MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

Convergent evolution of flightlessness in rails (Aves: Rallidae)

A thesis presented in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Evolutionary Ecology at Massey University, Manawatū, New Zealand

Julien Gaspar

2022

Abstract

Different species can independently evolve similar phenotypic traits in response to the same environmental challenges. The resulting convergence of traits shows how environmental circumstances that apply selective pressure on the genomes of different lineages can result in analogous adaptations. A remarkable example of this evolutionary process is the secondary loss of flight in birds which repeatedly occurs in avian diversification. Flightlessness in bird species have been encountered on many oceanic islands and is interpreted as an effect of the insular condition that often provides a habitat with few or no predators, reduced competition for resources, and the opportunity to forage without flying. The rails or Rallidae are an exceptional avian family to study the evolution of flightlessness as among the 130 extant species, 30 have independently lost the ability to fly.

In this research, genomic and morphological data were integrated to compare traits of volant and flightless species in a phylogenetic context and to investigate the evolutionary processes involved in the loss of flight. First, morphological and phylogenetic data were used to compare species with and without the ability to fly in order to determine whether major phenotypic effects of the transition from volant to flightless are shared among lineages. Second, genome assemblies were generated for representatives of four rails: two volant and two flightless species. Then, a genome-wide comparison of coding regions from volant and flightless rails was performed to detect genetic regions associated with the flightless trait.

The newly assembled and annotated genomes showed differences in heterozygosity between flightless and volant species with lower heterozygosity in flightless species that probably reflects their relatively small populations. I found statistical support for similar morphological responses among unrelated flightless lineages, characterised by a shift in energy allocation from the forelimbs to the hindlimbs. Flightless birds exhibited smaller sterna and wings than volant taxa in the same family along with wider pelves and more robust femora. Phylogenetic signal tests showed that those differences were independent of phylogeny and instead demonstrated convergent morphological adaptation associated with a walking ecology. Evidence of different selective pressures between species with and without the ability to fly was detected in hundreds of genes. This included relaxed,

intensified, and positive selection in flightless species as well as evolutionary rate differences of genes in volant and flightless taxa and proteins carrying function-altering amino acid changes in the flightless rails. Genes associated with flightlessness were enriched in biological functions that aligned well with the ability to fly such as muscle, bone, limb, and heart development. However, other enriched functions were not directly linked to flying and may result from the ecological consequences of flightlessness; these included the immune system, renal functions, lipid metabolic system, cognition, and the sensory system. Finally, many genes under selective pressures in flightless species were involved in gene regulation and post-translational modification. This suggests that genetic adaptations to flightlessness are not only found in important developmental genes but also in patterns of gene expression and protein modifications. This study presents reliable methods to generate genomic data and to use it to assess the selective pressures involved in a convergent phenotype.

Acknowledgements

First, I would like to thank my supervisors who gave me the support I needed to complete this project. Gillian was always available to help and guide me during these four years, she was very patient and showed me how to use the bioinformatic tools. Steve believed in me since the internship I did in 2016 and shared with me his passion for evolutionary ecology. I could not have expected better mentorship from them and I am grateful for that.

My sincere thanks to the Phoenix Lab group, Mary, Simon, Nim, Mathieu, David, Mari, Leo, Yuta, Amanda, Michelle, Leo, Shogo, Emily, Felix, Suman and Charlotte (and everyone else who has come and gone), our weekly meetings and occasional potlucks were always a great time.

Thanks to Rob Elshire for his help with the genome assemblies and everyone in the Palmy Bioinformatic meetings for their useful inputs.

Thanks to all the staff and students in the Ecology group for their help, support and companionship. Special thanks to Mathieu for all we have been through during 6 years of friendship, to Nim, the most joyful and caring person I know, to David for all the help with my project and our conversation in the office and to Caio because he is doing well and I hope he will never change.

To my flatmates, lunch buddies and party enthusiasts friends during these four years, Ricky, Donna, Dillon, Aniek, Merce, Joan, Carlos, Lucas, Claire, Micaiah, Paul, Danny, Daniel, Sarah, Dimitris, Rayén, thank you for being there. I will miss our trips, games, karaoke's, hikes, drinks, jokes and movie nights.

Finally, I would like to thank my family and my friends who sent me love and support from the other side of the world.

Table of Contents

List of figures

Chapter One

Chapter Two

Chapter Three

Chapter Four

Chapter Five

Chapter Six

List of tables

Chapter One

Chapter Two

Chapter Three

Chapter Four

Chapter Five

Chapter Six

List of abbreviations

AU test: approximately unbiased test

CDS: coding sequence

CMD: codons with multiple differences

d_N: relative rate of nonsynonymous substitutions

d_S: relative rate of synonymous substitutions

FDR: false discovery rate

GO: gene ontology

ILS: incomplete lineage sorting

L50: the smallest number of scaffolds whose total length makes up half of the genome size

LR: likelihood ratio

LRT: likelihood ratio tests

ML: maximum likelihood

MNM: multinucleotide mutations

N50: the sequence length of the shortest scaffold at 50% of the total genome length

PCA: principal component analysis

RER: relative evolution rate

ω: d_N/d_S

Chapter One

General Introduction

This research contributes to a growing body of knowledge in molecular ecology through reasoning, question/hypothesis evaluation, and analysis of evolutionary processes. My PhD thesis is divided into six chapters including a *General Introduction*, four research chapters and a *General Discussion* to achieve an integrative structure and to explain the evolution of a set of phenotypic traits. The reader should find in this thesis an opportunity to understand the morphological and molecular mechanisms involved in the independent evolution of a combination of characters in species from different evolutionary lineages. Below, I briefly describe evolutionary concepts and methods relevant for this study, report the current knowledge in the field of the evolution of avian flightlessness, and present the objectives of this study and a short overview of each chapter to summarise the content of my thesis.

Convergent evolution

Convergent evolution is the independent evolution of the same or similar phenotypes in phylogenetically independent lineages (Losos 2011) and has been a well-recognised concept since the emergence of modern evolutionary biology (e.g. Darwin (1859), Conway-Morris (1999)). This process produces striking examples of parallel adaptation that provide a glimpse of underlying evolutionary mechanisms. Different lineages in environments with similar selective pressures can exhibit convergent phenotypes in response to similar evolutionary circumstances (Rosenblum et al. 2014, Hao et al. 2019). These convergent phenotypic solutions can be the consequences of different levels of molecular variations such as nucleotides, amino acids, genes, pathways, and biological functions. Similarities at one level may not be linked with other levels. For instance, different mutations in the same genes can have different phenotypic outcomes (Rosenblum et al. 2010) or convergent biological functions may not be the result of the same underlying genes (Roda et al. 2013, Zhang et al. 2016). Genomic and transcriptomic analyses of adaptive convergent evolution enable us to identify genomic regions involved in phenotypic traits and further classify them according to their functions, pathways, and interaction with other regions (Hao et al. 2019).

Classic examples of convergent adaptive evolution include the capacity for flight in birds, insects bats, and pterodactyls (Cao and Jin 2020, Naish et al. 2021), echolocation in bats

and marine mammals (Liu et al. 2014), transition to the marine environment in mammals (Chikina et al. 2016), and C4 photosynthesis in plants (Christin et al. 2009).

Regarding convergent evolution, the meaning of the two terms, convergence and parallelism, have been debated for many years (Rosenblum et al. 2014, Hao et al. 2019). Two different usages have been used for these terms; one distinguishing them based on phylogenetic relatedness (parallelism being defined as an independently evolved phenotype in closely related taxa and convergence as an independently evolved phenotype in distant taxa)(Osborn 1905), and the other distinguishing them based on the underlying molecular mechanism (parallelism being defined as similarities resulting from the same molecular basis and convergence being defined as similarities resulting from different molecular basis)(Wake et al. 2011). In the present thesis, I define convergent evolution as any evolutionary process that leads to phenotypic similarities in different evolutionary taxa and do not distinguish between convergence and parallelism. Terms including phenotypic or molecular convergence are used for more clarity.

Selection in convergent phenotypes

Relaxed selection

Environmental changes can reduce or eliminate a source or intensity of selection that was formerly involved in the maintenance of a phenotypic trait or a set of associated traits; this phenomenon is termed relaxed selection (Lahti et al. 2009, Wertheim et al. 2015). The release from selective constraint can impact an organism's fitness in different ways; in the most extreme circumstances, selection might be completely removed, but more often relaxed selection would have a directional impact or simply have a weaker stabilizing impact on fitness (see Lahti et al. (2009) for a review).

Classic examples interpreted as resulting from relaxed selection include the loss of eyesight in animals living permanently in the dark (Jeffery 2005, Hinaux et al. 2013, Rétaux and Casane 2013) and altered antipredator behaviour in the absence of predators (Møller and Ibáñez-Álamo 2012, Geffroy et al. 2015, Edwards et al. 2021).

The source of selection can be eliminated or weakened for more than one species and result in similar adaptive phenotypes in each of them. In this case, the molecular basis can be investigated by identifying the genomic regions exhibiting signs of relaxed selection in species showing the convergent phenotypes. In the present research, the convergent evolutionary response to avian flightlessness was studied by identifying the genes under relaxed selection in flightless species.

