Phylogeny and evolution of pigeons and doves (Columbidae) at different space and time scales

Jade Bruxaux

Université Fédérale
Toulouse Midi-Pyrénées

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par

Institut National des Sciences Appliquées de Toulouse (INSA de Toulouse)

Présentée et soutenue par :
Jade Bruxaux
le jeudi 8 mars 2018

Titre :

Phylogeny and evolution of pigeons and doves (Columbidae)
at different space and time scales

École doctorale et discipline ou spécialité :
ED SEVAB : Écologie, biodiversité et évolution

Unité de recherche :
Laboratoire Évolution et Diversité Biologique - UMR 5174 (UPS, CNRS, IRD)

Directeurs de Thèse :

Christophe Thébaud, Professeur, Université Toulouse 3 Paul Sabatier
Guillaume Besnard, Chargé de recherche, CNRS

Jury :
Marianne Elias, Directrice de recherche, ISYEB - UMR 7205, Rapportrice
Emmanuel Douzery, Professeur, ISEM - UMR 5554, Rapporteur
Céline Brochier-Armanet, Professeur, LBBE - UMR 5558, Examinatrice
Alice Cibois, Chargée de recherche, Muséum de Genève, Examinatrice
Pierre-André Crochet, Directeur de recherche, CEFE - UMR 5175, Examinateur

Remerciements

Je voudrais d'abord remercier mes deux directeurs de thèse, Christophe et Guillaume. Merci de m'avoir laissé une chance de mettre un pied dans la recherche. Je ne sais pas si je réussirai à y faire ma place, mais j'espère car j'ai vraiment aimé ces trois années (et quelques). Il n'y aura bientôt plus personne pour me demander si ça avance, la thèse, et ça va forcément me manquer un peu…

Merci également à Marianne Elias, Emmanuel Douzery, Alice Cibois, Céline Brochier-Armanet et Pierre-André Crochet d'avoir accepté d'évaluer ce travail. Il n'est forcément pas aussi abouti que je l'aurais souhaité, mais j'espère que sa lecture ne sera pas trop désagréable.

Merci enfin à Hervé Philippe, Léa Jacquin et Frédéric Delsuc pour leur disponibilité et leurs commentaires pendant mais aussi en dehors des comités de thèse.

This work should not have been done without all our collaborators, in museums (your collections are essential, no matter what some could say: keep going!) and elsewhere. I would like to especially thank Lounès Chikhi and his team at IGC for their hospitality for the two months I spent their: you are all doing great and inspiring science and I hope to keep collaborating! Thanks also to Mikkel Sidding and Tom Gilbert for their help dealing with ancient DNA. It was a really great experience for me! Finally, thanks for all those people I exchanged emails with, regarding papers, softwares, samples… and that I have finally met for some of them a few months ago in Paris (Paul Sweet, Thomas Trombone, Alice Cibois, Robert Prys-Jones, Knud Jønsson, Jérôme Fuchs…). Some discussions we had there were really inspiring, both professionally and personally.

Merci à tous ceux qui m'ont tendu la main, peut-être à un moment où je ne pouvais pas, où j'étais seule chez moi, et qui ont fait qu'aujourd'hui j'en suis là. Par ordre chronologique (en espérant ne pas en oublier trop) : Philippe Orsini, directeur du muséum d'histoire naturelle de Toulon, qui avait eu la gentillesse de recevoir une petite collégienne pour lui donner des conseils sur son stage de 3ème ; Benjamin Kabouche et Matthieu Lascève pour m'avoir mis un pied à l'étrier en ornitho (avec notamment une coche de Garrot à œil d'or aux Pesquiers) ; Daniel Poisson à Masséna, naturaliste passionné qui ne manquait jamais de me parler d'oiseaux entre deux khôles ; Marc Artois, Philippe Berny, Emmanuelle Gilot-Fromont, Marie-Pierre Callait-Cardinal et l'ensemble de « l'équipe Faune Sauvage » de l'ENVL qui soutenaient contre vents et marées les moutons noirs qui voulaient faire de

la recherche et non de la clientèle ; François Moutou, Vincent Dedet et les membres du groupe VFS pour leur gentillesse, leur ouverture d'esprit et leur passion qui donne envie ; Raphaël Musseau, Valentine Herrmann, Charly Souc et toute l'équipe de Biosphère Environnement pour toute l'expérience qu'ils m'ont permis d'acquérir en ornitho (il y a encore beaucoup du boulot, mais je compte bien profiter des prochains mois pour m'y remettre !) ; Tatiana Giraud, Antoine Branca, Jeanne Ropars, pour avoir presque déclenché une passion pour les champignons, mais avoir déclenché une passion certaine pour l'évolution, la génomique et surtout la recherche en général, tout en montrant que oui, c'est possible d'avoir une vie professionnelle réussie et pourtant équilibrée : un modèle !

Merci enfin à tous ceux qui font que la vie est plus douce au quotidien. Mes parents pour leur soutien indéfectible, qui se font du souci sans trop oser le dire. Merci de supporter ces choix professionnels pas toujours évidents (mais sinon ce serait trop simple !). Merci à ma sœur pour… tout. Pas d'autre mot. Merci. Merci évidemment au reste de la famille. Exilée en Corse pour une partie (2018, l'année du renouveau !). Mathieu (Dieu), pas d'inquiétude, je supporte toujours les rouges et noirs (mais les bons, je n'ai pas changé), et non je suis pas encore mariée (merci Lucile de me soutenir !). J'espère bien revenir faire une partie de carte sous peu ! Un immense merci à Christine pour son accueil chaleureux et tous nos échanges passionnants. J'en arriverais à aimer Paris (mais Port-Cros a ses atouts !).

Merci aux Randonneurs Toulousains, mes parents et grands-parents d'adoption à Toulouse !

Merci à tous les copains du lycée (Benoit !) et de l'ENVL (Alex, Myriam, Lauriane, Pauline, Charlotte, Aline, Joséphine, Séverine, Pernelle, ma poulotte Claire) pour être toujours là. Je compte bien profiter des prochains mois pour faire un tour de France et faire la connaissance de toutes ces petites bouilles ! Un énorme merci à Valentine de venir du bout du monde pour moi, ça me touche beaucoup plus que je ne serai jamais capable de le dire.

Merci à tous les copains du labo, grâce à qui se lever tous les matins est un bonheur : Lucie qui a renié le bureau 30 pour le 33 (mais tu me fais tellement rire que je ne peux pas t'en vouloir) : je te souhaite le meilleur pour la suite en Oklahoma ! Jess et ses petits repas aux invitations étudiées (la maman de la moitié du labo !). Pierrick et sa bonne humeur infaillible (un jour j'arriverai à te faire sortir ! pour un karaoké ?). Seb, Félix, Jan et Céline pour toutes les randos (un jour j'arriverai peut-être à vous suivre), Marine pour ses petits mots, ses câlins et sa motivation à toute épreuve pour entendre un jour de l'accordéon (promis). Merci à tous les anciens qui sont partis vers d'autres horizons mais nous ont fait

passer de supers moments au labo (Aurore ! un dynamisme à toute épreuve !) et en dehors (Isa, Boris, Arthur et toute la bande) voire au bout du monde (Joss, Yann, Jacques : de supers souvenirs de fous-rires à la Réunion). Merci à toute la bande des petits nouveaux, qui apporte une énergie d'enfer au labo (Fabian, Luana, Maxime, Isa, Alex et les autres !) et un merci tout spécial à Maëva pour avoir été une stagiaire au top, et pour être une super thésarde. Merci pour ton aide dans toutes les analyses, et tu vois, on y arrive ! ;) Donc fonce, avec toutes tes qualités, ça marchera forcément ! Merci aussi aux nouveaux un peu plus vieux (mais pas beaucoup plus que moi) : Jordi, Nico (pour les discussions passionnantes), Emilie (juste pour être toi, aussi improbable que ça puisse paraître [tu veux la liste des jours fériés 2018 ?]). Et enfin, last but not least, un IMMENSE merci à Kévin. Pour l'ambiance du bureau 30 (Seb et Lucie ont rien compris, ne leur en veux pas), pour les hangouts, pour les baleines (ma réputation est ruinée mais ça valait la peine, mais stop quand même maintenant), pour les week-end jeux et les dimanches Baraka, pour la blanquette, les karaokés… et tout le reste. Je n'ai aucun doute que tu réussiras, et j'espère bien pouvoir te rendre visite au Japon !

J'espère enfin que ces quelques pages vous convaincront que les pigeons sont beaucoup plus fascinants qu'on pourrait le croire…

Table of content

Introduction

1. Understanding the species distribution: an historical perspective

"Biogeography is the study of the facts and the patterns of species distribution. It's the science concerned with where animals are, where plants are, and where they are not. [...] biogeography does more than ask *Which species?* and *Where*. It also asks *Why?* and, what is sometimes more crucial, *Why not?*" David Quammen (The Song of the Dodo: Island Biogeography in an Age of Extinctions)

With the intensification of naturalist expeditions during the 18th century, knowledge on species distribution has rapidly increased revealing that biodiversity is unevenly distributed on Earth, which has early stirred up biologists' curiosity. When Linné began his description and classification of species in *Systema Naturae* (Linné, 1735), he explained the existence of different altitudinal ecosystems by the fact that plants and animals colonized them gradually from the Noah's Ark when the Great Flood receded. A few decades later, Buffon observed that a similar climate could host different species under different parts of the world; e.g. between warm climate in Old and New world [what will later be called the Buffon's law (comte de Buffon, 1756)]. He therefore opposed to Linné's idea, which would have implied finding the same species in similar climates. He even suggested that Africa and South America were continuous before their separation by the Atlantic Ocean, which lead to the splitting of one pool of species (comte de Buffon, 1766). Forster confirmed the Buffon's views through the study of plants during the second Cook's voyage around the world (Forster, 1778). He also observed that tropical regions were more species-rich than temperate or polar areas (what will later be called a latitudinal gradient of species richness) and that islands were poorer in species than mainland due to their size, thus introducing first concepts on island biogeography. Therefore, at the end of the 18th century, most of the foundations of biogeography were already stated. However, an important dimension was still lacking: time. Due to their Christian educations, all those naturalists believed in one event of creation. A major advance in biogeography was the development of the theory of evolution during the next century.

When Lamarck stated that organisms were changing through time following the environment in which they were living (Lamarck, 1801), the idea did not convince the scientific community, and especially not Cuvier, the father of comparative anatomy. However, when both Wallace and Darwin travelled around the world a few years later, the concept had emerged and allowed them to explain

small differences they had observed between species on close islands (Darwin & Wallace, 1858). What was really new with their contribution was the underlying mechanism: the natural selection. They both stated that there were slight differences among descendants of two parents. Some of these differences could favor their host, which would therefore increase its probability to survive and then transmit these differences to its own progeny. Alongside his tentative to explain little differences between close islands, Wallace also faced a major phenomenon in Indonesia: a huge demarcation between the fauna of two groups of islands despite short distance between them (later called the Wallace's Line). To explain differences between fauna of different islands, he studied mechanisms such as climate, stratigraphy, dispersion or competition (Wallace, 1876). He was therefore among the first biologists who developed the biogeography of animal species (or zoogeography). The present distribution of all taxa was however difficult to explain only with dispersion from an ancestral area. In particular, the *Glossopteris* flora (Cox & Moore, 2010) in the southern hemisphere is among the most striking distribution patterns to explain (Figure 1).

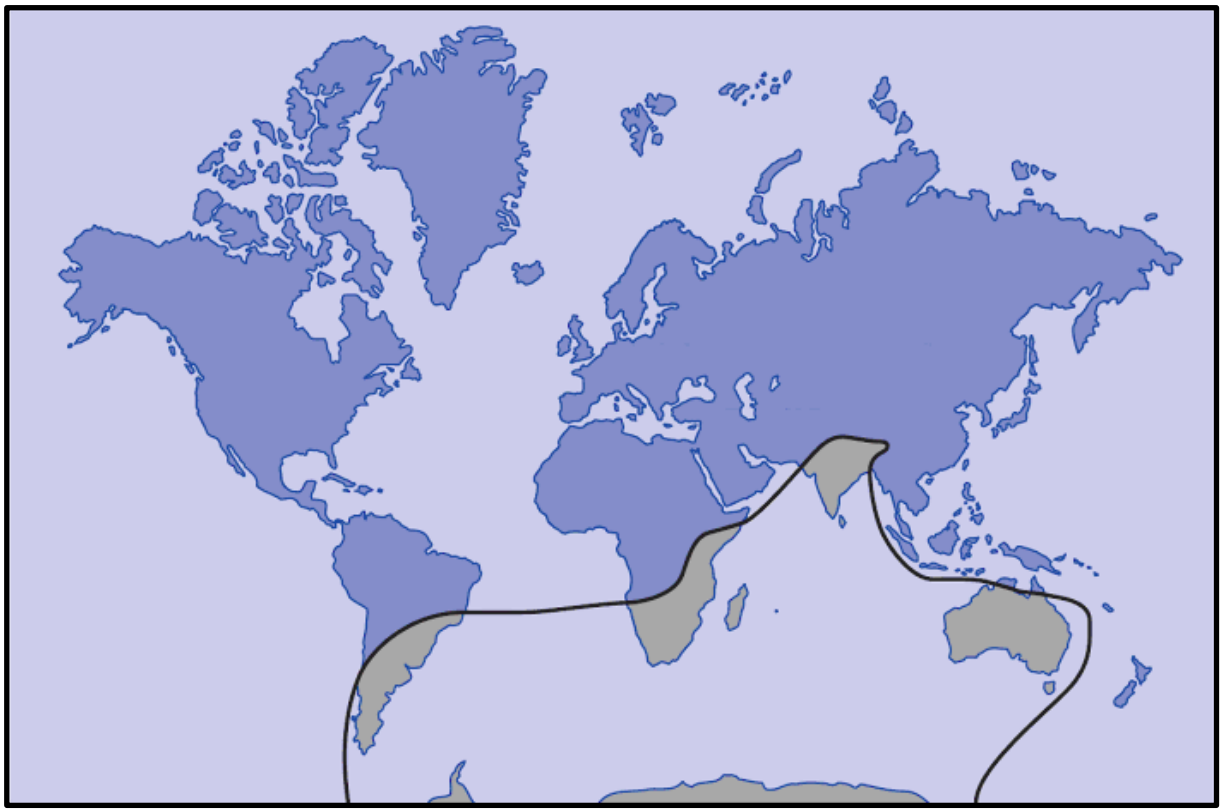


FIGURE 1: DISTRIBUTION OF THE *GLOSSOPTERIS* FLORA (COX & MOORE, 2010)

When the meteorologist Alfred Wegener proposed his continental drift theory (Wegener, 1912), it allowed a more parsimonious explanation for this type of distribution. He had however no mechanism to explain it, and did not manage to convince scientists. It was only widely accepted in the 1950s, after Arthur Holmes (Holmes, 1931) and later Samuel Carey (Carey, 1959) were able to clarify underlying mechanisms. Geology, alongside with ecology, climate changes and evolutionary biology are thus all involved to understand the current species distributions. However, if the species concept was rather clear when first biogeography studies were developed (based mainly on morphological similarities between individuals), it seems that the more scientist work on the subject, the less they are able to define it (Mayden, 1997)…

2. The different species concepts

The biological definition, known of everyone, states that a species is a group of individuals that are able to reproduce and whose offspring are themselves fertile. Horses and donkeys are therefore species, but mules and hinnies are not. It was clarified by Mayr (1942) as: "Species are groups of actually or potentially interbreeding natural populations, which are reproductively isolated from other such groups".

However, numerous situations break down this definition. First, it can only apply to sexually reproducing species, and therefore excludes many taxa such as bacteria and most of fungi. Second, it can only apply to extant species, as it is in many cases impossible to know from fossils how individuals could reproduce. Nevertheless, even for extant sexually reproducing species, the reproductive isolation has been recently found difficult to prove.

During the sexual reproduction, each parent transmits half of its characteristics to its offspring through the transmission of half of its genetic information (see Box 2 for more details). At the population scale, it allows a mixing of the whole information and the transmission of characteristics such as resistance against diseases or antibiotics, adaptation to altitude… This phenomenon is called "gene flow" and its absence has long been considered as a prerequisite for speciation (even if recent advances in genetics showed that it is probably more the exception than the rule; see 3.2). On the other side, two populations of a largely distributed species can have differentiated enough to be unable to reproduce anymore (examples of cline and ring species described in 3.2). Therefore, the biological definition of species based on reproduction presents some limits.

Another frequent view of the species is the phylogenetic or evolutionary concept, proposed by Wiley (Wiley, 1978): “a species is a lineage of ancestral descendant populations which maintains its identity from other such lineages and which has its own evolutionary tendencies and historical fate”. It is based on a more historical concept (all the descendants of a common ancestor). Its use has widely increased with the development of biogeographic studies that include genetic data, as these data compared together allow us to determine if individuals have a recent or older common ancestor (see Box 2 for more details). This species definition is at the basement of the barcoding approach (Hebert et al., 2003). The main idea of this concept is that genetic sequences can be considered as an identity card for each individual, including its affiliation to a species. If we compare sequences which evolve quickly enough to be different between species, but slowly enough to allow very little variation at the intraspecies level, we should be able to assign a species to each individual analyzed. Different markers were chosen depending on the kingdom: COI for animals (Hebert et al., 2003); *matK*, *rbcL*, *trnH-psbA* or ITS2 among others for plants. The number of potential markers in plants is indicative of a major limit of barcoding: one single marker can hardly fit the whole kingdom species delimitation. Even in animals where COI makes consensus, this mitochondrial gene is not suitable for all families (e.g. Meier et al., 2006). Other options have been proposed to delimitate species, using more markers at the population scale. They are based on the idea that each gene evolves through its own history, and therefore comparing different versions of this gene (called alleles) in different individuals allow us to find when the ancestral form occurred (coalescent theory; Kingman, 1982). With this method, even recent speciation events can be detected (Ence & Carstens, 2011), where barcodes struggled. Moreover, the use of different markers allows avoiding artifacts due to incomplete lineage sorting (where the two species do not recover the same sampling from the ancestral alleles, and artificially show a gene tree different from the species tree). Some particular cases like species radiations (where many speciation events occurred through a short period of time) still need specific caution to define species based on genetic data (Meiklejohn et al., 2016).

Both evolutionary and morphological / biological species concepts present their own limits, and their results do not always agree. While genetic data allows the discovery of cryptic (Bickford et al., 2007) and mimetic species (Brower, 1996), the approach sometimes suggests grouping different species into one. This is notably the case for sexually dimorphic species, for which male and female had been previously described separately by taxonomists (Nakahara et al., 2018). Finally, we also know today that genetic material can be exchanged at very large taxonomic distance via lateral gene transfer: from bacteria to nematodes (Danchin et al., 2016) or from bacteria to insects (Shelomi et al., 2016). The extent of such exchanges is not yet very well documented, but demonstrates that genomes are permeable, which may imply that usage of a single marker should be avoided.

As this work does not pretend to give an answer to the tough question of "what is a species?", we will use species names as they are stated in recognized databases such as the Handbook of the Birds of the World (HBW), the International Ornithological Congress (IOC) World Bird List or the International Union for Conservation of Nature and Natural Resources (IUCN) red list. Those definitions are mainly historical (and thus based on morphology and reproduction limitations), but are regularly updated following genetic analyses. They are useful for the biodiversity monitoring as species can be seen as units of conservation, but an absence of a consensual concept questions the relevance of such approaches (Fraser & Bernatchez, 2001). Understanding processes that generate biodiversity is thus essential in conservation biology, as well as in fundamental research.

The current biodiversity results both from historical processes of speciation, dispersal and extinction (Roy & Goldberg, 2007) and from ecological constraints (Currie et al., 2004). However, the relative importance of the different historical processes through space and time in shaping the richness distribution is still unclear (Ricklefs & Jønsson, 2014), as is the exact context of most speciation events. This works aims at shading light on the complex scenario that can underline a non-homogeneous current species richness. Diversification is the most studied process (*e.g.* Magallón & Sanderson, 2001; Stadler & Smrckova, 2016; part 3), but understanding the species relationships and dating the speciation events are key priors to this type of study (part 4).

3. Speciation mechanisms

3.1. The role of population isolation in speciation

When Wegener proposed his theory of continental drift (Wegener, 1912), two major mechanisms of isolation were in competition to explain species distribution. While they were supported by different intellectual currents and seemed therefore incompatible, it is now clear that these two mechanisms coexist. The first way to explain speciation is the dispersion followed by isolation. In this context (Figure 2), a few members of an established species manage to cross a barrier.

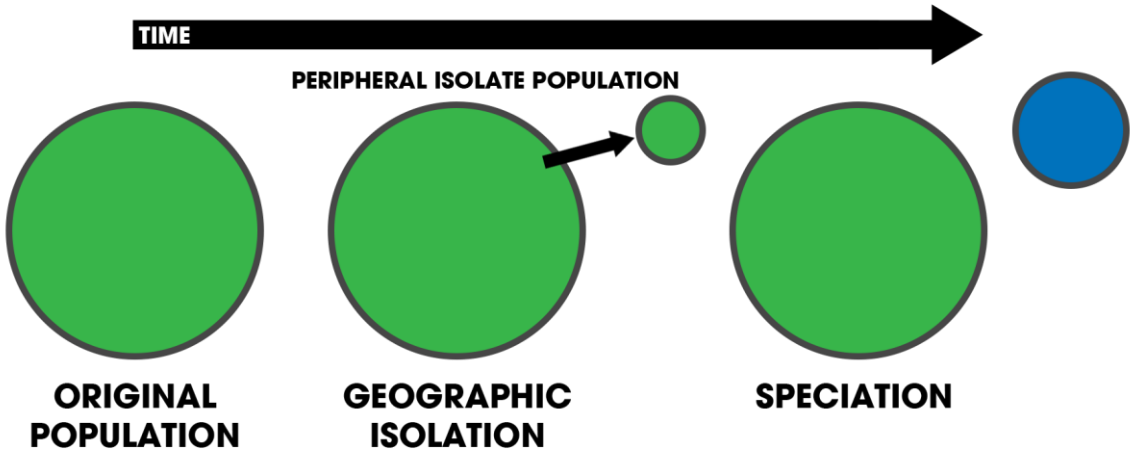


FIGURE 2: SPECIATION BY DISPERSION ACROSS A BARRIER (© ANDREW Z. COLVIN / CC-BY-SA-4.0)

The second mechanism, vicariance, is due to the appearance of a barrier in the species distribution, splitting the population into two (Figure 3).

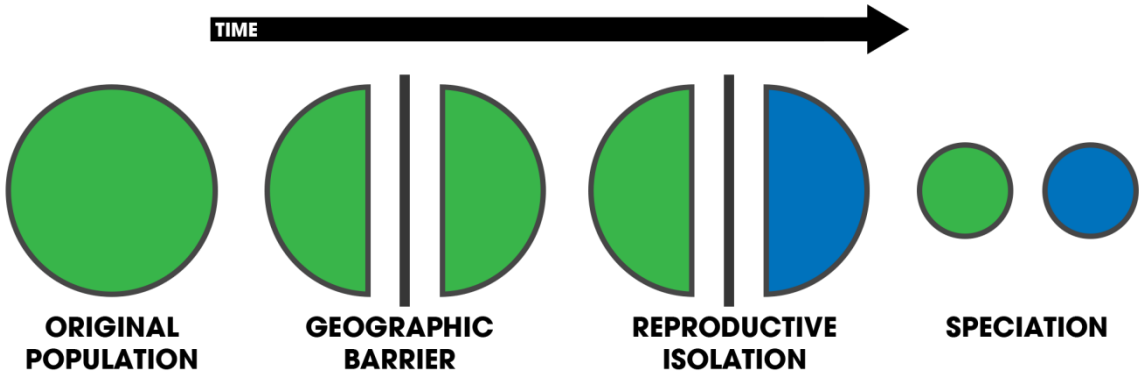


FIGURE 3: SPECIATION BY VICARIANCE DUE TO THE APPEARANCE OF A BARRIER IN THE INITIAL DISTRIBUTION (© ANDREW Z. COLVIN, CC-BY-SA-4.0)

In these two contexts, barriers can take various forms. The most obvious is the geographic barrier: high mountains or large oceans can be difficult to cross, even for birds (whose flight ability varies greatly between species). However, climatic and/or biotic barriers exist as well. During glacial periods, a mountain that could previously be crossed can host insurmountable glaciers (Figure 4).

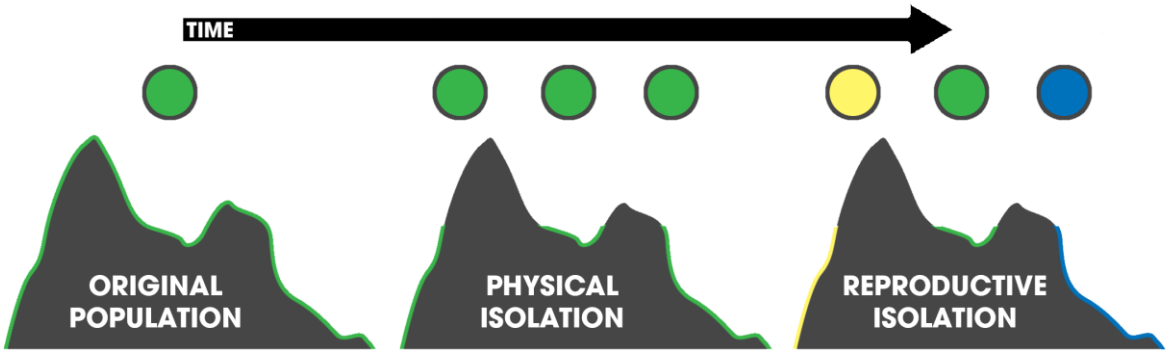


FIGURE 4: SPECIATION BY VICARIANCE DURING A GLACIAL PERIOD (BASED ON © ANDREW Z. COLVIN, CC-BY-SA-4.0)

On the other side, warming and drying periods can break up a previously continuous distribution through the appearance of unsuited environments like desert or savannas (Figure 5).

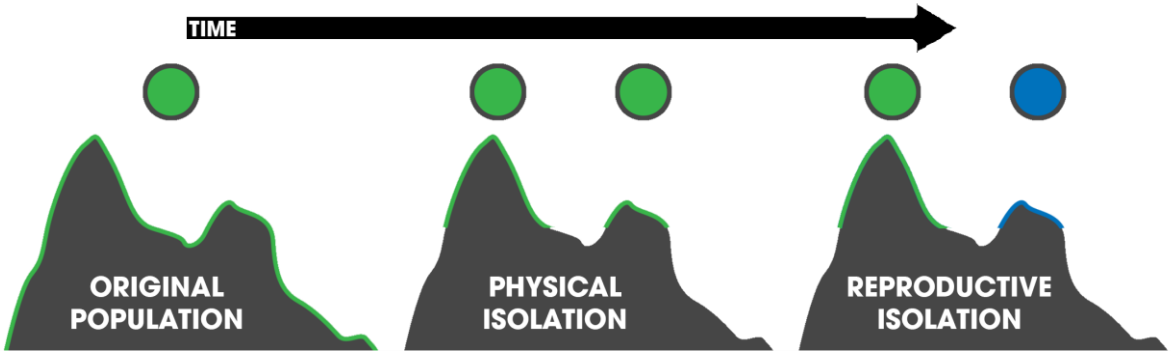


FIGURE 5: SPECIATION BY VICARIANCE DURING A WARM PERIOD (BASED ON © ANDREW Z. COLVIN, CC-BY-SA-4.0)

At the genetic scale, the presence of barriers in these two scenarios has a common consequence: sexual reproduction does not occur anymore and the gene flow between the two populations stops. Most of the genetic modifications that occur in the genome are due to non-corrected errors that occur before cell division. When the mutation happens in germinal cells, it can be transmitted to offspring. As these errors occur and are inherited by chance, the allele frequency changes through time. This stochastic phenomenon is called genetic drift (Wright, 1931) and can be responsible of differences occurring between two distinct populations. However, when two populations are separated by geographic barriers, they can also live in slightly different environments, and therefore be subject to evolutionary pressures which will select for different mutations depending on the group. In the absence of gene flow which mixes all the alleles, the two populations will gradually diverge with time, until phenotypic, behavioral and/or genetic differences does not allow reproduction between them anymore.

However, since the description of these two major speciation processes involving population isolation, others have been described, which could occur in a context of gene flow and/or in sympatry.

3.2. Speciation with gene flow

One case of speciation with gene flow is the cline speciation, whose famous example of ring species is often taught in textbooks (Mayr, 1942). Due to large distributions, some species can be present in an environment that progressively changes from one extremity to the other. Some variations are well described, like the Bergmann's rule, stating that species are larger in colder environments than in warmer ones (Bergmann, 1848), or the Gloger's rule, stating that endotherms are darker in more humid environments (Gloger, 1833). In such large distributions, even if geographic barriers are absent and gene flow occurs, individuals can only reproduce with closely sited relatives. It could give birth to really different populations at each extremity, populations which could ultimately be unable to interbreed. It can be observed practically when these two extremities meet after a range expansion on both sides of a barrier, forming a ring shaped distribution (Figure 6).

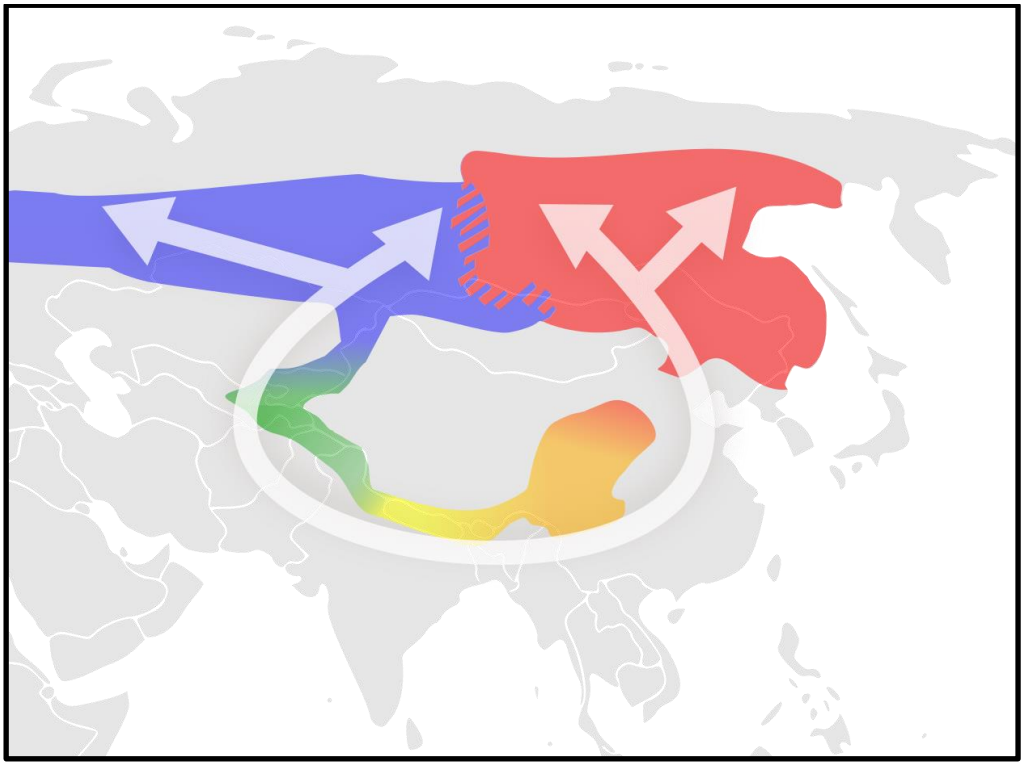


FIGURE 6: EXAMPLE OF THE GREENISH WARBLER *PHYLLOSCOPUS TROCHILOIDES* RING SPECIES (© G. AMBRUS, PUBLIC DOMAIN)

A few classical examples are famous, like the salamander *Ensatina eschscholtzii* in California, the herring gull *Larus argentatus* around the North Pole, or the greenish warbler *Phylloscopus trochiloides* around the Himalaya. In these three cases however, recent studies showed that the current absence of barrier is not representative of the whole species history, and that gene flow stopped for some time (Highton, 1998; Liebers et al., 2004; Alcaide et al., 2014). This has to be related with differences between time of speciation (often measured in millions of years) and ecological variations (with last glacial maximum "only" 26,500 years ago): it seems highly improbable that present distribution reflects what it was through the whole process of speciation.

Another example of speciation with gene flow is the sympatric speciation, where two species diverge one from another in a same place (Maynard Smith, 1966). Following Coyne & Orr (2004), a speciation event can only be considered as sympatric if (1) the two species are found at least partly in sympatry, (2) they are sister species, (3) reproductive isolation is proved, and (4) there is no reason to suppose that an allopatric period occurred. If the first three criteria can easily be proved, the last one is often struggled by the improvement of population genetics analyses (Grant & Grant, 2010), as it is the case with the ring species. However, despite discussions on the possibility of other scenarios (Stuessy, 2006), a few examples are still considered as sympatric speciation (Savolainen et al., 2006), including in birds (Sorenson et al., 2003). To clarify the existence or the absence of an allopatric period, recent methods have been developed, like the ones based on the Approximate Bayesian Computation (ABC) (Csilléry et al., 2010). Their use, eased by the increased computational power, has already nuanced previous results of sympatric speciation (Martin et al., 2015).

But without going as far as cline or sympatric speciation, it is now clear that gene flow can occur during what was previously considered an allopatric speciation process (e.g. Roux et al., 2013; Le Moan et al., 2016). In a context of ecological differences between two populations, if this gene flow does not concern the genomic parts subject to strong divergent selection, it should not delay the speciation process (Yang et al., 2017). However, if current methods allow to date and quantify the amount of gene flow, some results are still subject to discussion. First, the speciation process is not instantaneous and gene flow occurring at the beginning of this period, when the two populations are not yet different species, could be falsely detected as ancient gene flow (Yang et al., 2017). Moreover, both ancestral population structure and change of population size can be confounding factors and are not easy to take into account within the same model of gene flow presence/absence (Yang et al., 2017).

3.3. Instantaneous speciation through polyploidization or mutations

While previous mentioned mechanisms of speciation involve a gradual divergence of populations, faster processes have been described, even in sympatry but without (or much reduced) gene flow. In particular, the polyploidization, "the presence of three or more chromosome sets in an organism" (Grant, 1981), is a speciation process that is well known in plants where 15% of angiosperm and 31% of fern speciation events come with a ploidy increase (Wood et al., 2009). Polyploid animals are described as well [e.g. the African lungfish (*Protopterus dolloi*), the whole Salmonidae family, the North American treefrog *Hyla versicolor* and the red viscacha rat (*Tympanoctomys barrerae*) (Gregory & Mable, 2005)], but with the notable exception of birds. Their origin can be classified in two categories: autopolyploids resulting from genomic doubling, gametic non reduction or polyspermy; and allopolyploids that are the consequence of hybridization of two different species (Otto & Whitton, 2000). Regardless of the mechanism, the difference of chromosome numbers immediately prevents reproduction between diploids and polyploids. This speciation process seems to have played a major role through evolution, as it could be at the origin of a major part of gene redundancy observed today, and a transitional phase for many lineages (Soltis et al., 2004). Another mechanism of rapid speciation may involve sudden mutations with large effects (including genomic re-organizations) that led to the ecological and reproductive isolation of the mutant from its parental species. Such processes have been described in *Drosophila* (*e.g.* Masly et al., 2006) or in *Agrodiaetus* butterflies (Kandul et al., 2007).

With these few examples, we can already see that the speciation process can be studied at very different scales: from the genetic signature at the intraspecies level, to the worldwide dimension of species distribution, the first being partially responsible of the second. However, in all cases, it is essential to determine first the relationships between the individuals or species studied. To do so, we used phylogenies, based here on molecular data.

Box 1: What is a phylogenetic tree?

A phylogenetic tree is a type of diagram that shows relationships between entities called clades. The term “clade” is used to speak of all the descendents originating from a common ancestor and can include individuals, populations, species... A phylogenetic tree can be roughly compared to a family tree, with little differences. Each leaf corresponds to a clade and each intersection between two branches to a virtual ancestor, living at the time of the speciation event if the two resulting clades are different species (Figure 7).

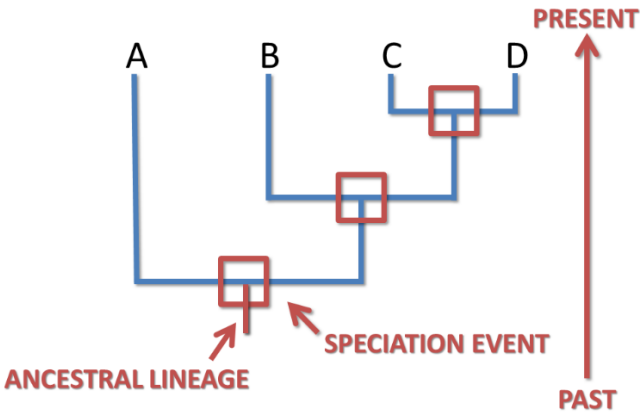


FIGURE 7: EXAMPLE OF PHYLOGENETIC TREE WITH FOUR DESCENDENTS, SHOWING ANCESTRAL LINEAGE AND SPECIATION EVENT

The farther this ancestor is from the leaves, the older it is. As in genealogic trees, there is a notion of time in phylogenetic trees. However, one of the biggest differences between these two concepts is the idea of evolution. On each branch, a few to many generations occur, corresponding to the unique history of the clade, or to a history shared with a sister clade (Figure 8). Since the most recent common ancestor of two clades, both of them evolved, meaning that characters transmitted from one generation to the following changed.

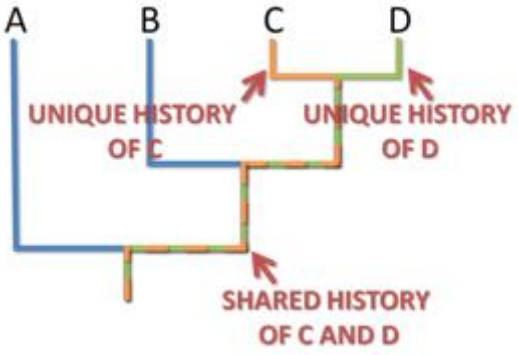


FIGURE 8: EXAMPLE OF PHYLOGENETIC TREE WITH FOUR DESCENDENTS, SHOWING SHARED AND UNIQUE HISTORY OF TWO DESCENDENTS

When different lineages group together all the descendents of a common ancestor (e.g. lineages B, C and D in Figure 8), the clade is called monophyletic. If it is not the case (e.g. lineages A, C and D in Figure 8), the group is called paraphyletic. Taxonomic groups such as genera or families should be monophyletic. If ancient group definitions do not correspond to monophyletic entities, they are

revised (for example, reptiles do not have any significance anymore, as birds are nested into this group).

Today, phylogenies are used to define species (see above), to describe communities (e.g. Matos et al., 2017) or host-parasites co-evolution (e.g. Sweet et al., 2017), to document the spread of an epidemic (e.g. Gire et al., 2014), or to reconstruct the evolutionary history of a clade (e.g. Cibois et al., 2017).

4. Phylogenies: an essential tool in biogeography

Tree diagrams were used for centuries to depict family pedigrees (genealogical trees) but became more popular to represent species relationships in the early 19th century (Augier, 1801), when scientists began to describe similarities between species to classify them. When Darwin drew his first famous tree in his Notebook (Darwin, 1837), later clarified in the Origin of Species (Darwin, 1859), he changed the meaning of the branches: affinities were inherited and subject to natural selection (see Box 1 for more details). However, formalizing rules to reliably construct these trees happened only in the middle of the 20th century, when the computers became accessible for scientists (Felsenstein, 2004).

The initial idea of grouping together the most similar species gave birth to both parsimony and distance matrix methods. Parsimony is based on a minimum number of events needed to explain differences between two species (Edwards & Cavalli-Sforza, 1963), when distance matrix methods measure directly the number of differences, grouping together the less different species (Cavalli-Sforza & Edwards, 1967; Fitch & Margoliash, 1967). Maximum likelihood and Bayesian statistics were also adapted for tree reconstruction (Edwards & Cavalli-Sforza, 1964; Rannala & Yang, 1996). All these methods can be used to every discrete character, either phenotypic or genetic one, postulating they are inherited (homologous).

Phylogenetic trees cannot only represent relationships between individuals or species (dendrogram), but also the evolutionary distance separating the studied entities (phylogram). This distance is shown through the branch length. When molecular data are used to reconstruct a phylogenetic hypothesis, the raw genetic distance has to be corrected to better reflect evolution, as different mutations can occur at the same position, masking the previous events. To consider this, different evolutionary models were proposed. The first and simplest one, the Jukes-Cantor model (Jukes & Cantor, 1969) assumes an equal probability for each type of substitution, no difference regarding the position, and bases in equivalent proportions. However, this model is rather unrealistic,

as bases frequencies are not equal, some substitution types are more frequent than others, and some regions are less prone to changes than others. Therefore, other models were proposed to evaluate more accurately the true genetic distance [e.g. the HKY model (Hasegawa et al., 1985); the general time-reversible model (GTR; Lanave et al., 1984; Tavaré, 1986)]. Finally, most models allow a slight variation around the mean substitution rate [gamma distributed rate (Jin & Nei, 1990)] and/or the existence of invariant sites (Hasegawa et al., 1987). To take into account evolutionary variation between regions, data can be partitioned following their nature (coding sequence, tRNA…) and different models can be applied to each.

Choosing the more representative model is not easy. Balancing realistic models, overparameterization and computing capacities requires model selection criteria (Sullivan & Joyce, 2005). This choice can be based on Likelihood-Ratio Tests (LRT), Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC), the last one being more suitable for large dataset, as it penalizes overparameterization and thus does not select the most complex models with a large number of sites (Sullivan & Joyce, 2005). Finally, two tests of model adequacy are often used to evaluate the reliability of the resulting tree: parametric bootstrap and Bayesian posterior probability. The first one uses the selected model and pseudoreplicates of the dataset (random draw of sites with replacement) to reconstruct trees, in which the number of occurrence of each node is counted (Felsenstein, 2004). The second uses all the possible trees reconstructed through the Bayesian inference to count the proportion which contains each node (Felsenstein, 2004). These two methods allow us to statistically evaluate the confidence we can have in the resulting tree.

With a reliable phylogram, whose branch length are proportional to evolutionary time separating two clades, it is possible to link speciation events with major climatic or geological events such as sea level fluctuation, sea opening, mountain raising, island emerging… It also allows comparing results with previous dates established by other studies. Two main methods are often used for dating. The first one is based on a molecular rate of evolution that is known from previous studies on this clade or on close relatives. The second one uses dated fossils or major geological events that give clues on appearance or disappearance of different clades. These two methods have their own advantages and limits.

Molecular dating is based on the idea that the number of changes in genomic sequences is mostly stable over time. This was initially demonstrated for mitochondrial genomes of primates and rodents (Brown et al., 1979) and extended to birds. One study focused especially on the mitochondrial cytochrome b gene (Weir & Schluter, 2008), where the molecular rate was evaluated at 2.1% (+/-0.1%). However, it is now clear that this rate can vary through time (Ho et al., 2005), across lineages and

among genes (Nabholz et al., 2016). This method is therefore particularly suited at low taxonomic scale, where improvements have been proposed to take taxonomic or ecological data into account (Nabholz et al., 2016; Quillfeldt, 2017). It is also an option when external calibration is impossible.

Calibrating trees with fossils or geological events is usually done by imposing minimum or maximum age for some speciation events (Benton et al., 2009). A fossil belonging to a genus gives a minimum age for this clade and repartition of fossils can statistically generate a maximum bound (Benton et al., 2009). As using erroneous calibration can have dramatic consequences, it is critical to choose which fossils can be used or not. A fossil should be used only if correctly documented (museum number, provenance…), described by comparison with extant species, precisely located in stratigraphy and dated (Parham et al., 2012). Fossils can also be used like other extant species, as dated tips, and not to calibrate speciation events (Heath et al., 2014). On the other side, geological events, which are often used as maximum bounds, should be considered with caution, as speciation does not always occur simultaneously with the geological event that splits or increases the initial distribution (Pillon & Buerki, 2017).

When a reliable dated phylogenetic tree is obtained, it allows the study of richness distribution through different ways. It permits to hypothesize a history of speciation and dispersal through different regions (*e.g.* Claramunt & Cracraft, 2015), to evaluate speciation and extinction rates (Paradis et al., 2013), to compare communities compositions (Graham & Fine, 2008)…

5. The system model: the Columbidae family

During my thesis, I investigated the biogeography of pigeons and doves (Aves: Columbidae), a worldwide distributed family of birds, as our knowledge on the diversification of this group is still incomplete and insufficient to understand how a family could become both widespread and species rich, with an uneven richness distribution. Intuitively, we could think that a worldwide family has reached its distribution thanks to good dispersing capacities. However, these capacities would allow mixing individuals at large scales. Therefore, gene flow should be high, and speciation lowered. Being both widespread and species-rich is thus surprising at first sight, and requires further study.

