightly shorter branches, suggesting a slower evolution, as expected in the nucleus compared to the mitochondrion (Gray et al., 1999). We also found a retrotransposon-like sequence at an extremity of the minor variant of GC88, sustaining the potential nuclear origin of the sequence.

Since the sequencing coverage of numts never exceeded 28% of reads matching to the mitogenome, and heteroplasmic sites were often found alone or in only a few sites along the genome with a rather constant proportion of minor/major variants (never exceeding 33% for the minor variant), using the majority-rule consensus sequence in our analyses likely excludes both numts and heteroplasmic variants from the data set.

**Supplementary methods 2:** Phylogenomic analyses with filtered nuclear data

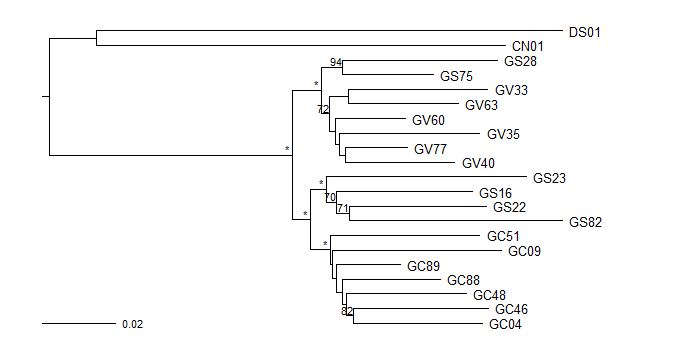
Phylogenomic analyses were performed after filtering the low copy conserved markers in order to minimize the influence of sequencing errors or deaminations that can have a strong impact at low coverage (Le and Durbin, 2011). These analyses were conducted on the same 19 *Goura* individuals as the analysis presented in the manuscript, and *Didunculus strigirostris* and *Caloenas nicobarica* as outgroups.

To do so, we used the bam files with quality weighted by mapDamage (Jónsson et al., 2013; see paragraph 2.4.2. for details before this step). These files were compared to the reference sequence with samtools mpileup (with default parameters) v.1.3.1 and bcftools call (with multiallelic caller) v.1.1-60-g3d5d3d9 (Li et al., 2009) to obtain a first list of SNPs (Single Nucleotide Polymorphism) for each individual. As nuclear data are expected to have either one or two alleles, SNPs with more than two different alleles were not considered and removed with vcftools v0.1.12a (Danecek et al., 2011). Data were also filtered with bcftools to keep only SNPs with a minimal depth of 2× to remove errors that are expected to produce false SNPs with low coverage. All resulting SNPs were then merged with vcf-merge (Danecek et al., 2011) and samtools mpileup was used to retrieve the genotypes for all individuals at all sites. Genotypes in each individual were called only if sites had at least a depth of 2x and a phred-scaled base score of 30 using bcftools filter (Li et al., 2009). Finally, vcf-merge (Danecek et al., 2011) was used to merge the resulting individual genotypes and apply additional filters on this dataset. First, SNPs were kept only when the minor allele was present at least 2 times in the site (corresponding to at least one homozygous individual or two heterozygous ones, option –mac 2). This step should remove most of the private SNPs due to sequencing or deamination errors. Second, a threshold of missing data was applied on each site, ranging from 50% to 80%. This final filtered dataset was converted from vcf to fasta format, where heterozygous sites were coded with IUPAC ambiguities thanks to a perl script adapted from Olofsson et al. (2016). A maximum-likelihood phylogenetic analysis was then performed with RAxML v.8.1.5 (Stamatakis, 2014) with eight alternative runs and 100 replicates of non-parametric bootstrapping, *Caloenas nicobarica* and *Didunculus strigirostris* being stated as outgroups.