Measuring selection

Synonymous and Nonsynonymous substitutions

Amino acids are specified by more than one mRNA codon, and as a consequence, not all nucleotide substitutions result in amino acid replacement. Synonymous nucleotide substitutions have no effect on the amino acid sequence of proteins whereas nonsynonymous substitutions cause amino acid changes. As natural selection operates mainly at the protein level, the relative rate of nonsynonymous (d_N) and synonymous (d_S) substitutions can be used to infer selective pressures (Kimura 1977). The ratio $\omega=d_N/d_S$ can be used to measure the balance between positive, neutral, and purifying selection. Functionally, neutral amino acid changes will be fixed at the same rate (as expected by chance) as synonymous substitutions resulting in a $\omega=1$. In the case of purifying selection, natural selective pressures constrain substitutions resulting in a relative reduction of amino acid changes. This yields a $\omega<1$ as the number of nonsynonymous substitutions is lower than the number of synonymous substitutions. Positive selection is inferred when an amino acid change provides a selective advantage. In that case, nonsynonymous substitutions are fixed at a higher rate than the background synonymous substitution resulting in a $\omega>1$ (Yang and Bielawski 2000).

The ω ratio can be combined with phylogenetic information and calculated for any codon site; the resulting branch-site models can be used to detect relaxed selection at the lineage level (Yang 2007). To study the molecular basis of a phenotype that is believed to be the product of relaxed selection, one can compare the taxa that exhibit the phenotype of interest and those that do not. Relaxed selection is characterized by a reduction in the efficiency of intensity of both purifying and positive selection (Lahti et al. 2009, Wertheim et al. 2015). This phenomenon is observed when codon sites with smaller ω values

(purifying selection, blue in Fig. 1) in the background lineages (the species that do not exhibit the trait of interest) increase toward one (neutral) in the foreground lineages (those exhibiting the trait of interest) whereas the ω values greater than one in the background lineages (positive selection, red in Fig. 1) decrease in the foreground lineages. On the other hand, if the sites under purifying selection have smaller values in the foreground lineages and the sites under positive selection have greater values in the foreground lineages the selection is considered intensified (Fig. 1).

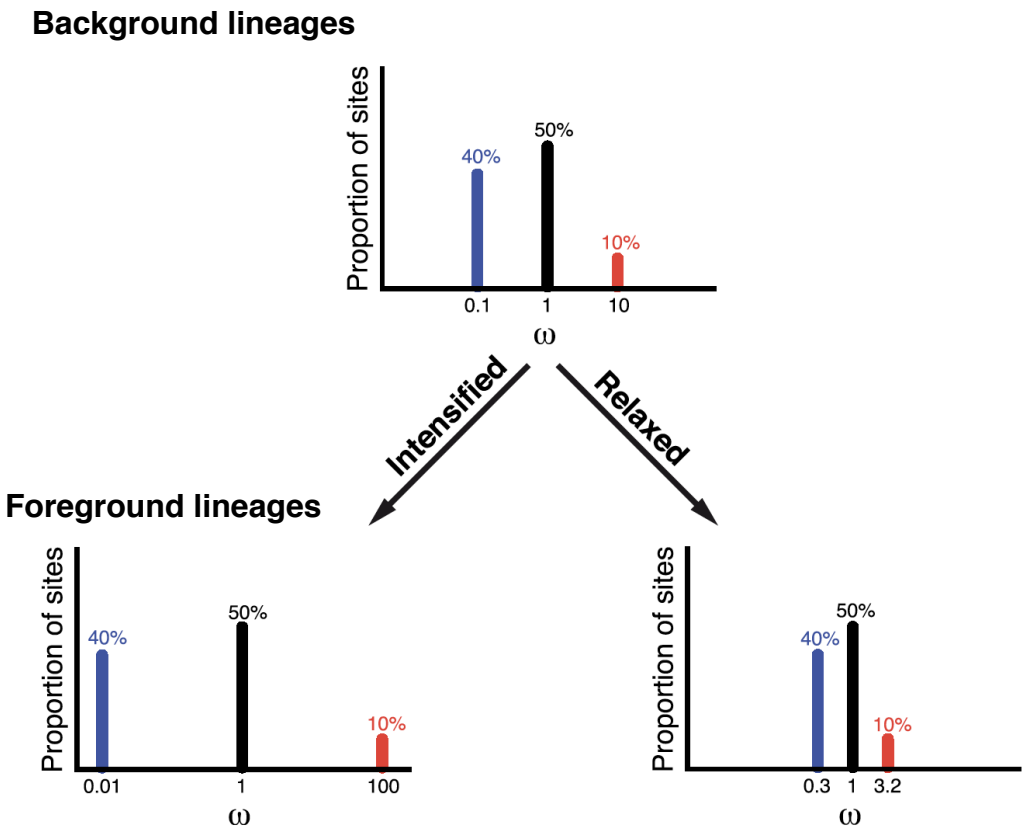


Figure 1: Relaxed and intensified selections. Sites under positive selection are shown in red and sites under purifying selection are shown in blue. Intensified selection results in all ω categories being pushed away from neutrality (ω=1) and relaxed selection results in all ω categories being pushed toward neutrality. (Redrawn after Wertheim et al. 2015).

Relative Evolutionary Rate

When lineages show adaptive convergent evolution toward the same phenotype it is expected that selection acts repeatedly upon the same genes in each lineage, leading to the signature of convergent evolutionary rate at these genes (Kowalczyk et al. 2019). Thus, the genes underlying the phenotypic change show a convergent shift in the number of

nucleotide or amino acid substitutions per unit of time caused by the reduced or increased selective constraints acting on them. Both adaptation and relaxation are predicted to result in a more rapid rate of sequence evolution (Chikina et al. 2016). Therefore, combining genome data with convergent phenotype information can be used to identify targets of convergent selection (Kowalczyk et al. 2019).

In practice, a list of gene trees (or protein trees or trees from non-coding regions) is generated and, for each tree, branch lengths (i.e. the number of substitutions) are used as a proxy for the rate of molecular evolution. If all the species showing the convergent phenotype exhibit an accelerated or decelerated evolutionary rate compared to the other species, the gene is considered a good candidate to be involved in the convergent phenotype.

In the present research, the relative evolutionary rates of volant and flightless bird lineages were compared in thousands of proteins and introns to identify genomic regions involved in avian flightlessness.

Functional impact of mutations

When organisms are released from selective constraints, it is expected that the purifying selection acting on the underlying genes will be lessened. Genes associated with that evolutionary process may therefore accumulate more amino acid replacements because nucleotide mutations will not be filtered through selective pressures. Therefore, identifying the genes carrying function-altering amino acid replacements (mutations) may help targeting the genomic regions impacted by that selective shift.

One way to assess the impact of a protein variant is to compare it with functionally homologous proteins. The assumption is that a variation that reduces similarities with the functional homologue is more likely to cause a damaging effect (Choi et al. 2012). Therefore, one can collect a set of homologous proteins related to a protein of interest, align them, and assess the functional impact of each variation based on the alignment.

In the present research, I hypothesize that transition to flightlessness results in reduced purifying selection acting on genes associated with flight which are therefore more likely

to contain function-altering amino acid changes. By assessing the functional impact of thousands of protein variants I identified genes that show evidence of reduced purifying selection in flightless species.

Convergent substitutions

Convergent phenotypes could plausibly be associated with amino acid positions that are inferred to have changed in species exhibiting the phenotype of interest (Besnard et al. 2009, Saribasak et al. 2012, Rey et al. 2018). Thus, looking for genomic regions that present sites with the same variation (or variations involving amino acids with similar biochemical properties) in species with convergent phenotypes can help to identify regions of interest.

In the present research, thousands of protein alignments were screened to find convergent amino acid substitutions that would impact the protein function in flightless species.

Challenges in measuring selection

With the rapid development of molecular biology, and the launch of multiple genome projects such as B10K (Bird 10,000 Genomes Project, https://b10k.genomics.cn/), EBP (the Earth BioGenome Project) and i5k (Sequencing Five Thousand Arthropod Genomes, http://i5k.github.io/), studies on the genetic mechanisms of convergent evolution have expanded considerably (Christin et al. 2010, Hao et al. 2019). With that expansion arose new methodological challenges to deal with the molecular basis of convergence at a large scale (Sackton and Clark 2019). For instance, Mendes et al. (2016) point out technical bias associated with gene tree discordance. Substitution rate of discordant genes may be misanalysed if considered in the context of a single, fixed species tree and lead to false positives in terms of selective pressure.

Venkat et. al (2018) showed that phylogenetic tests of adaptive evolution consider that each nucleotide substitution occurs independently. However, studies highlighted the fact that the DNA replication errors at adjacent sites happens more often than the background rate. This can result in multinucleotide mutations (MNMs) and lead to false support for positive selection in certain genes (Schrider et al. 2011, Saribasak et al. 2012).

Convergent evolution may occur through pseudogenization, an evolutionary phenomenon whereby a gene loses its function, accumulates mutations, and becomes a pseudogene. In the case of convergent relaxed selection in multiple species, a functional bias of gene loss can often be observed that is caused by the 'co-elimination' of genes. This means some genes lose their function because they are functionally linked in distinct pathways to a relaxed gene (Albalat and Cañestro 2016). This may cause confusion when targeting the genes associated with a specific environmental constraint.