Columbidae is a relatively large and diverse family, but it remains poorly known to the general audience. If most people thinking of pigeons imagine a medium-sized dull bird, this description does not pay tribute to the wide color diversity that can be observed within the family (Figure 9).

FIGURE 9: COLOR, SIZE AND MORPHOLOGICAL DIVERSITY IN THE COLUMBIDAE PHYLOGENY. IMAGES ARE SCALED AND REPRODUCED FROM DEL HOYO & COLLAR (2014) WITH PERMISSION OF LYNX EDICIONS. FROM LEFT TO RIGHT, TOP TO BOTTOM: *DREPANOPTILA HOLOSERICEA*, *CHRYSOENA VICTOR*, *GALLICOLUMBA RUFIGULA*, *COLUMBA LIVIA*, *CALOENAS NICOBARICA*, *ALECTROENAS PULCHERRIMUS*, *GOURA VICTORIA*, *DIDUNCULUS STRIGIROSTRIS*, *STARNOENAS CYANOCEPHALA*, *TRUGON TERRESTRIS*, *OTIDIPHAPS NOBILIS*.

The overall morphology is surprisingly conserved across the whole family (Gibbs et al., 2001). Only a few individuals differ markedly by their beak shape (the Tooth-billed Pigeon *Didunculus strigirostris*, the Thick-billed Ground-Pigeon *Trugon terrestris*), their long legs (the four pheasant-pigeons *Otidiphaps* spp.) or their very large size (the four crowned pigeons *Goura* spp., which weigh around 2.2 kg, when the Rock Pigeon *Columba livia* weighs 200-300 g).

The most studied pigeon is the rock pigeon (both wild and domesticated populations), a species sharing its history with humans for millennia. Wild rock pigeons could have been occasionally exploited for food since the Middle Pleistocene, with regular exploitation by Neanderthals and modern humans for over 40,000 years in a Gibraltar cave, the oldest evidence having been dated at 67,000 years ago (Blasco et al., 2014). The domestication by itself occurred probably in the Middle-East between 3,000 and 10,000 years ago, maybe several times in different places, but with at least proof of thousands of captive birds used as offering in Egypt 3,200 years ago (Johnston & Janiga, 1995). Since then, domestic pigeons are still bred for food, but also for esthetic purposes and sport competitions, where they exert their orientation capacity to cross large distances. This ability, which previously made them useful in communication, is now highly studied by biologists as a model in bird magnetic sense

(e.g. Mora et al., 2004). This species is also highly studied in genetics (Shapiro et al., 2013), physiology (Jacquin et al., 2011), or medicine (Levenson et al., 2015). Despite this, many other Columbidae are not as granted and a lot of knowledge is still lacking for most species.

From an ecological point of view, the understudying of pigeons is surprising, as their function of seed disperser is now clearly established (e.g. McConkey et al., 2005; Bucher & Bocco, 2009; Wotton & Kelly, 2012; Perea & Gutiérrez-Galán, 2016). Moreover, with already 17 extinct species (including the Dodo and the Rodrigues Solitaire from the Mascarene islands; IUCN red list, 2017) and more threatened species than expected by chance (Owens & Bennett, 2000), the Columbidae deserve more consideration. Major threats include both habitat loss and direct impact by humans or invasive species (Owens & Bennett, 2000). Understanding speciation and diversification of the family could therefore allow targeting regions and/or species as conservation priorities.

On the biogeographic side, they are interesting for several reasons. As previously stated, they are worldwide distributed, absent only from Antarctica, Hawaii and Easter island (Steadman, 1997), the last two islands having been only recently colonized by four and one species respectively (del Hoyo & Collar, 2014). They can be found in very different environments: dense forest, savannah, high mountains… The family is species rich (354 species; del Hoyo & Collar, 2014) and this richness is not equally distributed, attaining its maximum in New Guinea, especially in lowland forests with a maximum of 32 species per 0.5°x0.5° area (Figure 10) and up to 20 species which can co-occur (Pratt & Beehler, 2015). It makes them a good model to study diversification and dispersion, from the local scale to a worldwide perspective, and to understand better the non-homogeneous distribution of the species richness.

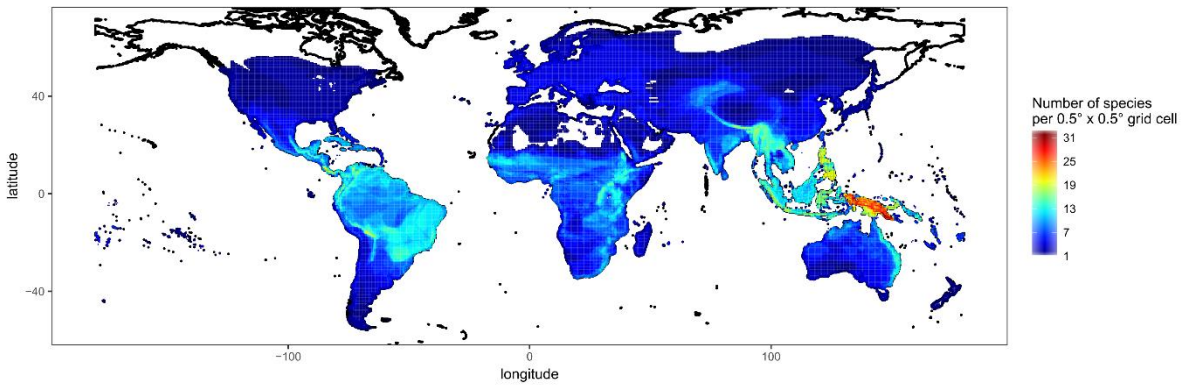


FIGURE 10: SPECIES RICHNESS OF THE COLUMBIDAE FAMILY, PER 0.5°X0.5° GRID CELL. DATA FROM BIRDLIFE INTERNATIONAL AND HANDBOOK OF THE BIRDS OF THE WORLD (2017).

6. Aims of the thesis

The central objective of this thesis is to describe the dispersion and diversification of a worldwide-distributed family, and to understand the non-homogeneous species-richness distribution. In a first chapter, we will focus on the global pattern by reconstructing a phylogeny for the whole family, with a representative sampling including at least one species per genus. This preliminary work will allow us to understand how the family dispersed to reach its current distribution and how it diversified despite good dispersal capacities. As Columbidae are especially present on islands (almost 60% are island restricted), we will question the importance of this geographical entity throughout the whole process. Are islands and especially archipelago responsible of recent and rapid speciation events through radiation? Do they play a role throughout the whole history by allowing speciation and/or dispersion? Moreover, the few morphologically divergent species seems to be mainly island species but also species from single-clade genera. Extreme morphologies result often from rapid radiation, with adaption to new ecological niches (*e.g.* Grant, 1999; Friedman, 2010). We will therefore look at these species in particular to see if they result from recent explosive diversification, as often described on islands.

To understand the process at a finer scale, we will focus in a second chapter on the speciation process that occurred in New Guinea, the richest part of the world for Columbidae, and resulted in particular in the four crowned pigeon species (*Goura* spp.). These species, which stand out from others due to their size and the crest they display on their head, have a non-overlapping distribution in a continuous lowland forest round the island. The aim of this chapter is to assess the relative importance of diversification within New Guinea or in places like Australia before the emergence of this young island, and to give a first explanation at the extreme observable richness on this island.

Finally, we will tackle a few ensuing question in discussion, including the whole Columbidae diversity in New Guinea, which could partly be the results of the same process as the crowned-pigeons'. We will especially discuss the importance of diversification on the island and of colonization.

Box 2: From the sample to the genome.

Birds, like all animals, are multicellular organisms. Each cell contains information inherited from both parents and responsible of observable characters (as known as phenotypic characters) such as size, color…This information is mainly transmitted via long molecules formed by succession of four different entities called nucleotides: the DNA (DeoxyriboNucleic Acid). Each nucleotide is composed of a sugar, a phosphate group and one of the four bases: Adenine (A), Thymine (T), Cytosine (C) or Guanine (G). Each base is the complement of another one with which it is linked through interaction of their hydrogen atoms. Adenine is the complement of thymine and cytosine is the complement of guanine. The DNA is thus a double-stranded molecule, and the sequencing step allows us to read one of these strains, especially in which order we can find the bases. These molecules are stored as chromosomes and protected within different part of the cell. The largest part is present in the nucleus whereas a smaller part but in many copies is stored in the mitochondria, the "powerhouse of the cell" (Siekevitz, 1957).

To analyze the information stored in each sample, we first need to extract DNA. This step, realized in our case with a commercial kit (Qiagen DNeasy Blood and Tissue kit, Qiagen Inc., Texas), aims at breaking cell and internal membranes and separating the DNA from other molecules such as proteins. In this thesis, the samples were analyzed through shotgun sequencing, meaning the whole genome was sequenced, without targeting special genomic regions of interest, allowing us to obtain DNA sequences from both nuclear and mitochondrial compartments. When purified, the DNA molecules are usually cut randomly in fragments shorter than 700 bases. In the special context of museum samples, due to age of samples and conservation treatment, the DNA is already present in short fragments (< 200-300 bp) in sample and this step is thus not necessary.

As the cut takes place more easily between adenine and his complement thymine (linked by only two hydrogen bounds, against three between cytosine and guanine), the cut does not occur at the same place on both strands (what is called sticky-end) (Figure 11). A phosphate is added on the longest strand at both ends and the shortest end is elongated to obtain blunt ends. An adenine is added on the strand that does not carry the phosphate group. This allows the adding on both sides of the fragment of adapters that carry specific index for each sample. Fragments carrying adapters on both sides are selectively replicated to increase their amount before the true sequencing step. The sequencing was realized through the Illumina technology (Illumina Inc., San Diego). Usually, 24 samples (currently up to 48) are mixed in equal concentrations to be sequenced together. The index will allow to assign each sequence to the originate sample.

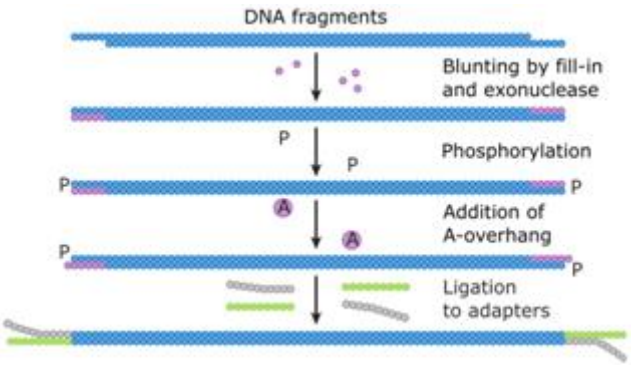


FIGURE 11: ILLUMINA'S LIBRARY-PREPARATION WORK FLOW (MARDIS, 2013)

When the mix of libraries is added on the flow cell, the adaptors hybridize with sequence anchored on the support (Figure 12). A complement sequence is elongated from each primer and the original fragments are removed. The anchored sequences fall over and the second adaptor link to the opposite primer for the elongation of the second strand. Thousands of copies are therefore synthesized to form a cluster of identical fragments. The reverse strands are washed away and sequence primers bind on adapters to initiate the reading. A mix of four fluorescent nucleotides is added and the complementary of the fragment base is incorporated. Thanks to the large number of copies, the fluorescence is strong enough to be read. When the maximum sequencing length is reached, the fragment synthesized (called read) is removed and the complement sequence is synthesized from the opposite primer. A second read is sequenced. We thus obtain two reads, each being sequenced from both ends of the original fragment, with converging directions. This type of sequencing is called “paired-end”.

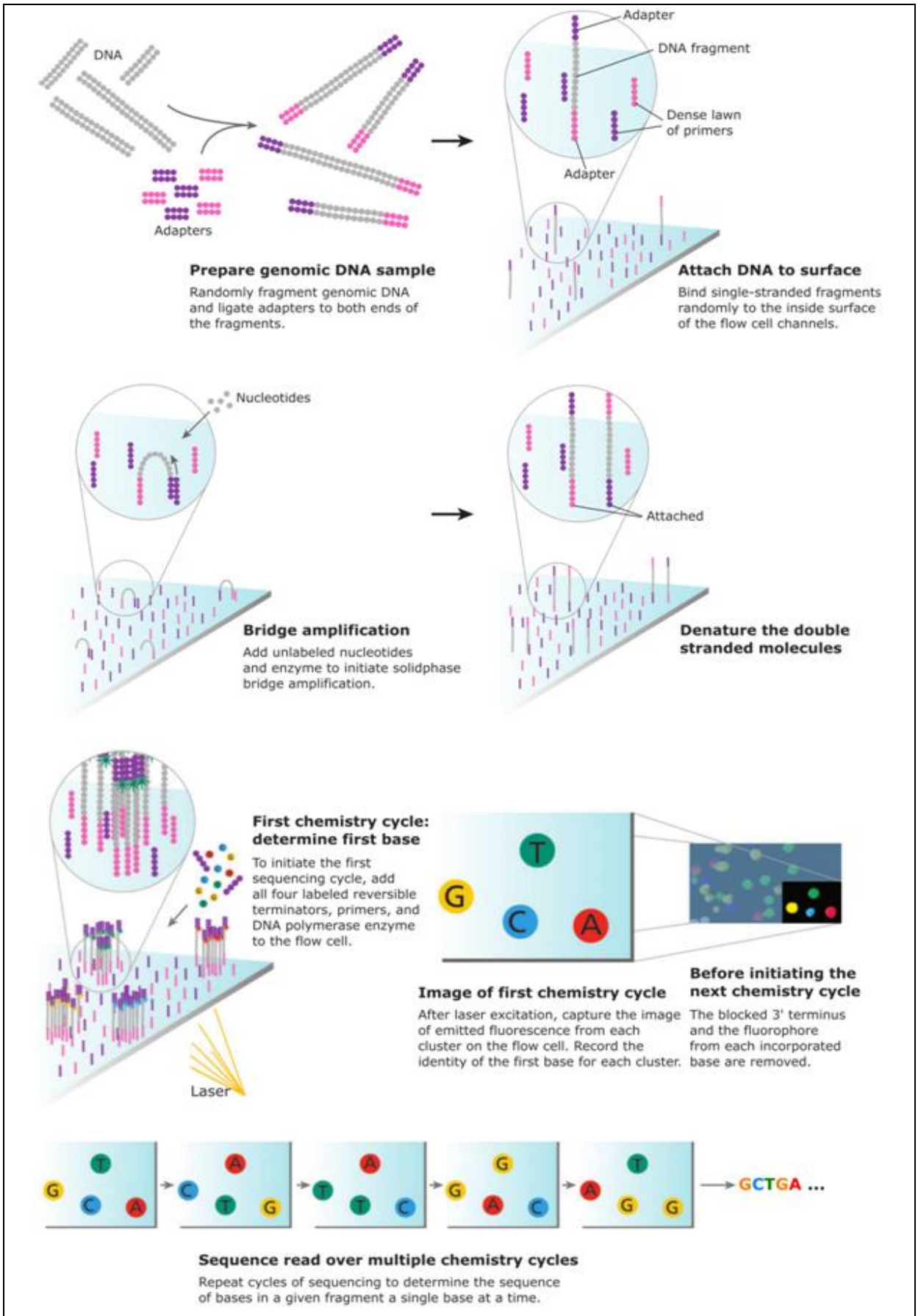


FIGURE 12: THE ILLUMINA SEQUENCING-BY-SYNTHESIS APPROACH (MARDIS, 2008)

We thus obtain a computer file containing millions of paired-end reads (short sequences composed of four letters: A, T, C, G). As the original sample contained many cells and therefore many genome copies, and because the cutting step is a random process, the resulting sequences can overlap. This property is used to reconstruct the longest possible sequence from the reads (Figure 13). If the overlapping is too short and then doubtful, we can use the fact that the reads are paired-end to increase the total length. The method that reconstructs the sequence from the reads only is called *de-novo*.

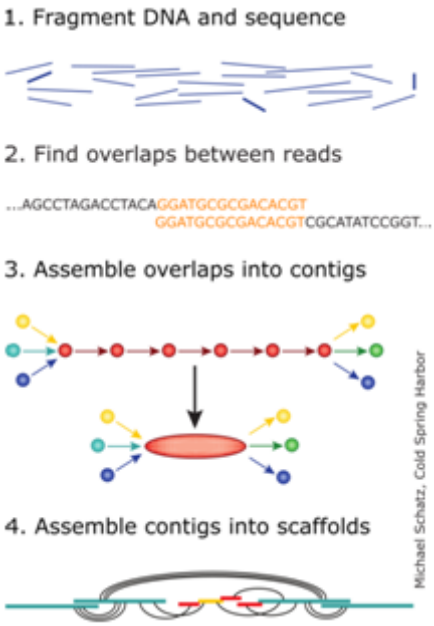


FIGURE 13: *DE-NOVO* GENOME ASSEMBLY (BAKER, 2012)

If we want to reconstruct only a special portion of the genome (e.g. the mitochondrial genome, specific nuclear markers…) or if the number of reads is too low to allow a *de-novo* assembly, it is possible to use a sequence from the same or a close species as reference. Each read is mapped against this reference and the final genome is derived from a consensus of these reads (Figure 14).

FIGURE 14: COMPARATIVE ASSEMBLY: READS ARE ALIGNED TO A REFERENCE SEQUENCE (ADAPTED FROM WAJID & SERPEDIN, 2016)

We finally obtain complete mitogenome and nuclear markers of interest for each sample analyzed, and use these data to reconstruct phylogenetic hypotheses.

References

Alcaide, M., Scordato, E.S.C., Price, T.D., & Irwin, D.E. (2014) Genomic divergence in a ring species complex. *Nature*, **511**, nature13285.

Augier, A. (1801) *Essai d'une nouvelle classification des végétaux conforme à l'ordre de la nature.* Bruyset, Lyon.

Baker, M. (2012) De novo genome assembly: what every biologist should know. *Nature Methods*, **9**, 333–337.

Benton, M., Donoghue, P., & Asher, R.J. (2009) Calibrating and constraining molecular clocks. *The timetree of life* pp. 35–86. Hedges, S.B., Kumar, S., Oxford, UK.

Bergmann, C. (1848) *Über die Verhältnisse der Wärmeökonomie der Thiere zu ihrer Grösse.* Vandenhoeck, Ruprecht, Göttingen.

Bickford, D., Lohman, D.J., Sodhi, N.S., Ng, P.K.L., Meier, R., Winker, K., Ingram, K.K., & Das, I. (2007) Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution*, **22**, 148–155.

BirdLife International and Handbook of the Birds of the World (2017) *Bird species distribution maps of the world. Version 2017.2. Available at http://datazone.birdlife.org/species/requestdis.*

Blasco, R., Finlayson, C., Rosell, J., Marco, A.S., Finlayson, S., Finlayson, G., Negro, J.J., Pacheco, F.G., & Vidal, J.R. (2014) The earliest pigeon fanciers. *Scientific Reports*, **4**, 5971.

Brower, A.V.Z. (1996) A new mimetic species of *Heliconius* (Lepidoptera: Nymphalidae), from southeastern Colombia, revealed by cladistic analysis of mitochondrial DNA sequences. *Zoological Journal of the Linnean Society*, **116**, 317–332.

Brown, W.M., George, M., & Wilson, A.C. (1979) Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America*, **76**, 1967–1971.

Bucher, E.H. & Bocco, P.J. (2009) Reassessing the importance of granivorous pigeons as massive, long-distance seed dispersers. *Ecology*, **90**, 2321–2327.

comte de Buffon, G.L.L. (1756) *Histoire naturelle, générale et particulière avec la description du cabinet du roy.* Pierre de Hondt, La Haye.

comte de Buffon, G.L.L. (1766) *Histoire naturelle, générale et particulière avec la description du cabinet du roy.* Pierre de Hondt, La Haye.

Carey, S.W. (1959) *The tectonic approach to continental drift.* Geology Dept., Univ. of Tasmania, Hobart.

Cavalli-Sforza, L.L. & Edwards, A.W.F. (1967) Phylogenetic analysis: Models and estimation procedures. *Evolution*, **21**, 550–570.

Cibois, A., Thibault, J.-C., Bonillo, C., Filardi, C.E., & Pasquet, E. (2017) Phylogeny and biogeography of the imperial pigeons (Aves: Columbidae) in the Pacific Ocean. *Molecular Phylogenetics and Evolution*, **110**, 19–26.

Claramunt, S. & Cracraft, J. (2015) A new time tree reveals Earth history's imprint on the evolution of modern birds. *Science Advances*, **1**, e1501005.

Cox, C.B. & Moore, P.D. (2010) *Biogeography: An Ecological and Evolutionary Approach.* John Wiley & Sons, Chichester.

Coyne, J.A. & Orr, H.A. (2004) *Speciation.* Sinauer Associates, Sunderland, MA.

Csilléry, K., Blum, M.G.B., Gaggiotti, O.E., & François, O. (2010) Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology & Evolution*, **25**, 410–418.

Currie, D.J., Mittelbach, G.G., Cornell, H.V., Field, R., Guégan, J.-F., Hawkins, B.A., Kaufman, D.M., Kerr, J.T., Oberdorff, T., O'Brien, E., & Turner, J.R.G. (2004) Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic richness. *Ecology Letters*, **7**, 1121–1134.

Danchin, E.G.J., Guzeeva, E.A., Mantelin, S., Berepiki, A., & Jones, J.T. (2016) Horizontal gene transfer from bacteria has enabled the plant-parasitic nematode *Globodera pallida* to feed on host-derived sucrose. *Molecular Biology and Evolution*, **33**, 1571–1579.

Darwin, C. (1837) Notebook B. *Charles Darwin's Notebooks, 1836-1844: Geology, Transmutation of Species, Metaphysical Enquiries (1987)* (ed. by P.H. Barrett, P.J. Gautrey, S. Herbert, and D. Kohn), Cambridge University Press, Ithaca, NY.

Darwin, C. (1859) *On the origin of species by means of natural selection: or the preservation of favoured races in the struggle for life.* John Murray, Albemarle Street, London.

Darwin, C. & Wallace, A. (1858) On the tendency of species to form varieties; and on the perpetuation of varieties and species by natural means of selection. *Zoological Journal of the Linnean Society*, **3**, 45–62.

Edwards, A.W. & Cavalli-Sforza, L.L. (1963) The reconstruction of evolution. *Heredity*, **18**, 553.

Edwards, A.W. & Cavalli-Sforza, L.L. (1964) Reconstruction of evolutionary trees. *Phenetic and Phylogenetic Classification* pp. 67–76. V. H. Heywood and J. McNeill, London.

Ence, D.D. & Carstens, B.C. (2011) SpedeSTEM: a rapid and accurate method for species delimitation. *Molecular Ecology Resources*, **11**, 473–480.

Felsenstein, J. (2004) *Inferring phylogenies.* Sinauer associates, Sunderland, MA.

Fitch, W.M. & Margoliash, E. (1967) Construction of Phylogenetic Trees. *Science*, **155**, 279–284.

Forster, J.R. (1778) *Observations made during a voyage round the world.* G. Robinson, London.

Fraser, D.J. & Bernatchez, L. (2001) Adaptive evolutionary conservation: towards a unified concept for defining conservation units. *Molecular Ecology*, **10**, 2741–2752.

Friedman, M. (2010) Explosive morphological diversification of spiny-finned teleost fishes in the aftermath of the end-Cretaceous extinction. *Proceedings of the Royal Society of London B: Biological Sciences*, rspb20092177.

Gibbs, D., Barnes, E., & Cox, J. (2001) *Pigeons and doves: a guide to the pigeons and doves of the world.* A&C Black, London.

Gire, S.K., Goba, A., Andersen, K.G., et al. (2014) Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science*, **345**, 1369–1372.

Gloger, C.L. (1833) *Das Abändern der Vögel durch Einfluss des Klima's: nach zoologischen, zunächst von den europäischen Landvögeln entnommenen Beobachtungen dargestellt, mit den entsprechenden Erfahrungen bei den europäischen Säugthieren verglichen, und durch Thatsachen aus dem Gebiete der Physiologie, der Physik und der physischen Geographie erläutert.* In Commission bei August Schulz, Breslau.

Graham, C.H. & Fine, P.V.A. (2008) Phylogenetic beta diversity: linking ecological and evolutionary processes across space in time. *Ecology Letters*, **11**, 1265–1277.

Grant, P.R. (1999) *Ecology and Evolution of Darwin's Finches.* Princeton University Press, Princeton, NJ.

Grant, P.R. & Grant, B.R. (2010) Sympatric speciation, immigration, and hybridization in island birds. *The Theory of Island Biogeography Revisited* pp. 326–357.

Grant, V. (1981) *Plant speciation.* Columbia University Press, New York.

Gregory, T.R. & Mable, B.K. (2005) Polyploidy in animals. *The evolution of the genome* pp. 427–517. Elsevier Academic Press, London.

Hasegawa, M., Kishino, H., & Yano, T. (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, **22**, 160–174.

Hasegawa, M., Kishino, H., & Yano, T. (1987) Man's place in Hominoidea as inferred from molecular clocks of DNA. *Journal of Molecular Evolution*, **26**, 132–147.

Heath, T.A., Huelsenbeck, J.P., & Stadler, T. (2014) The fossilized birth–death process for coherent calibration of divergence-time estimates. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, E2957–E2966.

Hebert, P.D.N., Cywinska, A., Ball, S.L., & deWaard, J.R. (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B: Biological Sciences*, **270**, 313–321.

Highton, R. (1998) Is *Ensatina eschscholtzii* a Ring-Species? *Herpetologica*, **54**, 254–278.

Ho, S.Y.W., Phillips, M.J., Cooper, A., & Drummond, A.J. (2005) Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Molecular Biology and Evolution*, **22**, 1561–1568.

Holmes, A. (1931) Radioactivity and earth movements. *Transactions of the Geological Society of Glasgow*, **18**, 559–606.

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.

Jacquin, L., Lenouvel, P., Haussy, C., Ducatez, S., & Gasparini, J. (2011) Melanin-based coloration is related to parasite intensity and cellular immune response in an urban free living bird: the feral pigeon Columba livia. *Journal of Avian Biology*, **42**, 11–15.

Jin, L. & Nei, M. (1990) Limitations of the evolutionary parsimony method of phylogenetic analysis. *Molecular Biology and Evolution*, **7**, 82–102.

Johnston, R.F. & Janiga, M. (1995) *Feral pigeons.* Oxford University Press, Oxford, UK.

Jukes, T. & Cantor, C. (1969) Evolution of protein molecules. *Mammalian protein metabolism* (ed. by M. Munro), pp. 21–132. Academic Press, New York.

Kandul, N.P., Lukhtanov, V.A., & Pierce, N.E. (2007) Karyotypic diversity and speciation in *Agrodiaetus* butterflies. *Evolution*, **61**, 546–559.

Kingman, J.F.C. (1982) The coalescent. *Stochastic Processes and their Applications*, **13**, 235–248.

Lamarck, J.-B. de M., Chevalier de (1801) *Système des animaux sans vertèbres.* Deterville, Paris.

Lanave, C., Preparata, G., Sacone, C., & Serio, G. (1984) A new method for calculating evolutionary substitution rates. *Journal of Molecular Evolution*, **20**, 86–93.

Le Moan, A., Gagnaire, P.-A., & Bonhomme, F. (2016) Parallel genetic divergence among coastal–marine ecotype pairs of European anchovy explained by differential introgression after secondary contact. *Molecular Ecology*, **25**, 3187–3202.

Levenson, R.M., Krupinski, E.A., Navarro, V.M., & Wasserman, E.A. (2015) Pigeons (*Columba livia*) as trainable observers of pathology and radiology breast cancer images. *PLoS ONE*, **10**, e0141357.

Liebers, D., De Knijff, P., & Helbig, A.J. (2004) The Herring Gull complex is not a ring species. *Proceedings of the Royal Society B: Biological Sciences*, **271**, 893.

Linné, C. von (1735) *Systema naturae: per regna tria naturae, secundum classes, ordines, genera, species cum characteribus, differentiis, synonymis, locis.* Leiden, The Netherlands.

Magallón, S. & Sanderson, M.J. (2001) Absolute diversification rates in angiosperm clades. *Evolution*, **55**, 1762–1780.

Mardis, E.R. (2008) Next-Generation DNA Sequencing Methods. *Annual Review of Genomics and Human Genetics*, **9**, 387–402.

Mardis, E.R. (2013) Next-Generation Sequencing Platforms. *Annual Review of Analytical Chemistry*, **6**, 287–303.

Martin, C.H., Cutler, J.S., Friel, J.P., Dening Touokong, C., Coop, G., & Wainwright, P.C. (2015) Complex histories of repeated gene flow in Cameroon crater lake cichlids cast doubt on one of the clearest examples of sympatric speciation. *Evolution*, **69**, 1406–1422.

Masly, J.P., Jones, C.D., Noor, M.A.F., Locke, J., & Orr, H.A. (2006) Gene transposition as a cause of hybrid sterility in *Drosophila*. *Science*, **313**, 1448–1450.

Matos, F.A.R., Magnago, L.F.S., Gastauer, M., Carreiras, J.M.B., Simonelli, M., Meira-Neto, J.A.A., & Edwards, D.P. (2017) Effects of landscape configuration and composition on phylogenetic diversity of trees in a highly fragmented tropical forest. *Journal of Ecology*, **105**, 265–276.

Mayden, R.L. (1997) A hierarchy of species concepts: The denouement in the saga of the species problem. *Species: The units of diversity,* (ed. by M.F. Claridge, H.A. Dawah, and M.R. Wilson), pp. 381–423. Chapman & Hall, London.

Maynard Smith, J. (1966) Sympatric Speciation. *The American Naturalist*, **100**, 637–650.

Mayr, E. (1942) *Systematics and the origin of species, from the viewpoint of a zoologist.* Harvard University Press, Harvard.

McConkey, K.R., Meehan, H.J., & Drake, D.R. (2005) Seed dispersal by Pacific pigeons (*Ducula pacifica*) in Tonga, western Polynesia. *Emu*, **104**, 369–376.

Meier, R., Shiyang, K., Vaidya, G., Ng, P.K.L., & Hedin, M. (2006) DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Systematic Biology*, **55**, 715–728.

Meiklejohn, K.A., Faircloth, B.C., Glenn, T.C., Kimball, R.T., & Braun, E.L. (2016) Analysis of a rapid evolutionary radiation using Ultraconserved Elements: evidence for a bias in some multispecies coalescent methods. *Systematic Biology*, **65**, 612–627.

Mora, C.V., Wild, J.M., Davison, M., & Walker, M.M. (2004) Magnetoreception and its trigeminal mediation in the homing pigeon. *Nature*, **432**, 508.

Nabholz, B., Lanfear, R., & Fuchs, J. (2016) Body mass-corrected molecular rate for bird mitochondrial DNA. *Molecular Ecology*, **25**, 4438–4449.

Nakahara, S., Zacca, T., Huertas, B., Neild, A.F., Hall, J.P., Lamas, G., Holian, L.A., Espeland, M., & Willmott, K.R. (2018) Remarkable sexual dimorphism, rarity and cryptic species: a revision of the 'aegrota species group' of the Neotropical butterfly genus *Caeruleuptychia* Forster, 1964 with the description of three new species (Lepidoptera, Nymphalidae, Satyrinae). *Insect Systematics & Evolution*, doi: 10.1163/1876312X-00002167.

Otto, S.P. & Whitton, J. (2000) Polyploid Incidence and Evolution. *Annual Review of Genetics*, **34**, 401–437.

Owens, I.P.F. & Bennett, P.M. (2000) Ecological basis of extinction risk in birds: Habitat loss versus human persecution and introduced predators. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 12144–12148.

Paradis, E., Tedesco, P.A., & Hugueny, B. (2013) Quantifying variation in speciation and extinction rates with clade data. *Evolution*, **67**, 3617–3627.

Parham, J.F., Donoghue, P.C.J., Bell, C.J., et al. (2012) Best practices for justifying fossil calibrations. *Systematic Biology*, **61**, 346–359.

Perea, R. & Gutiérrez-Galán, A. (2016) Introducing cultivated trees into the wild: Wood pigeons as dispersers of domestic olive seeds. *Acta Oecologica*, **71**, 73–79.

Pillon, Y. & Buerki, S. (2017) How old are island endemics? *Biological Journal of the Linnean Society*, **121**, 469–474.

Pratt, T.K. & Beehler, B.M. (2015) *Birds of New Guinea.* Princeton University Press, Princeton, NJ.

Quillfeldt, P. (2017) Body mass is less important than bird order in determining the molecular rate for bird mitochondrial DNA. *Molecular Ecology*, **26**, 2426–2429.

Rannala, B. & Yang, Z. (1996) Probability distribution of molecular evolutionary trees: A new method of phylogenetic inference. *Journal of Molecular Evolution*, **43**, 304–311.

Ricklefs, R.E. & Jønsson, K.A. (2014) Clade extinction appears to balance species diversification in sister lineages of Afro-Oriental passerine birds. *Proceedings of the National Academy of Sciences*, **111**, 11756–11761.

Roux, C., Tsagkogeorga, G., Bierne, N., & Galtier, N. (2013) Crossing the species barrier: genomic hotspots of introgression between two highly divergent *Ciona intestinalis* species. *Molecular Biology and Evolution*, **30**, 1574–1587.

Roy, K. & Goldberg, E.E. (2007) Origination, extinction, and dispersal: Integrative models for understanding present-day diversity gradients. *The American Naturalist*, **170**, S71–S85.

Savolainen, V., Anstett, M.-C., Lexer, C., Hutton, I., Clarkson, J.J., Norup, M.V., Powell, M.P., Springate, D., Salamin, N., & Baker, W.J. (2006) Sympatric speciation in palms on an oceanic island. *Nature*, **441**, nature04566.

Shapiro, M.D., Kronenberg, Z., Li, C., Domyan, E.T., Pan, H., Campbell, M., Tan, H., Huff, C.D., Hu, H., Vickrey, A.I., Nielsen, S.C.A., Stringham, S.A., Hu, H., Willerslev, E., Gilbert, M.T.P., Yandell, M., Zhang, G., & Wang, J. (2013) Genomic diversity and evolution of the head crest in the rock pigeon. *Science*, **339**, 1063–1067.

Shelomi, M., Danchin, E.G.J., Heckel, D., Wipfler, B., Bradler, S., Zhou, X., & Pauchet, Y. (2016) Horizontal gene transfer of pectinases from bacteria preceded the diversification of stick and leaf insects. *Scientific Reports*, **6**, srep26388.

Siekevitz, P. (1957) Powerhouse of the Cell. *Scientific American*, **197**, 131–144.

Soltis, D.E., Soltis, P.S., & Tate, J.A. (2004) Advances in the study of polyploidy since plant speciation. *New Phytologist*, **161**, 173–191.

Sorenson, M.D., Sefc, K.M., & Payne, R.B. (2003) Speciation by host switch in brood parasitic indigobirds. *Nature*, **424**, 928.

Stadler, T. & Smrckova, J. (2016) Estimating shifts in diversification rates based on higher-level phylogenies. *Biology Letters*, **12**, 20160273.

Steadman, D. (1997) The historic biogeography and community ecology of Polynesian pigeons and doves. *Journal of Biogeography*, **24**, 737–753.

Stuessy, T.F. (2006) Evolutionary biology: Sympatric plant speciation in islands? *Nature*, **443**, E12.

Sullivan, J. & Joyce, P. (2005) Model selection in phylogenetics. *Annual Review of Ecology, Evolution, and Systematics*, **36**, 445–466.

Sweet, A.D., Chesser, R.T., & Johnson, K.P. (2017) Comparative cophylogenetics of Australian phabine pigeons and doves (Aves: Columbidae) and their feather lice (Insecta: Phthiraptera). *International Journal for Parasitology*, **47**, 347–356.

Tavaré, S. (1986) Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on mathematics in the life sciences*, **17**, 57–86.

Wajid, B. & Serpedin, E. (2016) Do it yourself guide to genome assembly. *Briefings in Functional Genomics*, **15**, 1–9.

Wallace, A.R. (1876) *The geographical distribution of animals: with a study of the relations of living and extinct faunas as elucidating the past changes of the earth's surface.* Macmillan and Co., London.

Wegener, A. (1912) Die Entstehung der Kontinente. *Geologische Rundschau*, **3**, 276–292.

Weir, J.T. & Schluter, D. (2008) Calibrating the avian molecular clock. *Molecular Ecology*, **17**, 2321–2328.

Wiley, E.O. (1978) The evolutionary species concept reconsidered. *Systematic Biology*, **27**, 17–26.

Wood, T.E., Takebayashi, N., Barker, M.S., Mayrose, I., Greenspoon, P.B., & Rieseberg, L.H. (2009) The frequency of polyploid speciation in vascular plants. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 13875–13879.

Wotton, D.M. & Kelly, D. (2012) Do larger frugivores move seeds further? Body size, seed dispersal distance, and a case study of a large, sedentary pigeon. *Journal of Biogeography*, **39**, 1973–1983.

Wright, S. (1931) Evolution in Mendelian populations. *Genetics*, **16**, 97–159.

Yang, M., He, Z., Shi, S., & Wu, C.-I. (2017) Can genomic data alone tell us whether speciation happened with gene flow? *Molecular Ecology*, **26**, 2845–2849.

Chapter 1

Biogeography and diversification of a species-rich worldwide-distributed family, the Columbidae: evaluating the role of islands

1. Introduction

Biodiversity is unequally distributed on Earth, with some regions hosting higher number of species than others (Jenkins et al., 2013). However, disentangling the relative roles of present environmental constraints and historical patterns is not straightforward (Roy & Goldberg, 2007). If the current diversity pattern is regulated by environment variables (Currie et al., 2004), speciation, extinction, and dispersal also played a role in the uneven distribution of global biodiversity (Roy & Goldberg, 2007): species rich regions can thus either resulting from high speciation, low extinction and/or high immigration rates.

Islands play a prominent role in this context. Their endemic richness is a magnitude more important than the mainland one (Kier et al., 2009) despite covering only 5.3% of the total Earth land area (Tershy et al., 2015). They provide good conditions for allopatric speciation as their geographic isolation slows down gene flow between island and source populations (Mayr, 1963). But if islands are classically viewed as sinks (receiving colonizers from mainland; MacArthur & Wilson, 1967), they can be sometimes a center of diversification for species which can later disperse worldwide, especially when islands are large or old (Fjeldså, 2013). They can also be a route for dispersion through oceans (Warren et al., 2010; Katinas et al., 2013). Therefore, it is necessary to better evaluate the importance of islands throughout the whole diversification process of worldwide species-rich families.

Uneven global distribution of species diversity has been observed in numerous animal groups, even in lineages with supposedly high dispersal ability such as birds. Among them, Columbidae (Pigeons and doves) is the most speciose worldwide-distributed bird family among non-Passeriformes (Newton, 2003), with more than 340 described species [368 species for HBW (including 18 extinct or extinct in the wild species), 343 species for IOC 7.3 (including 13 extinct species)]. The group is present on all continents and archipelagos except Antarctica, Easter island and Hawaii, where species have been recently introduced by humans (Pyle & Pyle, 2017; Figure 15).

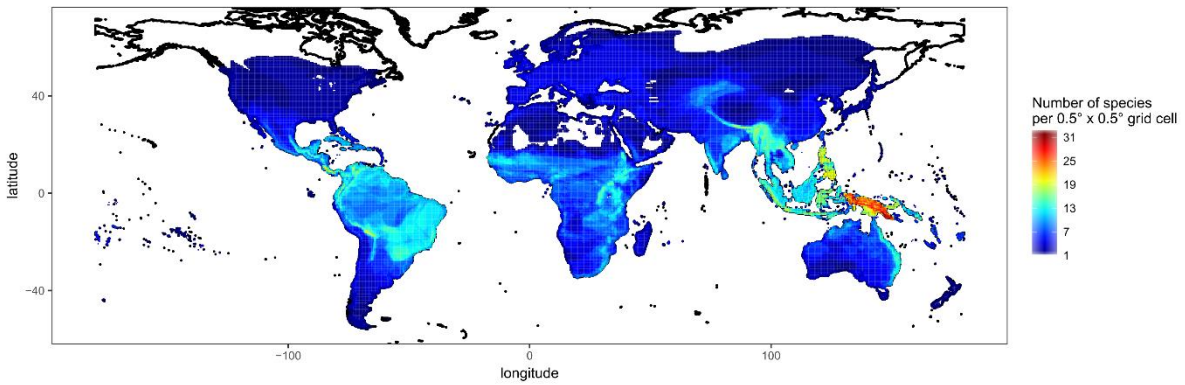


FIGURE 15 : SPECIES RICHNESS OF THE COLUMBIDAE FAMILY, PER 0.5°X0.5° GRID CELL. DATA FROM BIRDLIFE INTERNATIONAL AND HANDBOOK OF THE BIRDS OF THE WORLD (2017), MAPPED WITH THE RASTERSP PACKAGE ON R (BIBER, 2017).

The highest species richness is reached in New Guinea, where up to 20 species can co-occur (Pratt & Beehler, 2015; Figure 15). With surrounding regions, they share a rather large amount of diversity (Figure 16). On the other side, relatively poor species diversity is observed in large regions such as Palearctic where only 22 species of Columbidae are recorded.

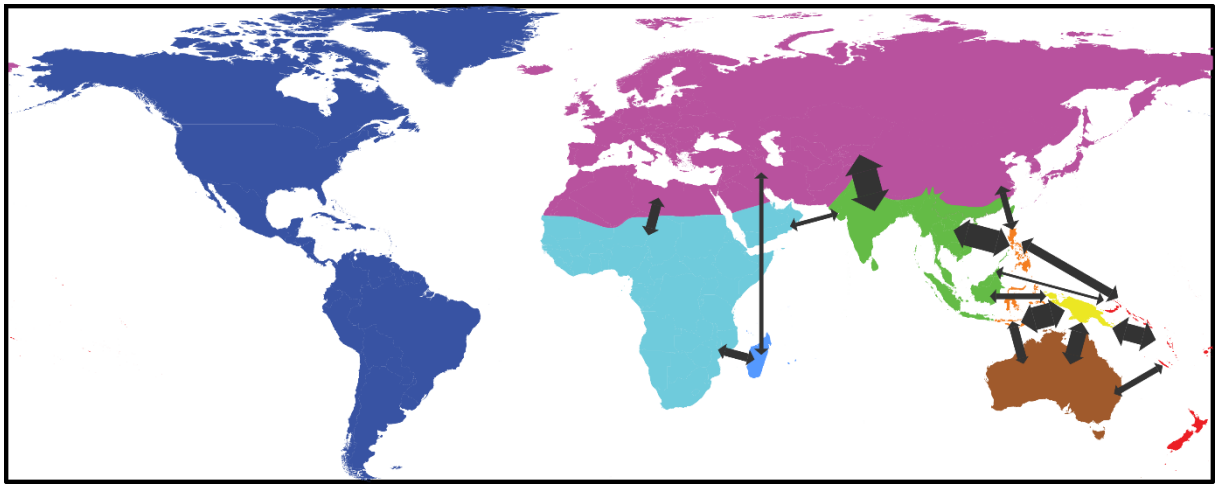


FIGURE 16: SHARED DIVERSITY BETWEEN PAIRS OF BIOGEOGRAPHICAL REGIONS DEPICTED BY THE ARROW THICKNESS. SHARED DIVERSITY IS CALCULATED FOLLOWING SIMPSON SIMILARITY INDEX (SIMPSON, 1947): THE NUMBER OF SHARED SPECIES BETWEEN TWO REGIONS IS DIVIDED BY THE LOWEST OF THE TWO REGION RICHNESSES. THIS INDEX IS ESPECIALLY WELL SUITED FOR DISSIMILAR REGION RICHNESSES.

When investigating patterns of species distribution over biogeographical regions (defined following Andersen et al., 2018), most species of Columbidae show a regional distribution, with only 43 species present in at least two biogeographic regions, and 17 present in more than two regions. The New World is an extreme case, with only endemic species. This high level of endemism is however marked in all regions (Figure 17), with a pattern slightly less marked when genera are considered.