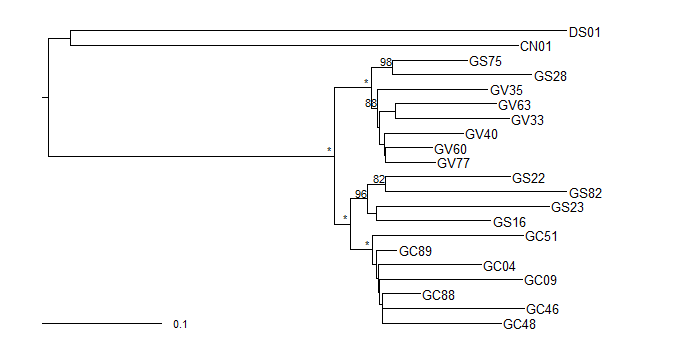
Most of the SNPs present in the initial dataset are either low covered or detected in only one individual (Table). Nevertheless, all the resulting trees recovered the same general topology as described in paragraph 3.4, with two clades of two sister species: *Goura cristata*/*Goura sclaterii* on one side, and *Goura victoria*/*Goura scheepmakeri* on the other side (Fig. S4 to S8). Applying more stringent thresholds does not seem to increase node support, only resulting in slightly shorter branches, especially for the most recent samples (GC88 and GC89).

**Table:** Number of sites included in each analysis. DP corresponds to the minimal depth, and mac to the minimal allele count.

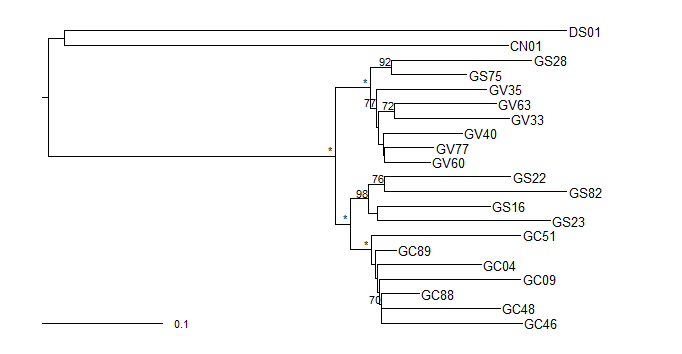
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Filtering thresholds | Only SNP, no filtering | DP2, mac2, missing data <80% | DP2, mac2, missing data <70% | DP2, mac2, missing data <60% | DP2, mac2, missing data <50% |
| Number of sites analyzed | 70,556 | 17,428 | 16,910 | 14,646 | 12,663 |



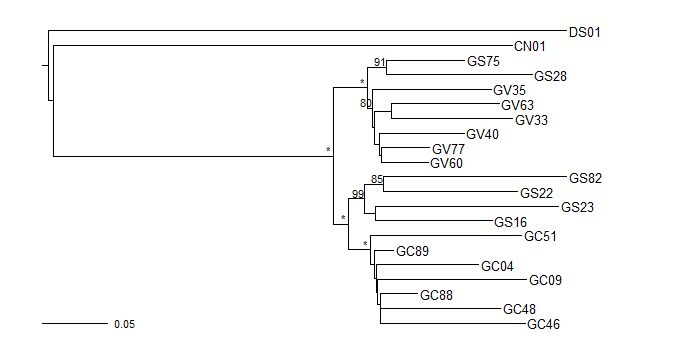
**Figure S4:** Phylogenetic tree obtained from all SNP of the conserved nuclear regions, without filtering. \* corresponds to a node support of 100, supports lower than 70 were removed.



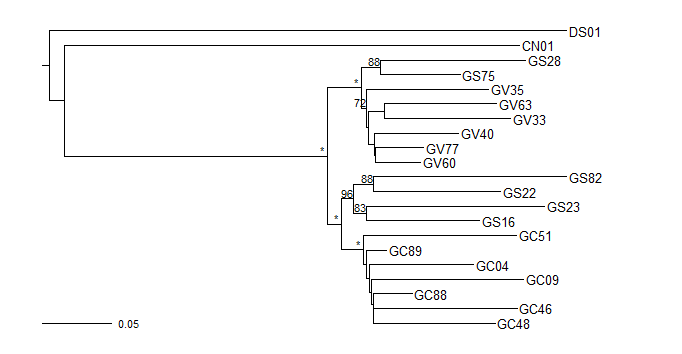
**Figure S5:** Phylogenetic tree obtained from SNP of the conserved nuclear regions filtered for a minimal depth of two, alleles present at least two times across individuals, and missing data representing less than 80% of the data at each position. \* corresponds to a node support of 100, supports lower than 70 were removed.