Flight and flightlessness

The origin of avian flight

Living neoaves include more than 10,000 extant species around the world in many different habitats (Brusatte et al. 2015). An almost universal feature of this diversity is a reliance on aerial flight. The origins of avian flight are still debated, but it is now widely accepted that neoaves are descended from maniraptoran theropod dinosaurs and that early birds and their closest relatives were small, winged, feathered and lightweight animals (Brusatte et al. 2015). Morphology associated with flight most probably originated in this group (Xu et al. 2014). The oldest fossils associated with this lineage date from the Middle-Late Jurassic (around 165 – 150 million years ago) (Godefroit et al. 2013).

Studies conducted on morphological evolution based on fossil evidence showed that birds developed laterally oriented large and robust forelimbs along with a large extension of the sternum (the keel) and powerful pectoral muscles that make flapping flight possible (Roots 2006, Xu et al. 2014). Their bodies also became smaller and streamlined (Turner et al. 2007) and their bones and muscles changed to offer power flight for reduced weight (Roots 2006). Birds were also subject to neurological adaptation required for flight including an enlarged forebrain (Balanoff et al. 2013).

Flight is an energetically demanding process which appears to be an explanation for the relatively high metabolic rate in birds compared to reptiles and mammals (Maina 2006, Møller 2009). For instance, a bird expends around 75% more energy per day than a terrestrial mammal of similar size (Maina 2006). In order to maintain a high metabolic rate birds possess highly efficient unidirectional, constant volume lungs in which oxygen

passes through gas exchange tissues during inhalation and exhalation linked to a complex system of air sacs (Brusatte et al. 2015). Birds also have a larger heart than mammals of similar size (Grubb 1983) to support oxygen demand during peak activities. It brings oxygen to the muscles and maintains a constant temperature which is, for instance, required for high altitude flights (Scott and Milsom 2006, Scott 2011).

Flightless birds

The demands of flight significantly impact the morphological and physiological characters in birds (Elliott et al. 2013). Indeed, this ability has a substantial energetic cost which leads to many constraints in terms of body size, weight, reproduction, shape etc (McNab 1994a). Such constraints demand ecological trade-off, so in an environment where flight does not provide a significant benefit, flightlessness can be positively selected.

Transitions to flightlessness are considered rapid and irreversible (McNab 1994a, Slikas et al. 2002, Kirchman 2009) and have occurred independently in more than 15 avian families (Kirchman 2012). For instance, the ratites (ostriches, kiwis, emus…), long thought to evolve from a single flightless ancestor, turned out to be a polyphyletic group characterized by multiple independent loss of flight (Harshman et al. 2008, Phillips et al. 2010). Flightlessness has been observed in many islands species and this is interpreted as an effect of the habitat that promotes the walking ecology with little or no predators, a limited set of available resources and the ability to forage without flight (Olson 1973, McNab 1994b, 2002, Wright et al. 2016). As a consequence, birds are released from the need for flight to get food or to escape predators. Wright et al. (2016) showed that islands with reduced raptor species richness and no mammalian predators have birds with smaller flight muscles, which is consistent with selection for flightlessness occurring in this environment. The loss of flight removes many constraints in terms of weight and body size leading to significant morphological changes (Livezey 2003). For instance, many flightless birds grew larger than their volant relatives. The most prominent examples are the ostriches which stand 2.5 meters and the recently extinct South Island giant moa (*Dinornis robustus*), up to 2 meters high. Nevertheless, flightless species have a great size range. Some of them are small compared to their flying relatives like the 12.5 cm-long Inaccessible Island rail (*Atlantisia rogersi*) (Roots 2006) and the tiny ‘modest’ rail of Chatham Island (Trewick 1997).

Molecular basis of flightlessness in birds

Molecular processes associated with the ability to fly have been studied in a number of avian lineages. For instance, Machado et al. (2016) investigated the association between bone-related genes and the origin of flight using DNA sequences from birds, mammals and reptiles. They found that most of the positively selected genes in birds are linked with functional pathways relevant to power flight like bones, muscles, and heart development. Two genes associated with bone mineral density and bone remodelling were found to play an important role in flight: *AHSG* (Alpha-2-HS-glycoprotein) and *P2RX7* (P2X purinoceptor 7). Burga et. al (2017) investigated the genetic pathway to flightlessness in the Galapagos cormorant (*Phalacrocorax harrisi*). They assessed the molecular consequences of genetic differences between volant and flightless species and detected deleterious substitution in genes that affect limb development in humans when mutated. They determined that a function-altering variant in *CUX1* (Cut-Like Homeobox 1) was a strong candidate to contribute to loss of flight in cormorant as a mutation in this gene results in severe wing truncation in chicken.

Farlie et. al (2017) performed transcriptome analysis of emu (*Dromaius spp.*) limb buds to investigate the genetic mechanisms regulating the development of the vestigial emu wing. The top-ranked gene identified through this analysis was the cardiac transcription factor *NKX2.5* (NK2 Homeobox 5).

In steamer ducks (*Tachyeres spp.*), genome-wide association analysis between flying and flightless birds showed that the gene *DYRK1A* (Dual Specificity Tyrosine Phosphorylation Regulated Kinase 1A) which is involved in growth and bone morphogenesis is associated with flightlessness (Campagna et al. 2019). Sternum morphology is highly correlated with flight and Bickley and Logan (2014) demonstrated that the modulation of *TBX5* (T-Box Transcription Factor 5) expression is involved in a decrease of the sternum and wing size in the emu. Moreover, they showed that a downstream target of *TBX5*, the gene *FGF10* (Fibroblast Growth Factor 10), when mutated forms a mouse without forelimbs. *TBX5* is however fully functional in the wingless moa (Huynen et al. 2014). Sackton et al. (2019) showed that the convergent evolution of regulatory regions plays an important role in the multiple independent loss of flight in ratites (ostriches, kiwis, tinamous and emus). They

also identified thousands of conserved non-coding regions that could be associated with gene expression in the developing forelimb.

The evolution of flightlessness in birds is highly associated with energy saving (McNab 1994a). As mitochondria provide the major part of the energy of the cells, they likely play an important role in bird locomotive ability. Shen et. al (2009) used 76 mitochondrial genomes to compare weakly flying and flightless birds to rapidly flying birds. They found that "flight degenerated" birds accumulated more nonsynonymous substitutions resulting in a greater d_N/d_S ratio. Mitochondrial DNA of weakly locomotive species experiences poor purifying selection to maintain efficient energy metabolism compared to mitochondrial DNA of strongly locomotive species. This phenomenon is explained by the fact that weakly locomotive individuals are more likely to survive a deleterious mutation than strongly locomotive individuals.

Rails (Aves: Rallidae)

The rails or Rallidae are a family of birds that includes around 130 species among which over 30 are (or were for extinct species) flightless (Steadman 1995, Trewick 1997, Kirchman 2012, Garcia–R et al. 2014). They have been described as a phenotypically diverse family of primarily terrestrial birds with bilaterally compressed bodies, relatively short tails, short, rounded wings and strong, variably elongated bills (Ripley et al. 1977, Taylor 1998, Livezey 2003). Most rails are known for their stealthy movement through dense vegetation and reluctance to take flight whereas a few species (especially *Fulica*) forage unwarily on open water (Livezey 2003). The origin of rails, based on analysis of mitochondrial genome data, goes back 40 million years during the Eocene (García–R et al. 2014). Despite the terrestrial lifestyle of most rails, with flight mostly associated with predator avoidance (Trewick 2011), they show a wide distribution around the world. Fossil records showed that human colonisations caused a large number of Pleistocene and Holocene extinction within this group (Steadman 2006). The majority of the flightless birds within this family are endemic to single islands (Livezey 2003). This means that, in most cases, their ancestors had to be volant to reach these habitats as most of the islands were never connected to continental lands (Kirchman 2009).

Garcia et al. (2014) showed, based on phylogenetic analysis of 5 genes (COI, cyt-b, 16S, FGB, RAG-1), that rails are separated into eight clades: *Fulica*, *Aramides*, *Porphyrio*, *Rallina*, *Porzana*, *Laterallus*, *Gallicrex* and *Rallus*. Most of these lineages include transitions to flightlessness, especially *Rallus* in which a great number of species are flightless.

Objectives and overview

The main objective of this research is to increase the overall knowledge of the molecular basis behind the evolution of organisms. To do so, this study focuses on investigating the evolutionary processes involved in the loss of flight in rails. This family includes hundreds of bird species, and among them, several independently lost the ability to fly which makes the rails a good model to study avian flightlessness. The convergent evolution of flightlessness provides unique opportunities for studying how species are affected by that transition and how genomes encode that phenotype. I want to use genomic and morphological data to compare flightless and volant species in a phylogenetic context and detect differences in the evolutionary processes acting on them. The main hypothesis is that the convergent phenotypic evolution of flightlessness in rails results from the molecular evolution of the same genes. Although, the hypothesis of a lineage specific molecular evolution is also investigated.

To study the molecular basis of flightlessness I mostly focus on single copy protein coding genes (see Chapters 3 to 6). It is therefore not a comprehensive analysis as the coding regions represent only a small fraction of the genome and a growing number of studies are showing the role of non-coding regions in phenotypic traits (Vavouri et al. 2007, Burga et al. 2017, Sackton et al. 2019). Studying the coding regions is however a good first step to uncover the genomic regions involved in flightlessness as these regions are straight forward to retrieve from assembled genomes and robust methods to test the selection acting on them are available (Yang 2007, Wertheim et al. 2015, Kowalczyk et al. 2019). Preliminary results about convergent evolution of non-coding regions and potential future research avenues are presented in the General Discussion.