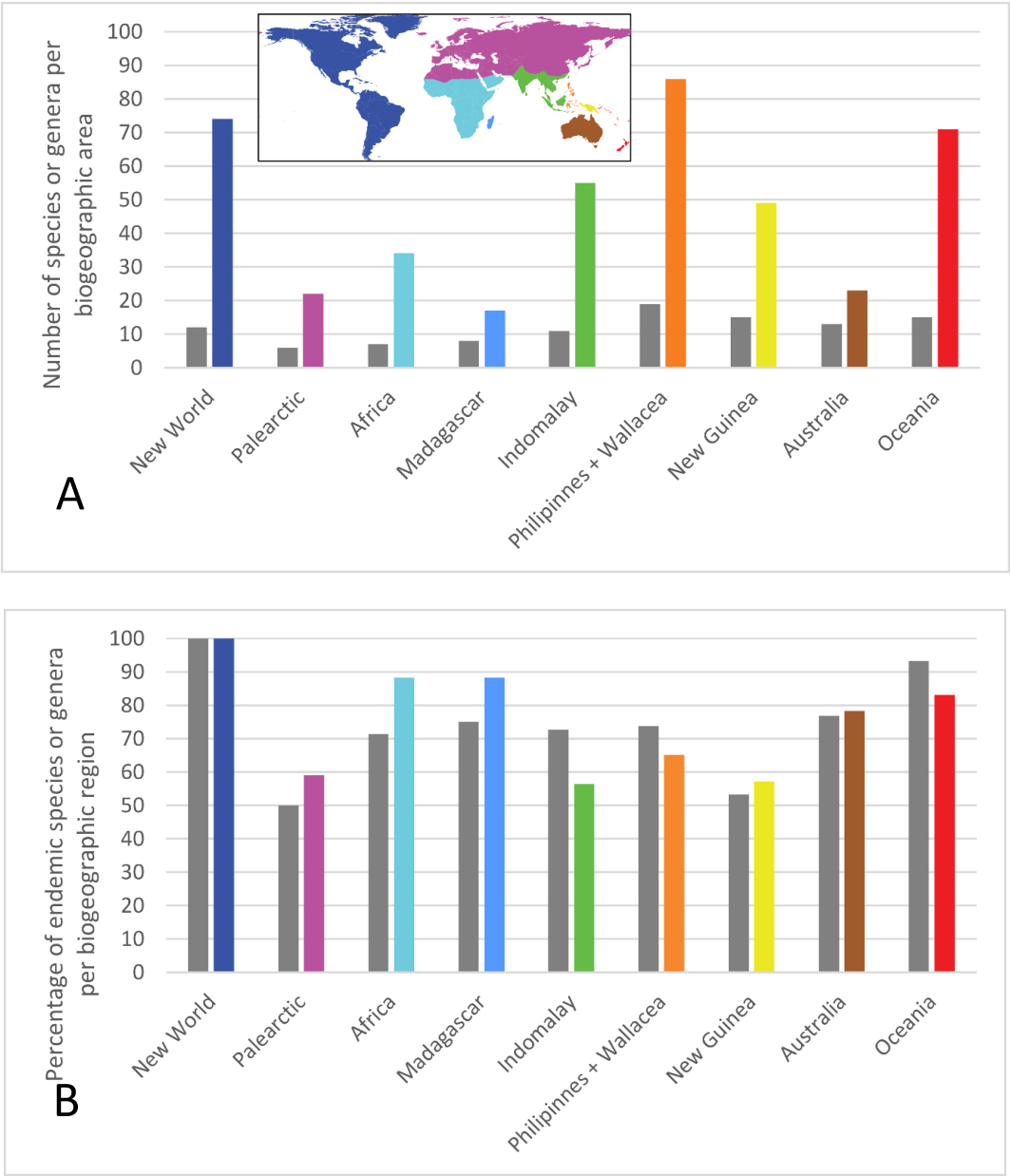


FIGURE 17: DISTRIBUTION AND ENDEMISM OF GENERA AND SPECIES PER BIOGEOGRAPHIC REGION, AS DEFINED BY ANDERSEN ET AL. (2018), WITH AUSTRALIA SEPARATED FROM NEW GUINEA, AND AFRICA FROM MADAGASCAR AND SURROUNDING ISLANDS. A: NUMBER OF GENERA (IN GREY, ON THE LEFT) AND SPECIES (IN COLOR, ON THE RIGHT) PER BIOGEOGRAPHIC REGION. B: PERCENTAGE OF GENERA (IN GREY, ON THE LEFT) AND SPECIES (IN COLOR, ON THE RIGHT) ENDEMIC TO EACH BIOGEOGRAPHIC REGION.

Almost 60% of the Columbidae species are insular (i.e. 58.6 and 59.7% according to the IOC and HBW taxonomies, respectively; Madagascar and New Guinea being considered here as the largest islands). Some lineages seem to have diversified on archipelago (e.g. *Ducula* and *Ptilinopus*; Gibb & Penny, 2010; Cibois et al., 2014, 2017).

Islands are also home of the few morphologically different species, which differ markedly from the general shape of pigeons by their size (Crowned-pigeons, *Goura* spp., Dodo, *Raphus cucullatus*; Solitaire, *Pezophaps solitaria*), beak shape (Tooth-billed pigeon, *Didunculus strigirostris*; Thick-billed Ground-pigeon, *Trugon terrestris*), leg length (Pheasant-pigeons, *Otidiphaps* spp.) or flightlessness (Dodo, *Raphus cucullatus*; Solitaire, *Pezophaps solitaria*). While morphologically distinct species are generally observed in explosive radiations for adaptive purpose (*e.g.* Grant, 1999; Friedman, 2010), most of these Columbidae species belong to monotypic genera.

The spread and diversification of Columbidae needs to be investigated in a phylogenetic framework. However, studies at a family scale are still partial, as many are based on an incomplete sampling (e.g. Johnson & Clayton, 2000; Shapiro et al., 2002; Johnson, 2004; Soares et al., 2016) or just focused on few lineages (Johnson et al., 2001; Johnson & Weckstein, 2011; Jønsson et al., 2011; Banks et al., 2013; Moyle et al., 2013; Cibois et al., 2014, 2017; Sweet & Johnson, 2015; Sweet et al., 2017). The phylogeny of the family is thus not yet comprehensive and completely resolved, with taxonomic revisions still recently proposed in different groups (e.g. *Patagioenas*: Johnson et al., 2001; *Gallicolumba*: Jønsson et al., 2011; Moyle et al., 2013; *Geotrygon*: Banks et al., 2013; *Ptilinopus*: Cibois et al., 2014; *Goura*: Bruxaux et al., 2018). The only study so far having a rather consequent sampling (39 among the 41 genera) and using a few nuclear and mitochondrial markers (Pereira et al., 2007) allowed inferring a biogeographical scenario involving an ancestral distribution in Gondwanaland with an initial radiation around 54.4 Ma (million years ago). This dating could have been overestimated as recent results place the earliest radiation into the Miocene (24.7 Ma; Soares et al., 2016), involving different contexts of dispersion of ancestral lineages. Moreover, making inferences about the origins of Columbidae is difficult both biogeographically and biologically due to the absence of closely related groups. The closest families could be the Mesitornithidae and Pteroclididae (Ericson et al., 2006; Hackett et al., 2008; Jarvis et al., 2014; Prum et al., 2015; McCormack et al., 2016 [little dataset]) but they are sometimes placed elsewhere in the phylogeny [sister to all Neoaves according to Glenn et al. (2008) and McCormack et al. (2016) [large dataset]; or sister to Cuculidae according to Suh et al.(2015)]. As pointed by Suh et al. (2015), these incongruent results could be due to a large amount of polymorphism in ancestral population, which persisted through incomplete lineage sorting.

Here we propose to clarify the phylogeny of Columbidae, a worldwide distributed family of birds, using the best sampling so far, which includes 111 species, all the genera being represented. To improve the reliability of the phylogeny, we use whole mitochondrial genomes (hereafter mitogenomes) for nearly all samples studied, and nuclear markers for most of the 75 samples sequenced specifically for this study. This representative sampling, along with studies published at the genus level, is used to investigate the process leading to a worldwide distribution, and especially the importance of islands in ancient and recent diversification events, but also as a route of dispersal at different stages of the evolutionary history of the family. To do so, we conduct here a biogeographical analysis and evaluate the variation of the speciation rate across the family. This phylogeny will later allow morphological comparisons of islands and single-species clades with the average space occupied by the family.

2. Material and methods

2.1. Phylogeny

2.1.1. Sampling, DNA extraction and sequencing

In this study, we sequenced 66 new tissue or toe-pad samples of Columbidae species and used nine previously sequenced samples (Bruxaux et al., 2018) (Supplementary Table 1). We took into account the already published mitogenomes (Supplementary Table 2) and established the sampling to obtain at least one species per genus, and at least one species per intra-genus clade when the monophyly of the genus was discussed. Extractions and sequencing were performed as previously described (Bruxaux et al., 2018; see Chapter 2), except for three samples. Two samples (*Microgoura meeki* and *Starnoenas cyanocephala* Scya063) were extracted and sequenced in other labs. [details to request here] Most samples were sequenced with 24 libraries per flow-cell lane, except *Starnoenas cyanocephala* Scya063 that was sequenced as 1/6 flow-cell lane, and *Pezophaps solitaria* that was extracted and sequenced following an ancient DNA protocol. [details to request here]

We used mitogenomes of *Pterocles gutturalis* and *Gallus gallus* (GenBank accessions: KX902237 and NC_001323, respectively) as outgroups for mitochondrial analyses. *Pterocles* belongs to Pteroclididae that is known as a putatively close relative family of Columbidae, while *Gallus* (Phasianidae) belongs to a more distant lineage (Prum et al., 2015).

Finally, two samples of *Melanocharis* (*M. nigra* and *M. striativentris*; Melanocharitidae) were used as distant outgroups in the nuclear genome analyses (Prum et al., 2015). Because their

sequencing was performed like the other samples, we were expecting the same type of data (including a rather large part of missing data), which could avoid potential biases linked with differential amounts of data across the phylogeny.

2.1.2. Mitogenome assembly and phylogenetic reconstruction

Mitogenomes were reconstructed as described in Bruxaux et al. (2018; Chapter 2), and MapDamage (Jónsson et al., 2013) was not used for mitochondrial DNA as it was found to not change consensus results in this study.

The 66 mitogenomes reconstructed for this study and the 48 obtained from GenBank were aligned using Muscle implemented in Geneious v.9.1.7 (Biomatter Ltd., Auckland, New Zealand). Annotations were transferred from previously published genomes to newly sequenced ones, and standardized across the 114-samples dataset. The data were then partitioned by gene and by codon position for coding gene. Each tRNA or ribosomal gene was proposed as a partition, and non-coding data were pulled together to form the last potential one. PartitionFinder v.2.1.1 (Lanfear et al., 2017) was used to find the best partitioning scheme along with molecular evolution models in order to perform both maximum likelihood and Bayesian analysis with RAxML 8.1.5 (Stamatakis, 2014) and MrBayes 3.2.2 (Ronquist et al., 2012), respectively.

Phylogenetic trees were rooted on *Pterocles gutturalis* and *Gallus gallus* (Soares et al., 2016). The maximum likelihood analysis was performed through 20 independent runs, with 1000 replicates of non-parametric bootstrapping. The Bayesian analysis was performed with eight independent runs of four Metropolis-coupled Markov chains (MCMC) for Monte Carlo simulations run for 5,000,000 generations, with parameters and trees sampled every 10,000 generations, and a burn-in of 1,250,000 generations. The convergence of the chains were checked with the package RWTY (Warren et al., 2017) run in R v.3.4.0 (R Core Team, 2017).

Due to the expected old ages of speciation events studied here, we could observe saturation due to different substitutions occurring at the same position and therefore masking previous genetic changes. It is especially the case at the 3rd codon position where the selection is relaxed due to genetic code redundancy. To evaluate the amount of saturation, we compared raw genetic distance with K80 genetic distance (Philippe et al., 1994) and trees based on coding sequences only, with and without the 3rd codon position.

2.1.3. Analyses based on conserved nuclear genome regions: assembly and phylogenetic reconstruction

Nuclear genomic regions were retrieved from the 66 samples sequenced for this study and for nine samples sequenced for a previous study (Bruxaux et al., 2018; Supplementary Table 1) using the same approach but a slightly different reference sequence. To avoid possible biases resulting from the usage of a reference sequence whose species is included in the sampling (involving different phylogenetic distances between species and reference across the sampling; Brandt et al., 2015), we replaced the *Treron vernans* reference (McCormack et al., 2013) by the *Pterocles exustus* data from the same article. The *Gallus gallus* reference (Prum et al., 2015) was kept unchanged. The new set of genes is composed of 1,130 ultra-conserved elements (UCE) with an average length of 351 bp, leading to a total of 1,520 independent loci, which were concatenated to create a unique reference sequence of ca. 890,000 bp. To root this tree, we used the two *Melanocharis* species.

As stated in Bruxaux et al. (2018), trimmed reads of each sample were mapped against the nuclear reference using GMAP-GSNAP v.2015-11-20 (Wu & Nacu, 2010), and PCR duplicates were removed with samtools v.1.3.1. (Li et al., 2009). We used mapDamage v2.0.2. (Jónsson et al., 2013) to lower the Phread base-quality score of the sites were deamination was highly likely. The consensus sequence was recovered with the doFasta command of ANGSD (Korneliussen et al., 2014). We only kept the samples for which at least 1000 reads had mapped against the reference to avoid too much missing data. Six samples were thus discarded, and the 69 remaining samples plus two outgroups were used for the maximum-likelihood phylogenetic analysis performed with RAxML (Stamatakis, 2014) with 20 alternative runs and 100 replicates of non-parametric bootstrapping.

2.2. Dating and biogeography

As the mitochondrial dataset has a better sequencing coverage and less missing data than the nuclear dataset, and because coalescent dates estimated from mitochondrial DNA data generally precede population splitting by only 0.2–0.3 million years for birds (Moore, 1995; Weir, 2006), we used the mitogenome sequences to date speciation events of the Columbidae. Following mitochondrial analysis conducted on the whole dataset, we kept only one individual per clade, reducing the dataset to 57 samples. Each clade corresponds to a genus, with two exceptions. The genus *Claravis* was split into two due to its paraphyly (Sweet & Johnson, 2015; Sweet et al., 2017), and the *Macropygia* and *Turacoena* spp. were lumped into one clade due to their overall similarity in morphology and calls

(Stresemann, 1941). The partitions previously defined were re-used as input for PartitionFinder (Lanfear et al., 2017) for dating analysis with Beast2 v.2.4.3 (Bouckaert et al., 2014) with one exception: the control region was analyzed with Gblocks v.0.91b (Castresana, 2000) to remove misalignment due to high divergence in this region.

2.2.1. Phylogenetic tree calibration

To date phylogenies in an absolute time, fossils have long been used to calibrate a few nodes (Donoghue & Benton, 2007). However, strategies to include these data are still discussed (dos Reis et al., 2016; Barba-Montoya et al., 2017). First, not all fossils are suitable for use as calibration points (Parham et al., 2012). Second, uncertainty in fossils dating and variable paleontological sampling effort must be taken into account (Ho & Phillips, 2009). Finally, if the oldest known fossils can be used as minimal bound, the method to establish maximal bound and probability densities is still discussed (Heath, 2012; Barba-Montoya et al., 2017).

Here, we used a classic strategy of node calibration. The separation of Columbidae and Pteroclidae has been dated by different methods and studies as occurring in the Paleocene (Jarvis et al., 2014; Claramunt & Cracraft, 2015; Prum et al., 2015) and a normal calibration at 55 Ma ±15 Ma was used by Soares et al. (2016).

Like other birds, Columbidae fossil remains are scarce and only a handful of them are suitable for phylogenetic analyses: *Arenicolumba prattae* ca. 18.5 Ma (Steadman, 2008), close relative of *Oena* spp. and *Turtur* spp.; *Rupephaps taketake* 16-19 Ma (Worthy et al., 2009), close relative of *Hemiphaga*; *Primophaps schoddei* 26-24 Ma (Worthy, 2012), close relative of *Petrophassa*, *Phaps*, *Geophaps* and *Ocyphaps*; *Ectopistes* spp. 3.7-4.8 Ma (Olson & Rasmussen, 2001); *Zenaida prior* (3.6-2.6 Ma) (Brodkorb, 1968). Most other remains date only from the Quaternary, or even from the Holocene (Balouet & Olson, 1987; Millener & Powlesland, 2001; Worthy, 2001; Worthy & Wragg, 2003; Wragg & Worthy, 2006; Worthy et al., 2008; Olson, 2011), making them unsuitable for old speciation event dating.

All these fossils could hardly be used by previous studies, as most of them were only recently described, while sampling was not always appropriate for the placement of these calibration points. Moreover, one often used fossil was recently identified as being a Pterocletidae instead of a Columbidae (*Gerandia calcaria*; Mlíkovsky, 2002), and the diversification of the family is probably more recent than previously thought (Soares et al., 2016).

We used nucleotide substitution models selected by PartitionFinder, and linked the partitions for the tree reconstruction. We applied a log normal relaxed molecular clock for each partition (Drummond et al., 2006; Lepage et al., 2007) and gamma priors for each substitution rate (α:2, β:0.25 for all rates but AG where we used α:2, β:0.5). The five uncorrelated log-normal relaxed clock means were given a gamma prior, and invariant proportion was given a uniform prior at 0.5 [0–1]. All the unstated priors were kept with default values. We ran the analysis with a birth-death Model (Stadler, 2009) for at least 100 million generations, logging every 5000 generations. We discarded the first 10% of trees as burn-in and manually checked for convergence using Tracer v1.6 (Rambaut et al., 2014). All analyses were performed using BEAST2 v2.4.3 (Bouckaert et al., 2014).

2.2.2. Biogeographic analysis

The dated tree was then used to infer a biogeographic hypothesis for the dispersion of the Columbidae family using the package BioGeoBEARS (Matzke, 2013a, 2013b) in R (R Core Team, 2017). Each tip was assigned the whole genus current distribution (recently extinct species included), described from Baptista et al. (2017) following the biogeographic regions from Andersen et al. (2018) in which we separated Madagascar from Africa and New Guinea from Australia, giving a total of nine regions. Localization of fossils can be included but, due to a possible sample bias resulting from an uneven prospecting effort among regions [see Cracraft & Claramunt (2017) and Mayr (2017)], the analysis was performed without this information [We will try to include them later]. The maximum range size was set at eight, as none of the genera are present worldwide, and all dispersals were allowed. Each species was assigned the distribution of its whole clade. We tested six different models: the dispersal–extinction–cladogenesis model (DEC; Ree, 2005; Ree & Smith, 2008), the dispersal-vicariance analysis (DIVA; Ronquist, 1997) and the Bayesian analysis of discrete biogeographic area (BAYAREA; Landis et al., 2013), as well as the same models with the possibility of founder-effect (+J; Matzke, 2014). All these models were compared through the Akaike second order information criterion (AICc; Akaike, 1987).

2.3.Speciation rates variations across the family

Finally, the dated tree was also used to evaluate variations of speciation rates through the family, and especially to observe possible impact of insularity on this process. Only a few methods do not need an exhaustive sampling and are able to deal with higher level phylogenies. We used here the methods described in Paradis et al. (2013) and compared the different models through AIC (Akaike, 1987) thanks

to their R script run with R v.3.4.0 (R Core Team, 2017). These models were fitted against intra-family data (outgroups excluded) of stem age speciation dates (age of split with the sister clade) and number of species per clade. We used the percentage of species strictly restricted to islands as a measure of clade insularity.

3. Results

3.1. Mitochondrial and nuclear phylogenies

We recovered complete circular mitogenomes for all but five samples [*Trugon terrestris* (MZB 6123), *Ptilinopus regina* (MZB 5773), *Phapitreron amethystinus* (ZMUC 113846), *Gallicolumba tristigmata* (MZB 33661), *Pezophaps solitaria* (NHMUK)] for which up to ca. 500 bp were missing in the expected ca. 17,000 bp (<3%).

Following PartitionFinder results (Lanfear et al., 2017), data were split into four partitions for analyses with RAxML (Stamatakis, 2014) and MrBayes (Ronquist et al., 2012), with small differences between the two repartitions (Supplementary Table 3). All these partitions were assigned a GTR+I+G model (Tavaré, 1986). Convergence of the Bayesian analysis was confirmed by RWTY (Warren et al., 2017) with all ESS value greater than 200, and no tree autocorrelation.

Saturation was observed at 3rd codon position for all coding genes except *ND2*, *ND3*, *ND4*, *ND5*, and *ND6*. However, the topology seems to be balanced without major differences in branch lengths between samples, signals coming from sequences without 3rd codon position and from 3rd codon position alone seem to be mostly congruent (Supplementary Figure 1), and GTR+I+G models allow to take a lot of rates variation into account. Therefore, we decided to use the complete dataset with partitions and different evolution models, as it allows a better support for the resulting tree, and to compare the results with a dataset of coding sequences without 3rd codon positions.

We obtained nuclear data for all the samples, with a number of reads mapping against the reference comprised between 98 and 53,451 (mean: 10,588.38, sd: 8,909.94), with one outlier individual more deeply sequenced (*Starnoenas cyanocephala* Scya063) allowing to recover 258,434 reads. We kept only the 71 samples (including the two *Melanocharis* samples) for which at least 1,000 reads matched (mean: 12,045.71, sd: 8,581.93; Scya063 excluded).

The topologies obtained from mitogenomic or nuclear genomic data are mostly congruent, except for a few deep nodes that are poorly supported by mitochondrial data, while generally well supported by nuclear data (Figure 2). In particular, the early diverging Columbidae lineage differs

between the two reconstructions: with mitochondrial data, the New-World ground doves clade (the clade including *Claravis* spp.) seems to be sister to all others, while nuclear data favor *Starnoenas cyanocephala* being sister to all other species, with the Australasian clade sister to the New-World ground-doves and *Columba* spp. clade. Coding mitochondrial data without 3rd codon positions confirm clades supported by the two previous analyses but does not allow to settle deep relationships as all these nodes are poorly supported (<80% for bootstrap, <0.9 for posterior probabilities) (Figure 19).

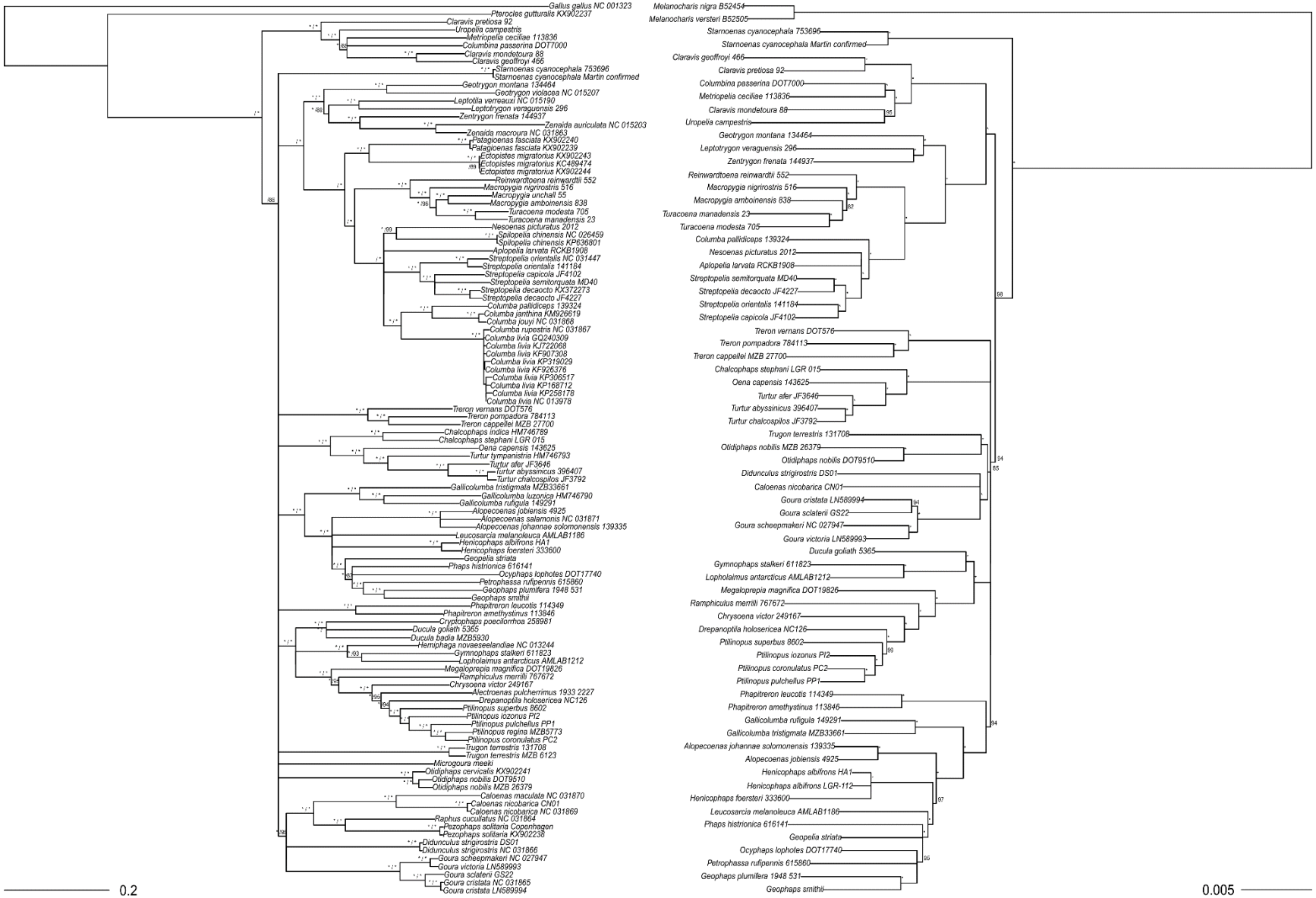

Gallus gallus NC 001323
Pterocles gutturalis KX902237
Claravis pretiosa 92
Uropelia campestris
Metriopelia ceciliae 113836
Columbina passerina DOT7000
Claravis mondetoura 88
Claravis geoffroyi 466
Starnoenas cyanocephala 753696
Starnoenas cyanocephala Martin confirmed
Geotrygon montana 134464
Geotrygon violacea NC 015207
Leptotila verreauxi NC 015190
Leptotrygon veraguensis 296
Zentrygon frenata 144937
Zenaida auriculata NC 015203
Zenaida macroura NC 031863
Patagioenas fasciata KX902240
Patagioenas fasciata KX902239
Ectopistes migratorius KX902243
Ectopistes migratorius KC489474
Ectopistes migratorius KX902244
Reinwardtoena reinwardtii 552
Macropygia nigrirostris 516
Macropygia unchall 55
Macropygia amboinensis 838
Turacoena modesta 705
Turacoena manadensis 23
Nesoenas picturatus 2012
Spilopelia chinensis NC 026459
Spilopelia chinensis KP636801
Aplopelia larvata RCKB1908
Streptopelia orientalis NC 031447
Streptopelia orientalis 141184
Streptopelia capicola JF4102
Streptopelia semitorquata MD40
Streptopelia decaocto KX372273
Streptopelia decaocto JF4227
Columba pallidiceps 139324
Columba janthina KM926619
Columba jouyi NC 031868
Columba rupestris NC 031867
Columba livia GQ240309
Columba livia KJ722068
Columba livia KF907308
Columba livia KP319029
Columba livia KF926376
Columba livia KP306517
Columba livia KP168712
Columba livia KP258178
Columba livia NC 013978
Treron vernans DOT576
Treron pompadora 784113
Treron cappellei MZB 27700
Chalcophaps indica HM746789
Chalcophaps stephani LGR 015
Oena capensis 143625
Turtur tympanistria HM746793
Turtur afer JF3646
Turtur abyssinicus 396407
Turtur chalcospilos JF3792
Gallicolumba tristigmata MZB33661
Gallicolumba luzonica HM746790
Gallicolumba rufigula 149291
Alopecoenas jobiensis 4925
Alopecoenas salamonis NC 031871
Alopecoenas johannae solomonensis 139335
Leucosarcia melanoleuca AMLAB1186
Henicophaps albifrons HA1
Henicophaps foersteri 333600
Geopelia striata
Phaps histrionica 616141
Ocyphaps lophotes DOT17740
Petrophassa rufipennis 615860
Geophaps plumifera 1948 531
Geophaps smithii
Phapitreron leucotis 114349
Phapitreron amethystinus 113846
Cryptophaps poecilorrhoa 258981
Ducula goliath 5365
Ducula badia MZB5930
Hemiphaga novaeseelandiae NC 013244
Gymnophaps stalkeri 611823
Lopholaimus antarcticus AMLAB1212
Megaloprepia magnifica DOT19826
Ramphiculus merrilli 767672
Chrysoena victor 249167
Alectroenas pulcherrimus 1933 2227
Drepanoptila holosericea NC126
Ptilinopus superbus 8602
Ptilinopus iozonus PI2
Ptilinopus pulchellus PP1
Ptilinopus regina MZB5773
Ptilinopus coronulatus PC2
Trugon terrestris 131708
Trugon terrestris MZB 6123
Microgoura meeki
Otidiphaps cervicalis KX902241
Otidiphaps nobilis DOT9510
Otidiphaps nobilis MZB 26379
Caloenas maculata NC 031870
Caloenas nicobarica CN01
Caloenas nicobarica NC 031869
Raphus cucullatus NC 031864
Pezophaps solitaria Copenhagen
Pezophaps solitaria KX902238
Didunculus strigirostris DS01
Didunculus strigirostris NC 031866
Goura scheepmakeri NC 027947
Goura victoria LN589993
Goura sclaterii GS22
Goura cristata NC 031865
Goura cristata LN589994
0.2
Melanocharis nigra B52454
Melanocharis versteri B52505
Starnoenas cyanocephala 753696
Starnoenas cyanocephala Martin confirmed
Claravis geoffroyi 466
Claravis pretiosa 92
Columbina passerina DOT7000
Metriopelia ceciliae 113836
Claravis mondetoura 88
Uropelia campestris
Geotrygon montana 134464
Leptotrygon veraguensis 296
Zentrygon frenata 144937
Reinwardtoena reinwardtii 552
Macropygia nigrirostris 516
Macropygia amboinensis 838
Turacoena manadensis 23
Turacoena modesta 705
Columba pallidiceps 139324
Nesoenas picturatus 2012
Aplopelia larvata RCKB1908
Streptopelia semitorquata MD40
Streptopelia decaocto JF4227
Streptopelia orientalis 141184
Streptopelia capicola JF4102
Treron vernans DOT576
Treron pompadora 784113
Treron cappellei MZB 27700
Chalcophaps stephani LGR 015
Oena capensis 143625
Turtur afer JF3646
Turtur abyssinicus 396407
Turtur chalcospilos JF3792
Trugon terrestris 131708
Otidiphaps nobilis MZB 26379
Otidiphaps nobilis DOT9510
Didunculus strigirostris DS01
Caloenas nicobarica CN01
Goura cristata LN589994
Goura sclaterii GS22
Goura scheepmakeri NC 027947
Goura victoria LN589993
Ducula goliath 5365
Gymnophaps stalkeri 611823
Lopholaimus antarcticus AMLAB1212
Megaloprepia magnifica DOT19826
Ramphiculus merrilli 767672
Chrysoena victor 249167
Drepanoptila holosericea NC126
Ptilinopus superbus 8602
Ptilinopus iozonus PI2
Ptilinopus coronulatus PC2
Ptilinopus pulchellus PP1
Phapitreron leucotis 114349
Phapitreron amethystinus 113846
Gallicolumba rufigula 149291
Gallicolumba tristigmata MZB33661
Alopecoenas johannae solomonensis 139335
Alopecoenas jobiensis 4925
Henicophaps albifrons HA1
Henicophaps albifrons LGR-112
Henicophaps foersteri 333600
Leucosarcia melanoleuca AMLAB1186
Phaps histrionica 616141
Geopelia striata
Ocyphaps lophotes DOT17740
Petrophassa rufipennis 615860
Geophaps plumifera 1948 531
Geophaps smithii
0.005

FIGURE 18: PHYLOGENETIC RECONSTRUCTIONS OF THE COLUMBIDAE FAMILY OBTAINED FROM THE ENTIRE MITOGENOME ON THE LEFT, AND CONSERVED NUCLEAR REGIONS ON THE RIGHT. BRANCH LENGTHS WERE OBTAINED FROM MAXIMUM-LIKELIHOOD ANALYSES. NUMBERS AT THE NODES REPRESENT BAYESIAN POSTERIOR PROBABILITIES (PP) ON THE LEFT, AND MAXIMUM LIKELIHOOD BOOTSTRAP VALUES ON THE RIGHT FOR THE MITOCHONDRIAL TREE. FOR THE NUCLEAR TREE, NUMBERS AT THE NODES REPRESENT MAXIMUM-LIKELIHOOD BOOTSTRAP VALUES. STARS CORRESPOND TO 1.0 IN PP, AND 100% IN BOOTSTRAP. NODES WITH SUPPORT VALUES LOWER THAN 80% IN BOOTSTRAP WERE COLLAPSED. NODES SUPPORT LOWER THAN 0.95 IN PP AND INTRA-SPECIFIC SUPPORT VALUES FOR *COLUMBA LIVIA* ARE REMOVED IN THE MITOCHONDRIAL TREE. NON COLLAPSED TREES ARE PRESENTED IN SUPPLEMENTARY FIGURE 2.

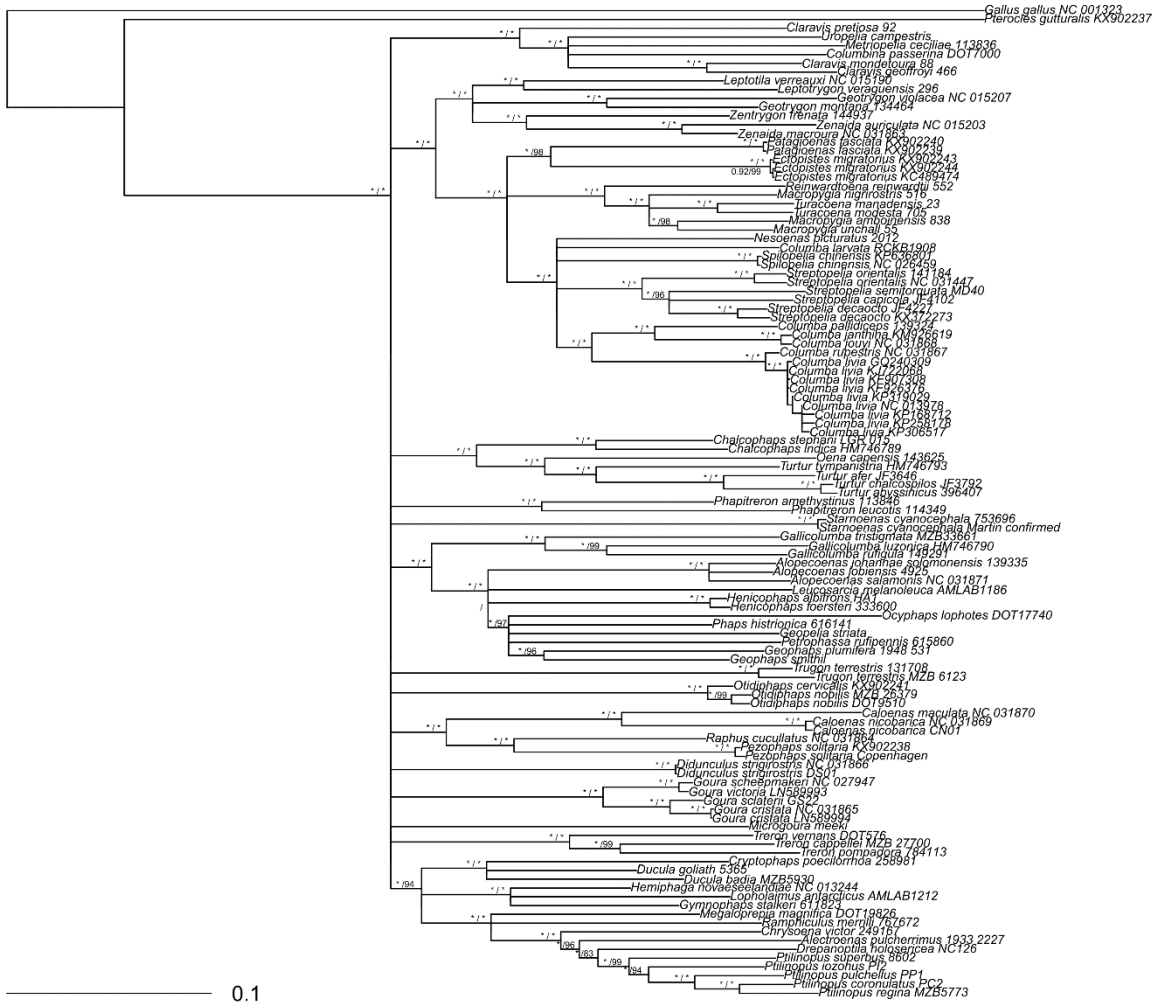


FIGURE 19: MAXIMUM-LIKELIHOOD PHYLOGENETIC TREE OBTAINED FROM MITOCHONDRIAL CODING GENES WHERE 3RD CODON POSITIONS WERE REMOVED. NODES WITH LOW SUPPORT (<80%) WERE COLLAPSED. VALUES AT NODES INDICATE BAYESIAN POSTERIOR PROBABILITIES ON THE LEFT AND MAXIMUM-LIKELIHOOD BOOTSTRAP ON THE RIGHT.

3.2. Dating and biogeography

Complete mitochondrial dataset was split into five partitions, with GTR+I+G+X models (Tavaré, 1986) for all of them (Supplementary Table 4), and mitochondrial coding genes without 3rd coding positions were split into two partitions with GTR+I+G+X models (Supplementary Table 5). We did not [yet] managed to obtain convergence for all parameters. However, large ESS values (>200) obtained for posterior probability, tree likelihoods and most recent common ancestors (MRCA) ages allow to consider the following results with rather good confidence.

We recovered a separation from the Pteroclidae around 70 Ma and a crown age occurring in the second half of Eocene (Figure 20). All the current genera were established before the mid-Miocene. The topology is roughly the same as the complete mitogenome dataset, with a few old nodes better supported. The New World ground-doves clade is here sister to the *Columba* spp. clade and the Indo-Pacific clade. Results with mitogenome coding sequences without 3rd codon positions are also comparable in ages. A few differences are present when comparing topologies of the whole dataset and the coding sequences without 3rd codon position, despite good node support (e.g. *Phapitreron leucotis*).

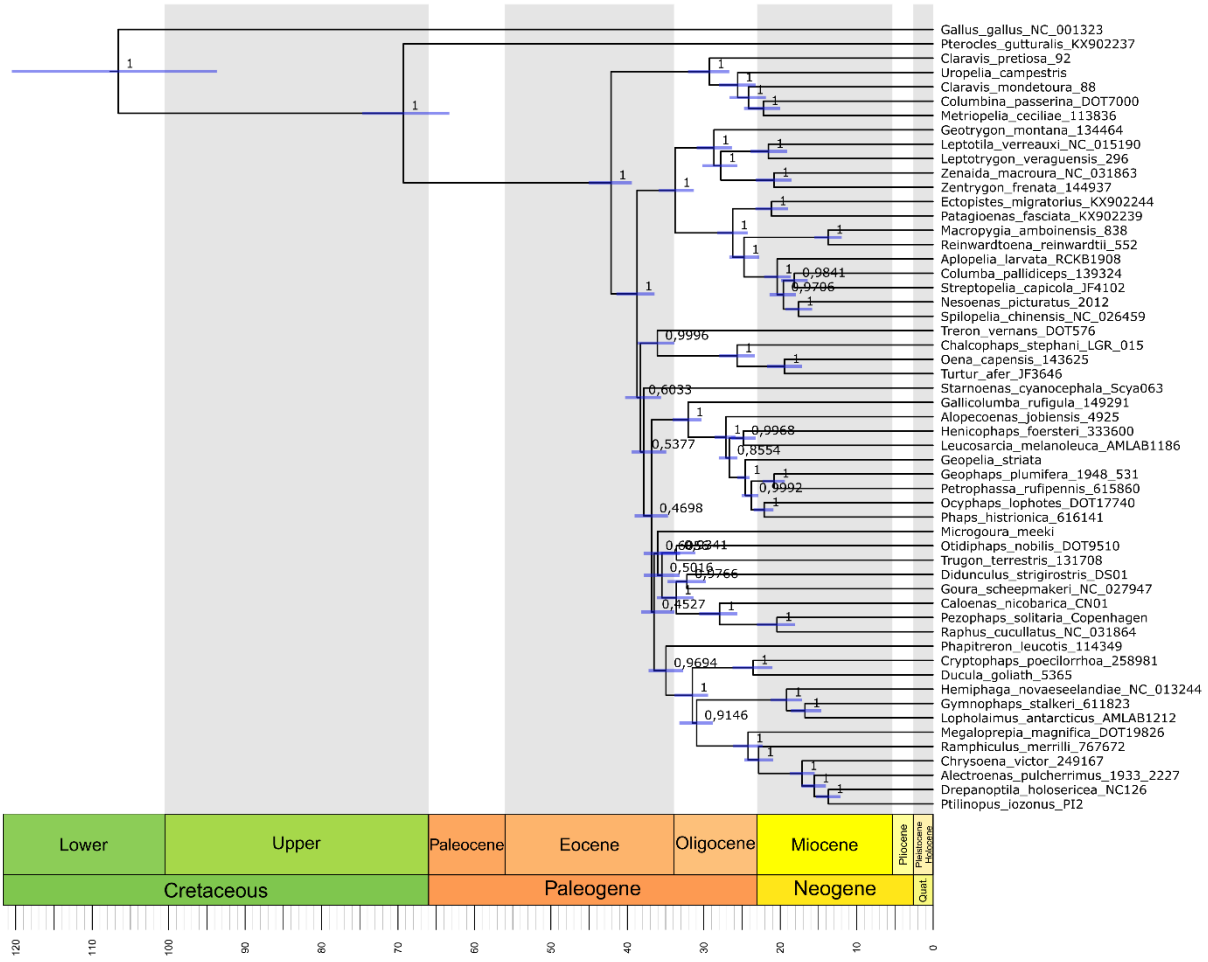


FIGURE 20: BAYESIAN PHYLOGENETIC TREE DATED BY A BEAST ANALYSIS. NUMBER AT NODES ARE POSTERIOR PROBABILITIES, AND NODES BARS REPRESENT 95% HEIGHT POSTERIOR DISTRIBUTION.

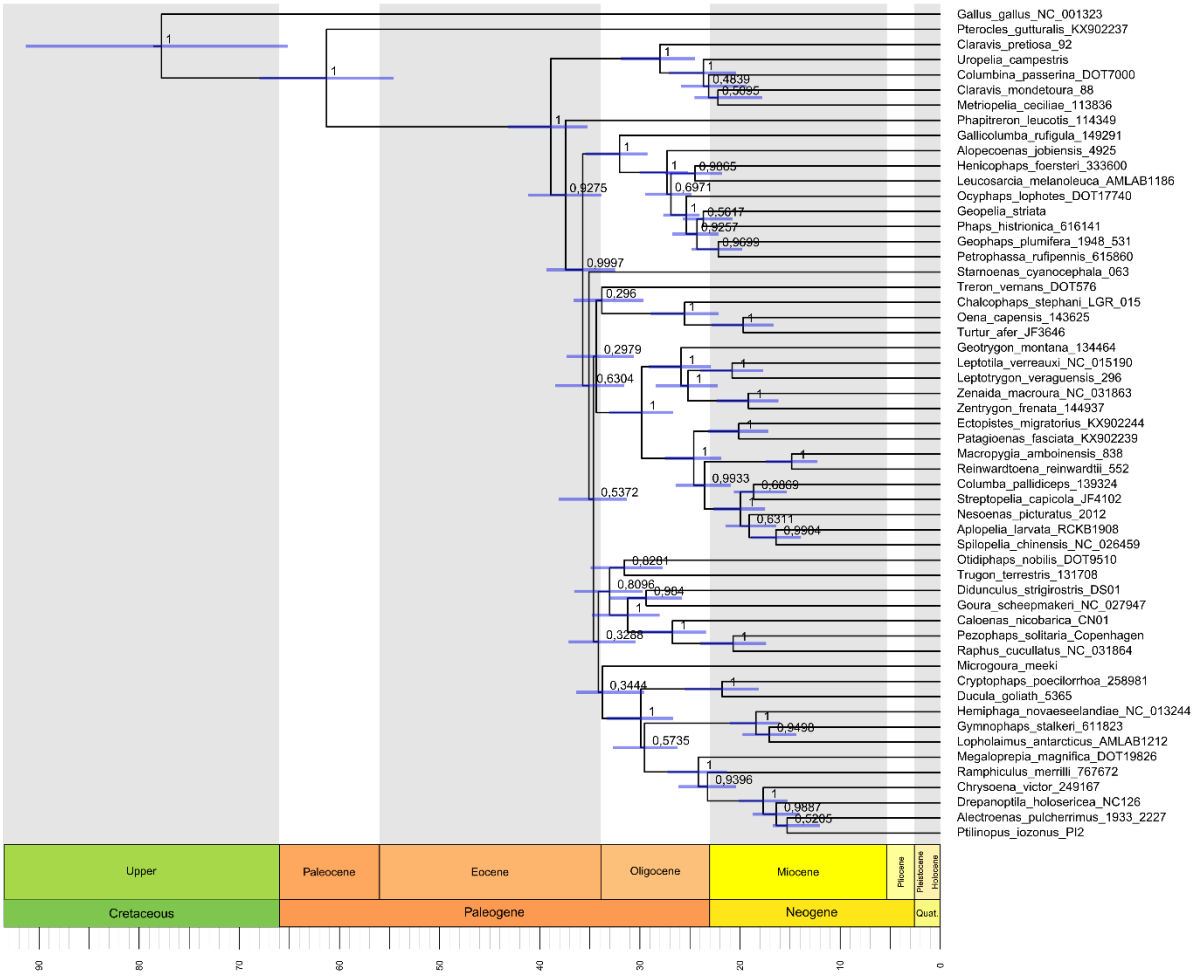


FIGURE 21: DATED PHYLOGENETIC TREE OBTAINED FROM MITOCHONDRIAL CODING GENES WITHOUT 3RD CODON POSITION. VALUES AT NODES ARE POSTERIOR PROBABILITIES, BARS ARE 95% HIGHEST POSTERIOR DENSITIES.