**Figure S6:** Phylogenetic tree obtained from SNP of the conserved nuclear regions filtered for a minimal depth of two, alleles present at least two times across individuals, and missing data representing less than 70% of the data at each position. \* corresponds to a node support of 100, supports lower than 70 were removed.



**Figure S7:** Phylogenetic tree obtained from SNP of the conserved nuclear regions filtered for a minimal depth of two, alleles present at least two times across individuals, and missing data representing less than 60% of the data at each position. \* corresponds to a node support of 100, supports lower than 70 were removed.



**Figure S8:** Phylogenetic tree obtained from SNP of the conserved nuclear regions filtered for a minimal depth of two, alleles present at least two times across individuals, and missing data representing less than 50% of the data at each position. \* corresponds to a node support of 100, supports lower than 70 were removed.

**Supplementary Table 1:** List and details of samples used in this study. Species follow del Hoyo and Collar (2014); latitude and longitude are expressed in decimal degrees (DD) and follow Beehler and Pratt (2016); regions are as defined by Mack and Dumbacher (2007). Institutional abbreviations are as follows: AMNH = American Museum of Natural History, New York; ANWC = Australian National Wildlife Collection, Canberra; EDB = Laboratoire Evolution et Diversité Biologique, Toulouse; MZB = Museum Zoological of Bogor; NHMUK = Natural History Museum, UK; ZMUC = Natural History Museum of Denmark, Copenhagen. BOU is used for the British Ornithological Union New Guinea expedition.