In short, this study has three main goals: (1) determine the phenotypic effects of the transition from volant to flightless, (2) generate sufficient genomic data to investigate the molecular basis of flightlessness in rails and (3) identify the genomic regions involved in the loss of flight and the selective pressure acting on these regions.

This study was designed as a thesis by publication. This means the four research chapters (Chapters 2 to 5) were written in the form of manuscripts that have been published (Chapter 2), submitted (Chapter 3) or prepared for submission in peer-review journals (Chapters 4 and 5). Therefore, some repetitions might be found, especially in the Introduction and Methods sections as some evolutionary concepts and protocols had to be described multiple times for each article to make sense on its own. The following is a short overview of each chapter to summarise the content of this thesis.

Chapter 2, *Convergent morphological responses to loss of flight in rails (Aves: Rallidae)* investigates the morphological consequences of flightlessness in the rail family where that condition has evolved repeatedly. Morphological and phylogenetic data are used to compare species with and without the ability to fly to determine major phenotypic effects of the transition from volant to flightless. I find statistical support for similar morphological responses among unrelated flightless lineages, characterised by a shift in energy allocation from the forelimbs to the hindlimbs. Flightless birds exhibit smaller sterna and wings than volant taxa in the same family along with wider pelves and more robust femora. Phylogenetic signal tests show that those differences are independent of phylogeny and instead demonstrate convergent morphological adaptation associated with a walking ecology. I also find that morphological variation was greater among flightless rails than volant ones, suggesting that relaxation of physiological demands during the transition to flightlessness frees morphological traits to evolve in response to more varied ecological opportunities.

Chapter 3, *De-novo genome assembly of four rails (Aves: Rallidae): a resource for comparative genomics* reports the protocol used to sequence, assemble, and annotate four new rail genomes: two native volant species, pūkeko *Porphyrio melanotus* and mioweka *Gallirallus philippensis,* and two endemic flightless species takahē *Porphyrio hochstetteri* and weka *Gallirallus australis*. The quality checks and comparison with existing rallid genomes show that the new assemblies are of high quality and that the annotations can be

trusted. Using the sequence read data, heterozygosity is found to be lowest in the endemic flightless species and this probably reflects their relatively small populations. These new assemblies double the number of available rallid genomes and are used in the following chapters to highlight the key genomic regions involved in rails flightlessness.

In Chapter 4, *Genetic pathway to flightlessness: evidence of convergent evolution in rails (Aves: Rallidae),* I use branch-site models to investigate the convergent evolution of flightlessness in rails. To do so, the coding regions from seven rails: three volant, *Porphyrio melanotus, Gallirallus philippensis* and *Fulica atra* and four flightless *Porphyrio hochstetteri, Gallirallus australis, Atlantisia rogersi,* and *Zapornia atra* are compared to identify genes under different selective pressure associated with flightlessness. Over 11,000 gene alignments for these seven species are used to infer relaxed, intensified, and positive selection in flightless species. Evidence of selection is detected in hundreds of genes, and I observe that genes under relaxed selection are enriched in GO (Gene Ontology) functional categories that are either directly associated with flight (muscle development and regulation of circulatory system) or with the ecological changes consequent to the loss of flight (immune response, memory).

Chapter 5, *Convergent accelerated rate of molecular evolution and functional impact of amino acid substitutions in flightless rails (Aves: Rallidae)*, reports how genomic data from rails and their distant relatives the cranes (Aves: Gruidae) is used to investigate the molecular basis of the loss of flight. Among more than 11,000 amino acid alignments, hundreds of genes exhibiting an accelerated evolutionary rate in flightless species are detected. These genes are enriched in biological functions associated with gene expression regulation or in functions that align well with the expected adaptation to flightlessness like the development of bones, limbs, muscles, and heart. Moreover, several proteins carrying function-altering mutations are found to be associated with flightlessness. These proteins are enriched in biological GO functional categories such as immune response, DNA conformation, cilium, muscle and circulatory system development and lipid metabolic process. Regarding the genetic pathway to flightlessness, there are signs of both convergent whole-gene evolution and lineage-specific pathways. This suggests that in converging towards the flightless phenotype, some genomic regions undergo similar selective pressures in all flightless species while other regions go through distinct evolutionary mechanisms in each flightless rail.

Chapter 6, *General discussion*, integrates the results from all the research chapters and reports the evolutionary mechanisms involved in the transition to flightlessness in rails at a larger scale. Morphological, phylogenetic, and genomic data are combined to describe that phenomenon from the single amino acid substitution, through the differential selective pressures in genes to the morphological changes. Genes associated with biological functions involved in flightlessness are listed and hypotheses to explain the link between these functions and the loss of flight are presented. Finally, I describe the potential changes, improvements and further analyses that could be done to increase the quality and robustness of the presented results.

Contributions

For each chapter, the research project was designed by myself (Julien Gaspar), Gillian C. Gibb and Steve A. Trewick from Massey University. I then carried out the lab work, data analysis and writing under the supervision of Gillian C. Gibb and Steve A. Trewick. Data from Bradley C. Livezey's book (Livezey 2003) were used in Chapter 2. In Chapter 3, the genome assemblies were generated using a docker container created by Gillian C. Gibb and Robert J. Elshire (https://hub.docker.com/r/gfanz/meraculous) with the help of Roger Moraga and Genomics for Aotearoa New Zealand (GFANZ, genomics.nz). Finally, in Chapter 5, analyses were run using New Zealand eScience Infrastructure (NeSI) high-performance computing facilities and consulting support (https://www.nesi.org.nz).

References

Albalat, R. and Cañestro, C. 2016. Evolution by gene loss. – Nat Rev Genet 17: 379–391.

Balanoff, A. M., Bever, G. S., Rowe, T. B. and Norell, M. A. 2013. Evolutionary origins of the avian brain. – Nature 501: 93–96.

Besnard, G., Muasya, A. M., Russier, F., Roalson, E. H., Salamin, N. and Christin, P.-A. 2009. Phylogenomics of C(4) photosynthesis in sedges (Cyperaceae): multiple appearances and genetic convergence. – Mol Biol Evol 26: 1909–1919.

Bickley, S. R. B. and Logan, M. P. O. 2014. Regulatory modulation of the T-box gene Tbx5 links development, evolution, and adaptation of the sternum. – PNAS 111: 17917–17922.

Brusatte, S. L., O'Connor, J. K. and Jarvis, E. D. 2015. The Origin and Diversification of Birds. – Current Biology 25: R888–R898.

Burga, A., Wang, W., Ben-David, E., Wolf, P. C., Ramey, A. M., Verdugo, C., Lyons, K., Parker, P. G. and Kruglyak, L. 2017. A genetic signature of the evolution of loss of flight in the Galapagos cormorant. – Science 356: eaal3345.

Campagna, L., McCracken, K. G. and Lovette, I. J. 2019. Gradual evolution towards flightlessness in steamer ducks*. – Evolution 73: 1916–1926.

Cao, T. and Jin, J.-P. 2020. Evolution of Flight Muscle Contractility and Energetic Efficiency. – Frontiers in Physiology 11: 1038.

Chikina, M., Robinson, J. D. and Clark, N. L. 2016. Hundreds of Genes Experienced Convergent Shifts in Selective Pressure in Marine Mammals. – Mol Biol Evol 33: 2182–2192.

Choi, Y., Sims, G. E., Murphy, S., Miller, J. R. and Chan, A. P. 2012. Predicting the functional effect of amino acid substitutions and indels. – PLoS One 7: e46688.

Christin, P.-A., Samaritani, E., Petitpierre, B., Salamin, N. and Besnard, G. 2009. Evolutionary Insights on C4 Photosynthetic Subtypes in Grasses from Genomics and Phylogenetics. – Genome Biology and Evolution 1: 221–230.

Christin, P.-A., Weinreich, D. M. and Besnard, G. 2010. Causes and evolutionary significance of genetic convergence. – Trends in Genetics 26: 400–405.

Darwin, C. 1859. On the Origin of the Species by Natural Selection. – Murray.

Edwards, M. C., Hoy, J. M., FitzGibbon, S. I. and Murray, P. J. 2021. Relaxed predation theory: size, sex and brains matter. – Biological Reviews 96: 153–161.

Elliott, K. H., Ricklefs, R. E., Gaston, A. J., Hatch, S. A., Speakman, J. R. and Davoren, G. K. 2013. High flight costs, but low dive costs, in auks support the biomechanical hypothesis for flightlessness in penguins. – PNAS 110: 9380–9384.

Farlie, P. G., Davidson, N. M., Baker, N. L., Raabus, M., Roeszler, K. N., Hirst, C., Major, A., Mariette, M. M., Lambert, D. M., Oshlack, A. and Smith, C. A. 2017. Co-option of the cardiac transcription factor Nkx2.5 during development of the emu wing. – Nature Communications 8: 132.

Garcia–R, J. C., Gibb, G. C. and Trewick, S. A. 2014. Eocene Diversification of Crown Group Rails (Aves: Gruiformes: Rallidae). – PLOS ONE 9: e109635.

García–R, J. C., Gibb, G. C. and Trewick, S. A. 2014. Eocene Diversification of Crown Group Rails (Aves: Gruiformes: Rallidae). – PLOS ONE 9: e109635.