The current family distribution was better explained by a Bayesian analysis of discrete biogeographic area (BAYAREA; Landis et al., 2013) in which founder effects were taken into account (+J; Matzke, 2014), with an AICc weight of 99% (Table 1). Figure 22 shows the most probable ancestral area at each node following this model. Two major dispersal events seem to have occurred between the New World and the Philippines / Wallacea region around 40 and 30 Ma, allowing later diversifications in this region and the surrounding ones. In contrast, our analysis suggests that the four African and Madagascan lineages have originated in Philippines / Wallacea or surrounding regions.

TABLE 1: PERFORMANCE OF SIX MODELS OF RANGE EVOLUTION TESTED ON THE COLUMBIDAE MAXIMUM CLADE CREDIBILITY TREE, FOR NINE REGIONS AND A MAXIMUM RANGE SIZE OF EIGHT. ESTIMATED PARAMETERS: LNL, LOG-LIKELIHOOD ; P, NUMBER OF PARAMETERS; D, RANGE EXPANSION RATE; E, EXTINCTION RATE; J, RELATIVE WEIGHT FOR FOUNDER-EVENT DISPERSAL AT NODE; AICC, AKAIKE SECOND ORDER INFORMATION CRITERION ; AICC_WT, AICC WEIGHT.

	LnL	P	d	e	j	AICc	AICc_wt
DEC	-227.7	2	0.0041	0.0023	0	459.6	6.8e-07
DEC+J	-217.1	3	0.0035	1.0e-12	0.029	440.7	0.0087
DIVALIKE	-232.1	2	0.0045	0.0026	0	468.3	8.7e-09
DIVALIKE+J	-220	3	0.0036	1.0e-12	0.030	446.6	0.0005
BAYAREALIKE	-242	2	0.0036	0.025	0	488.3	4.0e-13
BAYAREALIKE+J	-212.4	3	0.0026	1.0e-07	0.045	431.2	0.99

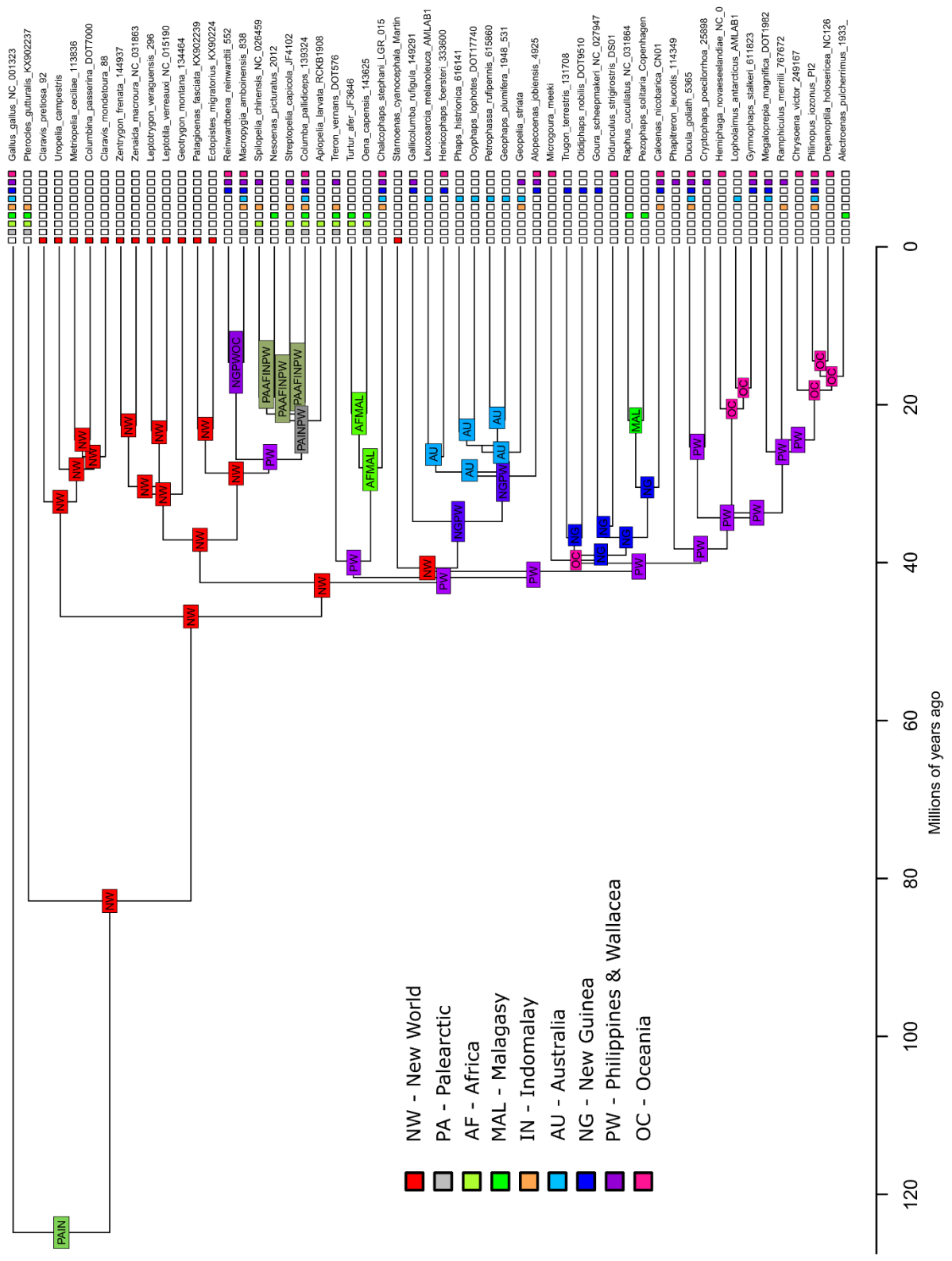


FIGURE 22: ESTIMATED ANCESTRAL RANGES OF COLUMBIDAE OBTAINED UNDER THE BAYAREA-LIKE + J MODEL OF RANGE EVOLUTION COMPUTED WITH BIOGEOBEARS, MAPPED ON THE BEAST MAXIMUM CLADE CREDIBILITY TREE. SQUARES ON THE RIGHT REPRESENT REGIONS CURRENTLY OR RECENTLY (FOR EXTINCT GENERA) OCCUPIED BY THE GENUS. LETTERS AT NODES REPRESENT MOST PROBABLE ANCESTRAL RANGES.

3.3. Diversification analyses

The first model tested assumed a constant speciation rate across the family (pure Yule model). We obtained an estimate $\hat{\lambda}$ of 0.0818 new species per million year (SE: 0.0054, AIC: 337) (Table 2). Like in Paradis et al. (2013), we did not manage to fit a birth-death model to the data. Most of the tests failed and the few results gave equal rates for speciation and extinction, which is unrealistic as the models assume an initial number of species of 1 (as analyses use stem ages).

TABLE 2: RESULTS OF FITTING MODELS OF SPECIATION VARIATION FROM PARADIS ET AL. (2013).

	ΔAIC	AIC	Estimates	Estimates standard-errors
Equal speciation rates	37	337	$\hat{\lambda}$=0.0818	0.0054
Two categories of insularity (threshold at 50%)	41	341	$\hat{\lambda}_1$=0.0815; $\hat{\lambda}_2$=0.0822	0.0075 ; 0.0201
Three categories of insularity (thresholds at 15 and 85%)	30	330	$\hat{\lambda}_1$=0.0828; $\hat{\lambda}_2$=0.1105; $\hat{\lambda}_3$=0.0587	0.0083; 0.0313; 0.0180
No a-priori (two categories)	4	304	$\hat{\lambda}_1$=0.1651; $\hat{\lambda}_2$=0.0539; $\hat{f}$=0.1393	0.0260; 0.0064; 0.0690
Normal distribution		300	g(λ)=ln(-lnλ) ~ $\mathcal{N}$(μ=1.0686; σ=0.2704)	0.0560; 0.0549

To test the impact of insularity, we first fitted a model assuming two categories of speciation rates. We separated clades as highly insular (≥50% of island restricted species) and slightly insular (<50%). We obtained speciation rates of 0.0815 (SE: 0.0075) and 0.0822 (SE: 0.0201) for highly and slightly insular clades respectively. This model did not explain better data than the previous one (AIC: 341). We therefore tried three categories, with highly insular clades (>85% of island restricted species), highly continental clades (<15%), and mixed ones (in-between). We obtained much contrasted speciation rates, with 0.0828 (SE: 0.0083), 0.1105 (SE: 0.0313) and 0.0587 (SE: 0.0180) for highly, intermediate and slightly insular clades respectively. This model explained better the data, with an AIC of 330.

Because insularity is probably not the only parameter differentiating clades, we also performed analyses with no a-priori on the number of categories and the distribution of clades among them. The best supported scenario included two categories (Table 3), with speciation rates of 0.1651 (SE: 0.0260) and 0.0539 (SE: 0.0064) and improved markedly the fitting to the data (AIC: 304). The first category included 13.93% of the clades and corresponds to *Columba* spp., *Ducula* spp., *Macropygia* and

Turacoena spp., *Patagioenas* spp., *Ptilinopus* spp. and *Streptopelia* spp. These six clades host 46% of the Columbidae species and have recent stem ages (*Ducula* spp. being the oldest at 25 Ma), implying a fast diversification process. Because ΔAIC between the models at two and three categories was only of four, we looked at the results in this second case. However, this model split the previous group of recent and species-rich clades into two, with equal probabilities for each to be assigned in one group or the second and the same speciation rate in both cases. It has therefore no biological meaning.

Finally, we tested a model of speciation rate variation across the family and following a normal distribution. This models explains the data as good as the previous one (AIC: 300). We obtained a normal distribution of λ, with a mean of 1.0686 (SE: 0.0560) and a standard deviation of 0.2704 (SE: 0.0549, Figure 23).

TABLE 3: RESULTS OF FITTING MODEL OF K CATEGORIES SPECIATION VARIATION FROM (PARADIS ET AL., 2013).

K	AIC
2	306
3	310
4	314
5	318
6	322
7	319

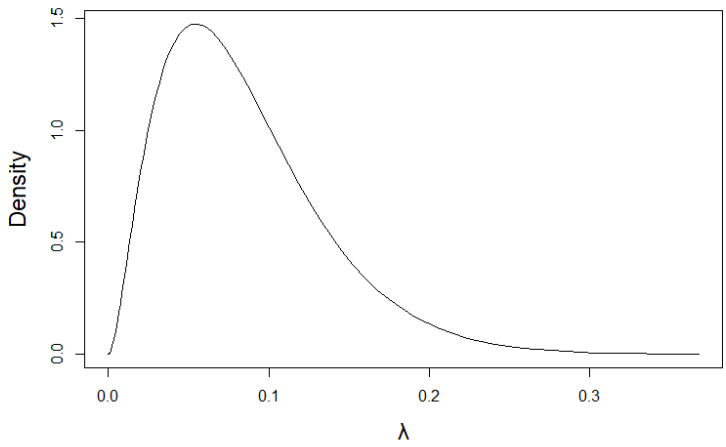


FIGURE 23: INFERRED DISTRIBUTION OF THE SPECIATION RATE Λ AMONG CLADES OF THE COLUMBIDAE FAMILY.

4. Discussion

4.1. Phylogenetic relationships

We recovered the three major clades defined by Pereira et al. (2007): the clade A, also described by Soares et al. (2016) as the Holarctic and New World clade; the clade B of New World ground-doves; and the clade C, formed by Indo-Pacific species. However, we did not recover the relationships proposed by Pereira at al. (2007) who found the clade A sister to both B and C. Here, mitochondrial data seem to favor B sister to A and C, and nuclear data support C sister to A and B. The placement of *Starnoenas cyanocephala*, which was never included before, seems also subject to discussion.

Incoherence between mitochondrial and nuclear data are frequent, due for example to introgression (Ballard & Whitlock, 2004) or resulting from near-simultaneous speciation events (*i.e.* Incomplete Lineage Sorting; Suh et al., 2015). However, the different topologies recovered for the same dataset between the two Bayesian analyses performed by MrBayes and BEAST may be due to methodological weakness. Even if these two softwares have a few methodological differences, they should converge to a common topology. Dating analyses are still in progress until we obtain convergence of all parameters. Moreover, the presence in the nuclear dataset of a clear outlier in terms of data amount could possibly unbalance the whole phylogeny, explaining the unexpected relationships found between the major clades, and especially the placement of *Starnoenas cyanocephala*.

Among the phylogenetic results obtained from the whole dataset, most of the relationships were expected from previous works (e.g. the *Claravis* spp. paraphyly), but some were slightly different (e.g. *Leucosarcia melanoleuca*). All of them are reviewed in the Supplementary results.

4.2. Dating and biogeographic analyses

The crown age of 42.12 Ma [95% HPD: 39.38-45.02] is intermediate between the estimation of Pereira et al. (2007) at 54.4 Ma [95% HPD: 46.1-63.6] and those of Soares et al. (2016) at 24.7 Ma [95% HPD: 18.9-31.3]. The difference with the latter is partially explained by the absence of the New World ground-doves clade in their sampling, but the corresponding speciation event in our results was still older 38.71 Ma [95% HPD: 36.41-41.43]. To estimate the possible impact of fossil calibration, we tried to use the only calibration Soares et al. (2016) used in their analysis: the normal prior at 55 ± 15 Ma for the Columbidae / Pteroclidae divergence. The analysis did not converge after more than 1 billion

generations, but dating results seem to be very similar to those recovered with fossils (Supplementary figure 3), and therefore slightly older than Soares' estimates.

Here, we used fossils to calibrate phylogenetic nodes. One recent method which is still under development is the use of fossils as tips, like any other sample (O'Reilly et al., 2015). However, their placement in the tree implies the use of morphological data whose recovering could be highly time-consuming. One intermediate solution is the use of the fossilized birth-death model (Heath et al., 2014) which can be used with or without morphological data if phylogenetic placement is already known. If we probably will not be able to use morphological data from all the fossils used here, their insertion in a fossilized birth-death model seems more realistic and would allow to take these species into account in the diversification process.

To estimate biases due to possible erroneous calibrations, it would be useful to use different calibration sets by removing each fossil to quantify its impact on the dating analysis. To avoid secondary calibrations using previous dating estimates which could potentially transmit errors, it would also be interesting to try using a maximum age for Columbidae diversification with the method developed by Claramunt & Cracraft (2015), who estimate a prior distribution of the clade age based on fossils occurrences. Moreover, a recent study found that molecular dating could be suitable at the family scale for the Columbidae, as the evolutionary rate seems to be constant (Quillfeldt, 2017). It would therefore allow comparing pure molecular dating with fossils calibrating. All these methods would allow evaluating the confidence we can have in the dating results.

Biogeographic results seem to favor first diversification events in the New World, which confirms previous hypothesis (Pereira et al., 2007), followed by two independent dispersal events to Philippines / Wallacea and surrounding regions. The absence of close relatives for this family makes it difficult to confidently infer an ancestral area, as current distributions could only be the result of dispersal from previous regions now unsuited for the species (*e.g.* Antarctica). However, dating analyses seem to place the crown age after the Gondwanaland breakup, and first long-distance dispersal detectable in the family seems to have occurred in Eocene, where continents were already well separated. At this time, islands were present in the South Atlantic, which could have allowed dispersal by a stepping-stone process, as previously described for Asteraceae (Katinas et al., 2013). Other islands were probably present between Africa and Madagascar, as well as between Madagascar and India, through the Seychelles Islands (Warren et al., 2010).

Other more recent long-distance dispersal events seem to have occurred in the opposite direction between New Guinea or Philippines / Wallacea and Africa, Mascareignes or Madagascar. South-East Asia was at that time more connected due to a lower see-level (Hall, 2002, 2009) and

stepping-stones dispersal are plausible as well. Moreover, winds and currents are present between Australia / Indonesia and Madagascar in summer (Warren et al., 2010), which could bring good flyers with them.

One current limit of this analysis is the absence of calibration for the New Guinea availability for colonization. This island is young, probably less than 15 million years (Hill & Hall, 2003; van Ufford & Cloos, 2005), and should not therefore be usable as an ancestral area before this date. Time-stratified analyses can be implemented in BioGeoBEARS and should lead to more realistic scenarios, even if changes should not be substantial.

4.3. Diversification analyses

The first diversification analyses conducted here did not favor scenarios where speciation rates would be higher in islands. This can be explained by the existence of species-rich genera which probably drive the models results. These genera can either be restricted to islands (e.g. *Ptilinopus* spp.), nearly absent from islands (e.g. *Streptopelia* spp.) or with an intermediate pattern of distribution (e.g. *Columba* spp.). As it was impossible here to incorporate extinction rates, it would be interesting to evaluate another method designed for higher-level phylogenies, such as described in Stadler & Smrckova (2016). Speciation and extinction rates could later be correlated with insularity.

5. Conclusion

In this second chapter, we reconstructed phylogenetic trees from a comprehensive sampling of Columbidae species. We recovered three major clades, largely consistent with previous studies (*e.g.* Shapiro et al., 2002; Pereira et al., 2007). Our dating results, which still need further improvement, tend to place the diversification of the family from the end of the Eocene to the middle of the Miocene, corresponding to intermediate results between previous studies (Pereira et al., 2007; Soares et al., 2016). Major dispersion events proposed by the most probable biogeographic scenario could have occurred thanks to stepping-stones movement on different islands (Warren et al., 2010; Katinas et al., 2013). Our current diversification analyses, on the other side, do not seem to favor scenarios with important role of islands on speciation rate variations.

This phylogeny will also allow studying the morphological diversity of the family, which, despite an important similarity in shape, show a few divergent species. Most of them live or lived on islands and belong to single-species clades, while aberrant morphologies are more often observed in fast

adaptive radiations. This should give us a more general understanding of the importance of islands in the process that led to the current Columbidae family.

References

Akaike, H. (1987) Factor analysis and AIC. *Psychometrika*, **52**, 317–332.

Andersen, M.J., McCullough, J.M., Mauck, W.M., Smith, B.T., & Moyle, R.G. (2018) A phylogeny of kingfishers reveals an Indomalayan origin and elevated rates of diversification on oceanic islands. *Journal of Biogeography*, doi:10.1111/jbi.13139.

Ballard, J.W.O. & Whitlock, M.C. (2004) The incomplete natural history of mitochondria. *Molecular Ecology*, **13**, 729–744.

Balouet, J.C. & Olson, S.L. (1987) A new extinct species of giant pigeon (Columbidae: *Ducula*) from archeological deposits on Wallis (Uvea) Island, South Pacific. *Proceedings of the Biological Society of Washington*, **100**, 769–775.

Banks, R.C., Weckstein, J.D., Remsen, J.J.V., & Johnson, K.P. (2013) Classification of a clade of New World doves (Columbidae: Zenaidini). *Zootaxa*, **3669**, 184–188.

Baptista, L.F., Trail, P.W., & Horblit, H.M. (2017) Pigeons, Doves (Columbidae). *Handbook of the Birds of the World Alive* (ed. by J. del Hoyo, A. Elliott, J. Sargatal, D.A. Christie, and E. de Juana), Lynx Edicions, Barcelona.

Barba-Montoya, J., dos Reis, M., & Yang, Z. (2017) Comparison of different strategies for using fossil calibrations to generate the time prior in Bayesian molecular clock dating. *Molecular Phylogenetics and Evolution*, **114**, 386–400.

Biber, M.F. (2017) *rasterSp - R Package for rasterizing species distribution data.*

BirdLife International and Handbook of the Birds of the World (2017) *Bird species distribution maps of the world. Version 2017.2. Available at http://datazone.birdlife.org/species/requestdis.*

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A., & Drummond, A.J. (2014) BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, **10**, e1003537.

Brandt, D.Y.C., Aguiar, V.R.C., Bitarello, B.D., Nunes, K., Goudet, J., & Meyer, D. (2015) Mapping bias overestimates reference allele frequencies at the HLA genes in the 1000 genomes project phase I data. *G3: Genes, Genomes, Genetics*, **5**, 931–941.

Brodkorb, P. (1968) An ancestral mourning dove from Rexroad, Kansas. *Quarterly Journal of the Florida Academy of Sciences*, **31**, 173–176.

Bruxaux, J., Gabrielli, M., Ashari, H., Prŷs-Jones, R., Joseph, L., Milá, B., Besnard, G., & Thébaud, C. (2018) Recovering the evolutionary history of crowned pigeons (Columbidae: *Goura*): Implications for the biogeography and conservation of New Guinean lowland birds. *Molecular Phylogenetics and Evolution*, **120**, 248–258.

Castresana, J. (2000) Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Molecular Biology and Evolution*, **17**, 540–552.

Cibois, A., Thibault, J.-C., Bonillo, C., Filardi, C.E., & Pasquet, E. (2017) Phylogeny and biogeography of the imperial pigeons (Aves: Columbidae) in the Pacific Ocean. *Molecular Phylogenetics and Evolution*, **110**, 19–26.

Cibois, A., Thibault, J.-C., Bonillo, C., Filardi, C.E., Watling, D., & Pasquet, E. (2014) Phylogeny and biogeography of the fruit doves (Aves: Columbidae). *Molecular Phylogenetics and Evolution*, **70**, 442–453.

Claramunt, S. & Cracraft, J. (2015) A new time tree reveals Earth history's imprint on the evolution of modern birds. *Science Advances*, **1**, e1501005.

Cracraft, J. & Claramunt, S. (2017) Conceptual and analytical worldviews shape differences about global avian biogeography. *Journal of Biogeography*, **44**, 958–960.

Currie, D.J., Mittelbach, G.G., Cornell, H.V., Field, R., Guégan, J.-F., Hawkins, B.A., Kaufman, D.M., Kerr, J.T., Oberdorff, T., O'Brien, E., & Turner, J.R.G. (2004) Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic richness. *Ecology Letters*, **7**, 1121–1134.

Donoghue, P.C.J. & Benton, M.J. (2007) Rocks and clocks: calibrating the Tree of Life using fossils and molecules. *Trends in Ecology & Evolution*, **22**, 424–431.

Drummond, A.J., Ho, S.Y.W., Phillips, M.J., & Rambaut, A. (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biology*, **4**, e88.

Ericson, P.G.., Anderson, C.L., Britton, T., Elzanowski, A., Johansson, U.S., Källersjö, M., Ohlson, J.I., Parsons, T.J., Zuccon, D., & Mayr, G. (2006) Diversification of Neoaves: integration of molecular sequence data and fossils. *Biology Letters*, **2**, 543–547.

Fjeldså, J. (2013) The global diversification of songbirds (Oscines) and the build-up of the Sino-Himalayan diversity hotspot. *Chinese Birds*, **4**, 132–143.

Friedman, M. (2010) Explosive morphological diversification of spiny-finned teleost fishes in the aftermath of the end-Cretaceous extinction. *Proceedings of the Royal Society of London B: Biological Sciences*, rspb20092177.

Gibb, G.C. & Penny, D. (2010) Two aspects along the continuum of pigeon evolution: A South-Pacific radiation and the relationship of pigeons within Neoaves. *Molecular Phylogenetics and Evolution*, **56**, 698–706.

Glenn, T.C., French, J.O., Heincelman, T.J., Jones, K.L., & Sawyer, R.H. (2008) Evolutionary relationships among copies of feather beta (β) keratin genes from several avian orders. *Integrative and Comparative Biology*, **48**, 463–475.

Grant, P.R. (1999) *Ecology and Evolution of Darwin's Finches.* Princeton University Press, Princeton, NJ.

Hackett, S.J., Kimball, R.T., Reddy, S., Bowie, R.C.K., Braun, E.L., Braun, M.J., Chojnowski, J.L., Cox, W.A., Han, K.-L., Harshman, J., Huddleston, C.J., Marks, B.D., Miglia, K.J., Moore, W.S., Sheldon, F.H., Steadman, D.W., Witt, C.C., & Yuri, T. (2008) A phylogenomic study of birds reveals their evolutionary history. *Science*, **320**, 1763–1768.

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. (2009) Southeast Asia's changing palaeogeography. *Blumea-Biodiversity, Evolution and Biogeography of Plants*, **54**, 148–161.

Heath, T.A. (2012) A Hierarchical Bayesian Model for Calibrating Estimates of Species Divergence Times. *Systematic Biology*, **61**, 793–809.

Heath, T.A., Huelsenbeck, J.P., & Stadler, T. (2014) The fossilized birth–death process for coherent calibration of divergence-time estimates. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, E2957–E2966.

Hill, K.C. & Hall, R. (2003) Mesozoic-Cenozoic evolution of Australia's New Guinea margin in a west Pacific context. *Geological Society of America Special Papers*, **372**, 265–290.

Ho, S.Y.W. & Phillips, M.J. (2009) Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Systematic Biology*, **58**, 367–380.

Jarvis, E.D., Mirarab, S., Aberer, A.J., et al. (2014) Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*, **346**, 1320–1331.

Jenkins, C.N., Pimm, S.L., & Joppa, L.N. (2013) Global patterns of terrestrial vertebrate diversity and conservation. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, E2602–E2610.

Johnson, K.P. (2004) Deletion bias in avian introns over evolutionary timescales. *Molecular Biology and Evolution*, **21**, 599–602.

Johnson, K.P. & Clayton, D.H. (2000) Nuclear and mitochondrial genes contain similar phylogenetic signal for pigeons and doves (Aves: Columbiformes). *Molecular Phylogenetics and Evolution*, **14**, 141–151.

Johnson, K.P., de Kort, S., Dinwoodey, K., Mateman, A.C., ten Cate, C., Lessells, C.M., Clayton, D.H., & Sheldon, F. (2001) A molecular phylogeny of the dove genera *Streptopelia* and *Columba*. *The Auk*, **118**, 874–887.

Johnson, K.P. & Weckstein, J.D. (2011) The Central American land bridge as an engine of diversification in New World doves. *Journal of Biogeography*, **38**, 1069–1076.

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.

Jønsson, K.A., Irestedt, M., Bowie, R.C.K., Christidis, L., & Fjeldså, J. (2011) Systematics and biogeography of Indo-Pacific ground-doves. *Molecular Phylogenetics and Evolution*, **59**, 538–543.

Katinas, L., Crisci, J.V., Hoch, P., Tellería, M.C., & Apodaca, M.J. (2013) Trans-oceanic dispersal and evolution of early composites (Asteraceae). *Perspectives in Plant Ecology, Evolution and Systematics*, **15**, 269–280.

Kier, G., Kreft, H., Lee, T.M., Jetz, W., Ibisch, P.L., Nowicki, C., Mutke, J., & Barthlott, W. (2009) A global assessment of endemism and species richness across island and mainland regions. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 9322–9327.

Korneliussen, T.S., Albrechtsen, A., & Nielsen, R. (2014) ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics*, **15**, 356.

Landis, M.J., Matzke, N.J., Moore, B.R., & Huelsenbeck, J.P. (2013) Bayesian analysis of biogeography when the number of areas is large. *Systematic Biology*, **62**, 789–804.

Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., & Calcott, B. (2017) PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution*, **34**, 772–773.

Lepage, T., Bryant, D., Philippe, H., & Lartillot, N. (2007) A general comparison of relaxed molecular clock models. *Molecular Biology and Evolution*, **24**, 2669–2680.

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.

MacArthur, R.H. & Wilson, E.O. (1967) *Theory of Island Biogeography.* Princeton University Press, Princeton, NJ.

Matzke, N.J. (2013a) *BioGeoBEARS: BioGeography with Bayesian (and likelihood) evolutionary analysis in R Scripts.*

Matzke, N.J. (2013b) Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. *Frontiers of Biogeography*, **5**, 242–248.

Matzke, N.J. (2014) Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. *Systematic Biology*, **63**, 951–970.

Mayr, E. (1963) *Animal species and evolution.* Belknap Press of Harvard University Press, Cambridge, Massachusetts.

Mayr, G. (2017) Avian higher level biogeography: Southern Hemispheric origins or Southern Hemispheric relicts? *Journal of Biogeography*, **44**, 956–958.

McCormack, J.E., Hird, S.M., Zellmer, A.J., Carstens, B.C., & Brumfield, R.T. (2013) Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution*, **66**, 526–538.

McCormack, J.E., Tsai, W.L.E., & Faircloth, B.C. (2016) Sequence capture of ultraconserved elements from bird museum specimens. *Molecular Ecology Resources*, **16**, 1189–1203.

Millener, P.R. & Powlesland, R.G. (2001) The Chatham Islands pigeon (Parea) deserves full species status; *Hemiphaga chathamensis* (Rothschild 1891); Aves: Columbidae. *Journal of the Royal Society of New Zealand*, **31**, 365–383.

Mlíkovsky, J. (2002) *Cenozoic birds of the world.* Ninox Press, Praha.

Moore, W.S. (1995) Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. *Evolution*, **49**, 718–726.

Moyle, R.G., Jones, R.M., & Andersen, M.J. (2013) A reconsideration of *Gallicolumba* (Aves: Columbidae) relationships using fresh source material reveals pseudogenes, chimeras, and a novel phylogenetic hypothesis. *Molecular Phylogenetics and Evolution*, **66**, 1060–1066.

Newton, I. (2003) *Speciation and biogeography of birds.* Academic Press, London.

Olson, S.L. (2011) The fossil record and history of doves on Bermuda (Aves: Columbidae). *Proceedings of the Biological Society of Washington*, **124**, 1–6.

Olson, S.L. & Rasmussen, P.C. (2001) Miocene and Pliocene birds from the Lee Creek Mine; North Carolina. *Geology and paleontology of the Lee Creek mine, North Carolina, III.* pp. 233–365.

O'Reilly, J.E., dos Reis, M., & Donoghue, P.C.J. (2015) Dating Tips for Divergence-Time Estimation. *Trends in Genetics*, **31**, 637–650.

Paradis, E., Tedesco, P.A., & Hugueny, B. (2013) Quantifying variation in speciation and extinction rates with clade data. *Evolution*, **67**, 3617–3627.

Parham, J.F., Donoghue, P.C.J., Bell, C.J., et al. (2012) Best practices for justifying fossil calibrations. *Systematic Biology*, **61**, 346–359.

Pereira, S.L., Johnson, K.P., Clayton, D.H., Baker, A.J., & Paterson, A. (2007) Mitochondrial and nuclear DNA sequences support a Cretaceous origin of Columbiformes and a dispersal-driven radiation in the Paleogene. *Systematic Biology*, **56**, 656–672.

Philippe, H., Sörhannus, U., Baroin, A., Perasso, R., Gasse, F., & Adoutte, A. (1994) Comparison of molecular and paleontological data in diatoms suggests a major gap in the fossil record. *Journal of Evolutionary Biology*, **7**, 247–265.

Pratt, T.K. & Beehler, B.M. (2015) *Birds of New Guinea.* Princeton University Press, Princeton, NJ.

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.

Pyle R.L. & Pyle P. (2017) Available at: http://hbs.bishopmuseum.org/birds/rlp-monograph/Default.htm.

Quillfeldt, P. (2017) Body mass is less important than bird order in determining the molecular rate for bird mitochondrial DNA. *Molecular Ecology*, **26**, 2426–2429.

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/.*

Ree, R.H. (2005) Detecting the historical signature of key innovations using stochastic models of character evolution and cladogenesis. *Evolution*, **59**, 257–265.

Ree, R.H. & Smith, S.A. (2008) Maximum Likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology*, **57**, 4–14.

dos Reis, M., Donoghue, P.C.J., & Yang, Z. (2016) Bayesian molecular clock dating of species divergences in the genomics era. *Nature Reviews Genetics*, **17**, 71–80.

Ronquist, F. (1997) Dispersal-vicariance analysis: A new approach to the quantification of historical biogeography. *Systematic Biology*, **46**, 195–203.

Ronquist, F., Teslenko, M., Mark, P. van der, Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., & Huelsenbeck, J.P. (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, **61**, 539–542.

Roy, K. & Goldberg, E.E. (2007) Origination, extinction, and dispersal: Integrative models for understanding present-day diversity gradients. *The American Naturalist*, **170**, S71–S85.

Shapiro, B., Sibthorpe, D., Rambaut, A., Austin, J., Wragg, G.M., Bininda-Emonds, O.R.P., Lee, P.L.M., & Cooper, A. (2002) Flight of the Dodo. *Science*, **295**, 1683–1683.

Simpson, G.G. (1947) Evolution, Interchange, and Resemblance of the North American and Eurasian Cenozoic Mammalian Faunas. *Evolution*, **1**, 218–220.

Soares, A.E.R., Novak, B.J., Haile, J., Heupink, T.H., Fjeldså, J., Gilbert, M.T.P., Poinar, H., Church, G.M., & Shapiro, B. (2016) Complete mitochondrial genomes of living and extinct pigeons revise the timing of the columbiform radiation. *BMC Evolutionary Biology*, **16**, 230.

Stadler, T. & Smrckova, J. (2016) Estimating shifts in diversification rates based on higher-level phylogenies. *Biology Letters*, **12**, 20160273.

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

Steadman, D. (2008) Doves (Columbidae) and Cuckoos (Cuclidae) from the Early Miocene of Florida. *Bulletin of the Florida Museum of Natural History*, **48**, 1–16.

Stresemann, E. (1941) Die Vögel von Celebes. *Journal für Ornithologie*, **89**, 1–102.

Suh, A., Smeds, L., & Ellegren, H. (2015) The dynamics of incomplete lineage sorting across the ancient adaptive radiation of neoavian birds. *PLoS Biology*, **13**, e1002224.

Sweet, A.D. & Johnson, K.P. (2015) Patterns of diversification in small New World ground doves are consistent with major geologic events. *The Auk*, **132**, 300–312.

Sweet, A.D., Maddox, J.D., & Johnson, K.P. (2017) A complete molecular phylogeny of *Claravis* confirms its paraphyly within small New World ground-doves (Aves: Peristerinae) and implies multiple plumage state transitions. *Journal of Avian Biology*, **48**, 459–464.

Tavaré, S. (1986) Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on mathematics in the life sciences*, **17**, 57–86.

Tershy, B.R., Shen, K.-W., Newton, K.M., Holmes, N.D., & Croll, D.A. (2015) The importance of islands for the protection of biological and linguistic diversity. *BioScience*, **65**, 592–597.

van Ufford, A.Q. & Cloos, M. (2005) Cenozoic tectonics of New Guinea. *AAPG Bulletin*, **89**, 119–140.

Warren, B.H., Strasberg, D., Bruggemann, J.H., Prys-Jones, R.P., & Thébaud, C. (2010) Why does the biota of the Madagascar region have such a strong Asiatic flavour? *Cladistics*, **26**, 526–538.

Warren, D.L., Geneva, A.J., & Lanfear, R. (2017) RWTY (R We There Yet): An R package for examining convergence of Bayesian phylogenetic analyses. *Molecular Biology and Evolution*, **34**, 1016–1020.

Weir, J.T. (2006) Divergent timing and patterns of species accumulation in lowland and highland neotropical birds. *Evolution*, **60**, 842–855.

Worthy, T., Wragg, G., Clark, G., & Leach, F. (2008) A new genus and species of pigeon (Aves: Columbidae) from Henderson Island, Pitcairn Group. *Islands of Inquiry - Colonisation, seafaring and the archaeology of maritime landscapes* (ed. by G. Clark, F. Leach, and S. O'Connor), pp. 510. Australian National University Press, Canberra.

Worthy, T.H. (2001) A giant flightless pigeon gen. et sp. nov. and a new species of *Ducula* (Aves: Columbidae), from Quaternary deposits in Fiji. *Journal of the Royal Society of New Zealand*, **31**, 763–794.

Worthy, T.H. (2012) A phabine pigeon (Aves : Columbidae) from Oligo-Miocene Australia. *Emu*, **112**, 23–31.

Worthy, T.H., Hand, S.J., Worthy, J.P., Tennyson, A.J.D., & Paul Scofield, R. (2009) A large fruit pigeon (Columbidae) from the early Miocene of New Zealand. *The Auk*, **126**, 649–656.

Worthy, T.H. & Wragg, G.M. (2003) A new species of *Gallicolumba*: Columbidae from Henderson Island, Pitcairn Group. *Journal of the Royal Society of New Zealand*, **33**, 769–793.

Wragg, G.M. & Worthy, T.H. (2006) A new species of extinct imperial pigeon (*Ducula*: Columbidae) from Henderson Island, Pitcairn Group. *Historical Biology*, **18**, 131–144.

Wu, T.D. & Nacu, S. (2010) Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics*, **26**, 873–881.

Supplementary material

The supplementary material includes:

Supplementary results: Taxonomic implications of the phylogenetic results.

Supplementary Figure 1: Comparison of the phylogenetic trees built from complete mitogenome on the left and coding sequence without the 3rd codon position (A) or the 3rd codon position only (B).

Supplementary Figure 2: Comparison of the non-collapsed phylogenetic trees built from complete mitogenome on the left and from nuclear genomic data on the right.

Supplementary Figure 3: Dated phylogenetic tree obtained from a BEAST analysis performed without fossils.

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

Supplementary Table 2: Published mitogenomes included in the phylogenetic analysis.

Supplementary Table 3: Description of each partition obtained for the whole mitogenome.

Supplementary Table 4: Description of each partition obtained for the whole mitogenome dating analysis.

Supplementary Table 5: Description of each partition obtained for the coding sequence mitogenome dating analysis from which 3rd codon positions were removed.

Supplementary results: Taxonomic implications of the phylogenetic results.

In both mitochondrial and nuclear tree, we recover the same three major clades as Pereira et al. (2007). However, we observe different relationships. Pereira found the clade A (*Columba* spp. clade) sister to both clades B and C, where we find clade B (the New-World ground-doves) sister to both clades A and C.

Starnoenas cyanocephala, which was not included in previous studies and therefore not assigned to any of the Pereira's clade, was found here sister to both clades A and C with mitochondrial data, and sister to the three clades in nuclear. It was previously considered as a close relative of other New-World Ground-doves (*Geotrygon* spp.) (Goodwin, 1983; Gibbs et al., 2001) or of Australasian genus *Geophaps* spp. based on behavioral data (Olson & Wiley, 2016).

In the clade B, we confirm the paraphyly of the genus *Claravis* (Sweet & Johnson, 2015; Sweet et al., 2017). We recover the same relationships as Sweet et al. (2017) for the mitochondrial data but not for the nuclear data, where *C. geoffroyi* clusters with *C. pretiosa*.

In the clade A, we confirm the relationships between New World doves (Johnson & Weckstein, 2011; Banks et al., 2013), and between *Ectopistes migratorius* and *Patagioenas fasciata* (Johnson et al., 2001; Soares et al., 2016). We find *Reinwardtoena reinwardti* sister to *Macropygia* spp. and *Turacoena* spp., as expected from Pereira et al. (2007) and Fulton et al. (2012). However, we found *Macropygia* spp. being paraphyletic, with *Turacoena* spp. nested inside (Stresemann, 1941). *Nesoenas picturatus* is found sister of *Spilopelia chinensis* with mitochondrial data, which confirms results from Johnson et al. (2001). *Aplopelia larvata*, which was never included in any phylogeny before, is found sister of all the *Streptopelia* spp. with mitochondrial data, with a congruent placement in the nuclear tree when taking into account different sampling and low support of the relation with *Nesoenas picturatus*.

In the clade C, we confirm the monophyletic group formed by *Chalcophaps* spp., *Oena capensis*, *Turtur* spp., and *Treron* spp. as sister of the rest of the clade (Shapiro et al., 2002). Despite their name, *Phapitreron* spp. are not related to the genus *Treron*, but its placement is still unclear: we find it sister to the clade belonging *Gallicolumba* spp. for nuclear data, and sister to the *Ducula* and *Ptilinopus* clade with mitogenomes. We confirm that all "big and/or peculiar pigeons" (*Goura* spp., *Didunculus strigirostris*, *Caloenas* spp., *Otidiphaps* spp., *Trugon terrestris*, *Raphus cuculatus*, *Pezophaps solitaria* and *Microgoura meeki*) cluster together, in line with Soares et al. (2016). In this group, the relative placement of *Didunculus* and *Caloenas* is still unclear due to sampling differences between mitochondrial and nuclear data. The placement of *Microgoura meeki* as sister of *Otidiphaps* spp. need to be confirmed with nuclear data.

We obtain one clade with *Gallicolumba* spp., *Alopecoenas* spp., *Henicophaps* spp., *Leucosarcia melanoleuca*, *Phaps histrionica*, *Geopelia striata*, *Ocyphaps lophotes*, *Geophaps* spp. and *Petrophassa rufipennis*. We agree with (Moyle et al., 2013) on *Gallicolumba* monophyly and *Alopecoenas* differentiation, but find *Henicophaps* sister to *Alopecoenas* and the other clade members, unlike Fulton et al. (2012). The placement of *Leucosarcia melanoleuca* is highly supported with nuclear data as sister of the last part of the clade, where Moyle et al. (2013) had found *Ocyphaps lophotes*. The

placement of *Geopelia striata* and *Phaps histrionica* is less clear, with different slightly supported relationships with mitochondrial and nuclear data [to be confirmed with last results].

We place for the first time *Cryptophaps poecilorrhoa*, as sister of *Ducula* spp. (as hypothesized by Goodwin (1983). *Hemiphaga novaeseelandiae*, *Gymnophaps stalkeri* and *Lopholaimus antarcticus* are found clustering together, as described by Gibb & Penny (2010). Finally, *Ptilinopus* spp. and their close relatives are found as described by (Cibois et al., 2014).

References

Banks, R.C., Weckstein, J.D., Remsen, J.J.V., & Johnson, K.P. (2013) Classification of a clade of New World doves (Columbidae: Zenaidini). *Zootaxa*, **3669**, 184–188.

Cibois, A., Thibault, J.-C., Bonillo, C., Filardi, C.E., Watling, D., & Pasquet, E. (2014) Phylogeny and biogeography of the fruit doves (Aves: Columbidae). *Molecular Phylogenetics and Evolution*, **70**, 442–453.

Fulton, T.L., Wagner, S.M., Fisher, C., & Shapiro, B. (2012) Nuclear DNA from the extinct Passenger Pigeon (*Ectopistes migratorius*) confirms a single origin of New World pigeons. *Annals of Anatomy - Anatomischer Anzeiger*, **194**, 52–57.

Gibb, G.C. & Penny, D. (2010) Two aspects along the continuum of pigeon evolution: A South-Pacific radiation and the relationship of pigeons within Neoaves. *Molecular Phylogenetics and Evolution*, **56**, 698–706.

Gibbs, D., Barnes, E., & Cox, J. (2001) *Pigeons and doves: a guide to the pigeons and doves of the world.* A&C Black, London.

Goodwin, D.H. (1983) *Pigeons and doves of the world.* Comstock Pub. Associates,

Johnson, K.P., de Kort, S., Dinwoodey, K., Mateman, A.C., ten Cate, C., Lessells, C.M., Clayton, D.H., & Sheldon, F. (2001) A molecular phylogeny of the dove genera *Streptopelia* and *Columba*. *The Auk*, **118**, 874–887.

Johnson, K.P. & Weckstein, J.D. (2011) The Central American land bridge as an engine of diversification in New World doves. *Journal of Biogeography*, **38**, 1069–1076.

Moyle, R.G., Jones, R.M., & Andersen, M.J. (2013) A reconsideration of *Gallicolumba* (Aves: Columbidae) relationships using fresh source material reveals pseudogenes, chimeras, and a novel phylogenetic hypothesis. *Molecular Phylogenetics and Evolution*, **66**, 1060–1066.

Olson, S.L. & Wiley, J.W. (2016) The Blue-headed Quail-Dove (*Starnoenas cyanocephala*): an Australasian dove marooned in Cuba. *The Wilson Journal of Ornithology*, **128**, 1–21.

Pereira, S.L., Johnson, K.P., Clayton, D.H., Baker, A.J., & Paterson, A. (2007) Mitochondrial and nuclear DNA sequences support a Cretaceous origin of Columbiformes and a dispersal-driven radiation in the Paleogene. *Systematic Biology*, **56**, 656–672.