| Museum | Museum ID | Code | Species | Sex | Year of collect | Collector | Type | Locality | Latitude (DD) | Longitude (DD) | Region |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| NHMUK | 1934.10.21.74 | **GC02** | *Goura cristata* | F | 1934 | W.G.C. Frost | Toe-Pad | Salawati Island | -0.8 | 130.92 | NW Islands |
| NHMUK | 1889.2.10.492 | **GC04** | *Goura cristata* | M | 1873 | A.B. Meyer | Toe-Pad | Rubi | -3.33 | 134.97 | Bird's Neck |
| NHMUK | 1873.5.12.2386 | **GC05** | *Goura cristata* | NA | 1860 | C. Allen | Toe-Pad | Misool Island | -1.9 | 130 | NW Islands |
| NHMUK | 1904.4.23.1 | **GC09** | *Goura cristata* | M | 1902 | J. Waterstradt | Toe-Pad | Waigeo Island | -0.23 | 131 | NW Islands |
| AMNH | SKIN 616600 | **GC45** | *Goura cristata* | NA | NA | NA | Toe-Pad | Arfak Mountains | -1.6 | 133 | Bird's Head |
| AMNH | SKIN 616602 | **GC46** | *Goura cristata* | NA | 1896 | C. Webster | Toe-Pad | Triton Bay | -3.81 | 134.18 | Bird's Neck |
| AMNH | SKIN 616604 | **GC48** | *Goura cristata* | NA | 1896 | C. Webster | Toe-Pad | Etna Bay | -3.96 | 134.71 | Bird's Neck |
| AMNH | SKIN 616612 | **GC51** | *Goura cristata* | M | 1925 | W.J. Frost | Toe-Pad | Waigeo Island (Majalibit Bay) | -0.35 | 130.93 | NW Islands |
| MZB | LGR-094 | **GC88** | *Goura cristata* | NA | 2014 | H. Ashari | Tissue | Kamaka, Triton Bay | -3.79 | 134.21 | Bird's Neck |
| MZB | LGR-095 | **GC89** | *Goura cristata* | NA | 2014 | H. Ashari | Feather | Lengguru River, Triton Bay | -3.75 | 134.12 | Bird's Neck |
| NHMUK | 1889.2.12.421 | **GS28** | *Goura scheepmakeri* | NA | [1879] | A. Goldie | Toe-Pad | Port Moresby | -9.45 | 147.2 | SE Peninsula |
| AMNH | SKIN 616626 | **GS54** | *Goura scheepmakeri* | M | 1893 | Lix | Toe-Pad | Yule Island, Nicura | -8.8 | 146.62 | SE Peninsula |
| AMNH | SKIN 819288 | **GS62** | *Goura scheepmakeri* | F | 1948 | E.T. Gilliard | Toe-Pad | Brown River road | -9.2 | 147.23 | SE Peninsula |
| ANWC | B03514 | **GS67** | *Goura scheepmakeri* | F | 1966 | NA | Toe-Pad | Lohiki/Vailala River junction | -7.78 | 145.49 | SE Peninsula |
| ANWC | B03686 | **GS70** | *Goura scheepmakeri* | M | 1966 | NA | Toe-Pad | Ravikevau, Purari delta | -7.77 | 145.17 | SE Peninsula |
| ANWC | B03986 | **GS75** | *Goura scheepmakeri* | F | 1966 | NA | Toe-Pad | Putei, near the Purari River | -7.8 | 146.13 | SE Peninsula |
| NHMUK | 1889.2.12.420 | **GS16** | *Goura sclaterii* | M | 1877 | V. D’Albertis | Toe-Pad | Fly River, southern entrance | -8.74 | 143.39 | Trans-Fly |
| NHMUK | 1911.12.20.44 | **GS22** | *Goura sclaterii* | F | 1910 | BOU | Toe-Pad | Mimika River | -4.42 | 136.55 | Southern Lowlands |
| NHMUK | 1911.12.20.47 | **GS23** | *Goura sclaterii* | M | 1912 | BOU | Toe-Pad | Utakwa River | -4.33 | 137.23 | Southern Lowlands |
| ANWC | B08343 | **GS82** | *Goura sclaterii* | M | 1964 | NA | Toe-Pad | Oriomo River, near the mouth | -9.03 | 143.18 | Trans-Fly |
| ANWC | B14301 | **GS83** | *Goura sclaterii* | NA | 1971 | NA | Toe-Pad | Cape Steenboon, Carstenz range | -4.92 | 136.83 | Southern Lowlands |