Garcia-R, J. C., Gibb, G. C. and Trewick, S. A. 2014. Deep global evolutionary radiation in birds: Diversification and trait evolution in the cosmopolitan bird family Rallidae. – Molecular Phylogenetics and Evolution 81: 96–108.

Geffroy, B., Samia, D. S. M., Bessa, E. and Blumstein, D. T. 2015. How Nature-Based Tourism Might Increase Prey Vulnerability to Predators. – Trends in Ecology & Evolution 30: 755–765.

Godefroit, P., Cau, A., Dong-Yu, H., Escuillié, F., Wenhao, W. and Dyke, G. 2013. A Jurassic avialan dinosaur from China resolves the early phylogenetic history of birds. – Nature 498: 359–362.

Grubb, B. R. 1983. Allometric relations of cardiovascular function in birds. – American Journal of Physiology-Heart and Circulatory Physiology 245: H567–H572.

Hao, Y., Qu, Y., Song, G. and Lei, F. 2019. Genomic Insights into the Adaptive Convergent Evolution. – Current Genomics 20: 81–89.

Harshman, J., Braun, E. L., Braun, M. J., Huddleston, C. J., Bowie, R. C. K., Chojnowski, J. L., Hackett, S. J., Han, K.-L., Kimball, R. T., Marks, B. D., Miglia, K. J., Moore,

W. S., Reddy, S., Sheldon, F. H., Steadman, D. W., Steppan, S. J., Witt, C. C. and Yuri, T. 2008. Phylogenomic evidence for multiple losses of flight in ratite birds. – PNAS 105: 13462–13467.

Hinaux, H., Poulain, J., Silva, C. D., Noirot, C., Jeffery, W. R., Casane, D. and Rétaux, S. 2013. De Novo Sequencing of Astyanax mexicanus Surface Fish and Pachón Cavefish Transcriptomes Reveals Enrichment of Mutations in Cavefish Putative Eye Genes. – PLOS ONE 8: e53553.

Huynen, L., Suzuki, T., Ogura, T., Watanabe, Y., Millar, C. D., Hofreiter, M., Smith, C., Mirmoeini, S. and Lambert, D. M. 2014. Reconstruction and in vivo analysis of the extinct tbx5 gene from ancient wingless moa (Aves: Dinornithiformes). – BMC Evolutionary Biology 14: 75.

Jeffery, W. R. 2005. Adaptive Evolution of Eye Degeneration in the Mexican Blind Cavefish. – Journal of Heredity 96: 185–196.

Kimura, M. 1977. Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. – Nature 267: 275–276.

Kirchman, J. J. 2009. Genetic tests of rapid parallel speciation of flightless birds from an extant volant ancestor. – Biol J Linn Soc 96: 601–616.

Kirchman, J. J. 2012. Speciation of Flightless Rails on Islands: A DNA-Based Phylogeny of the Typical Rails of the Pacific. – The Auk 129: 56–69.

Kowalczyk, A., Meyer, W. K., Partha, R., Mao, W., Clark, N. L. and Chikina, M. 2019. RERconverge: an R package for associating evolutionary rates with convergent traits. – Bioinformatics 35: 4815–4817.

Lahti, D. C., Johnson, N. A., Ajie, B. C., Otto, S. P., Hendry, A. P., Blumstein, D. T., Coss, R. G., Donohue, K. and Foster, S. A. 2009. Relaxed selection in the wild. – Trends in Ecology & Evolution 24: 487–496.

Liu, Z., Qi, F.-Y., Zhou, X., Ren, H.-Q. and Shi, P. 2014. Parallel Sites Implicate Functional Convergence of the Hearing Gene Prestin among Echolocating Mammals. – Molecular Biology and Evolution 31: 2415–2424.

Livezey, B. C. 2003. Evolution of Flightlessness in Rails. – American Ornithologists' Union.

Losos, J. B. 2011. Convergence, adaptation, and constraint. – Evolution 65: 1827–1840.

Machado, J. P., Johnson, W. E., Gilbert, M. T. P., Zhang, G., Jarvis, E. D., O'Brien, S. J. and Antunes, A. 2016. Bone-associated gene evolution and the origin of flight in birds. – BMC Genomics 17: 371.

Maina, J. N. 2006. Development, structure, and function of a novel respiratory organ, the lung-air sac system of birds: to go where no other vertebrate has gone. – Biological Reviews 81: 545–579.

McNab, B. K. 1994a. Energy Conservation and the Evolution of Flightlessness in Birds. – The American Naturalist 144: 628–642.

McNab, B. K. 1994b. Resource Use and the Survival of Land and Freshwater Vertebrates on Oceanic Islands. – The American Naturalist 144: 643–660.

McNab, B. 2002. Minimizing energy expenditure facilitates vertebrate persistence on oceanic islands. – Ecology Letters 5: 693–704.

Mendes, F. K. and Hahn, M. W. 2016. Gene Tree Discordance Causes Apparent Substitution Rate Variation. – Systematic Biology 65: 711–721.

Møller, A. P. 2009. Basal metabolic rate and risk-taking behaviour in birds. – J Evol Biol 22: 2420–2429.

Møller, A. P. and Ibáñez-Álamo, J. D. 2012. Escape behaviour of birds provides evidence of predation being involved in urbanization. – Animal Behaviour 84: 341–348.

Morris, S. C. 1999. The Crucible of Creation: The Burgess Shale and the Rise of Animals. – Oxford University Press.

Naish, D., Witton, M. P. and Martin-Silverstone, E. 2021. Powered flight in hatchling pterosaurs: evidence from wing form and bone strength. – Sci Rep 11: 13130.

Olson, S. L. 1973. Evolution of the rails of the South Atlantic islands (Aves: Rallidae). – Smithsonian Contributions to Zoology.

Osborn, H. F. 1905. The Ideas and Terms of Modern Philosophical Anatomy. – Science 21: 959–961.

Phillips, M. J., Gibb, G. C., Crimp, E. A. and Penny, D. 2010. Tinamous and Moa Flock Together: Mitochondrial Genome Sequence Analysis Reveals Independent Losses of Flight among Ratites. – Syst Biol 59: 90–107.

Rétaux, S. and Casane, D. 2013. Evolution of eye development in the darkness of caves: adaptation, drift, or both? – EvoDevo 4: 26.

Rey, C., Guéguen, L., Sémon, M. and Boussau, B. 2018. Accurate Detection of Convergent Amino-Acid Evolution with PCOC. – Molecular Biology and Evolution 35: 2296–2306.

Ripley, S. D., Lansdowne, J. F. and Olson, S. L. 1977. Rails of the World: A Monograph of the Family Rallidae. – M. F. Feheley Publishers.

Roda, F., Liu, H., Wilkinson, M. J., Walter, G. M., James, M. E., Bernal, D. M., Melo, M. C., Lowe, A., Rieseberg, L. H., Prentis, P. and Ortiz-Barrientos, D. 2013. Convergence and Divergence During the Adaptation to Similar Environments by an Australian Groundsel. – Evolution 67: 2515–2529.

Roots, C. 2006. Flightless Birds. – Greenwood Publishing Group.

Rosenblum, E. B., Römpler, H., Schöneberg, T. and Hoekstra, H. E. 2010. Molecular and functional basis of phenotypic convergence in white lizards at White Sands. – PNAS 107: 2113–2117.

Rosenblum, E. B., Parent, C. E. and Brandt, E. E. 2014. The Molecular Basis of Phenotypic Convergence. – Annual Review of Ecology, Evolution, and Systematics 45: 203–226.

Sackton, T. B. and Clark, N. 2019. Convergent evolution in the genomics era: new insights and directions. – Philosophical Transactions of the Royal Society B: Biological Sciences 374: 20190102.

Sackton, T. B., Grayson, P., Cloutier, A., Hu, Z., Liu, J. S., Wheeler, N. E., Gardner, P. P., Clarke, J. A., Baker, A. J., Clamp, M. and Edwards, S. V. 2019. Convergent regulatory evolution and the origin of flightlessness in palaeognathous birds. – Science 364: 74–78.

Saribasak, H., Maul, R. W., Cao, Z., Yang, W. W., Schenten, D., Kracker, S. and Gearhart, P. J. 2012. DNA polymerase ζ generates tandem mutations in immunoglobulin variable regions. – Journal of Experimental Medicine 209: 1075–1081.

Schrider, D. R., Hourmozdi, J. N. and Hahn, M. W. 2011. Pervasive Multinucleotide Mutational Events in Eukaryotes. – Current Biology 21: 1051–1054.

Scott, G. R. 2011. Elevated performance: the unique physiology of birds that fly at high altitudes. – Journal of Experimental Biology 214: 2455–2462.

Scott, G. R. and Milsom, W. K. 2006. Flying high: A theoretical analysis of the factors limiting exercise performance in birds at altitude. – Respiratory Physiology & Neurobiology 154: 284–301.

Shen, Y.-Y., Shi, P., Sun, Y.-B. and Zhang, Y.-P. 2009. Relaxation of selective constraints on avian mitochondrial DNA following the degeneration of flight ability. – Genome Res. 19: 1760–1765.

Slikas, B., Olson, S. L. and Fleischer, R. C. 2002. Rapid, independent evolution of flightlessness in four species of Pacific Island rails (Rallidae): an analysis based on mitochondrial sequence data. – Journal of Avian Biology 33: 5–14.