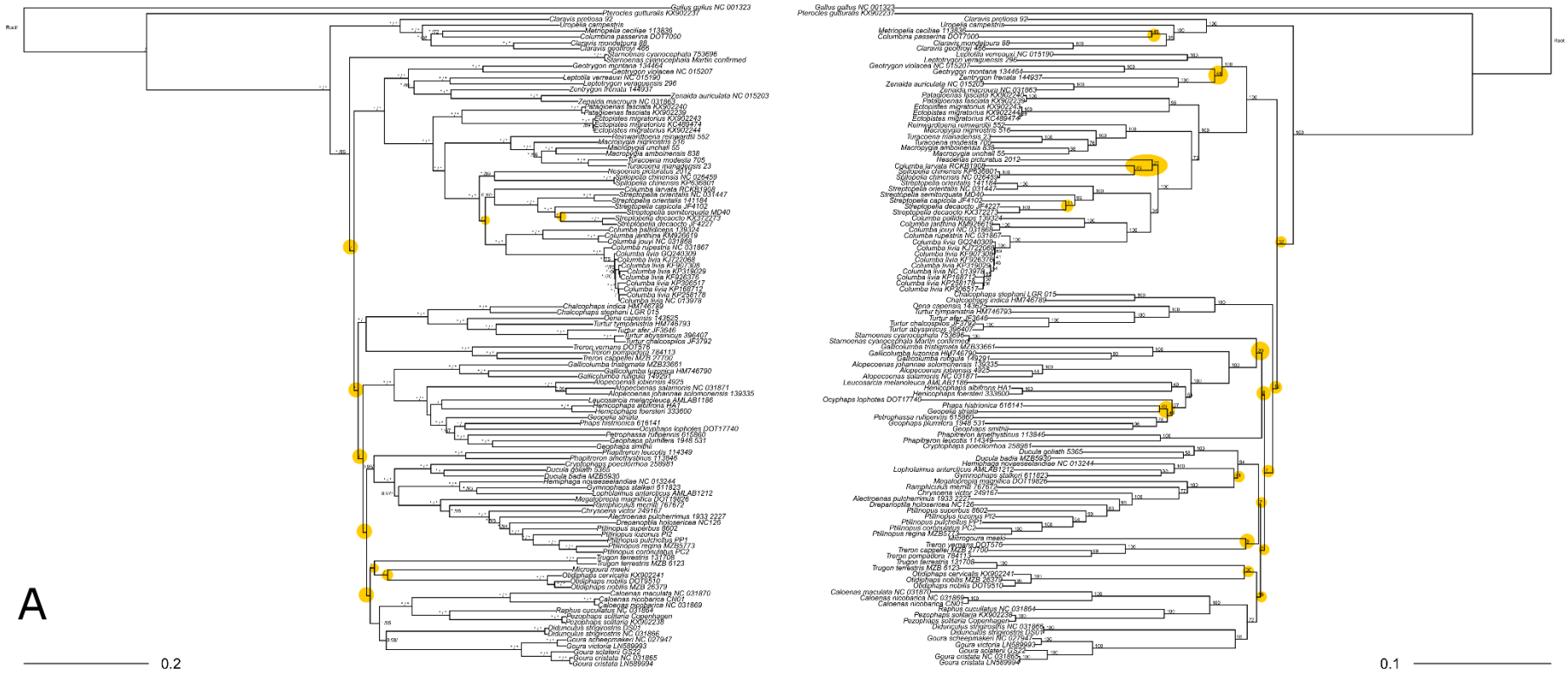
Shapiro, B., Sibthorpe, D., Rambaut, A., Austin, J., Wragg, G.M., Bininda-Emonds, O.R.P., Lee, P.L.M., & Cooper, A. (2002) Flight of the Dodo. *Science*, **295**, 1683–1683.

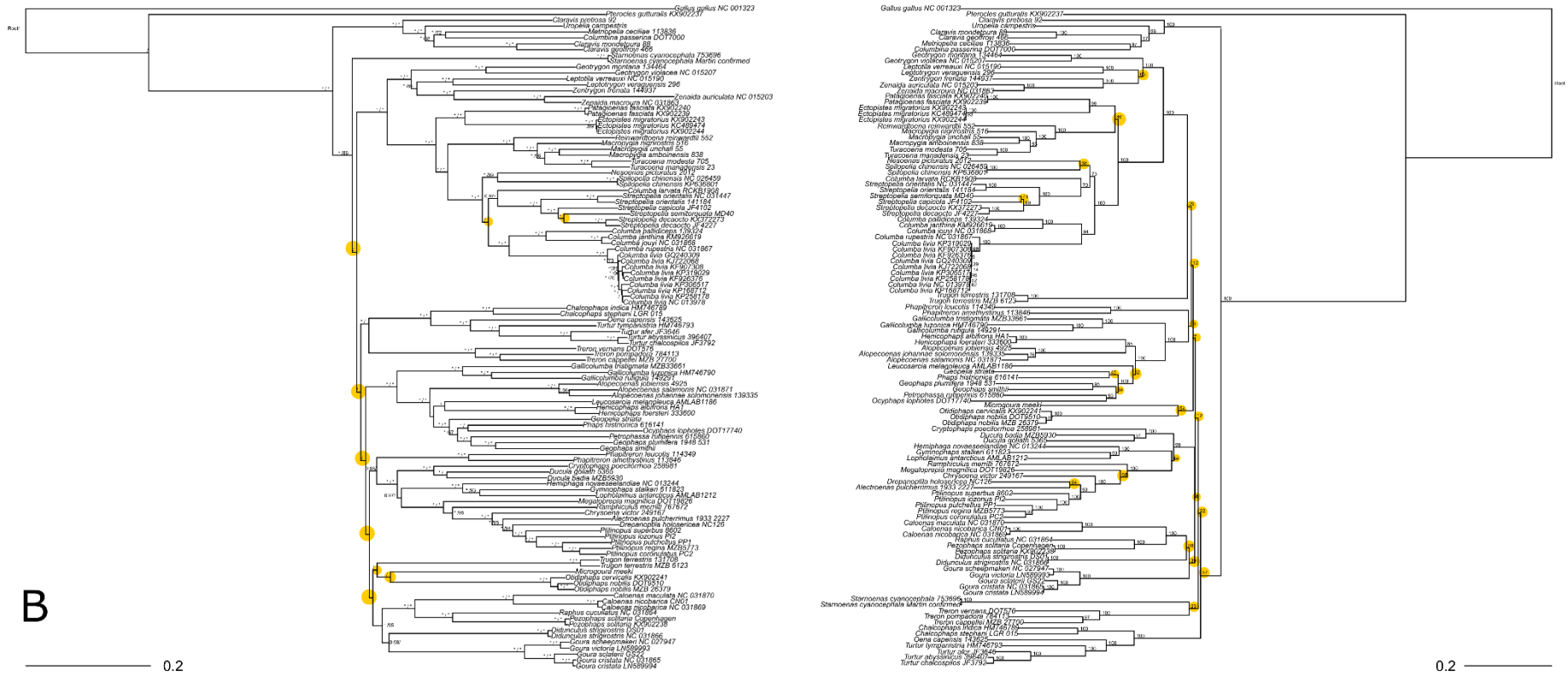
Soares, A.E.R., Novak, B.J., Haile, J., Heupink, T.H., Fjeldså, J., Gilbert, M.T.P., Poinar, H., Church, G.M., & Shapiro, B. (2016) Complete mitochondrial genomes of living and extinct pigeons revise the timing of the columbiform radiation. *BMC Evolutionary Biology*, **16**, 230.

Stresemann, E. (1941) Die Vögel von Celebes. *Journal für Ornithologie*, **89**, 1–102.

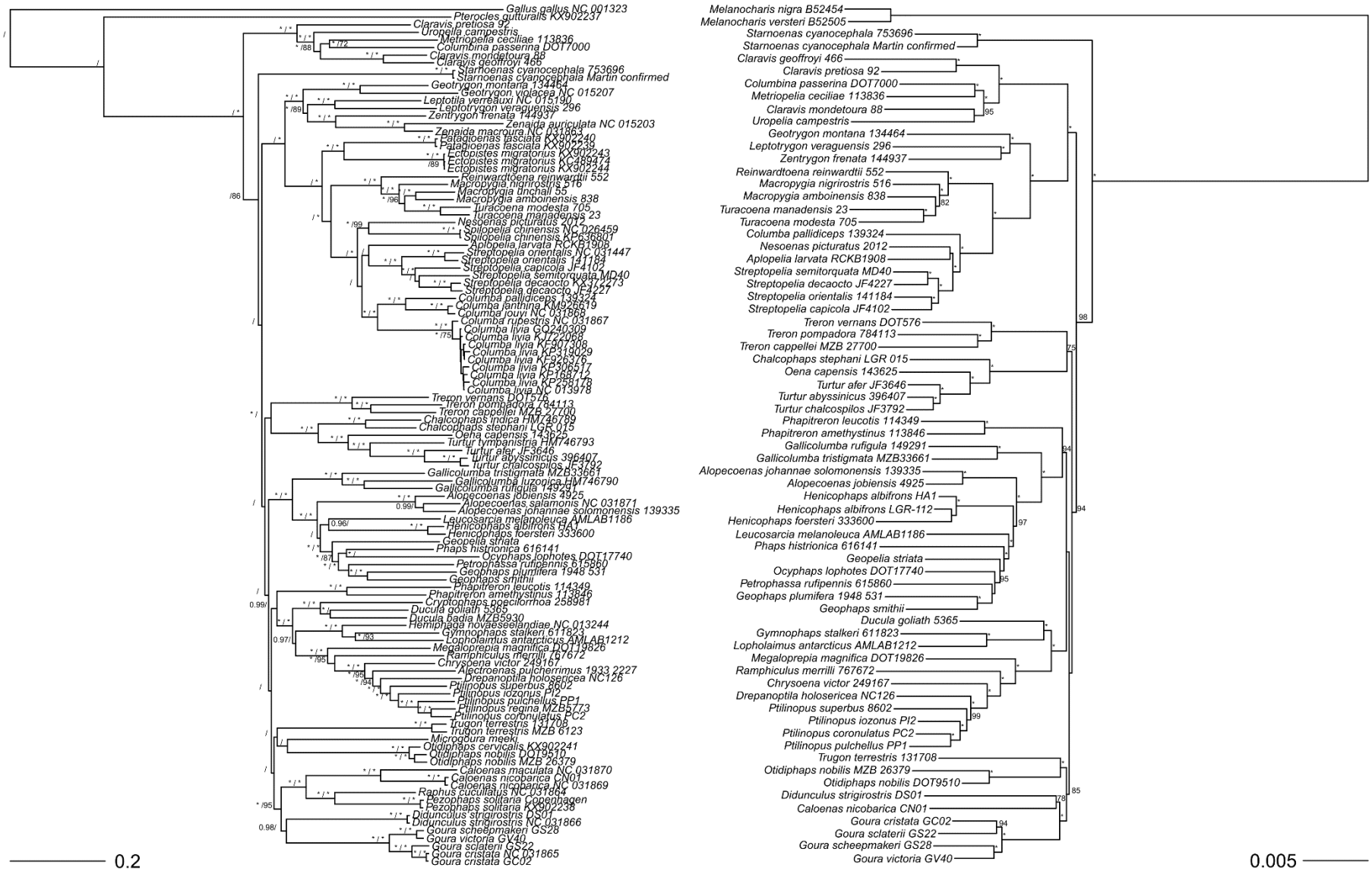
Sweet, A.D. & Johnson, K.P. (2015) Patterns of diversification in small New World ground doves are consistent with major geologic events. *The Auk*, **132**, 300–312.

Sweet, A.D., Maddox, J.D., & Johnson, K.P. (2017) A complete molecular phylogeny of *Claravis* confirms its paraphyly within small New World ground-doves (Aves: Peristerinae) and implies multiple plumage state transitions. *Journal of Avian Biology*, **48**, 459–464.

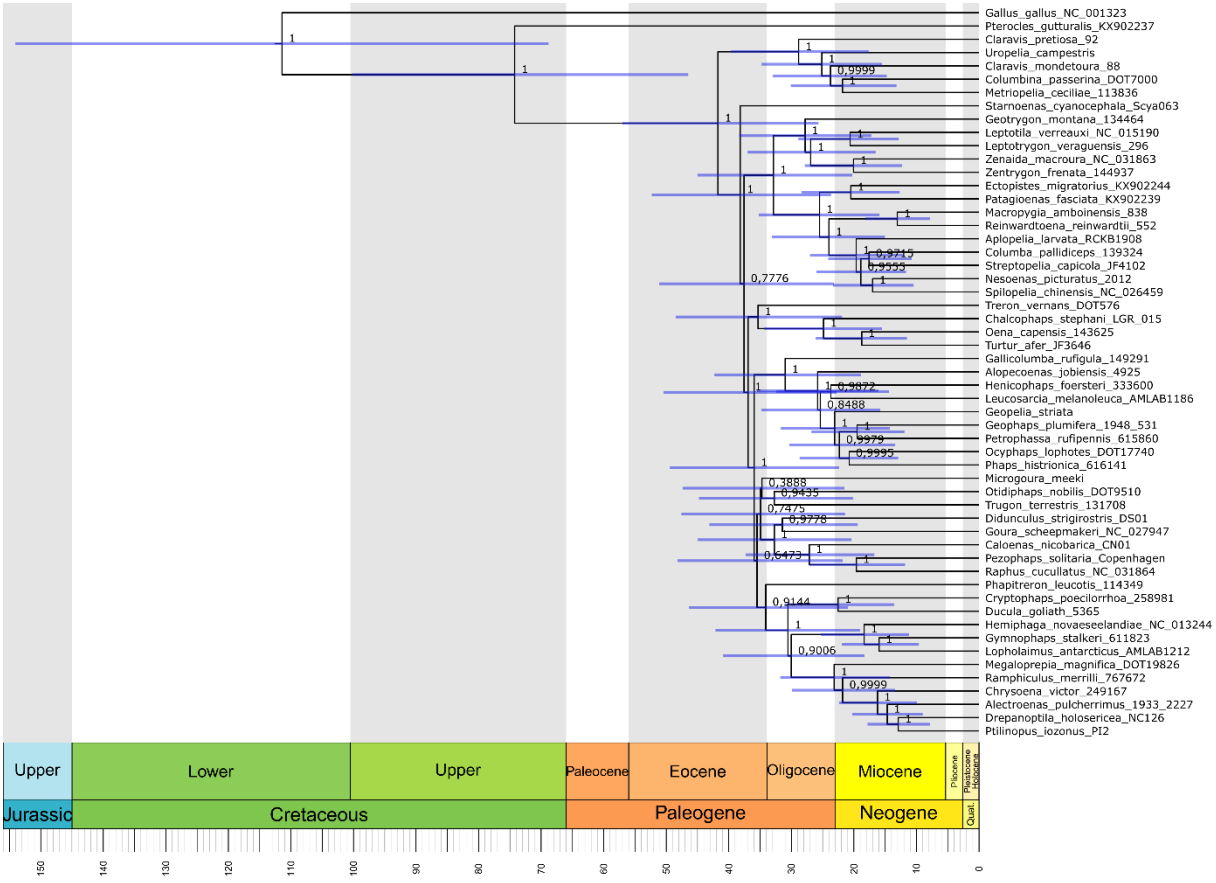

A
Root
Gallus gallus NC 001323
Pterocles gutturalis KX902237
Clarsiva protiosa 92
Urapelia camprestris
Metriopelia cecilise 113836
Columbina passerina DOT7000
Claravis mondetoura 88
Claravis geoffroyi 886
Starnoenas cyanocephala 753696
Starnoenas cyanocephala Martin confirmed
Geotrygon montana 134464
Geotrygon violacea NC 015207
Leptotila verreauxi NC 015190
Leptotrygon veraguensis 296
Zentrygon frenata 144937
Zenaida auriculata NC 015203
Zenaida macroura NC 031865
Patagioenas fasciata KX902240
Patagioenas fasciata KX902239
Ectopistes migratorius KX902243
Ectopistes migratorius KC488474
Ectopistes migratorius KX902244
Reinwardtoena reinwardtii 552
Macropygia nigrirostris 516
Macropygia unchall 55
Macropygia amboinensis 836
Turacoena modesta 705
Turacoena manadensis 23
Nesoenas picturatus 2012
Columba larvata RCKB1906
Streptopelia orientalis NC 031447
Streptopelia capicola JF4102
Streptopelia decaocto KX372273
Columba palliolaena 139324
Columba janthina RM929819
Columba jouyi NC 031868
Columba rupestris NC 031867
Columba livia GQ240309
Columba livia KJ722068
Columba livia KF907306
Columba livia KP318029
Columba livia KP926375
Columba livia KP906517
Columba livia KP168712
Columba livia KP926178
Columba livia KP926177
Columba livia NC 013978
Chalcophaps indica HM746789
Chalcophaps stephani LGR 015
Oena capensis 143625
Turtur tympanistria HM746793
Turtur afer JF3646
Turtur abyssinicus 396407
Turtur chalcospilos JF3792
Treron vernans DOT576
Treron sieboldii 784113
Treron calvus NMB 27702
Treron cappellei NMB 27702
Gallicolumba tristigmata MZB39661
Gallicolumba luzonica HM746790
Gallicolumba rufigula 149291
Alopecoenas xanthonura 4925
Alopecoenas salamonis NC 031871
Alopecoenas johannae solomonensis 139335
Leucosarcia melanoleuca AMLAB1186
Henicophaps albifrons HA1
Henicophaps foersteri 333600
Geopelia striata
Phaps histrionica 616141
Ocyphaps lophotes DOT17740
Petrophassa rufipennis 615860
Geophaps plumifera 1948 531
Geophaps smithii
Phapitreron leucotis 114349
Phapitreron amethystinus 113846
Cryptophaps poecilorrhoa 258981
Ducula goliath 5365
Ducula badia MZB5930
Henicophaps noveseelandiae NC 013244
Gymnophaps stalkeri 611823
Lopholaimus antarcticus AMLAB1212
Megaloprepia magnifica DOT19626
Ramphiculus melmli 767672
Chrysoena victor 249167
Alectroenas pulcherrimus 1933 2227
Drepanoptila holosericea NC126
Ptilinopus superbus 8602
Ptilinopus iozonus PI2
Ptilinopus pulchellus PP1
Ptilinopus regina MZB5773
Ptilinopus coronulatus PC2
Trugon terrestris 131708
Trugon terrestris MZB 6123
Microgoura meeki
Otidiphaps cervicalis KX902241
Otidiphaps nobilis DOT9510
Otidiphaps nobilis MZB 26379
Caloenas maculata NC 031870
Caloenas nicobarica CN01
Caloenas nicobarica NC 031869
Raphus cucullatus NC 031864
Pezophaps solitaria Copenhagen
Pezophaps solitaria KX902238
Didunculus strigirostris DS01
Didunculus strigirostris NC 031866
Goura scheepmakeri NC 027947
Goura victoria LN589993
Goura sclaterii GS25
Goura cristata NC 031865
Goura cristata LN589994
0.2
0.1



Supplementary Figure 1: Comparison of the phylogenetic trees built from complete mitogenome on the left and coding sequence without the 3rd codon position (A) or the 3rd codon position only (B). Branch lengths were obtained from maximum likelihood analyses. Numbers at the nodes represent Bayesian posterior probabilities (PP) on the left and maximum likelihood bootstrap values on the right for the complete mitogenome. For coding sequences, only maximum-likelihood are presented here. Stars correspond to 1.0 in PP, and 100% in bootstrap. In yellow are highlighted non-supported nodes.


Gallus gallus NC 001323
Pterocles gutturalis KX902237
Claravis pretiosa 92
Uropelia campestris
Metriopelia ceciliae 113836
Columbina passerina DOT7000
Claravis mondetoura 88
Claravis geoffroyi 466
Starnoenas cyanocephala 753696
Starnoenas cyanocephala Martin confirmed
Geotrygon montana 134464
Geotrygon violacea NC 015207
Leptotila verreauxi NC 015190
Leptotrygon veraguensis 296
Zentrygon frenata 144937
Zenaida auriculata NC 015203
Zenaida macroura NC 031863
Patagioenas fasciata KX902240
Patagioenas fasciata KX902539
Ectopistes migratorius KX902243
Ectopistes migratorius KC488473
Ectopistes migratorius KX902244
Reinwardtoena reinwardtii 552
Macropygia nigrirostris 516
Macropygia tenuirostris 55
Macropygia amboinensis 838
Turacoena modesta 705
Turacoena manadensis 23
Nesoenas picturatus 4015
Spilopelia chinensis NC 026459
Spilopelia chinensis KP636801
Aplopelia larvata RCKB1908
Streptopelia orientalis NC 031447
Streptopelia orientalis 141184
Streptopelia capicola JF4102
Streptopelia semitorquata MD40
Streptopelia decaocto KX377273
Streptopelia decaocto JF4227
Columba pallidiceps 139324
Columba janthina KM926619
Columba jouyi NC 031868
Columba rupestris NC 031867
Columba livia GQ240309
Columba livia KJ722068
Columba livia KF907308
Columba livia KP319029
Columba livia KF926316
Columba livia KP308517
Columba livia KP168712
Columba livia KP258178
Columba livia NC 013978
Treron vernans DOT576
Treron pompadora 784113
Treron cappellei MZB 27700
Chalcophaps indica HM746789
Chalcophaps stephani LGR 015
Oena capensis 143625
Turtur tympanistria HM746793
Turtur afer JF3646
Turtur abyssinicus 396407
Turtur chalcospilos JF3792
Gallicolumba tristigmata MZB33661
Gallicolumba luzonica HM746790
Gallicolumba rufigula 149291
Alopecoenas jobiensis 4925
Alopecoenas salamonis NC 031871
Alopecoenas johannae solomonensis 139335
Leucosarcia melanoleuca AMLAB1186
Henicophaps albifrons HA1
Henicophaps foersteri 333600
Geopelia striata
Phaps histrionica 616141
Ocyphaps lophotes DOT17740
Petrophassa rufipennis 615860
Geophaps plumifera 1948 531
Geophaps smithii
Phapitreron leucotis 114349
Phapitreron amethystinus 113846
Cryptophaps poecilorrhoa 258981
Ducula goliath 5365
Ducula badia MZB5930
Hemiphaga novaeseelandiae NC 013244
Gymnophaps stalkeri 611823
Lopholaimus antarcticus AMLAB1212
Megaloprepia magnifica DOT19826
Ramphiculus merrilli 767672
Chrysoena victor 249167
Alectroenas pulcherrimus 1933 2227
Drepanoptila holosericea NC126
Ptilinopus superbus 8602
Ptilinopus iozonus PI2
Ptilinopus pulchellus PP1
Ptilinopus regina MZB5773
Ptilinopus coronulatus PC2
Trugon terrestris 131708
Trugon terrestris MZB 6123
Microgoura meeki
Otidiphaps cervicalis KX902241
Otidiphaps nobilis DOT9510
Otidiphaps nobilis MZB 26379
Caloenas maculata NC 031870
Caloenas nicobarica CN01
Caloenas nicobarica NC 031869
Raphus cucullatus NC 031864
Pezophaps solitaria Copenhagen
Pezophaps solitaria KX902238
Didunculus strigirostris DS01
Didunculus strigirostris NC 031866
Goura scheepmakeri GS28
Goura victoria GV40
Goura sclaterii GS22
Goura cristata NC 031865
Goura cristata GC02
0.2
Melanocharis nigra B52454
Melanocharis versteri B52505
Starnoenas cyanocephala 753696
Starnoenas cyanocephala Martin confirmed
Claravis geoffroyi 466
Claravis pretiosa 92
Columbina passerina DOT7000
Metriopelia ceciliae 113836
Claravis mondetoura 88
Uropelia campestris
Geotrygon montana 134464
Leptotrygon veraguensis 296
Zentrygon frenata 144937
Reinwardtoena reinwardtii 552
Macropygia nigrirostris 516
Macropygia amboinensis 838
Turacoena manadensis 23
Turacoena modesta 705
Columba pallidiceps 139324
Nesoenas picturatus 2012
Aplopelia larvata RCKB1908
Streptopelia semitorquata MD40
Streptopelia decaocto JF4227
Streptopelia orientalis 141184
Streptopelia capicola JF4102
Treron vernans DOT576
Treron pompadora 784113
Treron cappellei MZB 27700
Chalcophaps stephani LGR 015
Oena capensis 143625
Turtur afer JF3646
Turtur abyssinicus 396407
Turtur chalcospilos JF3792
Phapitreron leucotis 114349
Phapitreron amethystinus 113846
Gallicolumba rufigula 149291
Gallicolumba tristigmata MZB33661
Alopecoenas johannae solomonensis 139335
Alopecoenas jobiensis 4925
Henicophaps albifrons HA1
Henicophaps albifrons LGR-112
Henicophaps foersteri 333600
Leucosarcia melanoleuca AMLAB1186
Phaps histrionica 616141
Geopelia striata
Ocyphaps lophotes DOT17740
Petrophassa rufipennis 615860
Geophaps plumifera 1948 531
Geophaps smithii
Ducula goliath 5365
Gymnophaps stalkeri 611823
Lopholaimus antarcticus AMLAB1212
Megaloprepia magnifica DOT19826
Ramphiculus merrilli 767672
Chrysoena victor 249167
Drepanoptila holosericea NC126
Ptilinopus superbus 8602
Ptilinopus iozonus PI2
Ptilinopus coronulatus PC2
Ptilinopus pulchellus PP1
Trugon terrestris 131708
Otidiphaps nobilis MZB 26379
Otidiphaps nobilis DOT9510
Didunculus strigirostris DS01
Caloenas nicobarica CN01
Goura cristata GC02
Goura sclaterii GS22
Goura scheepmakeri GS28
Goura victoria GV40
0.005

Supplementary Figure 2: Comparison of the non-collapsed phylogenetic trees built from complete mitogenome on the left and from nuclear genomic data on the right. Branch lengths were obtained from maximum likelihood analyses. Numbers at the nodes represent Bayesian posterior probabilities (PP) on the left and maximum likelihood bootstrap values on the right for the complete mitogenome. Stars correspond to 1.0 in PP, and 100% in bootstrap.



Supplementary Figure 3: Dated phylogenetic tree obtained from a BEAST analysis performed without fossils. Values at nodes are posterior probabilities, and node bars represent 95% High Posterior Density.

Supplementary Table 1: List and details of samples used in this study. Species names follow the Handbook of the Birds of the World (HBW; Baptista et al., 2017) and taxonomy of the International Ornithological Committee world bird list 7.3 (IOC, 2017) is provided when different. Institutional abbreviations are as follows: AMNH = American Museum of Natural History, New York; EDB = Laboratoire Evolution et Diversité Biologique, Toulouse; FMNH = Field Museum of Natural History, Chicago; MNHN = Museum National d'Histoire Naturelle, Paris; MZB = Museum Zoological of Bogor; NGBRC = New Guinea Binatang Research Center; 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	Project_ID	Current species name following HBW	Current species name following IOC if different	Locality	Date	Sex	Collector	Type	Sequenced for this study	Used for dating
MNHN	1933-2227	1933-2227	*Alectroenas pulcherrimus*		Seychelles				Toe-pad	Y	Y
NGBRC	3	4925	*Alopecoenas jobiensis*			2013			Blood	Y	Y
ZMUC	139335	139335	*Alopecoenas johannae solomonensis*	*Alopecoenas beccarii*	Nagashi ridge 550 m, Solomon Islands		NA	Makira	Tissue	Y	N
NHMUK	1956.60.566	1956.60.566	*Caloenas nicobarica*		Sapidan Island, W. Borneo	1956	NA	R.W. Sims & E. Banks	Toe-pad	N	Y
Lengguru	LGR-015	LGR-015	*Chalcophaps stephani*		Triton Bay	2014	M	B. Mila	Liver	Y	Y
AMNH	249167	249167	*Chrysoena victor*	*Ptilinopus victor*	Ngamia Island, Fiji	1924	M		Toe-pad	Y	Y
NHMUK	1895.4.1.466	466	*Claravis geoffroyi*						Toe-pad	Y	N
NHMUK	1914.11.26.88	88	*Claravis mondetoura*						Toe-pad	Y	Y
NHMUK	1953.68.92	92	*Claravis pretiosa*			1942			Toe-pad	Y	Y
MNHN	RCKB1908	RCKB1908	*Aplopelia larvata*	*Columba larvata*	South Africa				Tissue	Y	Y
ZMUC	139324	139324	*Columba pallidiceps*		Solomon Is.; Makira; Namaroi ridge	2006			Blood	Y	Y
AMNH	DOT 7000	DOT 7000	*Columbina passerina*		Dominican Republic; Prov. La Altagracia; Parque Nacional del Este, Guaraguao	1998	M	N.K. Klein	Tissue	Y	Y
AMNH	SKIN 258981	258981	*Cryptophaps poecilorrhoa*		Celebes: Minahassa	1884	F		Toe-pad	Y	Y
NHMUK	1939.12.9.2020	1939.12.9.2020	*Didunculus strigirostris*		Apia, Samoa	1896	NA		Toe-pad	N	Y
MNHN	NC126	NC126	*Drepanoptila holosericea*		New Caledonia				Tissue	Y	Y
MZB	MZB 5930	MZB 5930	*Ducula badia*		Borneo		F		Toe-pad	Y	N
MNHN	5365	5365	*Ducula goliath*		New Caledonia				Feather	Y	Y
ZMUC	149291	149291	*Gallicolumba rufigula*		Captivity (Erritzøe)		NA		Tissue	Y	Y
MZB	MZB 33661	MZB 33661	*Gallicolumba tristigmata*		Toli-toli, Sulawesi Tengah		F		Toe-pad	Y	N
EDB	286	286	*Geopelia striata*		St Leu, Reunion	2007	NA	B. Mila	Blood	N	Y
MNHN	1948-531	1948-531	*Geophaps plumifera*		Australia				Toe-pad	Y	Y
MNHN		Geo-smith	*Geophaps smithii*		Australia				Toe-pad	Y	N
ZMUC	134464	134464	*Geotrygon montana*		r'o Guaycuyacu Pichincha Ecuador	2004			Blood	Y	Y
NHMUK	1934.10.21.74	1934.10.21.74	*Goura cristata*		Salawati Island	1934	F	W.G.C. Frost	Toe-pad	N	N
NHMUK	1889.2.12.421	1889.2.12.421	*Goura scheepmakeri*		Port Moresby	[1879]	NA		Toe-pad	N	Y

Museum	Museum_ID	Project_ID	Current species name following HBW	Current species name following IOC if different	Locality	Date	Sex	Collector	Type	Sequenced for this study	Used for dating
NHMUK	1911.12.20.44	1911.12.20.44	*Goura sclaterii*	*Goura scheepmakeri sclaterii*	Mimika river	1910	F	BOU	Toe-pad	N	N
NHMUK	1921.12.30.40	1921.12.30.40	*Goura victoria*		Watam Sepik River	1920	M	W. Potter	Toe-pad	N	N
AMNH	SKIN 611823	611823	*Gymnophaps stalkeri*		Moluccas: G. Pinaia, Ceram	1911	M	E. Stresemann	Toe-pad	Y	Y
NGBRC		HA1	*Henicophaps albifrons*		Papua New Guinea: Madang Province; Wanang Conservation area	2013		K. Chmel	Blood	Y	N
Lengguru	LGR-112	LGR-112	*Henicophaps albifrons*		Arguni Bay	2014	F	B. Mila	Blood	Y	N
AMNH	SKIN 333600	333600	*Henicophaps foersteri*		Bangula Bay, New Britain	1952	F	W.F. Coultas	Toe-pad	Y	Y
NHMUK	1939.12.9.296	296	*Leptotrygon veraguensis*						Toe-pad	Y	Y
AMNH	DOT 2403	AM-LAB 1186	*Leucosarcia melanoleuca*		Australia: New South Wales; Sackville				Tissue	Y	Y
AMNH	DOT 2429	AM-LAB 1212	*Lopholaimus antarcticus*		Australia: New South Wales; 15 miles of Grafton				Tissue	Y	Y
NGBRC	838	838	*Macropygia amboinensis*		Wanang	2013		K. Chmel	Blood	Y	Y
NGBRC	516	516	*Macropygia nigrirostris*		Wanang	2013		K. Chmel	Blood	Y	Y
NHMUK	1937.1.17.55	55	*Macropygia unchall*						Toe-pad	Y	N
AMNH	DOT 19826	DOT 19826	*Megaloprepia magnifica*	*Ptilinopus magnificus*	Papua New Guinea: Madang Province; Karkar Island; Kevasob Bush Camp	2012		B.W. Benz	Tissue	Y	Y
ZMUC	113836	113836	*Metriopelia ceciliae*			1987			Blood	Y	Y
Beth Shapiro			*Microgoura meeki*							Y	Y
EDB	2012	2012	*Nesoenas picturatus*		Réunion Island			B. Mila	Blood	Y	Y
AMNH	DOT 17740	DOT 17740	*Ocyphaps lophotes*		Australia: Western Australia; Shire of Narembeen, ca 30 km N Hyden, near Anderson Rocks	2010	M	P. Sweet	Tissue	Y	Y
ZMUC	143625	143625	*Oena capensis*		Captivity (Erritzøe)		NA		Tissue	Y	Y
AMNH	DOT 9510	DOT 9510	*Otidiphaps nobilis*		Captivity: New Guinea		F?		Tissue	N	Y
MZB	MZB 26379	MZB 26379	*Otidiphaps nobilis*		Vogelkop, New Guinea		F		Toe-pad	Y	N
AMNH	SKIN 615860	615860	*Petrophassa rufipennis*		Australia: N.T.; S. Alligator R., Arnhem Land	1903	M	J.T. Tunney	Toe-pad	Y	Y
Julian Hume			*Pezophaps solitaria*						Toe-pad	Y	Y
ZMUC	113846	113846	*Phapitreron amethystinus*		Hamot, Philippines		NA	I. Luzon	Blood	Y	N
ZMUC	114349	114349	*Phapitreron leucotis*		Minuma Creek, Bintacan, Philippines		NA	I. Luzon	Tissue	Y	Y
AMNH	SKIN 616141	616141	*Phaps histrionica*		N.W. Australia: Derby	1886	M	Capt. Bowyer Bower	Toe-pad	Y	Y
NGBRC		PC2	*Ptilinopus coronulatus*		Papua New Guinea: Madang Province; Wanang Conservation area	2013		K. Chmel	Blood	Y	N
NGBRC		PI2	*Ptilinopus iozonus*		Papua New Guinea: Madang Province; Wanang Conservation area	2013		K. Chmel	Blood	Y	Y
NGBRC		PP1	*Ptilinopus pulchellus*		Papua New Guinea: Madang Province; Wanang Conservation area	2013		K. Chmel	Blood	Y	N

Museum	Museum_ID	Project_ID	Current species name following HBW	Current species name following IOC if different	Locality	Date	Sex	Collector	Type	Sequenced for this study	Used for dating
MZB	MZB 5773	MZB 5773	*Ptilinopus regina*		Banda Neira		M		Toe-pad	Y	N
NGBRC	7	8602	*Ptilinopus superbus*		Baitabag-Sec	2012			Blood	Y	N
AMNH	767672	767672	*Ramphiculus merrilli*	*Ptilinopus merrilli*	Luzon San Mariano, Sierra Mts, Philippines	1961	F		Toe-pad	Y	Y
NGBRC	552	552	*Reinwardtoena reinwardti*					K. Chmel	Blood	Y	Y
AMNH	SKIN 753696	753696	*Starnoenas cyanocephala*		Cuba: Algarrabo, Camaguey	1925	M		Toe-pad	Y	N
Martin Irestedt		Scya063	*Starnoenas cyanocephala*							Y	Y
MNHN	JF4102	JF4102	*Streptopelia capicola*		South Africa				Tissue	Y	Y
MNHN	JF4227	JF4227	*Streptopelia decaocto*		New Caledonia				Tissue	Y	N
ZMUC	141184	141184	*Streptopelia orientalis*		Kyrgyzstan; Kyrgyz ata; Kara Goi	2008			Blood	Y	N
MNHN	MD40	MD40	*Streptopelia semitorquata*		Guinea				Tissue	Y	N
MZB	MZB 27700	MZB 27700	*Treron capellei*		Kalimantan Tengah		F		Toe-pad	Y	N
AMNH	SKIN 784113	784113	*Treron pompadora*		Occ. Mindoro; Pamutusin, Paluan	1964	F		Toe-pad	Y	N
AMNH	DOT 576	DOT 576	*Treron vernans*		Malaysia: Sabah; Klias Forest Reserve, 8km W Beaufort	2004	F		Tissue	Y	Y
ZMUC	131708	131708	*Trugon terrestris*		Captivity (Erritzøe)		NA		Tissue	N	Y
MZB	MZB 6123	MZB 6123	*Trugon terrestris*		Mamberamo		M		Toe-pad	Y	N
NHMUK	1897.12.14.23	23	*Turacoena manadensis*						Toe-pad	Y	N
NHMUK	1881.5.1.3705	705	*Turacoena modesta*						Toe-pad	Y	Y
FMNH	396407	396407	*Turtur abyssinicus*		Ghana : Northern Region : Gonja Triange, 2 km W Buipe	2000	F		Tissue	Y	N
MNHN	JF3646	JF3646	*Turtur afer*		Guinea				Tissue	Y	Y
MNHN	JF3792	JF3792	*Turtur chalcospilos*		South Africa				Tissue	Y	N
MNHN		Uro-camp	*Uropelia campestris*		Brazil	1822			Toe-pad	Y	Y
ZMUC	144937	144937	*Zentrygon frenata*		Bolivia; Cochabamba; Tablas Montes	1991			Tissue	Y	Y

Supplementary Table 2: Published mitogenomes included in the phylogenetic and in the dating analysis. Species names follow GenBank taxonomy, with HBW and IOC correspondence if different (Baptista et al., 2017; IOC 7.3, 2017).

GenBank accession	GenBank species name	HBW / IOC species name if different	Used for dating
NC_031871	*Alopecoenas salamonis*		N
NC_031870	*Caloenas maculata*		N
NC_031869	*Caloenas nicobarica*		N
MG590264	*Caloenas nicobarica*		N
HM746789	*Chalcophaps indica*		N
KM926619	*Columba janthina*		N
NC_031868	*Columba jouyi*		N
GQ240309	*Columba livia*		N
KF926376	*Columba livia*		N
KJ722068	*Columba livia*		N
KP168712	*Columba livia*		N
KP258178	*Columba livia*		N
KP319029	*Columba livia*		N
NC_013978	*Columba livia*		N
KF907308	*Columba livia*		N
KP306517	*Columba livia*		N
NC_031867	*Columba rupestris*		N
NC_031866	*Didunculus strigirostris*		N
MG590266	*Didunculus strigirostris*		N
KC489474	*Ectopistes migratorius*		N
KX902243	*Ectopistes migratorius*		N
KX902244	*Ectopistes migratorius*		Y
HM746790	*Gallicolumba luzonica*		N
NC_001323	*Gallus gallus*		Y
MG590276	*Geopelia striata*		N
NC_015207	*Geotrygon violacea*		N
NC_031865	*Goura cristata*		N
LN589994	*Goura cristata*		N
NC_027947	*Goura scheepmakeri*		Y
MG590278	*Goura sclaterii*	*Goura scheepmakeri sclaterii* (IOC)	N
LN589993	*Goura victoria*		N
NC_013244	*Hemiphaga novaeseelandiae*		Y
NC_015190	*Leptotila verreauxi*		Y
KX902241	*Otidiphaps nobilis*	*Otidiphaps cervicalis* (HBW)	N
MG590265	*Otidiphaps nobilis nobilis*		N
KX902240	*Patagioenas fasciata*		N
KX902239	*Patagioenas fasciata*		N
KX902238	*Pezophaps solitaria*		N
KX902237	*Pterocles gutturalis*		N
NC_031864	*Raphus cucullatus*		N
NC_026459	*Streptopelia chinensis*	*Splilopelia chinensis* (IOC, HBW)	N
KP636801	*Streptopelia chinensis*	*Splilopelia chinensis* (IOC, HBW)	N
KX372273	*Streptopelia decaocto*		N
NC_031447	*Streptopelia orientalis*		N
MG590263	*Trugon terrestris*		N
HM746793	*Turtur tympanistria*		N
NC_015203	*Zenaida auriculata*		N
NC_031863	*Zenaida macroura*		N

Supplementary Table 3: Description of each partition obtained for the whole mitogenome. For coding sequences, first, second and third codon positions are treated separately. Each mitochondrial region is placed in a partition with a given evolutionary model in the Maximum-Likelihood (ML) and the Bayesian analyses. Note that a few genes were in different partitions in the ML and Bayesian analyses.

ML Analysis	Bayesian Analysis	List of mitochondrial genes
GTR+I+G	**GTR+I+G**	12S, 16S, ***ATP6*_1 (Bayesian)**, *ATP8*_1, *ATP8*_2, ***ATP8*_3 (ML)**, *ND2*_2, *ND2*_3, *ND4*_3, *ND5*_1, *ND5*_2, *ND5*_3, **tRNA-Lys (Bayesian)**, **tRNA-Val (Bayesian)**
GTR+I+G	**GTR+I+G**	***ATP6*_1 (ML)**, *ATP6*_2, *COX1*_1, *COX1*_2, *COX2*_1, *COX2*_2, *COX3*_1, *COX3*_2, *CYTB*_1, *CYTB*_2, *ND1*_1, *ND1*_2, *ND2*_1, *ND3*_2, *ND3*_3, *ND4*_1, *ND4L*_1, *ND4L*_2, *ND6*_1, *ND6*_2, *ND6*_3, tRNA-Ala, tRNA-Arg, tRNA-Asn, tRNA-Asp, tRNA-Cys, tRNA-Gln, tRNA-Glu, tRNA-Gly, tRNA-His, tRNA-Ile, tRNA-Leu_1, tRNA-Leu_2, **tRNA-Lys (ML),** tRNA-Met, tRNA-Phe, tRNA-Pro, tRNA-Ser_1, tRNA-Ser_2, tRNA-Thr, tRNA-Trp, tRNA-Tyr, **tRNA-Val (ML)**
GTR+I+G	**GTR+I+G**	*ATP6*_3, ***ATP8*_3 (Bayesian)**, *COX1*_3, *COX2*_3, *COX3*_3, *CYTB*_3, *ND1*_3, *ND3*_1, *ND4*_2, *ND4L*_3
GTR+I+G	**GTR+I+G**	Non_coding

Supplementary Table 4: Description of each partition obtained for the whole mitogenome dating analysis. For coding sequences, first, second and third codon positions are treated separately. Each mitochondrial region is placed in a partition with a given evolutionary model in the Bayesian analysis.

Evolutionary model	List of mitochondrial genes
GTR+I+G+X	12S, 16S, ATP8_1, ND2_2, ND2_3, ND5_1, ND5_2, ND5_3, tRNA-Lys, tRNA-Phe, tRNA-Thr
GTR+I+G+X	ATP6_1, ATP6_2, ATP8_2, COX1_1, COX1_2, COX2_1, COX2_2, COX3_1, COX3_2, CYTB_1, CYTB_2, ND1_1, ND1_2, ND2_1, ND3_2, ND3_3, ND4_1, ND4_3, ND4L_1, ND4L_2, tRNA-Ala, tRNA-Arg, tRNA-Asn, tRNA-Asp, tRNA-Cys, tRNA-Gln, tRNA-Glu, tRNA-Gly, tRNA-His, tRNA-Ile, tRNA-Leu_1, tRNA-Leu_2, tRNA-Met, tRNA-Pro, tRNA-Ser_1, tRNA-Ser_2, tRNA-Trp, tRNA-Tyr, tRNA-Val
GTR+I+G+X	ATP6_3, ATP8_3, COX1_3, COX2_3, COX3_3, CYTB_3, ND1_3, ND3_1, ND4_2, ND4L_3, Non_coding_1, Non_coding_4, Non_coding_5, Non_coding_8
GTR+I+G+X	ND6_1, ND6_2, ND6_3
GTR+I+G+X	Non_coding_2, Non_coding_3, Non_coding_6, Non_coding_7

Supplementary Table 5: Description of each partition obtained for the coding sequence mitogenome dating analysis from which 3rd codon positions were removed. First, second and third codon positions are treated separately. Each mitochondrial region is placed in a partition with a given evolutionary model in the Bayesian analysis.