| MZB | MZB 28858 | **GS91** | *Goura sclaterii* | F | NA | NA | Toe-Pad | Brazza River | -4.95 | 139.41 | Southern Lowlands |
| MZB | MZB 29747 | **GS92** | *Goura sclaterii* | F | NA | NA | Toe-Pad | Okaba | -8.1 | 139.7 | Trans-Fly |
| NHMUK | 1921.12.30.36 | **GV32** | *Goura victoria* | F | 1920 | W. Potter | Toe-Pad | Malala | -4.46 | 145.39 | Sepik - Ramu |
| NHMUK | 1921.12.30.38 | **GV33** | *Goura victoria* | M | 1920 | W. Potter | Toe-Pad | Ramu River (Botbot) | -4.04 | 144.68 | Sepik - Ramu |
| NHMUK | 1889.2.12.423 | **GV35** | *Goura victoria* | M | 1873 | A.B. Meyer | Toe-Pad | Kordo | -0.75 | 135.58 | Bay Islands |
| NHMUK | 1921.12.30.40 | **GV40** | *Goura victoria* | M | 1920 | W. Potter | Toe-Pad | Watam | -3.92 | 144.2 | Sepik - Ramu |
| NHMUK | 1921.12.30.35 | **GV41** | *Goura victoria* | M | 1920 | W. Potter | Toe-Pad | Dugumur Bay | -4.4 | 145.18 | Sepik - Ramu |
| AMNH | SKIN 339063 | **GV42** | *Goura victoria* | M | 1939 | NA | Toe-Pad | Bernhard Camp (along Idenburg River) | -3.48 | 139.22 | NW Lowlands |
| AMNH | SKIN 339080 | **GV44** | *Goura victoria* | NA | 1938 | NA | Toe-Pad | Humboldt Bay | -2.63 | 140.78 | NW Lowlands |
| AMNH | SKIN 616629 | **GV55** | *Goura victoria* | M | 1883 | H. Guillemard | Toe-Pad | Yapen Island, Geelvink Bay | -1.8 | 136.3 | Bay Islands |
| AMNH | SKIN 616631 | **GV56** | *Goura victoria* | F | 1896 | W.M. Doherty | Toe-Pad | Biak Island, Geelvink Bay | -1 | 136 | Bay Islands |
| AMNH | SKIN 616632 | **GV57** | *Goura victoria* | NA | NA | W.M. Doherty | Toe-Pad | Wonti, Waropen area | -2.26 | 136.66 | NW Lowlands |
| AMNH | SKIN 616636 | **GV58** | *Goura victoria* | NA | 1899 | E. Nyman | Toe-Pad | Stephansort | -5.42 | 145.72 | Sepik - Ramu |
| AMNH | SKIN 766222 | **GV60** | *Goura victoria* | M | 1954 | E.T. Gilliard | Toe-Pad | Yamanumba | -4.18 | 143.27 | Sepik - Ramu |
| AMNH | SKIN 791048 | **GV61** | *Goura victoria* | NA | 1959 | E.T. Gilliard | Toe-Pad | Oronga, Adelbert Mountains | -5.1 | 145.47 | Sepik - Ramu |
| AMNH | SKIN 828847 | **GV63** | *Goura victoria* | F | 1966 | J.M. Diamond | Toe-Pad | Utai, Bewani Mountains | -2.84 | 141.24 | Sepik - Ramu |
| ANWC | B07008 | **GV77** | *Goura victoria* | NA | 1966 | NA | Toe-Pad | Ambunti | -4.22 | 142.85 | Sepik - Ramu |
| MZB | MZB 14482 | **GV93** | *Goura victoria* | NA | NA | NA | Toe-Pad | Mamberamo River | -2 | 137.8 | NW Lowlands |
| NHMUK | 1956.60.566 | **CN01** | *Caloenas nicobarica* | NA | 1956 | R.W. Sims & E. Banks | Toe-Pad | Sapidan Island, W. Borneo | NA | NA | - |
| NHMUK | 1939.12.9.2020 | **DS01** | *Didunculus strigirostris* | NA | 1896 | NA | Toe-Pad | Apia, Upolu Island, Samoa | -13.87 | 171.73 | - |
| EDB | 286 | **GE01** | *Geopelia striata* | NA | 2007 | B. Mila | Blood | St Leu, Reunion | -21.14 | 55.3 | - |
| AMNH | DOT 9510 | **DOT9510** | *Otidiphaps nobilis* | F? | NA | NA | Tissue | From captive bird, New Guinea | NA | NA | - |
| ZMUC | 131708 | **131708** | *Trugon terrestris* | NA | NA | NA | Tissue | From captive bird (J. Erritzøe's collection) | NA | NA | - |