Steadman, D. W. 1995. Prehistoric Extinctions of Pacific Island Birds: Biodiversity Meets Zooarchaeology. – Science 267: 1123–1131.

Steadman, D. W. 2006. Extinction and Biogeography of Tropical Pacific Birds. – University of Chicago Press.

Taylor, B. 1998. Rails: A Guide to the Rails, Crakes, Gallinules and Coots of the World. – Bloomsbury Publishing.

Trewick, S. A. 1997. Sympatric flightless rails Gallirallus dieffenbachii and G. modestus on the Chatham Islands, New Zealand; morphometrics and alternative evolutionary scenarios. – Journal of the Royal Society of New Zealand 27: 451–464.

Trewick. 2011. Vicars & Vagrants: Assembly of the New Zealand avifauna. – Australasian Science 32: 24-27.

Turner, A. H., Pol, D., Clarke, J. A., Erickson, G. M. and Norell, M. A. 2007. A Basal Dromaeosaurid and Size Evolution Preceding Avian Flight. – Science 317: 1378–1381.

Vavouri, T., Walter, K., Gilks, W. R., Lehner, B. and Elgar, G. 2007. Parallel evolution of conserved non-coding elements that target a common set of developmental regulatory genes from worms to humans. – Genome Biol 8: R15.

Venkat, A., Hahn, M. W. and Thornton, J. W. 2018. Multinucleotide mutations cause false inferences of lineage-specific positive selection. – Nature Ecology & Evolution 2: 1280–1288.

Wake, D. B., Wake, M. H. and Specht, C. D. 2011. Homoplasy: From Detecting Pattern to Determining Process and Mechanism of Evolution. – Science 331: 1032–1035.

Wertheim, J. O., Murrell, B., Smith, M. D., Kosakovsky Pond, S. L. and Scheffler, K. 2015. RELAX: Detecting Relaxed Selection in a Phylogenetic Framework. – Mol Biol Evol 32: 820–832.

Wright, N. A., Steadman, D. W. and Witt, C. C. 2016. Predictable evolution toward flightlessness in volant island birds. – PNAS 113: 4765–4770.

Xu, X., Zhou, Z., Dudley, R., Mackem, S., Chuong, C.-M., Erickson, G. M. and Varricchio, D. J. 2014. An integrative approach to understanding bird origins. – Science 346: 1253293.

Yang, Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. – Mol Biol Evol 24: 1586–1591.

Yang, Z. and Bielawski, J. P. 2000. Statistical methods for detecting molecular adaptation. – Trends in Ecology & Evolution 15: 496–503.

Zhang, Z., Xu, D., Wang, L., Hao, J., Wang, J., Zhou, X., Wang, W., Qiu, Q., Huang, X., Zhou, J., Long, R., Zhao, F. and Shi, P. 2016. Convergent Evolution of Rumen Microbiomes in High-Altitude Mammals. – Curr Biol 26: 1873–1879.

Chapter Two

Convergent morphological responses to loss of flight in rails (Aves: Rallidae)

Abstract

The physiological demands of flight exert strong selection pressure on avian morphology. It is therefore expected that the evolutionary loss of flight capacity involves profound changes in phenotypic traits. Here, we investigated the morphological consequences of flightlessness in a bird family where that condition has evolved repeatedly. The Rallidae include more than 130 recognised species of which over 30 are flightless. Morphological and phylogenetic data were used to compare species with and without the ability to fly in order to determine major phenotypic effects of the transition from volant to flightless. We found statistical support for similar morphological responses among unrelated flightless lineages, characterised by a shift in energy allocation from the forelimbs to the hindlimbs. Flightless birds exhibited smaller sterna and wings than volant taxa in the same family along with wider pelves and more robust femora. Phylogenetic signal tests showed that those differences were independent of phylogeny and instead demonstrated convergent morphological adaptation associated with the walking ecology. We found too that morphological variation was greater among flightless rails than among volant ones, suggesting that relaxation of physiological demands during the transition to flightlessness frees morphological traits to evolve in response to more varied ecological opportunities.

Keywords: flightlessness, Rallidae, convergent evolution, island ecology

Note for the reader

This chapter was published in 2020 in Ecology and Evolution.

Reference: Gaspar, J., Gibb, G. C. and Trewick, S. A. 2020. Convergent morphological responses to loss of flight in rails (Aves: Rallidae). - Ecology and Evolution 10: 6186–6207.

Introduction

Living neoaves include more than 10,000 extant species around the world in many different habitats (Brusatte et al. 2015). An almost universal feature of this diversity is a reliance on aerial flight. Studies of morphological evolution based on fossil evidence showed that birds developed laterally wide and robust oriented forelimbs along with a large extension of the sternum called a keel and powerful pectoral muscles in order to make flapping flight possible (Roots 2006, Xu et al. 2014). Their bodies also became smaller and streamlined (Turner et al. 2007) and their bones and muscles evolved to generate powered flight for a reduced weight (Roots 2006). Flight is energetically demanding which appears to be one of the reasons for the relatively high metabolic rate in birds compared to reptiles and mammals (Maina 2006, Møller 2009). For instance, a bird expends around 75% more energy during one day than a terrestrial mammal of similar size (Maina 2006).

Flightlessness in birds

Flight demands significantly impact the morphological and physiological characters in birds (Elliott et al. 2013). Indeed, this ability has a substantial energetic cost which leads to many constraints in terms of body size, weight, reproduction, shape etc. (McNab 1994). Such constraints exert intense ecological trade-offs (Ricklefs 1973, Lighthill 1975, Rayner 1988, Alexander 1998). Therefore, flightlessness can be positively selected in an environment where flight does not provide a significant benefit. This can lead to conservation issues if the habitat changes swiftly and the flight is required again. For example, if predators are introduced, flightless birds could be unable to avoid them.

Transitions to flightlessness are considered rapid and irreversible (McNab 1994, Slikas et al. 2002, Kirchman 2009) and have occurred independently in more than 20 avian

families (Roff 1994). A notable example is the ratites (including ostriches, kiwi and emus), a polyphyletic group characterized by multiple independent loss of flight (Harshman et al. 2008, Phillips et al. 2010).

Flightlessness has been observed in many island species and is interpreted as an effect of the insular conditions which often provides a habitat with few or no predators and limited competition for resources (McNab 1994). Flightlessness evolves most frequently in island birds that belong to lineages for which flight is not essential for foraging, and are released from the need to escape predators (Olson 1973, McNab 1994). On islands with reduced raptor species richness and no mammalian predators, birds evolve smaller flight muscles, consistent with selection for flightlessness (Wright et al. 2016). The loss of flight removes many constraints in terms of weight and body size leading to significant morphological changes (Livezey 2003). For instance, many flightless birds are larger than their volant relatives (Roots 2006). The most prominent examples are the ostrich which stands 2.5 metres tall, and the recently extinct 2 metres high South Island giant moa (*Dinornis robustus*). Nevertheless, flightless species have a great size range. Some of them are small compared to their flying relatives like the 12.5 cm-long Inaccessible Island rail (*Atlantisia rogersi*) (Roots 2006).

The rails (Rallidae family) diversified during the Eocene around 40 million years ago (Garcia–R et al., 2014b) and include around 130 species among which over 30 are (or were, for recently extinct species) flightless (Steadman 1995, Kirchman 2012, Garcia-R et al., 2014a). Although rails have a terrestrial lifestyle (Taylor 1998) some lineages have a tendency to colonize oceanic islands (Olson 1973, Ripley et al. 1977) resulting in a wide representation around the world. Fossil records show that extensive late Quaternary extinction within this group resulted from the human colonisation of islands (Steadman 2006). The majority of the flightless birds within this family are endemic to single islands. This implies that, in most cases, their ancestors had to be volant to reach this habitat as most of the islands were never connected to continental landmasses (Trewick 1996, 1997a, b).

Qualitative and morphometric analyses of volant and flightless rails suggest that the transition to flightlessness often involves some common traits, but the phylogenetic hypothesis used to examine that evolutionary process relied on many of the same

morphological characters (Livezey, 2003). We now know that the morphological phenogram (Livezey, 2003) poorly represents many evolutionary relationships within the family, possibly reflecting morphological convergence associated with flightlessness (Garcia-R et al., 2014a).

Phylogenetic analyses based on 5 genes (three mitochondrial and two nuclear) show that rails are separated into eight clades: *Fulica, Aramides*, *Porphyrio, Rallina, Porzana, Laterallus, Gallicrex* and *Rallus* (Garcia-R et al., 2014a). Most of them include cases of transition to flightlessness, especially in the *Rallus* one in which a great number of birds are flightless. Here we use a modern and independent molecular phylogenetic hypothesis along with morphological data of over 90 rail species to investigate the convergent evolution of flightlessness in that bird family.

Methods

Datasets

Morphological data

We assembled a matrix that contains 10 morphological traits for 90 species (Appendix C2 Table 1) including extant taxa and those that went extinct after they were first described (Livezey, 2003). The selected traits were among the most commonly used in the literature associated with the morphological differences between volant and flightless birds (Livezey 1992, Trewick 1997b, Cubo and Arthur 2001, Roots 2006, Lambertz and Perry 2015). These data were supplemented by the standard body lengths of rails reported in the Handbook of the Birds of the World Alive Online (del Hoyo et al. 2015). Mean metric values were used when data from different individuals, or a range of values, were available. The number of missing values in the full dataset was close to 32%.