Evolutionary model	List of mitochondrial genes
GTR+I+G+X	ATP6bis_1, ATP6bis_2, ATP8bis_2, COX1bis_1, COX1bis_2, COX2bis_1, COX2bis_2, COX3bis_1, COX3bis_2, CYTBbis_1, CYTBbis_2, ND1bis_1, ND1bis_2, ND2bis_1, ND3bis_2, ND4bis_1, ND4Lbis_1, ND4Lbis_2, ND6bis_1, ND6bis_2
GTR+I+G+X	ATP8bis_1, ND2bis_2, ND3bis_1, ND4bis_2, ND5bis_1, ND5bis_2

Chapter 2

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

Chapter introduction

Within this chapter, we focus on the richest biogeographic region for Columbidae: New Guinea. If exploring the extraordinary diversity found on this island will be tackled in discussion, we consider here one special genus: the crowned pigeons, *Goura* spp.

This genus includes four species, all inhabiting the lowland forest of New Guinea, with non-overlapping distribution. Little is known about their ecology (even their exact distributions are discussed), and they are all endangered with two species classified as Vulnerable and two as Near Threatened. Famous for their size and the crest they carry on their head, they are also frequently found in zoos or in personal aviaries.

In this chapter, we wanted to clarify the recently revised taxonomy, currently based on morphological characters. We also tackled the diversification process, in particular regarding the young age of the island and the possibility of diversification in Australia that was proposed for songbirds. To do so, we sequenced 39 individuals (from which 37 were museum samples) and five related species. Through the bioinformatics process, we encountered unexpected intra-individual variation when reconstructing the mitochondrial genome sequences. We were able to distinguish between heteroplasmy and cases of mitochondrial genomic regions inserted in the nuclear genome. These analyses were described in supplementary material. We also used nuclear data at very low coverage which allowed us to obtain phylogenetic trees fully concordant with the one reconstructed from mitochondrial genomes. Finally, we used these results to discuss their implications for future conservation assessment.

Molecular Phylogenetics and Evolution 120 (2018) 248–258

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

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

Jade Bruxaux[a,*], Maëva Gabrielli[a], Hidayat Ashari[b], Robert Prŷs-Jones[c], Leo Joseph[d], Borja Milá[e], Guillaume Besnard[a,1], Christophe Thébaud[a,1,*]

[a] Laboratoire Evolution & Diversité Biologique (EDB, UMR 5174), CNRS/ENSFEA/IRD/Université Toulouse III, 118 route de Narbonne, 31062 Toulouse, France
[b] Research Center for Biology, Museum Zoologicum Bogoriense, Cibinong 16911, Indonesia
[c] Bird Group, Department of Life Sciences, Natural History Museum, Tring, Herts HP23 6AP, England, United Kingdom
[d] CSIRO, National Research Collections Australia, Australian National Wildlife Collection, Canberra, ACT, Australia
[e] Museo Nacional de Ciencias Naturales, Consejo Superior de Investigaciones Científicas (CSIC), Madrid 28006, Spain

ARTICLE INFO

Keywords:
New Guinea
Goura
Crowned pigeons
Molecular phylogeny
Biogeography

ABSTRACT

Assessing the relative contributions of immigration and diversification into the buildup of species diversity is key to understanding the role of historical processes in driving biogeographical and diversification patterns in species-rich regions. Here, we investigated how colonization, *in situ* speciation, and extinction history may have generated the present-day distribution and diversity of *Goura* crowned pigeons (Columbidae), a group of large forest-dwelling pigeons comprising four recognized species that are all endemic to New Guinea. We used a comprehensive geographical and taxonomic sampling based mostly on historical museum samples, and shallow shotgun sequencing, to generate complete mitogenomes, nuclear ribosomal clusters and independent nuclear conserved DNA elements. We used these datasets independently to reconstruct molecular phylogenies. Divergence time estimates were obtained using mitochondrial data only. All analyses revealed similar genetic divisions within the genus *Goura* and recovered as monophyletic groups the four species currently recognized, providing support for recent taxonomic changes based on differences in plumage characters. These four species are grouped into two pairs of strongly supported sister species, which were previously not recognized as close relatives: *Goura sclaterii* with *Goura cristata*, and *Goura victoria* with *Goura scheepmakeri*. While the geographical origin of the *Goura* lineage remains elusive, the crown age of 5.73 Ma is consistent with present-day species diversity being the result of a recent diversification within New Guinea. Although the orogeny of New Guinea's central cordillera must have played a role in driving diversification in *Goura*, cross-barrier dispersal seems more likely than vicariance to explain the speciation events having led to the four current species. Our results also have important conservation implications. Future assessments of the conservation status of *Goura* species should consider threat levels following the taxonomic revision proposed by del Hoyo and Collar (HBW and BirdLife International illustrated checklist of the birds of the world 1: non-passerines, 2014), which we show to be fully supported by genomic data. In particular, distinguishing *G. sclaterii* from *G. scheepmakeri* seems to be particularly relevant.

1. Introduction

New Guinea supports the richest lowland rainforest avifauna in Australasia, and a high proportion of species occurs nowhere else in the world (Beehler and Pratt, 2016; Mack and Dumbacher, 2007; Mayr, 1941). However, we still know little about the origin of New Guinean bird diversity, in particular how it evolved through immigration and/or diversification within the archipelago (Deiner et al., 2011; Jønsson et al., 2011; Moyle et al., 2016). While a large proportion of New Guinean avifauna shows close relationships to Australian lineages [*e.g.* Ptilonorhynchidae (Irestedt et al., 2015), Meliphagidae (Marki et al., 2017)], suggesting that Australia was the source area of many lineages (Moyle et al., 2016; Schodde, 2006; Heinsohn and Hope, 2006), the origin of a significant fraction of the region's avifauna remains unclear.

* Corresponding authors.
E-mail addresses: bruxaux.jade@gmail.com (J. Bruxaux), christophe.thebaud@univ-tlse3.fr (C. Thébaud).
[1] Co-senior authors.

https://doi.org/10.1016/j.ympev.2017.11.022
Received 7 June 2017; Received in revised form 28 November 2017; Accepted 29 November 2017
Available online 01 December 2017

Particularly enigmatic is the biogeographical and temporal diversification of lineages containing multiple species endemic to New Guinea, with no close relatives elsewhere, such as crowned pigeons (*Goura*), berrypeckers (*Melanocharis*), and true pitohuis (*Pitohui*) (Jønsson et al., 2011, 2016; Moyle et al., 2016; Schodde and Christidis, 2014).

Four important factors underlie the buildup of present day species diversity in this area. First, New Guinea is a geologically recent island that became available for colonization by terrestrial species about 5–15 million years ago (Ma) (Hill and Hall, 2003; van Ufford and Cloos, 2005), implying recent colonization and diversification. Second, much of New Guinea and Australia are part of the same continental shelf, the former being the tectonically deformed northern margin of the latter. Thus, they are separated by very shallow seas and have been intermittently connected during sea-level lowstands (Voris, 2000), making biotic interchange especially likely for lowland organisms and of potential significance to understand the colonization patterns of New Guinea (Beehler et al., 1986; Deiner et al., 2011; Jønsson et al., 2011; Moyle et al., 2016; Schodde, 2006; Toussaint et al., 2015; Heinsohn and Hope, 2006). By contrast, New Guinea is separated from the continental shelf of Asia by deep-water channels (Hall, 2013) that have acted as barriers to dispersal for most but not all Asian bird lineages, as noted early on by Alfred Russel Wallace (Wallace, 1869, 1905). Third, since the Miocene, Australia has undergone an extensive aridification that led to a dramatic decline of the tropical rainforest biota, now restricted to northeastern Australia in close proximity to New Guinea (Byrne et al., 2011). Such large-scale aridification may have caused the extinction of many rainforest-adapted species, some of which are now geographically restricted to New Guinea. Finally, in light of new data on the major geological features of the region, a number of recent studies have emphasized that mountain and island uplift in a proto-Papuan archipelago may have played the role of a "species pump" (*e.g.* Aggerbeck et al., 2014; Jønsson et al., 2011, 2017; Toussaint et al., 2014). Recently, however, in a study on diversification patterns in songbirds, Moyle et al. (2016) suggested that New Guinea may have served mainly as an "evolutionary refuge" for Australian lineages, with diversification taking place prior to immigration out of their ancestral range (see also Schodde, 2006; Heinsohn and Hope, 2006). Thus, a major challenge for understanding the evolution of New Guinea's avian diversity is to decipher the relative contributions of immigration into the region, *in situ* speciation, and extinction history. This can be done by testing hypotheses regarding both the spatial and temporal scales of diversification events in lineages that host multiple New Guinean endemics.

Crowned pigeons (Columbidae: *Goura*) are large, lowland forest-dwelling species endemic to New Guinea, that differ markedly from any other species of pigeons by their size, a spectacularly large laterally compressed fan-like crest held erect over the head, and no oil gland (Darwin, 1868; Gibbs et al., 2001; Wallace, 1876). They occupy the extensive alluvial basins covered by vast areas of lowland rainforests that surround the main mountain massifs (Pratt and Beehler, 2015) as well as several neighboring islands (Aru Islands, Misool, Salawati, Batanta, Waigeo, and Yapen) lying in shallow water on the New Guinean continental shelf and formerly connected to the New Guinea mainland during sea-level lowstands in the Pleistocene (Beehler and Pratt, 2016; Voris, 2000). Populations found on the oceanic islands of Biak and Supiori in Cenderawasih (Geelvink) Bay and on the large Moluccan island of Seram were almost certainly introduced by humans (Beehler and Pratt, 2016; Gibbs et al., 2001). The genus *Goura* is distributed across lowland habitats forming a continuous ring around the island, and four allopatric or parapatric species have been recently recognized (del Hoyo and Collar, 2014) (Fig. 1). This raises the possibility that discontinuities among lowland basins due to presently eroded mountain ranges, inland bays, or more complex geological processes may have promoted speciation within New Guinean lowlands, as suggested by Mack and Dumbacher (2007). However, no comprehensive species-level phylogeny is currently available, and published evidence on genetic divergence (Besnard et al., 2016) is too incomplete to recognize taxa with a unique evolutionary history. In addition, two species recently recognized by del Hoyo and Collar (2014), i.e. Sclater's Crowned Pigeon *G. sclaterii* and Scheepmaker's Crowned Pigeon *G. scheepmakeri,* have long been considered as subspecies on the basis of plumage color similarities (Mayr, 1941; Gibbs et al., 2001; Rand and Gilliard, 1968), but this species split based on overall phenotypic divergence awaits confirmation. A phylogenomic and phylogeographic framework is needed to clarify the relationships among crowned pigeons and to document the biogeographical and temporal diversification of this lineage. In particular, such a phylogenomic and temporal framework should allow relevant comparisons with the major geological and climatic features that could have played a role in shaping the New Guinean biota in space and time.

A major obstacle to phylogenomic studies of taxa restricted to large and remote areas such as New Guinea is the difficulty of obtaining comprehensive geographic and taxonomic sampling of materials for DNA extraction and analysis. With all species of *Goura* becoming increasingly scarce in all but the most remote lowland rainforests (Gibbs et al., 2001), museum collections provide today the main source of DNA for addressing their evolutionary history while accounting for the range of variation found across New Guinea. However, most specimens are > 50 years old, have not been maintained so as to prevent DNA damage, and are therefore expected to contain degraded DNA that consists mostly of 100–200 base-pair fragments (Irestedt et al., 2006). Fortunately, many limitations due to low quality DNA can now be overcome using next generation sequencing (NGS) (*e.g.* Besnard et al., 2016; McCormack et al., 2016). In particular, genomic regions with high sequencing coverage (> 30×) such as complete mitogenomes can now be confidently assembled from old museum specimens (*e.g.* Besnard et al., 2016; Guschanski et al., 2013). However, phylogenetic reconstruction based on mitochondrial data alone may only partly reflect the evolutionary history of a group of species/populations, so that a multi-locus approach based on mitochondrial as well as nuclear genes is often necessary to increase the likelihood that the true species tree has been recovered (*e.g.* Ballard and Whitlock, 2004). Following recent advances in NGS analyses of museum samples (*e.g.* McCormack et al., 2016; Olofsson et al., 2016), it has become possible to recover nuclear markers that can be used in phylogenetic analyses even from relatively old specimens, thus providing excellent opportunities to assess and overcome the potential bias introduced by the use of mitochondrial markers alone.

In this study, we use a "genome skimming" approach (Straub et al., 2012) to retrieve complete mitogenomes and nuclear DNA sequences, including ribosomal regions, 391 independent nuclear loci, and 1336 ultraconserved elements (UCEs) previously used for reconstructing global avian phylogenies (McCormack et al., 2013; Prum et al., 2015). This yields one mitochondrial and two nuclear datasets that allow us to examine the relationships among recognized species of *Goura* and to reconstruct the diversification history of the group in space and time, in order to infer the processes that may have shaped its evolution. Finally, we discuss the implications of our results for the taxonomy and conservation of *Goura* species.

2. Material and methods

2.1. Sampling

We sampled a total of 39 *Goura* individuals, comprising 37 museum specimens (toe-pads) and two samples of fresh tissues obtained in the field, distributed over most of the known distribution range of each recognized species (Fig. 1; Supplementary Table 1). Samples include ten individuals of Western Crowned Pigeon *Goura cristata*, six of *G. scheepmakeri*, seven of *G. sclaterii* and 16 of Victoria Crowned Pigeon *G. victoria* [following the taxonomy in del Hoyo and Collar (2014)]. Thirteen specimens were more than 100-years old, including one *G. cristata* sample from Alfred Russel Wallace's collection obtained by

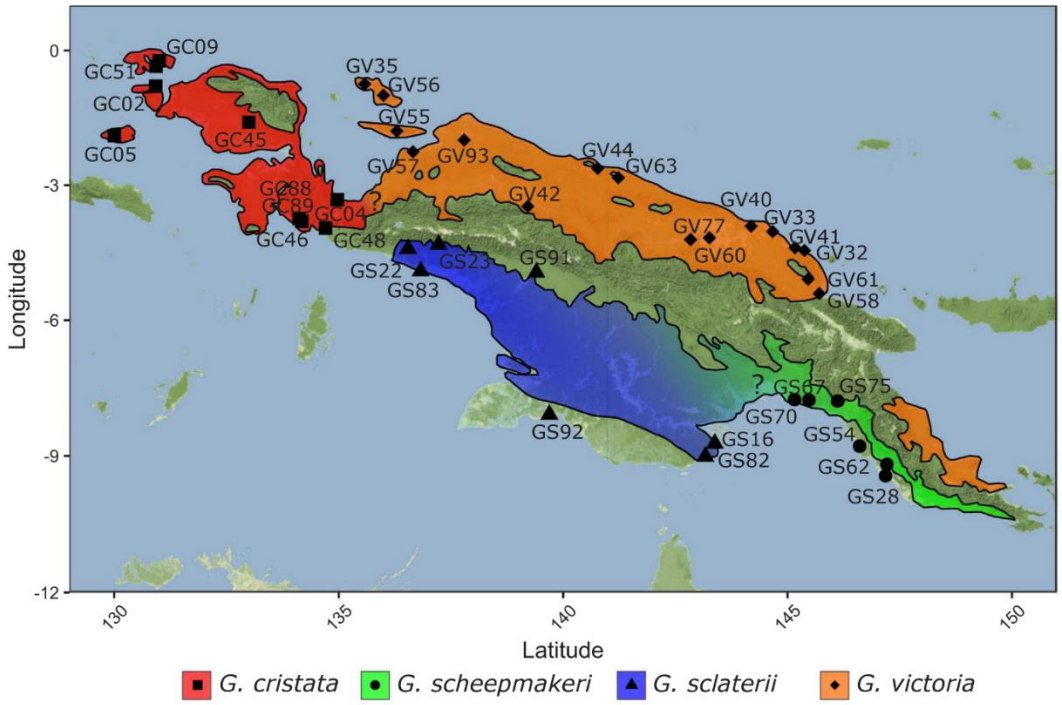


Fig. 1. Sampling localities of *Goura* crowned pigeons in New Guinea. Species distribution ranges are based on del Hoyo and Collar (2014) and Pratt and Beehler (2015) and are distinguished by specific colors and symbols. Question marks are added where distribution limits are unclear. Fading colors indicate uncertainty in distribution limits between species.

Charles Allen on the island of Misool, most probably in 1860 (van Wyhe and Rookmaaker, 2013). We also included four species belonging to the extant genera reported as the most closely related to crowned pigeons (Shapiro et al., 2002): *Didunculus strigirostris* (Tooth-billed Pigeon from Samoa; toe-pad museum sample), *Caloenas nicobarica* (Nicobar Pigeon from islands off south-east Asia to Solomon Islands; toe-pad museum sample), *Trugon terrestris* (Thick-billed Ground-Pigeon from New Guinea; tissue sample), and *Otidiphaps nobilis* (Pheasant Pigeon from New Guinea; tissue sample). In addition, *Geopelia striata* (Zebra Dove from Indonesia and the Malay peninsula; blood sample obtained on Reunion where the species has been introduced) was used as a more distant relative (Pereira et al., 2007). We finally added to this sampling the mitogenome data from two extinct relatives that were recently released in GenBank (Soares et al., 2016): *Raphus cucullatus* (the Dodo from the island of Mauritius; GenBank accession NC_031864.1) and *Pezophaps solitaria* (the Rodrigues Solitaire from the island of Rodrigues; GenBank accession KX902238.1).

Museum samples were kindly provided by the Natural History Museum, UK, (NHMUK), the American Museum of Natural History (AMNH), the Australian National Wildlife Collection (ANWC), the Museum Zoologicum Bogoriense (MZB), and the Zoological Museum University of Copenhagen (ZMUC). Details about specimens and collection localities are available in Supplementary Table 1.

2.2. DNA extraction and sequencing

Precautions were taken to avoid contamination of museum samples with modern DNA (*i.e.* use of a room where no fresh bird samples had ever been manipulated, laminar flow hood and sterilized materials; Besnard et al., 2016). In addition, DNA extractions and the first steps of library preparation were all performed in rooms where no amplified DNA was handled.

For toe-pad samples, we used around 1 mm^3 of tissue, as described in Besnard et al. (2016). When possible, we tried to use internal parts of the sample to avoid contamination with microbial or human DNA present on the surface of the sample. We used the Qiagen DNeasy Blood and Tissue kit (Qiagen Inc., Texas) following manufacturer recommendations, adding more proteinase K and DTT when tissue lysis was not easily obtained. The extracted DNA was eluted in a 120-µL volume of Buffer AE. Double stranded DNA concentration was measured using a Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Delaware). For fresh samples (blood, tissue or feather), extractions were performed in a separate room following manufacturer's recommendations.

Fifty-four µL of each sample (9.2–929 ng of double stranded DNA) were used to prepare the sequencing library with the Illumina TruSeq Nano DNA Sample Prep kit following the instructions of the supplier (Illumina Inc., San Diego). DNA extracted from museum specimens was not sonicated since mean DNA fragment size was expected to be below 200 bp (Besnard et al., 2016). Libraries were multiplexed (24 per flow cell lane), and inserts were then sequenced from both ends on a HiSeq 2000, 2500 or 3000 sequencer (Illumina Inc., San Diego). Library preparation and sequencing were performed at the GeT-PlaGe core facility, hosted by INRA (Toulouse, France).

2.3. Mitogenome assembly

For each sample we first tried to assemble the mitogenome *de novo* using the organelle assembler Org.Asm v0.2.03 (http://pythonhosted.org/ORG.asm/). We used all the mitochondrial protein sequences of a rock pigeon (*Columbia livia*) as seeds (GenBank accession NC_013978.1). Read trimming was done automatically at the onset of the assembly to remove adaptor ends when the insert size was shorter than the read sequence. When the sequence obtained was incomplete, complete mitogenomes were obtained by mapping the individual reads against a good quality reference successfully built by Org.Asm from

another individual of the same species (GC02, GS22, GS28 and GV32 were used as references for *G. cristata*, *G. sclaterii*, *G. scheepmakeri* and *G. victoria*, respectively). At this stage, we first used Trimmomatic v0.32 (Bolger et al., 2014) to remove the adaptors and all reads shorter than 36 bp. GMAP-GSNAP v2015-11-20 (Wu and Nacu, 2010) and Samtools v1.1 (Li et al., 2009) were then used to map the reads against the reference mitogenome, remove PCR duplicates, reconstruct the consensus sequence and obtain alignment statistics. We filtered the results in relation to the quality of the alignment, keeping only alternative alleles called with an error probability of less than 1% (Phred quality score > 20 in the samtools vcf output). We used mapDamage v2.0.2 (Jónsson et al., 2013) to evaluate the number of errors likely due to the age of samples (via post-mortem deamination). The quality score was thus modified accordingly before the construction of the consensus sequence with Samtools (Li et al., 2009). We also manually checked final assemblies and scrutinized all positions showing polymorphisms that were due to the presence of two alternative sequences [*e.g.* heteroplasmy or nuclear mitochondrial copies (numts); more details in Supplementary Methods 1].

2.4. Nuclear data retrieval

Two strategies were applied to reconstruct nuclear DNA datasets from a subsample of specimens representative of each *Goura* species, *Didunculus* and *Caloenas*. First, we assembled the nuclear ribosomal cluster, and we then retrieved low copy DNA regions using read mapping on reference sequences.

2.4.1. Nuclear ribosomal cluster assembly

Nuclear ribosomal DNA (nrDNA) regions have rarely been used to reconstruct phylogenies in animals and particularly in birds (*e.g.* Chikuni et al., 1996; Paško et al., 2011). This is mainly because the Internal Transcribed Spacer (ITS) is too long (*i.e.* 4–5 kb) to be amplified with PCR methods. However, the repeated nature of the nrDNA cluster is putatively an advantage to assemble this nuclear genomic region from NGS data, even with shallow sequencing of highly fragmented DNA from museum specimens. Here, we aimed to assemble nrDNA sequences (ca. 10 kb) that included the complete 18S gene, ITS1, 5.8S gene, ITS2, and 28S gene, plus 5′ and 3′ ends of the external transcribed spacer (ETS). To achieve this, we first used the 18S gene of *Columba livia* (1,737 bp; GenBank accession AF173630.1) as a seed to start the nrDNA cluster assembly of *G. cristata* sample GC02 [*Goura* sample used as reference in Besnard et al. (2016)], *C. nicobarica,* and *D. strigirostris*. The 18S sequence was extended using Geneious v. 9.0.5 (Biomatters Ltd., Auckland, New Zealand) by mapping reads with reiterations until sequence ends could not be properly elongated (due to numerous indels in the ETS). The consensus sequence of GC02 was then used to reconstruct the nrDNA sequence of other *Goura* individuals by mapping reads onto this reference (as described for the mitogenome reconstruction). A consensus sequence was reconstructed for each accession, and a site was considered to be polymorphic (using IUPAC codes) when two nucleotides or indels were supported by at least 25% of reads, such ambiguities likely resulting from the presence of different nrDNA paralogues in the genome. Sites covered with only 1× were considered as unknown bases (N). Since this marker did not provide true (unambiguous) intra-specific polymorphisms (see below), we chose to analyze 18 *Goura* samples only (four or five accessions per species, among those that rendered the best nuclear DNA/mitochondrial DNA ratio; see Supplementary Methods).

2.4.2. Construction of the reference sequence for low copy nuclear DNA regions and mapping of reads

A common strategy to recover nuclear markers is to target particular regions of the genome and to enrich and sequence them using specific probes (*e.g.* McCormack et al., 2013; Prum et al., 2015). As no nuclear genome is currently available to design probes for any of the *Goura* species or close relatives, we used an alternative strategy consisting in mapping the Illumina reads to a reference sequence. While we could have used the Rock Pigeon *Columba livia* reference genome (Shapiro et al., 2013), this species is not a close relative to *Goura* and we preferred an alternative strategy based on using two large sets of markers that were assembled recently for producing large-scale phylogenetic hypotheses in modern birds (McCormack et al., 2013; Prum et al., 2015). These two sets comprise markers which correspond to highly conserved regions in birds, increasing the probability of finding reads corresponding to homologous sequences.

The first set of genes is composed of 1,336 ultra-conserved elements (UCE) with an average length of 410 bp that were characterized in the Pink-necked Green Pigeon *Treron vernans* (McCormack et al., 2013), a pigeon presumed to be more closely related to *Goura* than *Columba livia* (Fulton et al., 2012; Johnson and Clayton, 2000; Pereira et al., 2007). The second set of genes is composed of 391 nuclear loci with an average length of 1,263 bp, which were defined from the Red Junglefowl *Gallus gallus* genome and used to draw genetic data from 198 species of birds, including five species of Columbidae (Prum et al., 2015). Both sets of genes were aligned in Geneious to check for shared genes between the two datasets with the "Map to reference" option and default parameters. One gene shared by both datasets was deleted, leading to a total of 1726 independent loci, which were concatenated to create a unique reference sequence of ca. 1 Mbp. The Illumina reads were first trimmed with Trimmomatic (Bolger et al., 2014), allowing removal of the adapters, removal of the bases at the beginning or the end of reads when their quality was below three, and removal of the reads that after those manipulations, were shorter than 36 bp (default parameters). Only the reads where both forward and reverse reads had survived the trimming were kept for the alignment. The remaining Illumina reads from the 39 *Goura* individuals and two outgroups (*D. strigirostris* and *C. nicobarica*) were aligned to the reference sequence using GMAP-GSNAP v2015-11-20 (Wu and Nacu, 2010). Finally, Samtools v1.1 (Li et al., 2009) was used to build the bam files from the read alignments after removing PCR duplicates and to obtain statistics on this process (*i.e.* number of mapped reads, sequencing depth). This alignment was analyzed with mapDamage v2.0.2. (Jónsson et al., 2013), so that we could lower the Phread base quality score of the sites were deamination was highly likely. Following the read mapping, fasta files were generated for each of the 39 *Goura* individuals and the two outgroups using the doFasta command of ANGSD (Korneliussen et al., 2014) and keeping the most frequent base at each site.

2.5. Phylogenomic analyses

2.5.1. Mitogenome data

All mitogenomes were annotated as reported in Besnard et al. (2016) using Geneious v9.0.5 (Biomatters Ltd., Auckland, New Zealand). We removed the repeated sequences in the control region and aligned each protein-coding gene, each transfer RNA gene (tRNA), each ribosomal RNA gene (rRNA) and each non-coding region separately using the Geneious aligner with default parameters. We concatenated these alignments with SeqCat.pl v1.0 (Bininda-Emonds, 2005) and obtained a final dataset of 15,694 bp.

We used PartitionFinder v2.1.1 (Lanfear et al., 2017) to find the best partition schemes and nucleotide substitution models for both RAxML (Stamatakis, 2014) and MrBayes (Ronquist et al., 2012) analyses, with the greedy algorithm and considering 64 potential partitions: the 12S rRNA, the 16S rRNA, each tRNA, all non-coding sequences, and the three codon positions in each of the 13 protein-coding genes. We unlinked branch lengths and used the Bayesian information criterion (BIC) to select the best-fitting model and partition scheme (Malé et al., 2014), which was then used for tree reconstructions.

The maximum-likelihood (ML) analyses were performed using RAxML v8.1.5 (Stamatakis, 2014). We used 20 alternative runs and 500 replicates of non-parametric bootstrapping to evaluate the reliability of

the resulting tree. The Bayesian tree reconstruction was performed with MrBayes v3.2.2 (Ronquist et al., 2012). Two independent runs of four Metropolis-coupled Markov chains (MCMC) for Monte Carlo simulations were run for 20,00,000 generations, with parameters and trees sampled every 100 generations, and a burn-in of 50,000 generations. Convergence of the chains was manually checked using Tracer v1.6 (Rambaut et al., 2014).

2.5.2. Nuclear data

2.5.2.1. Ribosomal cluster. We annotated our nrDNA clusters in Geneious using a complete nuclear ribosomal DNA sequence of *G. gallus* (GenBank accession KT445934.2) as a reference. The complete alignment was 11,347-bp long. To determine the best partition scheme and nucleotide substitution model, we used PartitionFinder v2.1.1 (Lanfear et al., 2017) after separating our data in seven potential partitions [*i.e.* three ribosomal genes (18S, 5.8S, and 28S) and four intergenic spacers (5′ETS, ITS1, ITS2, and 3′ETS)] and used the same parameters as described for mitochondrial analyses. Similarly, we used RAxML v8.1.5 (Stamatakis, 2014) with 20 alternative runs and 1000 replicates of non-parametric bootstrapping, and MrBayes v3.2.2 (Ronquist et al., 2012) with two independent runs of four MCMC for 2,00,000 generations sampled every 100 generations and a burn-in of 50,000 generations, and manually checked the chain convergence using Tracer v1.6 (Rambaut et al., 2014).

2.5.2.2. Low copy conserved regions. Since we obtained low quality nuclear DNA sequences for some individuals, reflected by a low number of mapped reads, we decided to keep data only from individuals which had at least 5,000 mapped reads to minimize any potential bias related to sequencing errors or DNA degradation. Thus, our dataset contained 19 *Goura* individuals, including seven *G. cristata*, four *G. sclaterii*, two *G. scheepmakeri* and six *G. victoria.*

We then performed a ML reconstruction using RAxML v8.1.5 (Stamatakis, 2014) with 20 alternative runs and 100 replicates of non-parametric bootstrapping, using a General Time Reversible model (Tavaré, 1986) with invariable sites and a gamma distributed rate variation among sites (GTR + I + G model).

2.6. Dating divergence events

Dating species divergences on an absolute time scale is challenging in groups such as *Goura* for which fossil data is lacking. One solution is to date mitochondrial phylogenies by using a mean divergence rate of 2.1% per million years (Weir and Schluter, 2008). While this approach is widely used (*e.g.* Carmi et al., 2016; Shipham et al., 2015), it is often criticized (*e.g.* Lovette, 2004) because gene divergence often precedes population divergence (Edwards and Beerli, 2000) and variation in divergence rate has been observed among lineages in relation to variation in life-history traits, such as body mass (Nabholz et al., 2016). For birds, coalescent dates for mitochondrial sequences have been found to generally precede population splitting by 0.2–0.3 million years (Moore, 1995; Weir, 2006); hence, it seems reasonable to approximate the date of species divergence with the coalescent date. To date the splits within the genus *Goura*, we used the standard mitochondrial substitution rate corrected for body mass, following recommendations in Nabholz et al. (2016) and using a value of 2200 g for all *Goura* species since they all have a similar body mass (Gibbs et al., 2001). Since the corrected substitution rate is derived from coding-gene positions, we only kept coding-genes for dating analyses and removed *ATP8* and *ND4L* due to their short length, and *ND6* due to its reverse sense, as done by Nabholz et al. (2016). We partitioned the data according to the results of a new PartitionFinder analysis run on this new dataset (including *Goura* sequences only), and linked partitions for the tree reconstruction. We used a strict clock for each partition, fixing the third codon position partition to a substitution rate of 0.00942646 substitutions per site per million years, following Nabholz et al. (2016), and used a uniform prior for all the substitution rates. The substitution rate of first and second codon positions was given a uniform prior of 0.01 [0.005–0.015], while others priors were kept at their default values. We ran the analysis with a birth-death model (Stadler, 2009) for 10 million generations, logging every 1000 generations. We discarded the first 10% of trees as burn-in and manually checked for convergence using Tracer v1.6 (Rambaut et al., 2014).

To date the stem age of the *Goura* lineage, we used two individuals per species: GC05 and GC45, GS16 and GS82, GS67 and GS70, GV42 and GV56 for *G. cristata*, *G. sclaterii*, *G. scheepmakeri* and *G. victoria*, respectively. Here we included all the outgroup sequences (*D. strigirostris*, *C. nicobarica*, *T. terrestris*, *O. nobilis*, *R. cucullatus*, *P. solitaria*, *G. striata*) in the analysis. Only coding genes were used but *ATP8*, *ND4L* and *ND6* were excluded as described above. We used the age of the oldest divergence event within the *Goura* lineage found in the previous dating analysis as a prior to calibrate the phylogenetic tree of the entire group. We added a normal calibration point at 24.7 ± 3.5 Ma for the diversification of Columbidae following Soares et al. (2016), whose dating result is in agreement with the one obtained by Prum et al. (2015) and with the calibration point used in Claramunt and Cracraft (2015). We used nucleotide substitution models selected by PartitionFinder, and linked the partitions for the tree reconstruction. We applied a log normal relaxed molecular clock for each partition (Drummond et al., 2006; Lepage et al., 2007) and uniform priors for each substitution rate. The three uncorrelated log-normal relaxed clock means were given an exponential prior, and invariant proportion, when applicable, was given a uniform prior at 0.5 [0–1]. All the unstated priors were kept with default values. We ran the analysis with a birth-death Model (Stadler, 2009) for 100 million generations, logging every 1000 generations. We discarded the first 20% of trees as burn-in and manually checked for convergence using Tracer v1.6 (Rambaut et al., 2014). All analyses were performed using BEAST2 v2.4.3 (Bouckaert et al., 2014).

3. Results

3.1. Sequencing data

We obtained on average 11,317,165 paired-end reads per individual (range: 5,112,052–20,116,808; Supplementary Table 2) that were used to assemble the whole mitogenome and nuclear ribosomal DNA cluster, and to retrieve highly conserved nuclear regions spread across the genome.

3.2. Mitogenome assembly, and phylogenetic analyses

A complete mitogenome was reconstructed for all individuals. Despite evidence for some deamination, the consensus sequences were not modified by weighting the quality scores. This reflects the high mean sequencing depth of all mitogenomes, ranging from 16.74 to 7759.56×, with only one individual (GS91) showing a mean sequencing depth < 30× (Supplementary Table 2). Seventeen mitogenomes presented at least one site with two alternative nucleotides. In individuals with the lowest mean sequencing depth, we found some minor variants represented by very few reads (less than 10% of the mapped reads) and likely due to sequencing errors. Other minor variants were found (supported by 14–33% of the mapped reads) and were carefully examined in order to understand their origin, after the possibility of cross contamination had been ruled out (see Supplementary Methods 1). Two categories of individuals could be clearly differentiated, according to both the number of variable sites and the ratio of nuclear/mitochondrial DNA sequencing depth. This ratio, with an average of 0.06 in individuals without minor variant, ranged from 0.20 to 0.78 for all individuals with at least two variable sites except GC09 (see Supplementary Methods 1). A high ratio, which implies a higher nuclear coverage, allows the detection of mitochondrial genome copies

included in the nuclear genome [numts (Lopez et al., 1994)]. Therefore, the minor variants in this first case were likely numts. In contrast, a few individuals with one variable site and also GC09 did not show a particularly high ratio of nuclear/mitochondrial DNA sequencing depths. In such cases, we considered that apparent polymorphism was likely due to heteroplasmy. Phylogenetic relationships of major and minor variants also confirmed this difference: while minor variants considered as numts were sister to the whole species and showed short branches (pattern of "fossil" genome), supposed minor heteroplasmic sequences clustered with other mitochondrial sequences (see Supplementary Methods 1). For the rest of the analysis, we thus used majority-rule consensus sequences. All mitogenomes are deposited in GenBank (see Supplementary Table 2 for more details).

Following PartitionFinder results, we used a GTR + I + G model for the ML analysis, with the four partitions described in Supplementary Table 3. For the Bayesian analysis, PartitionFinder proposed the same partitions apart from the second codon position of *ATP8*. A GTR + I + G model was used for the first and third partitions, a Hasegawa-Kishino-Yano model (Hasegawa et al., 1985) with invariant sites and a gamma distributed rate variation (HKY + I + G model) was used for the second partition, and a GTR + G model (without invariant sites) was used for the fourth partition (Supplementary Table 3). All the Effective Sample Size (ESS) values were higher than 200, suggesting a good convergence of the Bayesian analysis.

We obtained very similar results with both the ML and Bayesian methods (Fig. 2A). These analyses highlighted four well supported monophyletic groups corresponding to the four recognized species, with reciprocal monophyly between *G. cristata/G. sclaterii* and *G. victoria/G. scheepmakeri* that were both recovered as well-supported monophyletic groups.

3.3. Nuclear ribosomal cluster assembly, and phylogenetic analyses

A partial nrDNA cluster of 10,397–10,738 bp was reconstructed for 18 *Goura* individuals, and two outgroups, *D. strigirostris* and *C. nicobarica*. The mean sequencing depth ranged from 13.39 to 356.62× (Supplementary Table 2). Four GC-rich regions in the ITS1 and 28S gene showed very low sequencing depth, and one to four of these parts were not covered in eight *Goura* individuals (for a total missing part ranging from 19 to 406 bp). In addition, a few sites (from 3 to 36 per individual) were considered to be polymorphic due to the presence of different nrDNA copies in the genome (see Material and Methods). Therefore, we checked visually all poorly covered regions and polymorphic sites. Manual adjustments (*e.g.* alignments in regions with insertions or deletions) were made before deciding which sites were polymorphic in the assembly. In the end, very low levels of polymorphism were observed within each recognized species, and no polymorphism was recovered from poorly covered regions, making us confident that the presence of different nrDNA copies in the genome did not lead to systematic errors in our analyses.

Following PartitionFinder results, ribosomal genes (*i.e.* 18S, 5.8S, and 28S) and intergenic spacers (*i.e.* ETS and ITS) were separately treated in two partitions. The selected nucleotide substitution models differ slightly between the ML and Bayesian analyses, probably due to the low number of models available in RAxML (Supplementary Table 4). In the ML analysis, we implemented a GTR + I + G model for the ribosomal gene partition, and a GTR + G model for the intergenic spacers. The Bayesian analysis was performed using a HKY + I model for ribosomal genes, and a HKY + G model for the intergenic spacers. All the ESS values were higher than 200, suggesting a good convergence of the Bayesian analysis.

The phylogenetic trees obtained with the ML and Bayesian methods are fully congruent with four different monophyletic groups corresponding to the four currently recognized species, with the two pairs of species, *G. cristata/G. sclaterii* and *G. victoria/G. scheepmakeri*, being recovered as well-supported monophyletic groups (Fig. 2B).

3.4. Low copy conserved regions assembly, and phylogenetic analyses

Among the 19 *Goura* individuals that were used in this analysis, Illumina reads effectively mapped against all 1,726 loci previously characterized by Prum et al. (2015) and McCormack et al. (2013). Therefore, our approach was successful in retrieving nuclear sequences from a large number of markers spread across the genome.

The number of mapped reads per individual ranged from 5,608 to 24,712, with an average number of 13,295 reads per individual. This corresponds to sequence lengths of 466,245 bp per individual on average (range: 240,318–804,043 bp). Given the size of the reference sequence (1,043,293 bp), the sequences generated showed a high level of missing data, as expected from shallow sequencing data (55.31% on average; range: 22.93–76.97%). The depth across the 1,726 genetic markers averaged over individuals was very low (3.11×) and ranged from 2.51 to 3.92× per individual. Despite this obvious limitation, our approach demonstrates that nuclear data can be relatively easily recovered from museum samples, as already suggested in a previous study (Besnard et al., 2016).

The nuclear phylogeny (Fig. 2C) supports the existence of four monophyletic groups with a topology which is fully congruent with those based on both mitochondrial and nuclear ribosomal DNA, with the two pairs of species, *G. cristata/G. sclaterii* and *G. victoria/G. scheepmakeri*, being again recovered as well-supported monophyletic groups.

3.5. Estimation of divergence times

For the mitochondrial dating analysis, we defined two partitions with PartitionFinder, the first one corresponding to the first two codon positions, and the second corresponding to the third position (Supplementary Table 5). A Tamura and Nei (1993) model with a gamma distributed rate variation among sites and estimated frequencies (TRN + G + X) was selected for the first partition, and a GTR model with estimated frequencies was selected for the second one (GTR + X). All the ESS values were higher than 200, as expected from a Bayesian analysis which had converged.

Taking into account body mass, we dated the crown age of the *Goura* genus at 5.90 Ma [median value with a 95% Highest Probability Density (HPD) of 5.32–6.50 Ma]. This result was used as a prior for the second calibration point of the larger phylogeny where we then found a slightly younger crown age at 5.73 Ma [HPD: 5.15–6.29] (Fig. 3). The split between *G. cristata* and *G. sclaterii* was dated at 1.94 Ma [HPD: 1.42–2.49], while the split between *G. scheepmakeri* and *G. victoria* was dated at 0.69 Ma [HPD: 0.46–0.95]. The stem age of the *Goura* genus was dated at 21.39 Ma [HPD: 17.32–25.44].

4. Discussion

4.1. On the use of museum specimens to obtain mitochondrial and nuclear genomic datasets

In this study, we used successfully museum study skins to acquire genomic data from both the nuclear and the mitochondrial genomes of several species of Columbidae. Our approach confirms once again that museum specimens are a valuable source of materials to obtain genomic information that can be used for phylogenetic analyses (Guschanski et al., 2013). In the course of this study, we developed new strategies to recover nuclear genetic data from a large number of markers spread across the genome. This should be particularly useful for forthcoming studies on the diversification of lineages for which recent collections are few or difficult to achieve, and existing museum collections are the only available source of DNA.

We found that the nuclear ribosomal DNA cluster can be assembled from the sequences obtained from poorly preserved DNAs. The long length of the internal transcribed spacer that appears to have limited

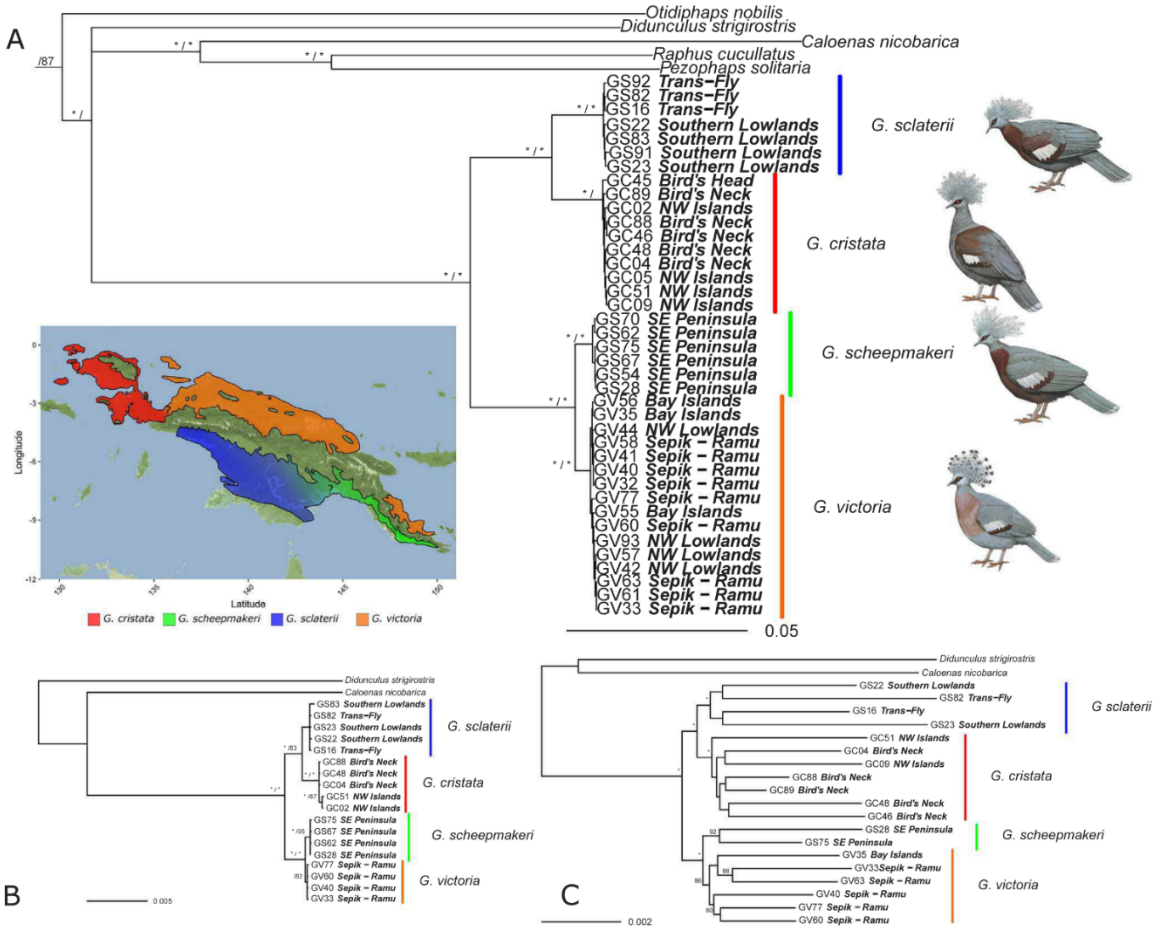


Fig. 2. Phylogenetic trees obtained from the entire mitogenome (A), nuclear ribosomal cluster (B), and conserved nuclear regions (C). Branch lengths were obtained from maximum likelihood analyses. Numbers at the nodes represent Bayesian posterior probabilities (PP) on the left, and maximum likelihood bootstrap values on the right. Stars correspond to 1.0 in PP, and 100% in bootstrap. Support values lower than 0.95 in PP and 70% in bootstrap are not included. Intra-specific support values are removed for clarity in (A). All images are reproduced from del Hoyo and Collar (2014) with permission of Lynx Edicions.

the use of this marker in avian phylogenetics (Chikuni et al., 1996; Paśko et al., 2011) can be overcome with a shotgun sequencing approach. However, variable intra-genomic sites require caution during the assembly process. We also showed that it is possible to recover low copy nuclear regions by mapping reads onto reference sequences, as expected from our previous work (Besnard et al., 2016) and already demonstrated in other taxa like plants (*e.g.* Olofsson et al., 2016). However, we found that not all individuals were amenable to nuclear genome analyses. For some specimens for which a very high quality mitogenome was recovered, we obtained a very low sequencing depth for the nuclear genome (*e.g.* GS70, GV41, GV42, GV56, more details in Supplementary methods 1) that prevented their inclusion in our datasets. This highly variable ratio could be linked to sample storage conditions before and after being accessioned at a museum (Binladen et al., 2006).