**Supplementary Table 2:** Number of reads obtained per sample with Hi-Seq, results summary of mitochondrial and nuclear ribosomal DNA mapping (number of reads and mean coverage) and GenBank accessions of mitogenomes and nrDNA clusters.

| Code | Species | Reads sequenced | Reads mapped against the mitochondrial reference | Mean coverage | Mitochondrial GenBank accession | Reads mapped against the nuclear ribosomal reference | Mean coverage | Nuclear ribosomal GenBank accession |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **GC02** | *Goura cristata* | 20,656,792 | 128,108 | 509.82× | LN589994 | 15,465 | 93.78× | MG590306 |
| **GC04** | *Goura cristata* | 17,194,178 | 137,738 | 513.52× | MG590267 | 19,141 | 109.42× | MG590307 |
| **GC05** | *Goura cristata* | 17,160,678 | 267,589 | 1003.92× | MG590268 | NA | NA | - |
| **GC09** | *Goura cristata* | 18,954,564 | 115,102 | 427.98× | MG590269 | NA | NA | - |
| **GC45** | *Goura cristata* | 19,797,226 | 313,143 | 1572.30× | MG590270 | NA | NA | - |
| **GC46** | *Goura cristata* | 18,736,478 | 14,849 | 72.15× | MG590271 | NA | NA | - |
| **GC48** | *Goura cristata* | 30,813,326 | 36,603 | 183.83× | MG590272 | 16,012 | 125.08× | MG590308 |
| **GC51** | *Goura cristata* | 22,881,668 | 37,219 | 184.44× | MG590273 | 4865 | 38.10× | MG590309 |
| **GC88** | *Goura cristata* | 31,479,818 | 38,362 | 240.33× | MG590274 | 22,642 | 216.50× | MG590310 |
| **GC89** | *Goura cristata* | 38,370,918 | 70,497 | 450.12× | MG590275 | NA | NA | - |
| **GS16** | *Goura sclaterii* | 22,995,404 | 139,419 | 498.22× | MG590277 | 19,031 | 106.37× | MG590311 |
| **GS22** | *Goura sclaterii* | 17,975,616 | 127,339 | 480.78× | MG590278 | 30,839 | 173.62× | MG590312 |
| **GS23** | *Goura sclaterii* | 17,553,786 | 91,202 | 307.09× | MG590279 | 2797 | 15.69× | MG590313 |
| **GS82** | *Goura sclaterii* | 29,111,282 | 1,739,745 | 7759.56× | MG590285 | 5013 | 47.02× | MG590318 |
| **GS83** | *Goura sclaterii* | 18,911,380 | 88,898 | 482.85× | MG590286 | 8092 | 63.52× | MG590319 |
| **GS91** | *Goura sclaterii* | 14,696,262 | 3543 | 16.74× | MG590287 | NA | NA | - |
| **GS92** | *Goura sclaterii* | 10,224,104 | 12,314 | 59.52× | MG590288 | NA | NA | - |
| **GS28** | *Goura scheepmakeri* | 18,271,590 | 67,737 | 230.89× | LN589995 | 6343 | 35.03× | MG590314 |
| **GS54** | *Goura scheepmakeri* | 18,501,598 | 575,064 | 2821.68× | MG590280 | NA | NA | - |
| **GS62** | *Goura scheepmakeri* | 27,803,068 | 280,044 | 1411.31× | MG590281 | 1518 | 13.39× | MG590315 |
| **GS67** | *Goura scheepmakeri* | 28,865,030 | 1,594,477 | 7709.31× | MG590282 | 7639 | 75.47× | MG590316 |
| **GS70** | *Goura scheepmakeri* | 34,625,736 | 1,584,541 | 7440.93× | MG590283 | NA | NA | - |
| **GS75** | *Goura scheepmakeri* | 21,579,692 | 873,080 | 4681.55× | MG590284 | 3809 | 31.82× | MG590317 |
| **GV32** | *Goura victoria* | 22,669,096 | 311,843 | 1226.65× | MG590289 | NA | NA | - |
| **GV33** | *Goura victoria* | 20,037,350 | 53,712 | 193.99× | MG590290 | 11,516 | 67.20× | MG590320 |
| **GV35** | *Goura victoria* | 18,787,684 | 60,595 | 219.54× | MG590291 | NA | NA | - |
| **GV40** | *Goura victoria* | 17,234,258 | 34,427 | 132.00× | LN589993 | 15,355 | 93.59× | MG590321 |
| **GV41** | *Goura victoria* | 18,723,496 | 1,168,706 | 4862.62× | MG590292 | NA | NA | - |
| **GV42** | *Goura victoria* | 18,245,126 | 987,073 | 5112.58× | MG590293 | NA | NA | - |
| **GV44** | *Goura victoria* | 13,551,122 | 192,782 | 978.74× | MG590294 | NA | NA | - |
| **GV55** | *Goura victoria* | 19,955,304 | 312,486 | 1625.22× | MG590295 | NA | NA | - |
| **GV56** | *Goura victoria* | 40,233,616 | 1,302,139 | 6822.75× | MG590296 | NA | NA | - |
| **GV57** | *Goura victoria* | 16,641,594 | 450,888 | 2421.54× | MG590297 | NA | NA | - |
| **GV58** | *Goura victoria* | 36,175,222 | 119,544 | 610.25× | MG590298 | NA | NA | - |
| **GV60** | *Goura victoria* | 34,752,218 | 34,445 | 186.43× | MG590299 | 41,649 | 356.62× | MG590322 |
| **GV61** | *Goura victoria* | 15,211,852 | 50,474 | 235.31× | MG590300 | NA | NA | - |
| **GV63** | *Goura victoria* | 22,029,804 | 443,776 | 2483.29× | MG590301 | NA | NA | - |
| **GV77** | *Goura victoria* | 29,079,904 | 637,338 | 3659.16× | MG590302 | 27,426 | 232.82× | MG590323 |
| **GV93** | *Goura victoria* | 22,250,994 | 7733 | 33.33× | MG590303 | NA | NA | - |
| **CN01** | *Caloenas nicobarica* | 27,983,732 | 20,070 | 71.01× | MG590264 | 14,472 | 80.86× | MG590304 |
| **DS01** | *Didunculus strigirostris* | 25,830,726 | 69,276 | 238.57× | MG590266 | 7696 | 40.68× | MG590305 |
| **DOT9510** | *Otidiphaps nobilis* | 29,542,534 | 55,201 | 368.06× | MG590265 | NA | NA | - |
| **131708** | *Trugon terrestris* | 34,956,334 | 270,672 | 2080.26× | MG590263 | NA | NA | - |
| **GE01** | *Geopelia striata* | 20,370,544 | 41,682 | 225.98× | MG590276 | NA | NA | - |