The taxonomy used in this study follows the “Clements Checklist 2018” (Clements et al. 2018), so some of the names presented in Livezey (2003) have been modified accordingly.

Each species was characterised as a volant or flightless species according to Taylor (1998) and Garcia-R et al. (2014a). Other information including the distribution and habitat was added to the dataset based on Garcia-R et al. (2014a). We also generated a subset of the data that included only the species for which both morphological and phylogenetic information were available (Appendix C2 Table 2). This subset included 52 species and 11 morphological traits: body length, wing length (chord of the flattened wing), body mass, cranial length, cranial depth, cranial width, sternum length (the length of the extension of the sternum called keel or carina), sternum depth (perpendicular depth of the keel), pelvis width (interacetabular width), femur length and femur width (the width of femoral head or caput) and contained only 11% missing values.

We treated the purple swamphens (genus *Porphyrio*): *P. bellus, P. melanopterus, P. melanotus, P. melanotus ellioti, P. poliocephalus, P. porphyrio, P. pulverulentus, P. samoensis,* considered by Livezey (2003) as different species, as a single taxon: *Porphyrio porphyrio* (Garcia-R et al., 2014c). Mean metric values (when data were available) were used to determine *P. porhyrio* morphological data.

Molecular data

Molecular data were available for 88 rail species and 7 outgroup species. Five genetic markers were used including 3 mitochondrial genes (Cytochrome c oxidase subunit 1, Cytochrome b, 16S rRNA) and 2 nuclear genes (Fibrinogen Beta Chain, Recombination activating gene 1) from Garcia-R et. al (2014a) (NCBI accession numbers available in Appendix C2 Table 3). The number of available sequences per gene varies between 64 (FGB) and 85 (cyt-b).

Analysis

Phylogenetics

For each of the five genes (COI, cyt-b, 16S, FGB, RAG-1), the sequences were independently aligned (Geneious Alignment, free gaps, 65% similarity) using the software Geneious 11.1.4 (https://www.geneious.com) then concatenated into a single alignment (see supplementary data). The alignment was processed using PartitionFinder2 (Lanfear et al. 2017) via the Cipres Science Gateway (Miller et al.

2010). The selected partitioning scheme and associated models of molecular evolution were as follows: 16S: GTR+I+G; COI first codon positions: GTR+I+G; COI second codon positions: TVM+I+G; COI third codon positions: TIM+G; cyt-b first codon positions: TVM+I+G; cyt-b second and third codon positions: GTR+I+G; FGB7: TVM+G; RAG1 first codon positions: GTR; RAG1 second codon positions: HKY+I; RAG1 third codon positions: SYM+G. Maximum Likelihood (ML) analyses were implemented in RAxML v8.2.10 (Stamatakis 2014) via the CIPRES Science Gateway (Miller et al. 2010) with bootstrapping automatically stopped employing the majority rule criterion. The consensus tree (including 88 rails plus 7 outgroup species, Appendix C2 Figure 1) was then pruned down to the subset of 52 rail species for which morphological data was available. This 52 taxa tree was used for all downstream analyses. Discrete traits (e.g. habitat and the ability to fly) were mapped to that tree using R package *phytools* (Revell 2012). The same tree was used in association with the results of the principal component analysis (PCA, see the section below) to generate a graph of phylomorphospace depicting the projection of a phylogenetic tree within the two first dimensions of a PCA.

Statistics

Statistical analyses were performed in R (R Core Team 2014) (the script is available in supplementary data) using the following packages: *FactorMineR* (Lê S, Husson 2008), *car* (Fox and Weisberg 2018), *phytools* (Revell 2012), *ggplot2* (Wickham 2011) and *phylosignal* (Keck et al. 2016). A first principal component analysis (PCA) on 90 species (65 volant and 25 flightless) was performed to visualise the variation within the rail group and to determine the importance of each morphological trait (Appendix C2 Figure 2). This analysis revealed a high level of correlation between all themorphological traits. After detecting a significant correlation between the trait "Body length" and the first dimension of the PCA (that covers 75.6% of the variance) using a linear model ($F(1, 67)= 244.7$, $p < .000$, $R^2=0.78$), a correction was applied to the dataset by dividing each trait by the body length of the relevant species. This standardisation of the dataset allowed us to analyse the differences in the overall body shape between volant and flightless rails rather than to compare the actual size of each body part. Thus, the corrected dataset represents a ratio of each trait compared to the body length of each

species. The body mass was log-transformed as the distribution of that trait was not normally distributed.

A subset of the data for the 52 species with phylogenetic information was generated and contained a lower frequency of missing values (11% compared to 32% in the 90 species dataset). For each trait, a phylogenetic hypothesis was obtained by pruning the full phylogeny to represent only the species for which the trait values were available. The phylogenetic signal was quantified using Blomberg's K statistic (Blomberg et al. 2003), which estimates the phylogenetic signal using the morphological trait variance relative to an expectation under a Brownian motion null model of evolution. A K value less than one implies that relatives resemble each other less than would be expected under Brownian motion evolution across the candidate tree.

A PCA on the 52 species dataset was performed after replacing the remaining missing values within the matrix with the average value of the available data for each trait. The coordinates from the three first dimensions were used to evaluate the variance differences between groups. We used a Shapiro-Wilk test to determine the normality of each distribution and then performed an F test when the distribution was normal and Levene's test when it was not.

Bivariate correlation plots were then used to visualise patterns associated with flight ability. All the species for which the "Body length" value was available (75 species) were included in that analysis.

Major differences between flightless and volant species were observed in the correlations involving traits associated with flight and traits associated with walking. To investigate this phenomenon, a 52 species dataset of ratios was created by dividing the trait values from the upper part of the body (sternum depth and wing length) by the trait values from the lower part of the body (pelvis width and femur length) and body mass. Body length divided by body mass and sternum depth divided by sternum length were also investigated. T-tests were performed to compare volant and flightless birds on different trait ratios.

Binary logistic regression was performed on the 52 species dataset to evaluate the influence of each trait on the ability to fly. In order to minimize the loss of information

resulting from missing values, this analysis was performed independently for each of the 10 traits.

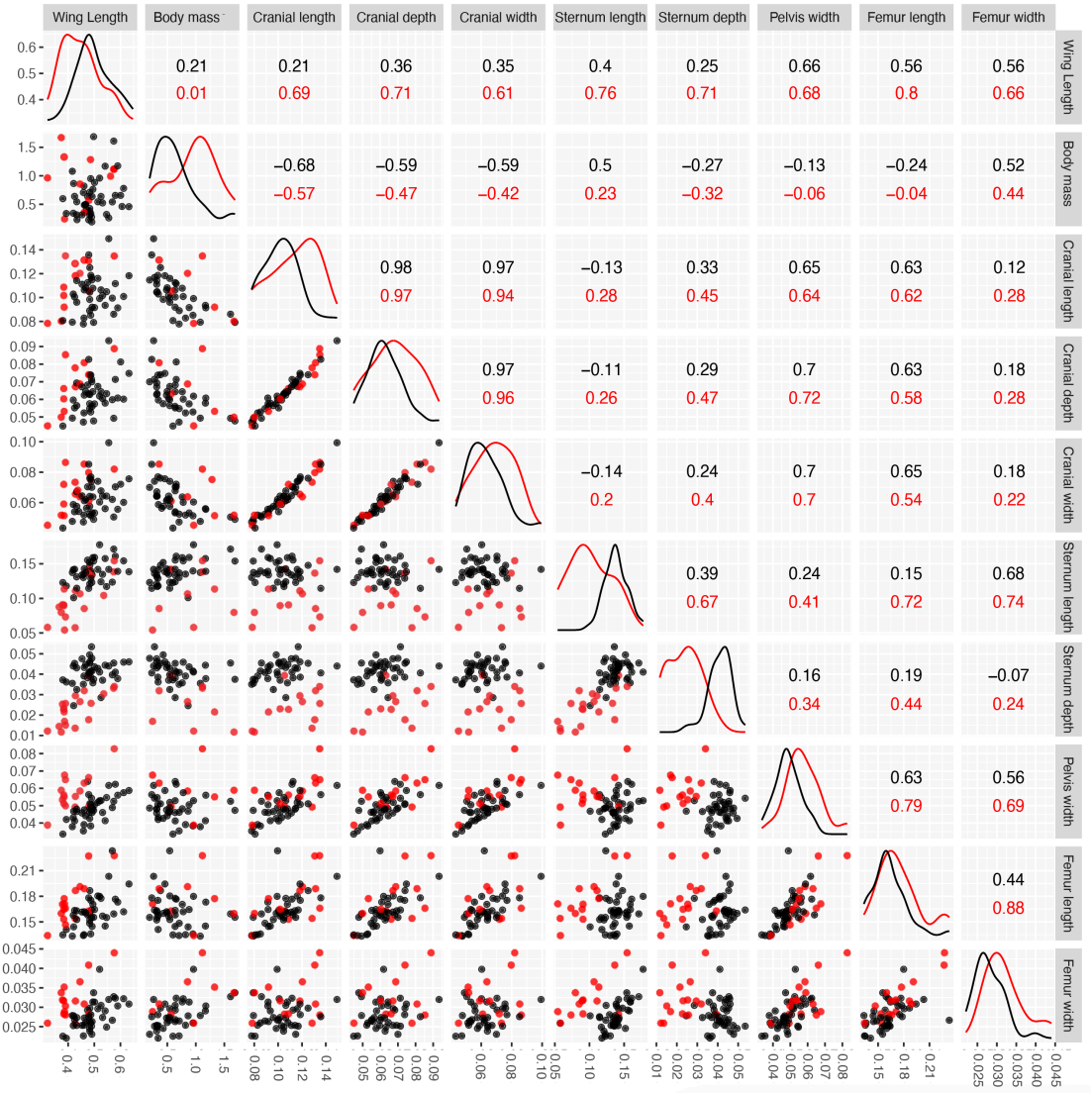


Figure 1: Scatter plot matrix of 10 traits (corrected by body length) from 75 rail species, 49 volant and 26 flightless. The upper part of the diagonal shows the coefficient of determination (r^2) for flightless (red) and volant (black) species. Two traits are considered highly correlated when the coefficient is close to one. The lower part of the diagonal shows the scatter plots for each pair of traits and the diagonal shows the distribution of the values for each group.