Phylogenetic analyses revealed entirely congruent topologies among the three genomic datasets, but terminal branches are longer and nodal support is lower in the analysis based on low copy conserved nuclear regions, especially at the intra-species level (see Fig. 2). This pattern on nuclear single-copy regions is likely due to post-mortem damages like cytosine deamination, even if the levels of deamination are much lower in museum samples than those typically found in ancient DNA samples (Sawyer et al., 2012). To minimize the impact of sequencing errors that result from this process, all sites that showed signs of degradation received a lower quality value [through mapDamage (Jónsson et al., 2013)] prior to analysis. To further minimize the impact of sequencing errors, we could also have removed all low coverage sites (*e.g.* less than 2–3×) from the dataset prior to analysis. We chose not to do so since this would have drastically reduced the dataset (28% for a threshold at 2×, 72% at 3×; see Supplementary Methods 2 for filtered data phylogenies). Higher nuclear genome coverage may alleviate some of the limitations we encountered but our analyses suggest that even low coverage data from conserved regions can prove very useful to infer phylogenetic relationships.

4.2. Unexpected phylogenetic relationships

All three phylogenetic analyses (*i.e.* based on mitogenome, nrDNA, and low-copy nuclear markers) resolved *Goura* into four main monophyletic groups, corresponding to the four species recently recognized by del Hoyo and Collar (2014). They are further grouped into two pairs of strongly supported sister species: the first pair comprises *Goura cristata* (western New Guinea) and *G. sclaterii* (southern New Guinea), while the second pair includes *G. victoria* (northern New Guinea) and *G.*

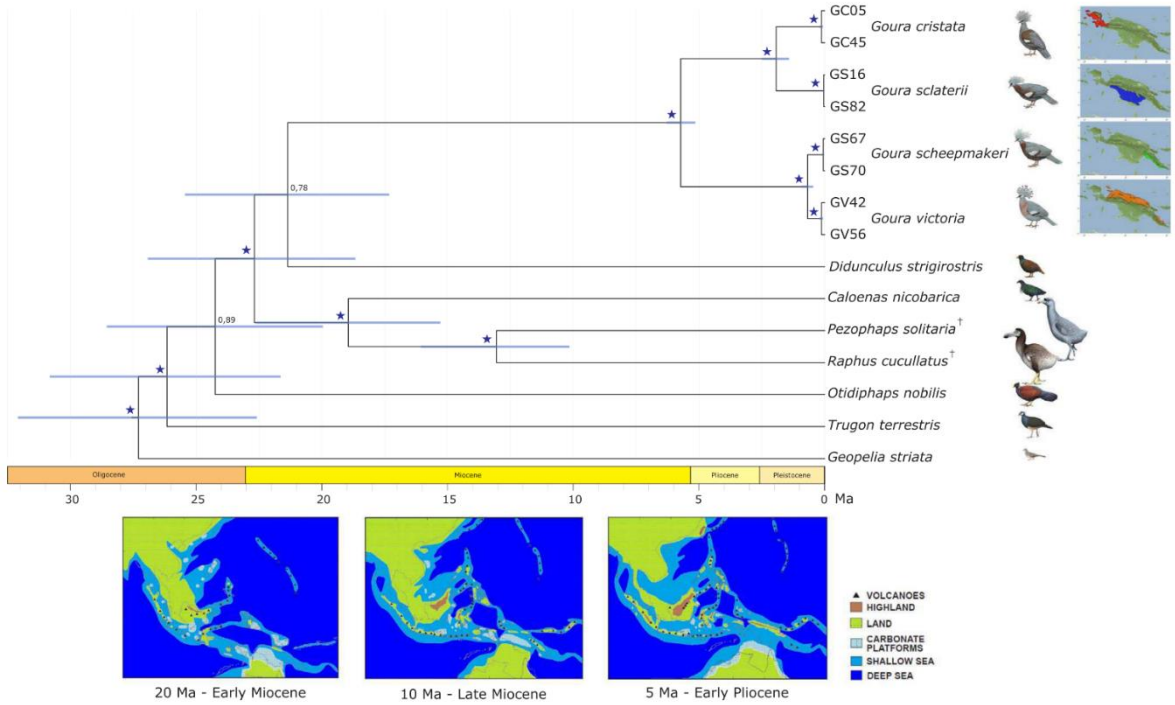


Fig. 3. Dated tree of *Goura* and its closest known relatives. Nodes are placed at median ages with 95% Highest Probability Density in blue. Values at nodes show posterior probabilities, and stars represent a probability of 1. Geological maps of the region [based on Hall (1998)] are shown at different ages (20, 10 and 5 million years ago). All Columbidae images are reproduced from del Hoyo and Collar (2014) with permission of Lynx Edicions.

scheepmakeri (southeastern New Guinea).

Our research confirms the distinction between *G. scheepmakeri* and *G. sclaterii* that was first proposed by del Hoyo and Collar (2014) mainly on the basis of morphology and plumage coloration, rather than previous taxonomic treatments that considered both taxa as subspecies of *G. scheepmakeri* [see Beehler and Pratt (2016) for discussion]. Moreover, *G. cristata* and *G. victoria*, whose distributions seem to abut along the Siriwo River on the eastern side of Cenderawasih (Geelvink) Bay, where they have been described as possibly hybridizing (Mayr, 1941), are also not each other's closest relatives.

Species with known or apparent contact zones are therefore not each other's closest relatives, which suggests a non-trivial scenario of speciation. Distributional contact zones among non-sister taxa such as those observed in *Goura* have been characterized in other New Guinea bird species complex (*e.g.* Deiner et al., 2011; Heads, 2001; Irestedt et al., 2015) and also in other taxa such as turtles (Georges et al., 2014), plants, mammals or snakes (reviewed in Heads, 2001), possibly implying a series of shared historical events responsible for the diversification of New Guinean fauna and flora. Investigating the speciation process in relation to past or recent gene flow between the diverging species would thus be of great interest. Unfortunately, in this study the low sequencing depth of nuclear data (leading to a high amount of missing data and a heterogeneous sequencing error rate among samples) prevented us from using population genomic analyses to evaluate alternative demographic models. Intraspecific analyses would also be interesting regarding possible geographic patterns in *G. sclaterii* for mitogenomes (Fig. 2A), and in *G. cristata* for nrDNA data (Fig. 2B). This second clustering, separating island and mainland individuals, is consistent with the proposal that *G. cristata* may comprise two subspecies (del Hoyo and Collar, 2014; Mayr, 1941; Rand and Gilliard, 1968). It is, however, weakly supported and since these subspecies do not clearly differ phenotypically (Beehler and Pratt, 2016), this warrants further exploration. Interestingly, the species with the largest distribution (*G. victoria*) does not seem to present any geographic clustering in spite of also containing two described subspecies, one smaller restricted to the islands of Cenderawasih (Geelvink) Bay, and the other larger and widely distributed across northern New Guinea from the Siriwo River to Astrolabe Bay (Beehler and Pratt, 2016).

4.3. Biogeographical and temporal diversification of Goura in lowland rainforests of New Guinea

New Guinea has long been seen as a diversification center for some lineages, such as corvoid passerine birds (Jønsson et al., 2011, 2017). However, the fact that New Guinea has been available for dispersal and colonization by terrestrial organisms for less than 15 Ma, and perhaps as little as 5 Ma, makes this view difficult to generalize. Schodde (2006), Schodde and Christidis (2014), Moyle et al. (2016) and Heinsohn and Hope (2006) proposed that the island could have played a refugial role for tropical rainforest birds that 'escaped' from Australia when elements of their biota progressively disappeared. Early diversification could have thus occurred in Australia, possibly followed by more recent events in New Guinea, after it became available for colonization.

Our results, which highlight a first diversification event in the *Goura* lineage around 5.73 Ma [HPD: 5.32–6.50], are consistent with the view that current species diversity is best explained by speciation occurring within New Guinea, similarly to what has been described in other bird groups, such as lories and lorikeets (Schweizer et al., 2015), and woodland kingfishers and kookaburras (Andersen et al., 2017).

The continuity of lowland rainforests around the island of New Guinea implies a possibility of contact and gene flow between diverging species comprising this biota. However, distributional limits between congeneric species or subspecies, such as those observed in *Goura*,

located either in the Bird's Neck [the so called Zoogeographers' Gap (Hartert et al., 1936)] or in the south-eastern part of the central cordillera, have also been found in other species complexes (Mack and Dumbacher, 2007). Further, these limits appear to coincide with possible ancient and now disappeared physical barriers such as the Aure Trough, an inland sea filled by 3–4 Ma (Deiner et al., 2011), that now separates the distribution of *G. sclaterii* and *G. scheepmakeri*. Recent physical barriers could also be involved, such as the Bird's Neck isthmus that was formed between 3 and 5 Ma (Charlton, 2000) and now separates *G. cristata* from both *G. sclaterii* and *G. victoria*. The southeastern Peninsula that separates *G. victoria* and *G. scheepmakeri* may have existed for more than 5 Ma (Hill and Hall, 2003; van Ufford and Cloos, 2005) and be considered as another limit. While these barriers, in particular the Bird's Neck isthmus or the Aure Trough, might have led to the splitting of an ancestral population into the two main *Goura* lineages through vicariance, our dating results show divergence times < 2 Ma and strongly suggest that cross-barrier dispersal has been the primary event leading to speciation within these two lineages.

Geographically separated ranges abutting along common boundaries, as is the case between *G. cristata* and *G. victoria*, *G. cristata* and *G. sclaterii*, and *G. scheepmakeri* and *G. sclaterii*, can be associated with spatial changes in environmental factors, species interactions in areas of contact, or dispersal limitation even in the absence of physical barriers (Bull, 1991). Unfortunately, identifying the processes that may have led to differentiation in the first place, as well as the mechanisms that prevent range overlap in zones of contact, requires some knowledge of environmental niche differentiation among species. Such knowledge is currently not available, even though the different species of *Goura* are seen as ecologically similar (Gibbs et al., 2001), and will be difficult to acquire without obtaining much more data on the precise geographic distribution of the different species.

We estimated the stem age of the *Goura* lineage at 21.39 Ma [HPD: 17.32–25.44]. The interpretation of stem age is contingent upon taxon sampling and is highly sensitive to recently extinct taxa that were not sampled. The long temporal gap between stem and crown ages suggests a possible role of extinction history, especially since two of the closest relatives of *Goura* (*R. cucullatus* and *P. solitaria*) are indeed extinct (Shapiro et al., 2002). Known or presumed relatives live or lived mainly on islands from Mascarene Islands in the west to Samoan Islands in the east (Gibbs et al., 2001; Worthy, 2001). These islands, mainly of volcanic origin, are geologically unstable and that could have caused extinctions among close relatives (Heupink et al., 2014), in addition to the massive prehistoric and historic human-caused extinctions that have affected many landbirds, including pigeons, and are still and may remain largely undocumented (Steadman, 2006, 1995). Alternatively, it is also plausible that the *Goura* lineage originated and perhaps diversified in Australia before dispersing across to New Guinea, and then went extinct in Australia (Schodde, 2006); the absence of any known close relative of *Goura* in Australia and the lack of fossil data does not allow us to formally test the hypothesis, however.

4.4. Implications for conservation

Our results confirm the recent decision of del Hoyo and Collar (2014) to consider four species rather than three in the genus *Goura*. Current *ex situ* conservation measures such as breeding programs do not take this taxonomic change into account yet (*e.g.* EAZA, 2017). Most of the individuals kept in zoos as *G. scheepmakeri* probably belong to *G. sclaterii*. The distribution of *G. sclaterii* is much larger than that of *G. scheepmakeri*, for which only a few recent records are available, all in Lakekamu Basin where it may be rather common but is still hunted (B.M. Beehler, pers. comm.), and which may have already been extirpated long ago throughout most of its range in the southeastern peninsula (King and Nijboer, 1994). A review of digital images available on Google Images using MORPHIC (Leighton et al., 2016) using either "*Goura scheepmakeri*" or "Scheepmaker's Crowned Pigeon" as keywords did not yield a single picture corresponding to *G. scheepmakeri*. This raises important concerns regarding the conservation status of this species. Our study thus highlights that a rapid assessment of the populations throughout the known and presumed distribution range of *G. scheepmakeri* is of critical importance for the conservation of that species.

5. Conclusions

In this study, we confirmed the view that the *Goura* genus comprises four well separated lineages (del Hoyo and Collar, 2014), and two pairs of species that are sister groups (*i.e.*, *G. cristata* and *G. sclaterii* being the sister group to *G. victoria* and *G. scheepmakeri*). Interestingly, species with abutting geographic ranges are not each other's closest relatives, except *G. cristata* and *G. sclaterii*, which may meet somewhere between Etna Bay (Bird's Neck) and Mimika River (westernmost southern lowlands) (Mayr, 1941). Species diversification might have begun at the Miocene-Pliocene boundary, which suggests that it could have taken place within New Guinea. If so, our dating results show that an initial range expansion prior to barrier formation could have led to an early divergence due to vicariance associated with orogeny, but subsequent rounds of speciation were more likely due to cross-barrier dispersal events. Similar patterns appear to emerge from other recent studies of New Guinea complex lowland avian lineages, *e.g.* the Little Shrike-thrush *Colluricincla megarhyncha* (Deiner et al., 2011) and the White-eared Catbird *Ailuroedus buccoides* (Irestedt et al., 2015). This suggests that complicated phylogeographic processes have been at play in lowland regions of New Guinea, even though the vast expanses of forests occupying these regions form a continuous ring around the island with few apparent geographical and ecological isolating barriers (Mack and Dumbacher, 2007). We also found that the age of the *Goura* lineage itself accords with the hypothesis that Australia has been an important centre of diversification for Australasian bird lineages that dispersed to the island of New Guinea prior to extinction in Australia owing to post Miocene contraction of lowland rainforests (Schodde, 2006; Heinsohn and Hope, 2006). However, proving this connection may remain a difficult task in groups that are now restricted to New Guinea and for which the fossil record is scarce or even absent.

Acknowledgments

We are very grateful to Mark Adams (NHMUK), Paul Sweet and Thomas Trombone (AMNH), Robert Palmer (ANWC), and Jan Bolding Kristensen and Jon Fjeldså (ZMUC), for their invaluable contribution to this work through tissue loans. We are also grateful to Kadarusman, Gono Semiadi, Laurent Pouyaud, Régis Hocdé and Amos for encouraging us to conduct this study and for their invaluable help in the field, to Boris Delahaie, Joris Bertrand, and Yann Bourgeois for helping with toe-pad sampling at NHMUK, and to Jill Olofsson for enlightening discussions. We thank Hélène Holota for help with the laboratory work and Genotoul bioinformatics platform Toulouse Midi-Pyrenees for providing access to computing resources. We further thank Lynx Edicions for allowing us to reproduce paintings from the Handbook of the Birds of the World, and Staffan Widstrand and Fondation Iris for providing the picture used in the graphical abstract. We are also grateful to the reviewers for comments and suggestions that greatly improved the manuscript. The first author was supported by a Ministère de l'Enseignement Supérieur et de la Recherche PhD scholarship. Fieldwork and part of the museum work were supported by the Lengguru 2014 Project (www.lengguru.org), conducted by the French Institut de Recherche pour le Développement (IRD), the Indonesian Institute of Sciences (LIPI) with the Research Center for Biology (RCB) and the Research Center for Oceanography (RCO), the University of Papua (UNIPA), the University of Cendrawasih (UNCEN), the University of Musamus (UNMUS) and the Polytechnic School of Sorong (POLTEK) with corporate sponsorship from COLAS

Group, Veolia Water and the Total Foundation. Sequencing was supported by the 'Laboratoire d'Excellence' TULIP (ANR-10-LABX-41; ANR-11-IDEX-0002-02).

Data accessibility

In the NCBI nucleotide database, all newly assembled sequences are available under accessions MG590263 to MG590323.

Authors' contributions

C.T. and G.B. initiated, coordinated and supervised the project; J.B., CT. and G.B. conceived the study and designed the analyses; C.T., R.P-J., H.A., B.M., and L.J. coordinated the collection of field and museum samples; molecular data were obtained by J.B. with contributions from G.B; J.B., M.G. and G.B. analyzed the data, with contributions from C.T; J.B., M.G., G.B., and C.T. wrote the paper with comments from other authors.

Appendix A. Supplementary materials

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2017.11.022.

References

Aggerbeck, M., Fjeldså, J., Christidis, L., Fabre, P.-H., Jønsson, K.A., 2014. Resolving deep lineage divergences in core corvoid passerine birds supports a proto-Papuan island origin. Mol. Phylogenet. Evol. 70, 272–285. http://dx.doi.org/10.1016/j.ympev.2013.09.027.

Andersen, M.J., McCullough, J.M., Mauck, W.M., Smith, B.T., Moyle, R.G., 2017. A phylogeny of kingfishers reveals an Indomalayan origin and elevated rates of diversification on oceanic islands. J. Biogeogr. http://dx.doi.org/10.1111/jbi.13139. (in press).

Ballard, J.W.O., Whitlock, M.C., 2004. The incomplete natural history of mitochondria. Mol. Ecol. 13, 729–744. http://dx.doi.org/10.1046/j.1365-294X.2003.02063.x.

Beehler, B.M., Pratt, T.K., 2016. Birds of New Guinea: Distribution, Taxonomy, and Systematics. Princeton University Press, Princeton, NJ, USA.

Beehler, B.M., Pratt, T.K., Zimmerman, D.A., 1986. Birds of New Guinea. Princeton University Press, Princeton, NJ, USA.

Besnard, G., Bertrand, J.A.M., Delahaie, B., Bourgeois, Y.X.C., Lhuillier, E., Thébaud, C., 2016. Valuing museum specimens: high-throughput DNA sequencing on historical collections of New Guinea crowned pigeons (*Goura*). Biol. J. Linn. Soc. 117, 71–82. http://dx.doi.org/10.1111/bij.12494.

Bininda-Emonds, O.R.P., 2005. SeqCat.pl version 1.0. Program distributed by the author. AG Systematik und Evolutionsbiologie, IBU - Fakultät V, Carl von Ossietzky Universität Oldenburg.

Binladen, J., Wiuf, C., Gilbert, M.T.P., Bunce, M., Barnett, R., Larson, G., Greenwood, A.D., Haile, J., Ho, S.Y.W., Hansen, A.J., Willerslev, E., 2006. Assessing the fidelity of ancient DNA sequences amplified from nuclear genes. Genetics 172, 733–741. http://dx.doi.org/10.1534/genetics.105.049718.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. http://dx.doi.org/10.1093/bioinformatics/btu170.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 10, e1003537. http://dx.doi.org/10.1371/journal.pcbi.1003537.

Bull, C.M., 1991. Ecology of parapatric distributions. Annu. Rev. Ecol. Syst. 22, 19–36. http://dx.doi.org/10.1146/annurev.es.22.110191.000315.

Byrne, M., Steane, D.A., Joseph, L., Yeates, D.K., Jordan, G.J., Crayn, D., Aplin, K., Cantrill, D.J., Cook, L.G., Crisp, M.D., Keogh, J.S., Melville, J., Moritz, C., Porch, N., Sniderman, J.M.K., Sunnucks, P., Weston, P.H., 2011. Decline of a biome: evolution, contraction, fragmentation, extinction and invasion of the Australian mesic zone biota. J. Biogeogr. 38, 1635–1656. http://dx.doi.org/10.1111/j.1365-2699.2011.02535.x.

Carmi, O., Witt, C.C., Jaramillo, A., Dumbacher, J.P., 2016. Phylogeography of the Vermilion Flycatcher species complex: Multiple speciation events, shifts in migratory behavior, and an apparent extinction of a Galápagos-endemic bird species. Mol. Phylogenet. Evol. 102, 152–173. http://dx.doi.org/10.1016/j.ympev.2016.05.029.

Charlton, T.R., 2000. Tertiary evolution of the eastern Indonesia collision complex. J. Asian Earth Sci. 18, 603–631. http://dx.doi.org/10.1016/S1367-9120(99)00049-8.

Chikuni, K., Minaka, N., Ikenaga, H., 1996. Molecular phylogeny of some Passeriformes, based on cytochrome b sequences. J. Yamashina Inst. Ornithol. 28, 1–8. http://dx.doi.org/10.3312/jyio1952.28.1.

Claramunt, S., Cracraft, J., 2015. A new time tree reveals Earth history's imprint on the evolution of modern birds. Sci. Adv. 1, e1501005. http://dx.doi.org/10.1126/sciadv.1501005.

Darwin, C., 1868. The Variation of Animals and Plants under Domestication. In: John Murray, London, UK.

Deiner, K., Lemmon, A.R., Mack, A.L., Fleischer, R.C., Dumbacher, J.P., 2011. A passerine bird's evolution corroborates the geologic history of the island of New Guinea. PLoS ONE 6. http://dx.doi.org/10.1371/journal.pone.0019479.

del Hoyo, J., Collar, N.J., 2014. HBW and BirdLife International Illustrated Checklist of the Birds of the World 1: Non-passerines. In: Lynx Edicions, Barcelona.

Drummond, A.J., Ho, S.Y.W., Phillips, M.J., Rambaut, A., 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4, e88. http://dx.doi.org/10.1371/journal.pbio.0040088.

EAZA, 2017. Specialist Programmes [WWW Document]. Eur. Assoc. Zoos Aquar, URL http://www.eaza.net/conservation/programmes/ (accessed 04.14.17).

Edwards, S.V., Beerli, P., 2000. Perspective: gene divergence, population divergence, and the variance incoalescence time in phylogeographic studies. Evolution 54, 1839–1854. http://dx.doi.org/10.1554/0014-3820(2000) 054[1839:PGDPDA]2.0.CO;2.

Fulton, T.L., Wagner, S.M., Fisher, C., Shapiro, B., 2012. Nuclear DNA from the extinct Passenger Pigeon (*Ectopistes migratorius*) confirms a single origin of New World pigeons. Ann. Anat. - Anat. Anz. Special Issue: Ancient DNA 194, 52–57. http://dx.doi.org/10.1016/j.aanat.2011.02.017.

Georges, A., Zhang, X., Unmack, P., Reid, B.N., Le, M., McCord, W.P., 2014. Contemporary genetic structure of an endemic freshwater turtle reflects Miocene orogenesis of New Guinea. Biol. J. Linn. Soc. 111, 192–208. http://dx.doi.org/10.1111/bij.12176.

Gibbs, D., Barnes, E., Cox, J., 2001. Pigeons and Doves: A Guide to the Pigeons and Doves of the World. Bloomsbury Publishing, London, UK.

Guschanski, K., Krause, J., Sawyer, S., Valente, L.M., Bailey, S., Finstermeier, K., Sabin, R., Gilissen, E., Sonet, G., Nagy, Z.T., Lenglet, G., Mayer, F., Savolainen, V., 2013. Next-generation museomics disentangles one of the largest primate radiations. Syst. Biol. 62, 539–554. http://dx.doi.org/10.1093/sysbio/syt018.

Hall, R., 2013. The palaeogeography of Sundaland and Wallacea since the Late Jurassic. J. Limnol. 72, 1.

Hall, R., 1998. The plate tectonics of Cenozoic SE Asia and the distribution of land and sea. In: Hall, Robert, Holloway, Jeremy D. (Eds.), Biogeography and Geological Evolution of SE Asia. Backhuys Publishers, Leiden, The Netherlands, pp. 99–131.

Hartert, E., Paludan, K., Rothschild, Lord, Stresemann, E., 1936. Die Vögel des Weyland-Gebirges und seines Vorlandes. Mitteilungen Aus Dem Zool. Mus. Berl. 21, 165–240.

Hasegawa, M., Kishino, H., Yano, T., 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 22, 160–174. http://dx.doi.org/10.1007/BF02101694.

Heads, M., 2001. Birds of paradise, biogeography and ecology in New Guinea: a review. J. Biogeogr. 28, 893–925. http://dx.doi.org/10.1046/j.1365-2699.2001.00600.x.

Heinsohn, T., Hope, G., 2006. The Torresian connections: zoogeography of New Guinea. In: Merrick, J.R., Archer, M., Hickey, G.M., Lee, M.S.Y. (Eds.), Evolution and Biogeography of Australian Vertebrates. Australian Scientific Publishing, Oatlands, Australia.

Heupink, T.H., van Grouw, H., Lambert, D.M., 2014. The mysterious Spotted Green Pigeon and its relation to the Dodo and its kindred. BMC Evol. Biol. 14, 136. http://dx.doi.org/10.1186/1471-2148-14-136.

Hill, K.C., Hall, R., 2003. Mesozoic-Cenozoic evolution of Australia's New Guinea margin in a west Pacific context. Geol. Soc. Am. Spec. Pap. 372, 265–290. http://dx.doi.org/10.1130/0-8137-2372-8.265.

Irestedt, M., Batalha-Filho, H., Roselaar, C.S., Christidis, L., Ericson, P.G.P., 2015. Contrasting phylogeographic signatures in two Australo-Papuan bowerbird species complexes (Aves: *Ailuroedus*). Zool. Scr. 45, 365–379. http://dx.doi.org/10.1111/zsc.12163.

Irestedt, M., Ohlson, J.I., Zuccon, D., Källersjö, M., Ericson, P.G.P., 2006. Nuclear DNA from old collections of avian study skins reveals the evolutionary history of the Old World suboscines (Aves, Passeriformes). Zool. Scr. 35, 567–580. http://dx.doi.org/10.1111/j.1463-6409.2006.00249.x.

Johnson, K.P., Clayton, D.H., 2000. Nuclear and mitochondrial genes contain similar phylogenetic signal for pigeons and doves (Aves: Columbiformes). Mol. Phylogenet. Evol. 14, 141–151. http://dx.doi.org/10.1006/mpev.1999.0682.

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. http://dx.doi.org/10.1093/bioinformatics/btt193.

Jønsson, K.A., Borregaard, M.K., Carstensen, D.W., Hansen, L.A., Kennedy, J.D., Machac, A., Marki, P.Z., Fjeldså, J., Rahbek, C., 2017. Biogeography and biotic assembly of indo-pacific corvoid passerine birds. Annu. Rev. Ecol. Evol. Syst. 48, 231–253. http://dx.doi.org/10.1146/annurev-ecolsys-110316-022813.

Jønsson, K.A., Fabre, P.-H., Kennedy, J.D., Holt, B.G., Borregaard, M.K., Rahbek, C., Fjeldså, J., 2016. A supermatrix phylogeny of corvoid passerine birds (Aves: Corvides). Mol. Phylogenet. Evol. 94, 87–94. http://dx.doi.org/10.1016/j.ympev.2015.08.020.

Jønsson, K.A., Fabre, P.-H., Ricklefs, R.E., Fjeldså, J., 2011. Major global radiation of corvoid birds originated in the proto-Papuan archipelago. Proc. Natl. Acad. Sci. U. S. A. 108, 2328–2333. http://dx.doi.org/10.1073/pnas.1018956108.

King, C.E., Nijboer, J., 1994. Conservation considerations for crowned pigeons, genus *Goura*. Oryx 28, 22–30. http://dx.doi.org/10.1017/S0030605300028258.

Korneliussen, T.S., Albrechtsen, A., Nielsen, R., 2014. ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinf. 15, 356. http://dx.doi.org/10.1186/s12859-014-0356-4.

Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B., 2017. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol. Biol. Evol. 34, 772–773. http://dx.doi.org/10.

1093/molbev/msw260.
Leighton, G.R.M., Hugo, P.S., Roulin, A., Amar, A., 2016. Just Google it: assessing the use of Google Images to describe geographical variation in visible traits of organisms. Methods Ecol. Evol. 7, 1060–1070. http://dx.doi.org/10.1111/2041-210X.12562.
Lepage, T., Bryant, D., Philippe, H., Lartillot, N., 2007. A general comparison of relaxed molecular clock models. Mol. Biol. Evol. 24, 2669–2680. http://dx.doi.org/10.1093/molbev/msm193.
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, pp. 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. http://dx.doi.org/10.1007/BF00163806.
Lovette, I.J., 2004. Mitochondrial dating and mixed support for the "2% rule" in birds. Auk 121, 1–6. http://dx.doi.org/10.1642/0004-8038(2004) 121[0001:MDAMSF]2.0.CO;2.
Mack, A., Dumbacher, J., 2007. Birds of Papua. In: Marshall, Andrew J., Beehler, Bruce M. (Eds.), The Ecology of Papua, Part I. The Ecology of Indonesia Series VI. Periplus, Singapore, pp. 654–688.
Malé, P.-J.G., Bardon, L., Besnard, G., Coissac, E., Delsuc, F., Engel, J., Lhuillier, E., Scotti-Saintagne, C., Tinaut, A., Chave, J., 2014. Genome skimming by shotgun sequencing helps resolve the phylogeny of a pantropical tree family. Mol. Ecol. Resour. 14, 966–975. http://dx.doi.org/10.1111/1755-0998.12246.
Marki, P.Z., Jønsson, K.A., Irestedt, M., Nguyen, J.M.T., Rahbek, C., Fjeldså, J., 2017. Supermatrix phylogeny and biogeography of the Australasian Meliphagides radiation (Aves: Passeriformes). Mol. Phylogenet. Evol. 107, 516–529. http://dx.doi.org/10.1016/j.ympev.2016.12.021.
Mayr, E., 1941. List of New Guinea Birds: A Systematic and Faunal List of the Birds of New Guinea and Adjacent Islands. In: American Museum of Natural History, New-York, USA.
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. http://dx.doi.org/10.1371/journal.pone.0054848.
McCormack, J.E., Tsai, W.L.E., Faircloth, B.C., 2016. Sequence capture of ultraconserved elements from bird museum specimens. Mol. Ecol. Resour. 16, 1189–1203. http://dx.doi.org/10.1111/1755-0998.12466.
Moore, W.S., 1995. Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. Evolution 49, 718–726. http://dx.doi.org/10.1111/j.1558-5646.1995.tb02308.x.
Moyle, R.G., Oliveros, C.H., Andersen, M.J., Hosner, P.A., Benz, B.W., Manthey, J.D., Travers, S.L., Brown, R.M., Faircloth, B.C., 2016. Tectonic collision and uplift of Wallacea triggered the global songbird radiation. Nat. Commun. 7, 12709. http://dx.doi.org/10.1038/ncomms12709.
Nabholz, B., Lanfear, R., Fuchs, J., 2016. Body mass-corrected molecular rate for bird mitochondrial DNA. Mol. Ecol. 25, 4438–4449. http://dx.doi.org/10.1111/mec.13780.
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. http://dx.doi.org/10.1111/mec.13914.
Paśko, Ł., Ericson, P.G.P., Elzanowski, A., 2011. Phylogenetic utility and evolution of indels: A study in neognathous birds. Mol. Phylogenet. Evol. 61, 760–771. http://dx.doi.org/10.1016/j.ympev.2011.07.021.
Pereira, S.L., Johnson, K.P., Clayton, D.H., Baker, A.J., 2007. Mitochondrial and nuclear DNA sequences support a Cretaceous origin of Columbiformes and a dispersal-driven radiation in the Paleogene. Biol. Syst. http://dx.doi.org/10.1080/10635150701549672.
Pratt, T.K., Beehler, B.M., 2015. Birds of New Guinea, second ed. Princeton University Press, Princeton, NJ, USA.
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. http://dx.doi.org/10.1038/nature15697.
Rambaut, A., Suchard, M., Xie, D., Drummond, A., 2014. Tracer v1.6, Available from http://beast.bio.ed.ac.uk/Tracer.
Rand, A.L., Gilliard, E.T., 1968. Handbook of New Guinea Birds. Weidenfeld & Nicolson, London, UK.
Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539–542. http://dx.doi.org/10.1093/sysbio/sys029.
Sawyer, S., Krause, J., Guschanski, K., Savolainen, V., Pääbo, S., 2012. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. PLoS ONE 7, e34131. http://dx.doi.org/10.1371/journal.pone.0034131.
Schodde, R., 2006. Australia's bird fauna today–origins and evolutionary development. In: Merrick, J.R., Archer, G.M., Hickey, G.M., Lee, M.S.Y. (Eds.), Evolution and Biogeography of Australasian Vertebrates. Auscipub, Oatlands, Australia, pp. 413–458.
Schodde, R., Christidis, L., 2014. Relicts from Tertiary Australasia: undescribed families and subfamilies of songbirds (Passeriformes) and their zoogeographic signal. Zootaxa 3786, 501–522. https://doi.org/10.11646/zootaxa.3786.5.1.
Schweizer, M., Wright, T.F., Peñalba, J.V., Schirtzinger, E.E., Joseph, L., 2015. Molecular phylogenetics suggests a New Guinean origin and frequent episodes of founder-event speciation in the nectarivorous lories and lorikeets (Aves: Psittaciformes). Mol. Phylogenet. Evol. 90, 34–48. http://dx.doi.org/10.1016/j.ympev.2015.04.021.
Shapiro, B., Sibthorpe, D., Rambaut, A., Austin, J., Wragg, G.M., Bininda-Emonds, O.R., Lee, P.L., Cooper, A., 2002. Flight of the dodo. Science 295. http://dx.doi.org/10.1126/science.295.5560.1683.
Shapiro, M.D., Kronenberg, Z., Li, C., Domyan, E.T., Pan, H., Campbell, M., Tan, H., Huff, C.D., Hu, H., Vickrey, A.I., Nielsen, S.C.A., Stringham, S.A., Hu, H., Willerslev, E., Gilbert, M.T.P., Yandell, M., Zhang, G., Wang, J., 2013. Genomic diversity and evolution of the head crest in the rock pigeon. Science 339, 1063–1067. http://dx.doi.org/10.1126/science.1230422.
Shipham, A., Schmidt, D.J., Joseph, L., Hughes, J.M., 2015. Phylogenetic analysis of the Australian rosella parrots (*Platycercus*) reveals discordance among molecules and plumage. Mol. Phylogenet. Evol. 91, 150–159. http://dx.doi.org/10.1016/j.ympev.2015.05.012.
Soares, A.E.R., Novak, B.J., Haile, J., Heupink, T.H., Fjeldså, J., Gilbert, M.T.P., Poinar, H., Church, G.M., Shapiro, B., 2016. Complete mitochondrial genomes of living and extinct pigeons revise the timing of the columbiform radiation. BMC Evol. Biol. 16, 230. http://dx.doi.org/10.1186/s12862-016-0800-3.
Stadler, T., 2009. On incomplete sampling under birth–death models and connections to the sampling-based coalescent. J. Theor. Biol. 261, 58–66. http://dx.doi.org/10.1016/j.jtbi.2009.07.018.
Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313. http://dx.doi.org/10.1093/bioinformatics/btu033.
Steadman, D.W., 2006. An extinct species of tooth-billed pigeon (*Didunculus*) from the Kingdom of Tonga, and the concept of endemism in insular landbirds. J. Zool. 268, 233–241. http://dx.doi.org/10.1111/j.1469-7998.2005.00010.x.
Steadman, D.W., 1995. Prehistoric extinctions of pacific island birds: biodiversity meets zooarchaeology. Science 267, 1123–1131. http://dx.doi.org/10.1126/science.267.5201.1123.
Straub, S.C.K., Parks, M., Weitemier, K., Fishbein, M., Cronn, R.C., Liston, A., 2012. Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics. Am. J. Bot. 99, 349–364. http://dx.doi.org/10.3732/ajb.1100335.
Tamura, K., Nei, M., 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. 10, 512–526. http://dx.doi.org/10.1093/oxfordjournals.molbev.a040023.
Tavaré, S., 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. Lect. Math. Life Sci. 17, 57–86.
Toussaint, E.F.A., Hall, R., Monaghan, M.T., Sagata, K., Ibalim, S., Shaverdo, H.V., Vogler, A.P., Pons, J., Balke, M., 2014. The towering orogeny of New Guinea as a trigger for arthropod megadiversity. Nat. Commun. 5, 4001. http://dx.doi.org/10.1038/ncomms5001.
Toussaint, E.F.A., Shaverdo, H., Hendrich, L., Balke, M., 2015. Mosaic patterns of diversification dynamics following the colonization of Melanesian Islands. Sci. Rep. 5, 16016. http://dx.doi.org/10.1038/srep16016.
van Ufford, A.Q., Cloos, M., 2005. Cenozoic tectonics of New Guinea. AAPG Bull. 89, 119–140. http://dx.doi.org/10.1306/08300403073.
van Wyhe, J., Rookmaaker, K., 2013. Alfred Russel Wallace: Letters from the Malay Archipelago. Oxford University Press, Oxford, UK.
Voris, H.K., 2000. Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. J. Biogeogr. 27, 1153–1167. http://dx.doi.org/10.1046/j.1365-2699.2000.00489.x.
Wallace, A.R., 1905. My Life: A Record of Events and Opinions. Dodd, Mead & Company, New-York, USA.
Wallace, A.R., 1876. The Geographical Distribution of Animals: with a Study of the Relations of Living and Extinct Faunas as Elucidating the Past Changes of the Earth's Surface. Macmillan and Co, London, UK.
Wallace, A.R., 1869. The Malay Archipelago: The Land of the Orang-utan, and the Bird of paradise. A Narrative of Travel, with Studies of Man and Nature. Macmillan and Co, London, UK.
Weir, J.T., 2006. Divergent timing and patterns of species accumulation in lowland and highland neotropical birds. Evolution 60, 842–855. http://dx.doi.org/10.1111/j.0014-3820.2006.tb01161.x.
Weir, J.T., Schluter, D., 2008. Calibrating the avian molecular clock. Mol. Ecol. 17, 2321–2328. http://dx.doi.org/10.1111/j.1365[HYPHEN]294X.2008.03742.x.
Worthy, T.H., 2001. A giant flightless pigeon gen. et sp. nov. and a new species of Ducula (Aves: Columbidae), from Quaternary deposits in Fiji. J. R. Soc. New Zealand 31 (4), 763–794. http://dx.doi.org/10.1080/03014223.2001.9517673.
Wu, T.D., Nacu, S., 2010. Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics 26, 873–881. http://dx.doi.org/10.1093/bioinformatics/btq057.

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 (χ^2= 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).

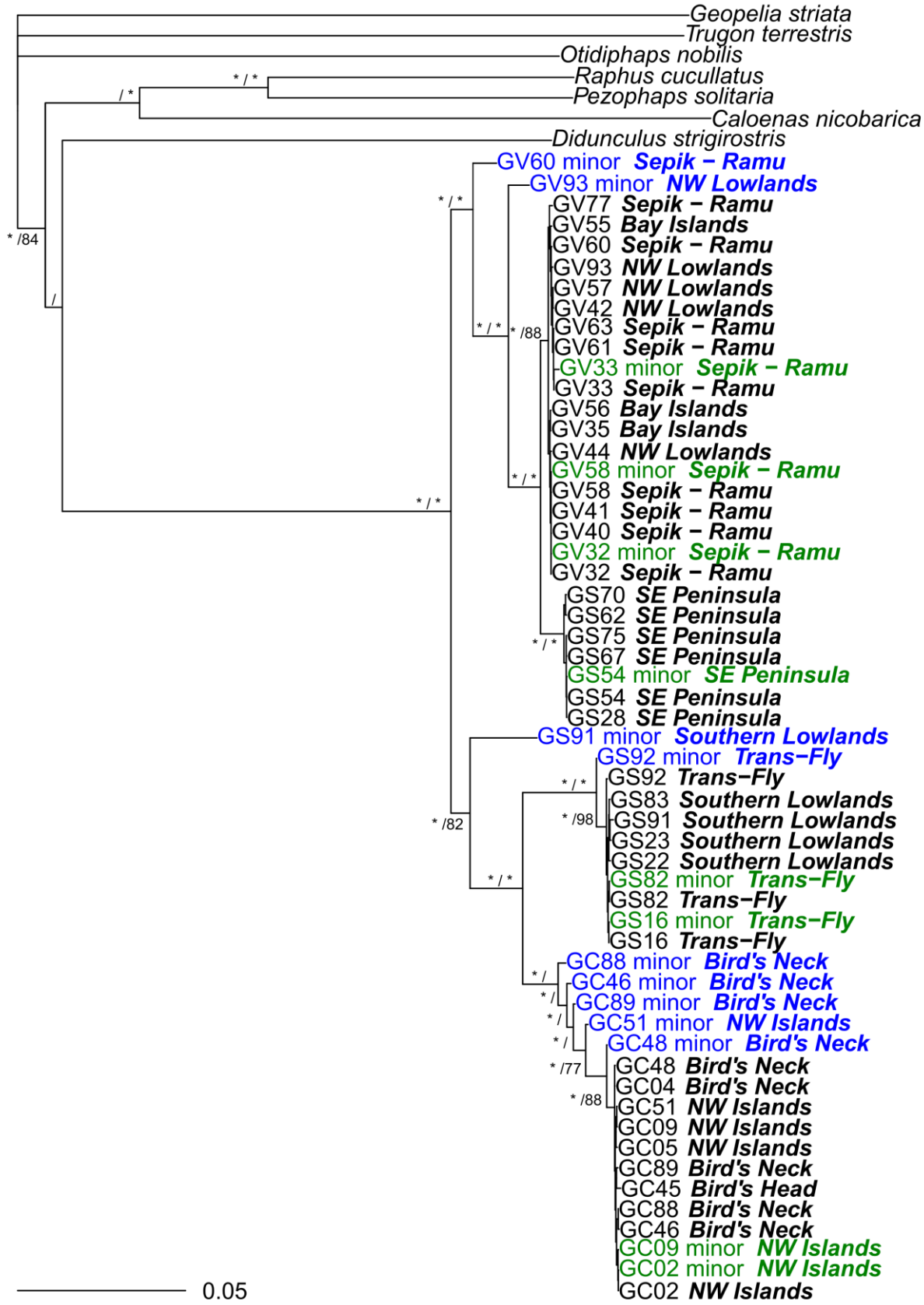


Figure S1: Phylogenetic tree including all the mitogenome and minor variants for the 17 individuals involved. On major nodes are mentioned Bayesian posterior probabilities on the left and bootstrap value on the right. Minor variants of group I are written in green, those of group II are in blue.