**Supplementary Table 3:** Description of each partition obtained for the whole mitogenome excluding the repeated sequence of the control region. For coding sequences, first, second and third codon positions are treated separately. Each mitochondrial regionis placed in a partition with agiven evolutionary model in the Maximum-Likelihood (ML) and the Bayesian analyses. Note that the second codon position of *ATP8* was in a different partition in the ML and Bayesian analyses.

|  |  |  |
| --- | --- | --- |
| ML | Bayesian | List of mitochondrial regions |
| GTR+I+G | GTR+I+G | 12S, 16S, *ATP6*\_1, *ATP8*\_1, ***ATP8*\_2 (ML)**, *COX1*\_1, *COX2*\_1, *COX3*\_1, *CYTB*\_1, *ND1*\_1, *ND2*\_1, *ND3*\_1, *ND4*\_1, *ND4L*\_1, *ND5*\_1, *ND6*\_1, *ND6*\_2, tRNA-Ala, tRNA-Arg, tRNA-Asn, tRNA-Asp, tRNA-Cys, tRNA-Gln, tRNA-Glu, tRNA-His, tRNA-Ile, tRNA-L1, tRNA-L2, tRNA-Lys, tRNA-Met, tRNA-Phe, tRNA-Pro, tRNA-S1, tRNA-S2, tRNA-Thr, tRNA-Trp, tRNA-Tyr, tRNA-Val |
| GTR+I+G | HKY+I+G | *ATP6*\_2, ***ATP8*\_2 (Bayesian)**, *COX1*\_2, *COX2*\_2, *COX3*\_2, *CYTB*\_2, *ND1*\_2, *ND2*\_2, *ND3*\_2, *ND4*\_2, *ND4L*\_2, *ND5*\_2, tRNA-Gly |
| GTR+I+G | GTR+I+G | *ATP6*\_3, *ATP8*\_3, *COX1*\_3, *COX2*\_3, *COX3*\_3, *CYTB*\_3, *ND1*\_3, *ND2*\_3, *ND3*\_3, *ND4*\_3, *ND4L*\_3, *ND5*\_3 |
| GTR+I+G | GTR+G | *ND6*\_3, Non\_coding |

**Supplementary Table 4:** Description of each partition obtained for the nuclear ribosomal DNA cluster. Each ribosomal DNA regionis placed in a partition with agiven evolutionary model in the Maximum-Likelihood (ML) and Bayesian analyses.

|  |  |  |  |
| --- | --- | --- | --- |
| Partition | ML | Bayesian | Ribosomal DNA regions |
| 1 | GTR+I+G | HKY+I | 5.8S, 18S, 28S |
| 2 | GTR+G | HKY+G | 3’ETS, 5’ETS, ITS2, ITS1 |

**Supplementary Table 5:** Description of the two partitions obtained for the mitochondrial coding genes (excluding *ATP8*, *ND4L* and *ND6*) in the dating analysis (withBeast2). First, second and third codon positions are treated separately for each gene.

|  |  |
| --- | --- |
| Bayesian | List of mitochondrial genes |
| TRN+G+X | *ATP6*\_1, *ATP6*\_2, *COX1*\_1, *COX1*\_2, *COX2*\_1, *COX2*\_2, *COX3*\_1, *COX3*\_2, *CYTB*\_1, *CYTB*\_2, *ND1*\_1, *ND1*\_2, *ND2*\_1, *ND2*\_2, *ND3*\_1, *ND3*\_2, *ND4*\_1, *ND4*\_2, *ND5*\_1, *ND5*\_2 |
| GTR+X | *ATP6*\_3, *COX1*\_3, *COX2*\_3, *COX3*\_3, *CYTB*\_3, *ND1*\_3, *ND2*\_3, *ND3*\_3, *ND4*\_3, *ND5*\_3 |