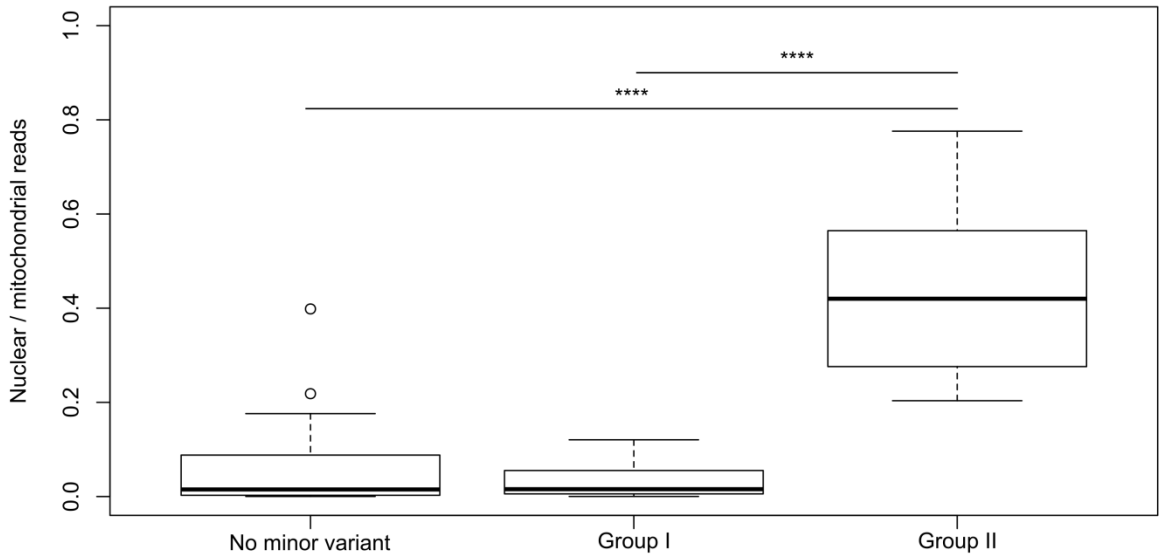


Figure S2: Comparison of the ratio nuclear / mitochondrial reads between three groups of samples: no minor variant, Group I, and Group II. The difference is statistically significant between the first and last groups (Z=-4.273, *p*-value = 2.89e-05) and between the last two groups (Z=-3.389, *p*-value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

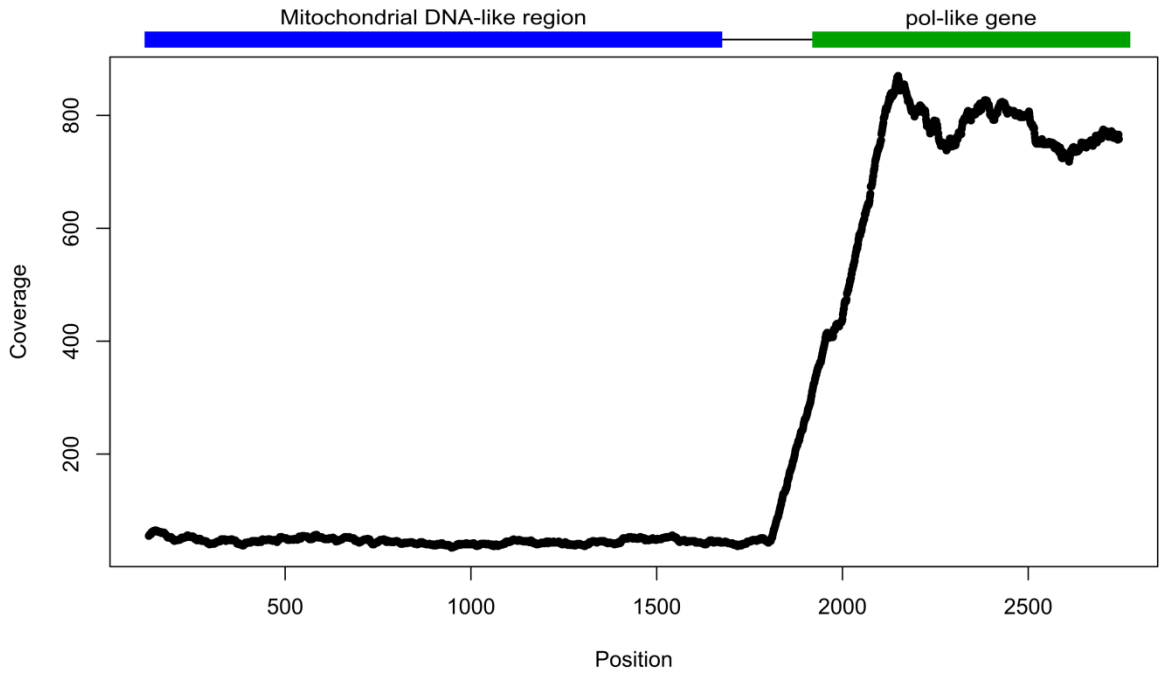


Figure S3: Sequencing depth along the GC88 minor mitochondrial DNA-like variant extended. The first part (1611 bp) shows a high homology with the *Goura* mitogenome, whereas the second part shows a high homology with a nuclear gene encoding a pol-like protein of retrotransposon (LTR).

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* species clades 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 get rid 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

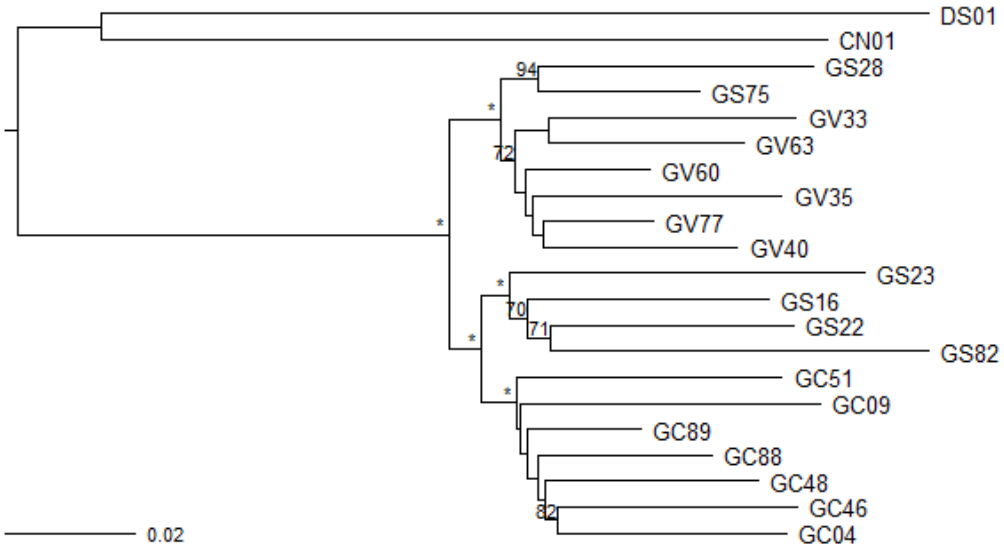


Figure S4: Phylogenetic tree obtained from all SNP of the conserved nuclear regions, without filtering. * corresponds to a node support of 100, supports lower than 70 were removed.

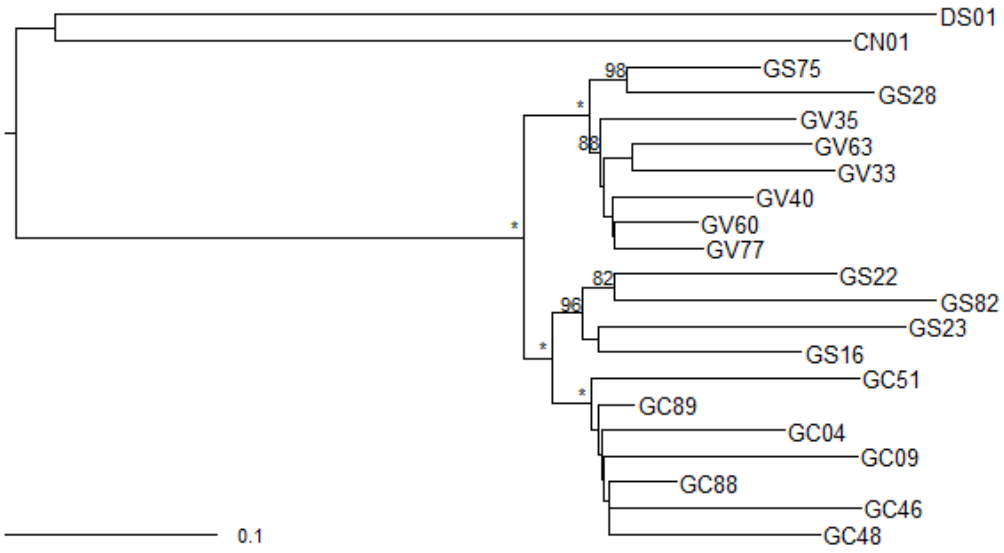


Figure S5: Phylogenetic tree obtained from SNP of the conserved nuclear regions filtered for a minimal depth of two, alleles present at least two times across individuals, and missing data representing less than 80% of the data at each position. * corresponds to a node support of 100, supports lower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

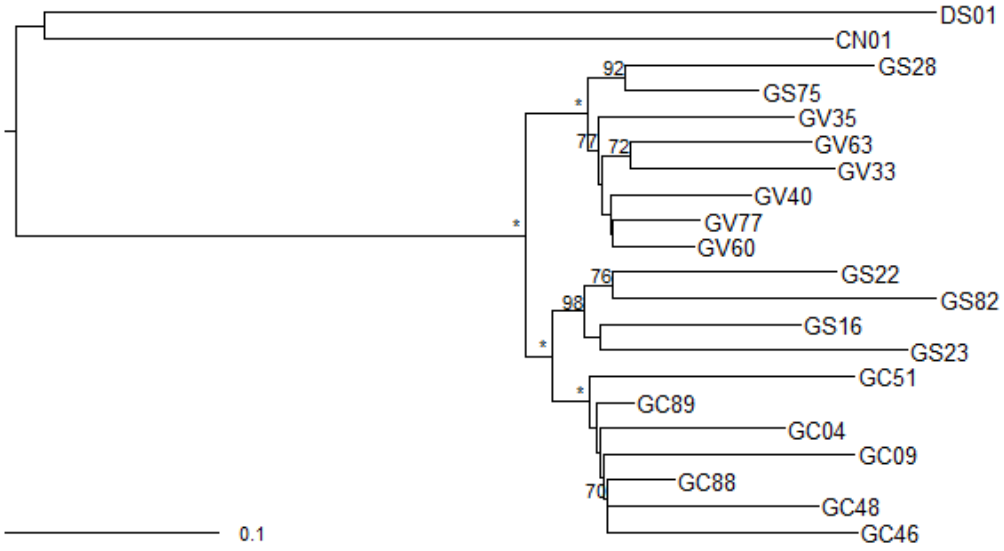


Figure S6: Phylogenetic tree obtained from SNP of the conserved nuclear regions filtered for a minimal depth of two, alleles present at least two times across individuals, and missing data representing less than 70% of the data at each position. * corresponds to a node support of 100, supports lower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

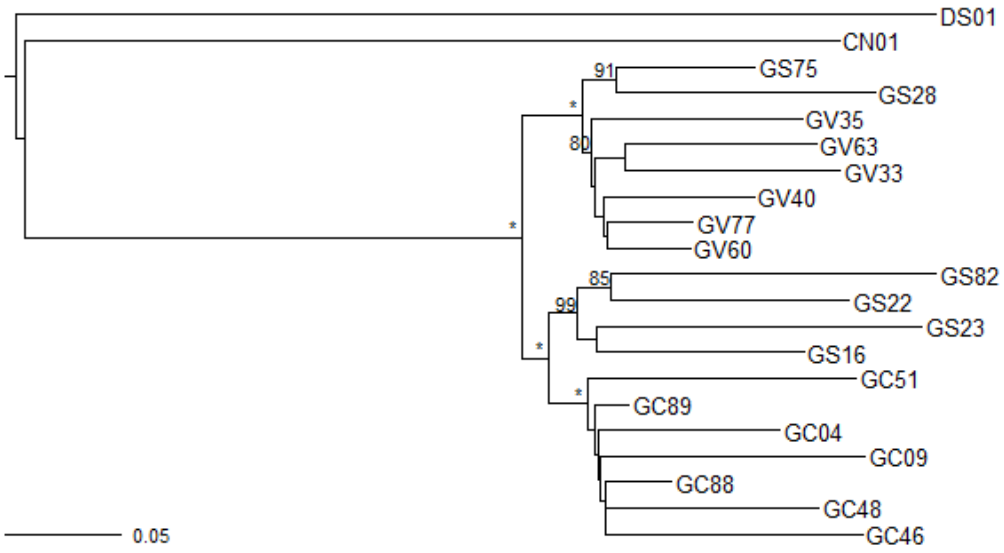


Figure S7: Phylogenetic tree obtained from SNP of the conserved nuclear regions filtered for a minimal depth of two, alleles present at least two times across individuals, and missing data representing less than 60% of the data at each position. * corresponds to a node support of 100, supports lower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

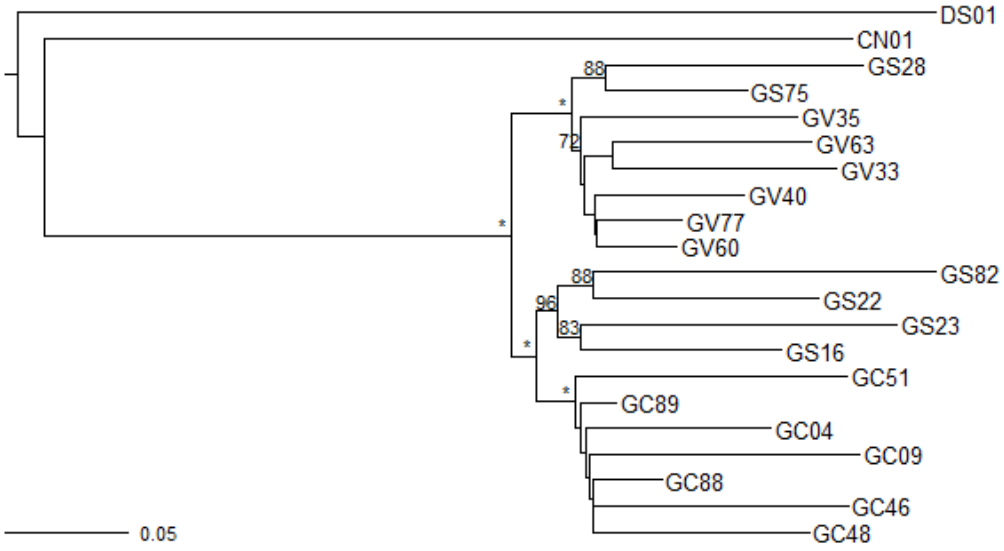


Figure S8: Phylogenetic tree obtained from SNP of the conserved nuclear regions filtered for a minimal depth of two, alleles present at least two times across individuals, and missing data representing less than 50% of the data at each position. * corresponds to a node support of 100, supports lower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

Museum	Museum ID	Code	Species	Sex	Year of collect	Collector	Type	Locality	Latitude (DD)	Longitude (DD)	Region
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, 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	-

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
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.

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. 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. 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

General discussion

The general objective of this thesis was to explain the diversification process that gave birth to a species-rich worldwide-distributed family. With ca. 350 species assigned to 50 genera, the Columbidae family forms an interesting model for biogeographical studies of birds. They are distributed through all biogeographic area, except polar regions and extremely isolated islands, which they colonized only since human settlement. Often present on islands (almost 60% of species are restricted to these entities), they are especially diverse on New Guinea, where up to 20 species can co-occur.

Throughout this whole thesis, we used phylogenies to describe speciation from the worldwide perspective to the intra-island level. To obtain representative sampling that allow us to address these questions, we mostly used museum samples (toe-pad from skins) and sequenced them by genome skimming. This method allows obtaining good to excellent coverage for regions such as mitogenomes or nuclear ribosomal cluster, but allows also the study of nuclear single-copy regions at low coverage.

In the following pages, we will briefly summarize major results we obtained in the two previous chapters and highlight a few important points that deserve discussion. We will then tackle different perspectives of this work, once the biogeography analysis of the whole family will be refined.

1. Summary and general conclusions

Diversification of the Columbidae family probably took place between mid-Eocene and mid-Miocene, where all extant clades were already established. Two independent dispersal events seem to have occurred between New World and Philippines / Wallacea before surrounding dispersals. These dispersions would have taken place thanks to stepping-stones movement through South Atlantic islands to Africa, and subsequently to India. These islands could also explains reverse movements from New Guinea and Wallacea to Africa, Madagascar and Mascareignes, which are supported by winds and currents. On the other side, islands do not seem to have played a so important role in diversification, with rates varying following a normal distribution, more than a model depending of insularity.

Nevertheless, the highest Columbidae species-richness is reached in Philippines, Wallacea and Australasia, regions especially rich in islands. New Guinea, in particular, hosts an impressive diversity. The biggest extant Columbidae, belonging to the genus *Goura* (crowned pigeons), are endemic to this island. Their present diversity probably arose within New Guinea since the last 5-6 million years. Geological events such as the cordillera orogeny probably played a major role by splitting ancestral distribution, while more recent dispersals over barriers may have also led to the settlement of isolated populations that promoted diversification.

1.1. Natural history collections importance

All these results were obtained thanks to museum collections, which are of critical importance not only for the study of ancient or ongoing phenomenon (DuBay & Fuldner, 2017), but also for sampling of large, remote areas. Technological development of next-generation sequencing has allowed access to increasingly ancient genetic data (Orlando et al., 2013), giving an additional value to these samples. However, general audience and part of the scientific community display an increasing opposition to modern collections (Kaplan & Moyer, 2015), and museums face some difficulties to extend their collections (Kaplan, 2017). Our results demonstrate the feasibility of phylogenomic analyses of large groups without field works, and we can barely imagine what future technologies will allow us to gather from centuries-old samples. Natural history collections thus deserve conservation and extension (Webster, 2017).

1.2. Sequencing strategies

Many strategies can be adopted to recover genetic data, whether samples are from fresh or ancient samples. If the study focuses on a few genes or regions of interest, they can be recovered by PCR (Polymerase Chain Reaction) or by capture (McCormack et al., 2016). This allows optimizing the data sequenced and the money spent to answer a specific question. On the other side, new questions arising from preliminary or previous analyses can imply resequencing the same sample for different markers.

Here, we used another classical method: genome skimming. The sample is sequenced as a whole, allowing primarily the recovery of mitogenome data due to their better conservation in old samples (Pääbo et al., 1989). The major constraint of this method with historical sample is the sequencing of all the contaminants found on its surface. However, we show here that, even when mitochondrial data are the main goal of this sequencing (as with old samples), nuclear data can be exploited, allowing an optimization of the sample usage and a second analysis of older data from which only mitochondrial data had been studied. The amount of missing data does not seem to affect the topology (Wiens, 2003; Wiens & Moen, 2008) and using nuclear data avoids the limitations of mitochondrial analysis only (Zink & Barrowclough, 2008; Galtier et al., 2009; Toews & Brelsford, 2012).

2. Perspectives

2.1. Explaining the current extraordinary diversity of Columbidae in New Guinea

New Guinea and its surrounding islands (Rajat Ampat, Geelvink Bay islands, d'Entrecasteaux and Louisiade archipelagos and Aru islands; Bismarck archipelago, New Britain and Solomon islands being not considered here) are inhabited by 49 of the 367 Columbidae species considered in the Handbook of the Birds of the World (Baptista et al., 2017). Among them, 28 are endemic to this region. The New Guinea mainland itself hosts 39 species, including seven endemics.

These species are especially diverse, ranging from the smallest Columbidae species (*Ptilinopus nainus*, weighing 49 g) to the biggest extant ones (*Goura* spp., weighing around 2 kg). They belong to 15 genera, from single-species genera to species-rich ones (Figure 24).

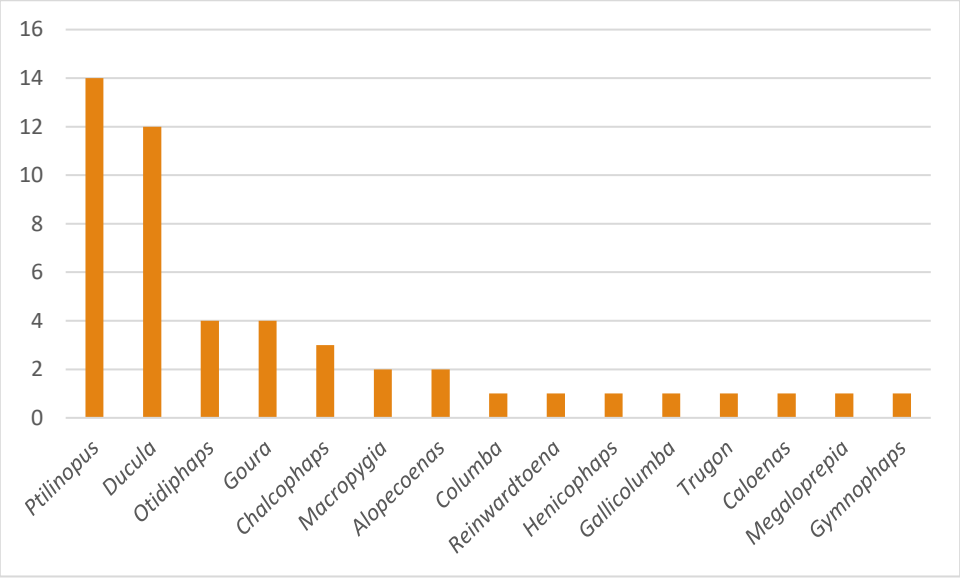


FIGURE 24: SPECIES RICHNESS PER GENUS IN THE NEW GUINEA REGION *LATO SENSU*. DATA ARE FROM HBW (BAPTISTA ET AL., 2017).

Surprisingly, all these genera are not closely related (Figure 25) and for some of them, all the species are present in the region (e.g. *Goura* spp., *Otidiphaps* spp.), when for others only a little proportion of the genus richness is present in this region (e.g. one *Columba* spp. among 35). This pattern probably implies complex scenarios of immigration, extinction and *in-situ* diversification that could be tested statistically thanks to programs like the package DAISIE (Etienne et al., 2017) developed on R (R Core Team, 2017).

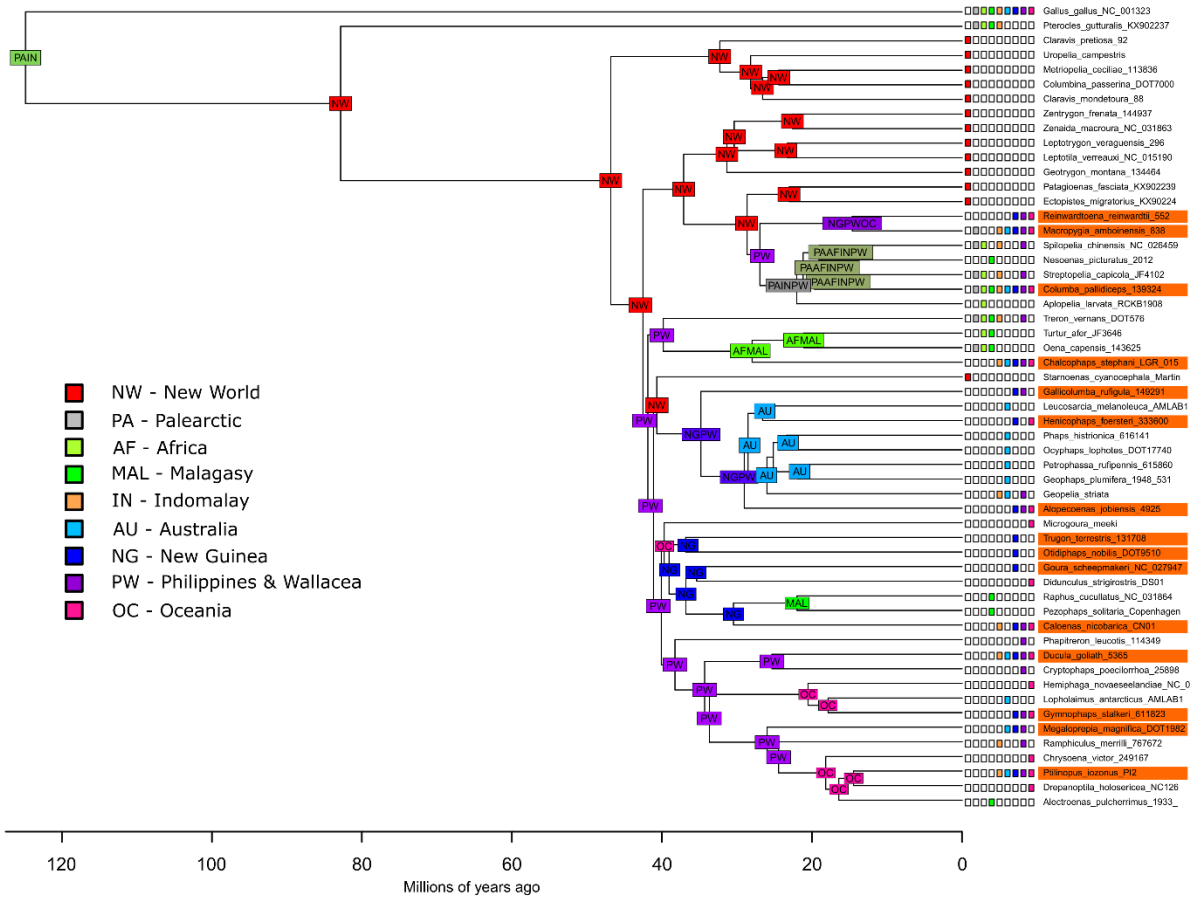


FIGURE 25: MITOCHONDRIAL DNA PHYLOGENY OF THE COLUMBIDAE FROM CHAPTER 1, WITH NEW GUINEAN LINEAGES HIGHLIGHTED IN ORANGE.

2.2. Comparing the crowned pigeons diversification with other New Guinean clades

In the second chapter, we explained the current crowned pigeons distributions by past events such as orogeny of the central cordillera and crossing of ancient barriers like the Aure Trough or the Bird's Neck isthmus. Interestingly, these barriers are common to many New Guinean species distribution (Mack & Dumbacher, 2007) and could therefore have played a major role in many lineages diversification. To evaluate this role at a larger biological scale, it would be interesting to compare these results with those recovered from other clades.

One of these other examples could be the endemic family of Melanocharitidae, which includes four genera and 11 species. Within the *Melanocharis* genus, one species seems to be especially suited for this comparison, *M. nigra* (Black Berrypecker). This species, which has four sub-species recognized, present common distributional patterns with the *Goura* genus.

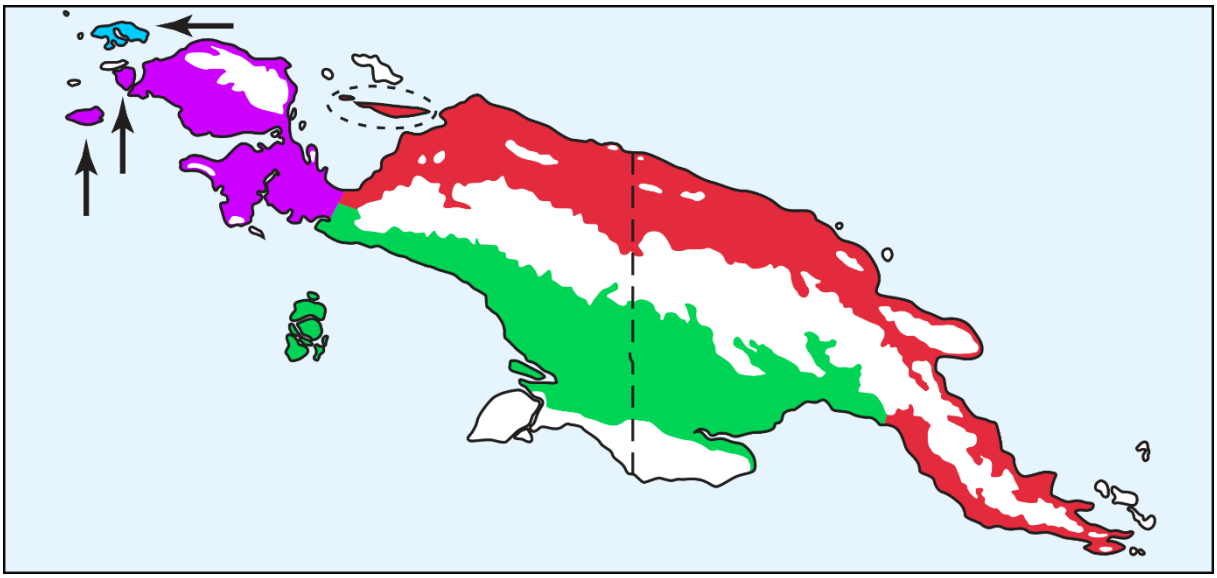


FIGURE 26: *MELANOCHARIS NIGRA* DISTRIBUTION MAP, BASED ON (PRATT & BEEHLER, 2015). SUB-SPECIES ARE PRESENTED BY DIFFERENT COLORS: *M. N. PALLIDA* IN BLUE, *M. N. NIGRA* IN PURPLE, *M. N. UNICOLOR* IN RED, AND *M. N. CHLOROPTERA* IN GREEN.

I worked as a side-project on this genus phylogeny. Data are still preliminary but show more deeper divergence than the one observed with the crowned pigeons data (around a level of magnitude higher). This is maybe a consequence of a faster evolutionary rate in *Melanocharis* spp. than in Goura spp., possibly due to size differences (Nabholz et al., 2016). A dating analysis will be performed to compare diversification of this genus with the one observed in crowned pigeons.

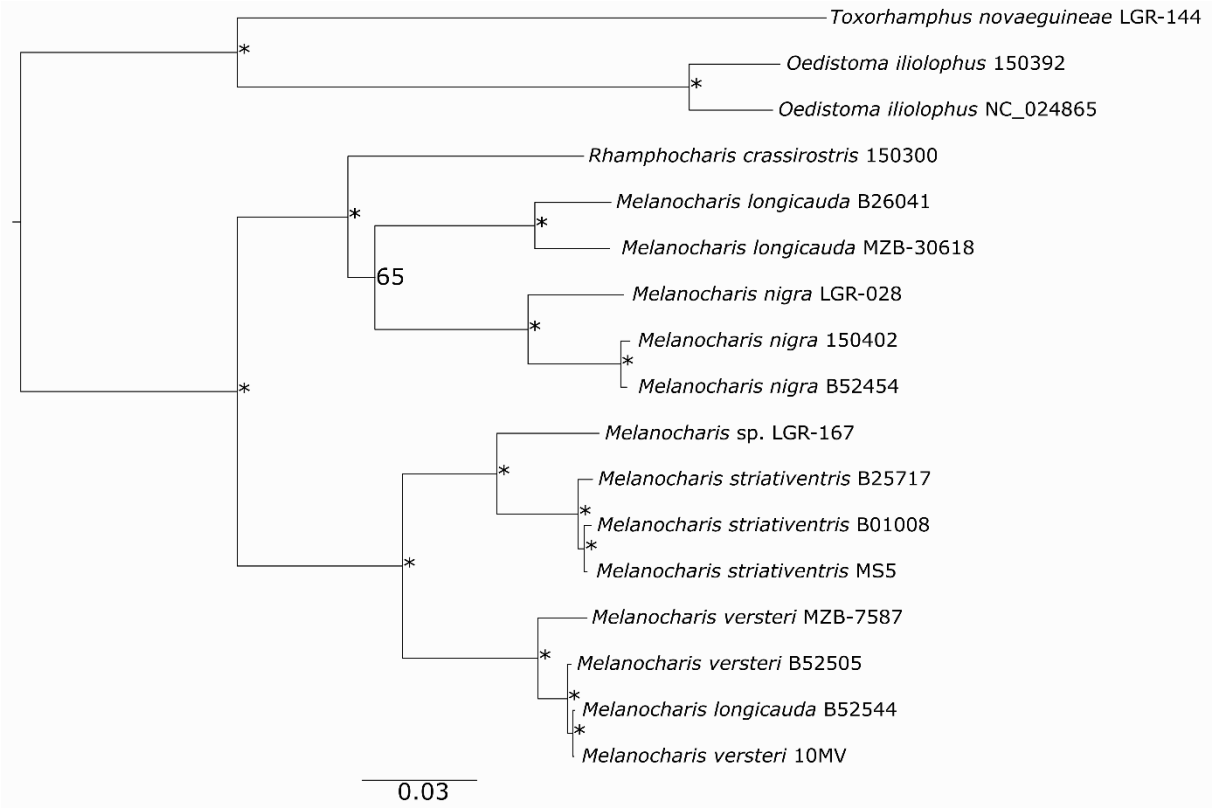


FIGURE 27: PHYLOGENETIC TREE OBTAINED THROUGH MAXIMUM-LIKELIHOOD ANALYSIS WITH A GTR+I+G MODEL AND 100 BOOTSTRAPS. VALUES AT NODES INDICATE BOOTSTRAP SUPPORT AND STARS ARE EQUIVALENT OF 100.

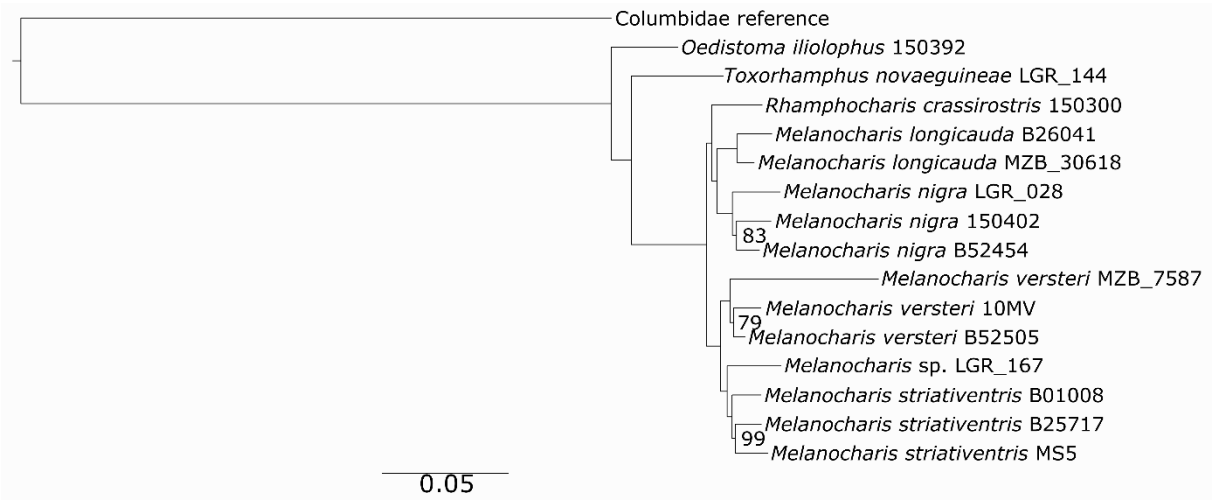


FIGURE 28: MAXIMUM-LIKELIHOOD PHYLOGENETIC TREE OF THE MELANOCHARITIDAE FAMILY, RECONSTRUCTED FROM NUCLEAR CONSERVED LOCI WITH A GTR+I+G MODEL. NODE SUPPORT ARE SHOWN ONLY WHEN DIFFERENT THAN 100.

2.3.Clarifying the speciation process at the population scale

As already explained in the introduction, speciation can occur in different contexts and some methods have been developed to clarify the exact underlying process. Both population genomics and coalescent analysis from a few genomes may help to test different scenario of speciation in New Guinea and surrounding archipelagos.

2.3.1. Population genomic analyses

At the population scale, using genomic data from a few individuals/populations of the species studied, it is now possible to compare different scenario of population isolation. Two major advances were recently proposed.

- The first one, by Beaumont et al. (2002), is the Approximate Bayesian Computation. Its objective was to improve the computational efficiency of classic Bayesian analysis for complex scenarios such as those of population genomics. To do so, rather than calculating the exact value of all parameters, they approximate values of summary statistics by comparing them to those obtained from simulations. It has the advantage that simulations, performed under the coalescent theory, can represent highly complex scenarios. However, testing multiple complex scenarios is computationally intensive and is therefore not always suitable for genome-wide datasets.
- With the same idea of reducing computationally limitation, Gutenkunst et al. (2009) proposed another method: comparing Allele Frequency Spectrum (AFS) obtained from dataset with those obtained from simulated scenarios. This method uses Maximum-Likelihood analyses rather than Bayesian ones and is based on independent SNP datasets.

These two methods allow dating speciation events (in a number of generations), to evaluate if gene flow occurred and when, if population sizes changed through time… However, it requires including several individuals per species with sufficient nuclear coverage to recover diploids data with confidence, which can be costly and time consuming, especially when working with degraded samples. The usage of a lot of missing data in very low coverage genomes for these studies would require specific simulations to validate the methods for this specific type of data. To do this, we tried to simulate trees following a known speciation process (strict isolation, ancient or recent gene flow, with constant or variable population sizes) with ms (Hudson, 2002) and seq-gen (Rambaut & Grass, 1997), and used the sequences obtained to reconstruct reads with gargammel (Renaud et al., 2017). These

reads were then used as input in the population genetics analysis. Unfortunately, scenario selected by the analysis was not the one simulated, even for the simplest model (strict isolation with constant population size) and with complete data. Digging into this unexpected result would have required too much time for a side-project and we did not go further.

2.3.2. Coalescent analysis from a single genome

Another realistic alternative at rather low cost is to use one well-covered and phased genome per species. This is possible using Pairwise Sequentially Markovian Coalescent (PSMC; (Li & Durbin, 2011). This analysis allows recovering population size history of the individual by coalescent analyses of the two copies of each of its genes. The major advantage is that this analysis does not need any *a-priori* scenario, as it is the case for the two previous methods. However, it is highly sensible to population structure (Mazet et al., 2015; Orozco-terWengel, 2016).

References

Baptista, L.F., Trail, P.W., & Horblit, H.M. (2017) Pigeons, Doves (Columbidae). *Handbook of the Birds of the World Alive* (ed. by J. del Hoyo, A. Elliott, J. Sargatal, D.A. Christie, and E. de Juana), Lynx Edicions, Barcelona.

Beaumont, M.A., Zhang, W., & Balding, D.J. (2002) Approximate Bayesian Computation in Population Genetics. *Genetics*, **162**, 2025–2035.

DuBay, S.G. & Fuldner, C.C. (2017) Bird specimens track 135 years of atmospheric black carbon and environmental policy. *Proceedings of the National Academy of Sciences*, **114**, 11321–11326.

Etienne, R.S., Valente, L.M., Phillimore, A.B., & Haegeman, B. (2017) *DAISIE: Dynamical Assembly of Islands by Speciation, Immigration and Extinction.*

Galtier, N., Nabholz, B., Glémin, S., & Hurst, G.D.D. (2009) Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular Ecology*, **18**, 4541–4550.

Gutenkunst, R.N., Hernandez, R.D., Williamson, S.H., & Bustamante, C.D. (2009) Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data. *PLoS Genetics*, **5**, e1000695.

Hudson, R.R. (2002) Generating samples under a Wright–Fisher neutral model of genetic variation. *Bioinformatics*, **18**, 337–338.

Kaplan, S. (2017) A university is eliminating its science collection — to expand a running track. *Washington Post*, https://www.washingtonpost.com/news/speaking-of-science/wp/2017/03/29/a-university-is-eliminating-its-science-collection-to-expand-a-running-track/.

Kaplan, S. & Moyer, J.W. (2015) A scientist found a bird that hadn't been seen in half a century, then killed it. Here's why. *Washington Post*, https://www.washingtonpost.com/news/morning-mix/wp/2015/10/12/a-scientist-found-a-bird-that-hadnt-been-seen-in-half-a-century-then-killed-it-heres-why/.

Li, H. & Durbin, R. (2011) Inference of human population history from individual whole-genome sequences. *Nature*, **475**, 493.

Mack, A. & Dumbacher, J. (2007) Birds of Papua. *The Ecology of Papua, Part I. The Ecology of Indonesia Series VI. Periplus, Singapore* pp. 654–688.

Mazet, O., Rodríguez, W., & Chikhi, L. (2015) Demographic inference using genetic data from a single individual: Separating population size variation from population structure. *Theoretical Population Biology*, **104**, 46–58.

McCormack, J.E., Tsai, W.L.E., & Faircloth, B.C. (2016) Sequence capture of ultraconserved elements from bird museum specimens. *Molecular Ecology Resources*, **16**, 1189–1203.

Nabholz, B., Lanfear, R., & Fuchs, J. (2016) Body mass-corrected molecular rate for bird mitochondrial DNA. *Molecular Ecology*, **25**, 4438–4449.

Orlando, L., Ginolhac, A., Zhang, G., et al. (2013) Recalibrating *Equus* evolution using the genome sequence of an early Middle Pleistocene horse. *Nature*, **499**, 74.

Orozco-terWengel, P. (2016) The devil is in the details: the effect of population structure on demographic inference. *Heredity*, **116**, 349–350.

Pääbo, S., Higuchi, R.G., & Wilson, A.C. (1989) Ancient DNA and the polymerase chain reaction. The emerging field of molecular archaeology. *The Journal of Biological Chemistry*, **264**, 9709–9712.

Pratt, T.K. & Beehler, B.M. (2015) *Birds of New Guinea.* Princeton University Press, Princeton, NJ.

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/*.

Rambaut, A. & Grass, N.C. (1997) Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Bioinformatics*, **13**, 235–238.

Renaud, G., Hanghøj, K., Willerslev, E., & Orlando, L. (2017) gargammel: a sequence simulator for ancient DNA. *Bioinformatics*, **33**, 577–579.

Toews, D.P.L. & Brelsford, A. (2012) The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, **21**, 3907–3930.

Webster, M.S. (2017) *The Extended Specimen: Emerging Frontiers in Collections-Based Ornithological Research.* CRC Press, Boca Raton, FL.

Wiens, J.J. (2003) Missing Data, Incomplete Taxa, and Phylogenetic Accuracy. *Systematic Biology*, **52**, 528–538.

Wiens, J.J. & Moen, D.S. (2008) Missing data and the accuracy of Bayesian phylogenetics. *Journal of Systematics and Evolution*, **46**, 307–314.

Zink, R.M. & Barrowclough, G.F. (2008) Mitochondrial DNA under siege in avian phylogeography. *Molecular Ecology*, **17**, 2107–2121.

AUTHOR: Jade Bruxaux

TITLE: Phylogeny and evolution of pigeons and doves (Columbidae) at different space and time scales

PHD SUPERVISORS: Christophe Thébaud and Guillaume Besnard

PLACE AND DATE OF DEFENCE: Toulouse University, 8 March 2018

ABSTRACT:

Diversification of the Columbidae family probably took place between mid-Eocene and mid-Miocene, where all extant clades were already established. Two independent dispersal events seem to have occurred between New World and Philippines / Wallacea before surrounding dispersals. These dispersions would have taken place thanks to stepping-stones movement through South Atlantic islands to Africa, and subsequently to India. These islands could also explain movements from New Guinea and Wallacea to Africa, Madagascar and Mascareignes, which are supported by winds and currents. On the other side, islands do not seem to have played a so important role in diversification, with speciation rates varying following a normal distribution, more than a model depending of insularity. Nevertheless, the highest Columbidae species-richness is reached in Philippines, Wallacea and Australasia, regions especially rich in islands. New Guinea, in particular, hosts an impressive diversity, with up to 20 species co-occurring in some places. The biggest Columbidae, belonging to the genus *Goura* (crowned pigeons), are endemic to this island. Our analyses suggest that their diversification probably started within New Guinea around 5.73 Ma. Geological events such as the cordillera orogeny probably played a major role by splitting ancestral distribution, while more recent dispersals over barriers may have also led to the settlement of isolated populations that promoted diversification. This work allowed the generation of phylogenetic hypotheses from museum samples, which will allow answering numerous questions regarding Columbidae biogeography and diversification.

KEYWORDS: Evolution, Biogeography, Genomics, Phylogeny, Columbidae

AUTEUR : Jade Bruxaux

TITRE : Phylogénie et évolution des pigeons et tourterelles (Columbidae) à différentes échelles de temps et d'espace

DIRECTEURS DE THÈSE : Christophe Thébaud et Guillaume Besnard

LIEU ET DATE DE SOUTENANCE : Université de Toulouse, le 8 mars 2018

RÉSUMÉ :

La diversification de la famille des Columbidae s'est probablement déroulée entre le milieu de l'Eocène et le milieu du Miocène, où tous les clades actuels étaient déjà apparus. Deux événements indépendants de dispersion depuis les Amériques vers les Philippines et la Wallacea semblent avoir précédé la dispersion des espèces dans les régions adjacentes. Ces dispersions à longue distance peuvent avoir eu lieu vers l'Afrique par l'intermédiaire d'îles présentes dans l'Atlantique Sud, puis vers l'Inde par des îles présentes dans l'Océan Indien. Celles-ci pourraient avoir également permis des mouvements depuis la Nouvelle-Guinée et la Wallacea vers l'Afrique, Madagascar et les Mascareignes, poussés par des vents et courants suivant cette direction. A l'inverse, les îles ne semblent pas avoir joué un rôle majeur dans le processus de diversification de la famille, avec des taux de spéciation non proportionnels au taux d'insularité. Malgré tout, la diversité maximale au sein de la famille est atteinte dans la région des Philippines, de la Wallacea et de l'Australasie, régions particulièrement riches en îles. La Nouvelle-Guinée, en particulier, héberge une diversité considérable, avec 20 espèces pouvant coexister dans certaines localités. Les columbidés les plus gros, qui appartiennent au genre *Goura* (pigeons couronnés), sont endémiques de cette île. Leur diversification a probablement commencé au sein même de l'île, il y a environ 5.73 millions d'années. Des événements géologiques tels que l'orogénie de la cordillère centrale ont probablement joué un rôle important en partageant la distribution de l'espèce ancestrale. Des dispersions par-delà des barrières géographiques établies peuvent ensuite avoir permis l'établissement de population isolées et entrainé leur diversification. Les hypothèses phylogénétiques produites ici à partir d'échantillons de musée permettront de répondre à de nombreuses questions sur la biogéographie et la diversification des Columbidae.

MOTS-CLES : Évolution, Biogéographie, Génomique, Phylogénie, Columbidae

---

DISCIPLINE ADMINISTRATIVE : Écologie, biodiversité et évolution

---

LABORATOIRE : Évolution et diversité biologique (UMR 5174 - Université Toulouse 3 Paul Sabatier, CNRS, IRD) - Université Toulouse III Paul Sabatier - Bâtiment 4R1 - 118, route de Narbonne - 31062 Toulouse cedex 9 - France