**References**

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Ballenghien, M., Faivre, N., Galtier, N., 2017. Patterns of cross-contamination in a multispecies population genomic project: detection, quantification, impact, and solutions. BMC Biol. 15, 25. https://doi.org/10.1186/s12915-017-0366-6

Beehler, B.M., Pratt, T.K., 2016. Birds of New Guinea: distribution, taxonomy, and systematics. Princeton University Press, Princeton, UK.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., 2011. The variant call format and VCFtools. Bioinformatics 27, 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Dawid, I.B., Blackler, A.W., 1972. Maternal and cytoplasmic inheritance of mitochondrial DNA in *Xenopus*. Dev. Biol. 29, 152–161. https://doi.org/10.1016/0012-1606(72)90052-8

del Hoyo, J., Collar, N.J., 2014. HBW and BirdLife International illustrated checklist of the birds of the world 1: non-passerines. Lynx Edicions, Barcelona.

Dunn, O.J., 1964. Multiple Comparisons Using Rank Sums. Technometrics 6, 241–252. https://doi.org/10.2307/1266041

Gray, M.W., Burger, G., Lang, B.F., 1999. Mitochondrial Evolution. Science 283, 1476–1481. https://doi.org/10.1126/science.283.5407.1476

Jónsson, 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. https://doi.org/10.1093/bioinformatics/btt193

Kruskal, W.H., Wallis, W.A., 1952. Use of ranks in one-criterion variance analysis. J. Am. Stat. Assoc. 47, 583–621. https://doi.org/10.2307/2280779

Kvist, L., Martens, J., Nazarenko, A.A., Orell, M., 2003. Paternal leakage of mitochondrial DNA in the great tit (*Parus major*). Mol. Biol. Evol. 20, 243–247. https://doi.org/10.1093/molbev/msg025

Le, S.Q., Durbin, R., 2011. SNP detection and genotyping from low-coverage sequencing data on multiple diploid samples. Genome Res. 21, 952–960. https://doi.org/10.1101/gr.113084.110

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., Subgroup, 1000 Genome Project Data Processing, 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lopez, J.V., Yuhki, N., Masuda, R., Modi, W., O’Brien, S.J., 1994. Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. J. Mol. Evol. 39, 174–190. https://doi.org/10.1007/BF00163806

Mack, A., Dumbacher, J., 2007. Birds of Papua. In: The Ecology of Papua, Part I. The Ecology of Indonesia Series VI. Andrew J. Marshall and Bruce M. Beehler (eds.). Periplus, Singapore, pp. 654–688.

McCormack, J.E., Harvey, M.G., Faircloth, B.C., Crawford, N.G., Glenn, T.C., Brumfield, R.T., 2013. A phylogeny of birds based on over 1,500 loci collected by target enrichment and high-throughput sequencing. PLoS ONE 8, e54848. https://doi.org/10.1371/journal.pone.0054848

Olofsson, J.K., Bianconi, M., Besnard, G., Dunning, L.T., Lundgren, M.R., Holota, H., Vorontsova, M.S., Hidalgo, O., Leitch, I.J., Nosil, P., Osborne, C.P., Christin, P.-A., 2016. Genome biogeography reveals the intraspecific spread of adaptive mutations for a complex trait. Mol. Ecol. 25, 6107–6123. https://doi.org/10.1111/mec.13914

Prum, R.O., Berv, J.S., Dornburg, A., Field, D.J., Townsend, J.P., Lemmon, E.M., Lemmon, A.R., 2015. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. Nature 526, 569–573. https://doi.org/10.1038/nature15697

R Core Team, 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Wessler, S.R., 2006. Transposable elements and the evolution of eukaryotic genomes. Proc. Natl. Acad. Sci. U. S. A. 103, 17600–17601. https://doi.org/10.1073/pnas.0607612103