Data Deposition

Data available from the Dryad Digital Repository:
https://doi.org/10.5061/dryad.dz08kprsz

Results

Trait correlations

A scatter plot matrix of ten traits was used to visualise patterns associated with flight (Fig. 1), although the number of species for each correlation was not constant due to missing values. For some traits, differences between the volant (black) and the flightless (red) group were readily apparent from the density plots (Fig. 1 on the diagonal); the most obvious being body mass and sternum depth. Scatterplots of the three cranial measurements showed that they were correlated with one another despite no difference between volant and flightless taxa. Among other traits, scatterplot clustering and group differences were mostly observed where sternum depth and, to a lesser extent, sternum length were involved. Wing length, when compared with leg traits (pelvis width, femur length and width), also exhibited differences between volant and flightless groups. Broadly speaking, group differences were observed in plots involving traits associated with flight (wing length and sternum depth) and traits associated with walking (pelvis width, femur length and femur width). Finally, we observe clear group clustering in the evolution of sternum depth relative to sternum length.

Principal Component Analysis

A principal component analysis (PCA) was performed on the 52 species dataset including 14 flightless and 38 volant rails using the 10 morphological traits (Fig. 2). The two first principal components (Fig. 2) explained 41.8% and 23.4% of the variance respectively (Table 1). PC1 was mostly influenced by cranial length, depth and width contributing 21%, 21% and 20% of the variance respectively. PC2 was influenced by sternum depth (36%), sternum length (17%) and femur width (16%) (see Table 2). Volant and flightless species were clustered separately with volant taxa mostly in the upper part of the plot and most flightless species in the lower part. The distinction between these groups was therefore mainly explained by the second principal component (the vertical dimension on the plot).

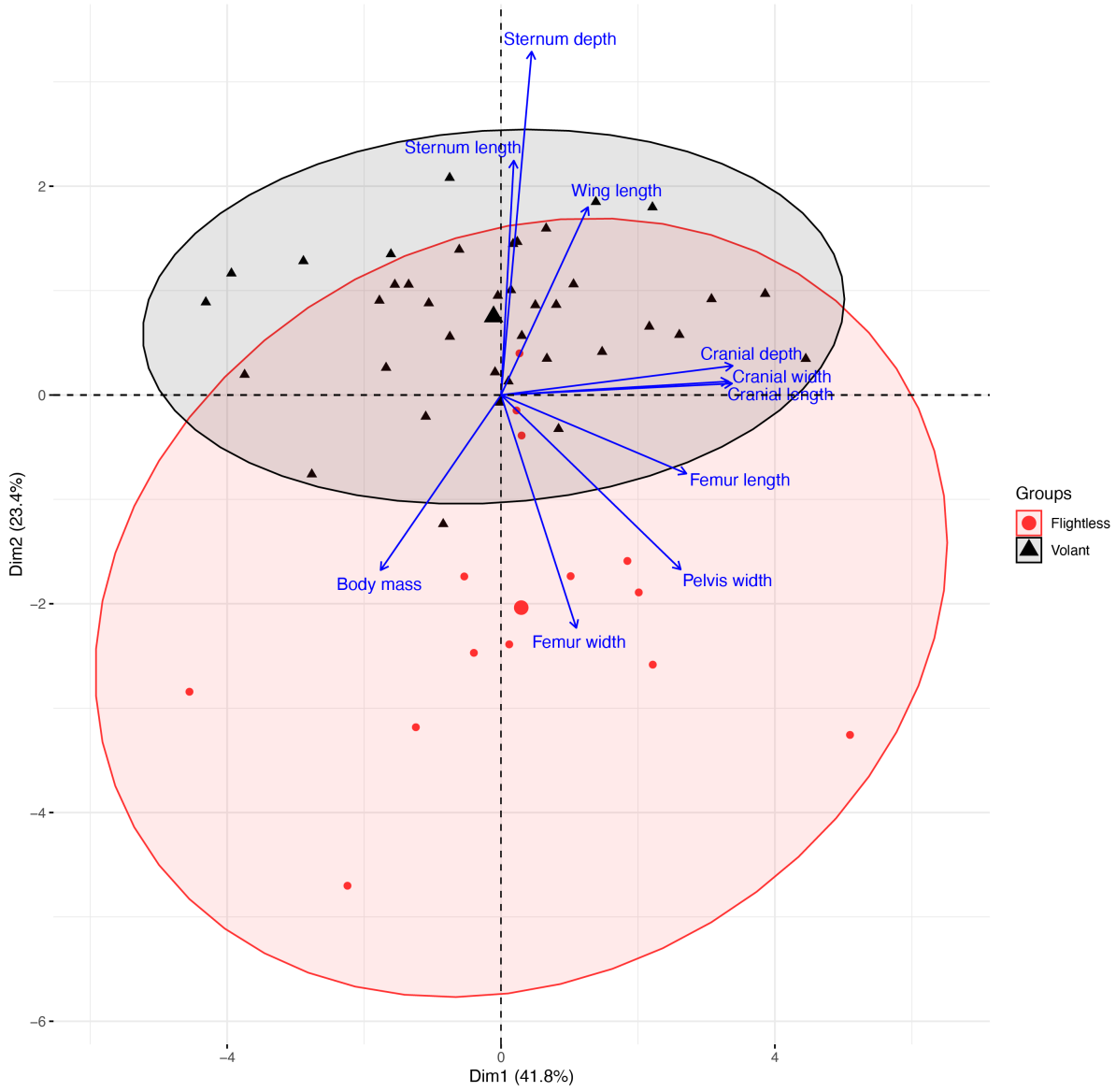


Figure 2: Principal component analysis (PCA) plot showing the two first dimensions of the multivariate variation among 52 rail species in terms of morphological traits. Vectors indicate the direction and strength of each trait contribution to the overall distribution. Black triangles represent volant species and red dots are flightless species. 95% confidence ellipses are displayed (red for flightless rails, black for volant rails), a larger ellipse is associated with a high group variance.

Generally, the ability to fly was positively correlated with the sternum depth and length and with the wing length. The flightless rails had wider femora and pelves and a heavier body. Cranial traits did not seem to be discriminant variables. Although the flightless group had fewer species, its variance and the 95% confidence ellipse appeared larger than the volant group. To test that, a variance test was run on each of the three first dimensions. Variances in volant and flightless groups were not significantly different in

dimensions 1 and 3 but in dimension 2 the variance of the flightless group was significantly higher than in the volant one (Table 1).

Table 1: Variance explained by each of the first three dimensions in the principal component analysis. The variance test implies a null hypothesis that the two groups (volant and flightless) have the same variance. An F-test is used when the distribution is normal (Dimensions 1 and 3) and Levene's test when the distribution is not normal (Dimension 2)

		Dimension 1	Dimension 2	Dimension 3
Variance		4.177	2.338	2.217
Percentage of variances explained		41.767	23.381	22.166
Cumulative percentage		41.767	65.148	87.314
Variance test	Statistic	1.26	7.16	1.084
	Significance	0.56	0.010 *	0.803

Table 2: Variance explained by each variable in the principal component analysis. "Coord." Is the coordinate that indicates (from 0 to 1) the correlation between the variable and the principal component; "Contrib." is the contribution and indicate the percentage of how much each trait explains the variance; cos2 (= Coord. * Coord.) is used to estimate the quality of the representation.

	Dimension 1			Dimension 2			Dimension 3		
Trait	Coord.	Contrib.	cos2	Coord.	Contrib.	cos2	Coord.	Contrib.	cos2
Wing length	0.352	2.974	0.124	0.5	10.673	0.25	0.631	17.952	0.398
Body mass	-0.488	5.7	0.238	-0.466	9.27	0.217	0.642	18.616	0.413
Cranial length	0.936	20.989	0.877	0.031	0.04	0.001	-0.282	3.598	0.08
Cranial depth	0.939	21.104	0.881	0.078	0.258	0.006	-0.253	2.89	0.064
Cranial width	0.922	20.357	0.85	0.036	0.056	0.001	-0.212	2.036	0.045
Sternum length	0.052	0.064	0.003	0.623	16.607	0.388	0.684	21.13	0.468
Sternum depth	0.124	0.369	0.015	0.913	35.637	0.833	0.258	3.002	0.067
Pelvis width	0.728	12.699	0.53	-0.464	9.199	0.215	0.341	5.249	0.116
Femur length	0.751	13.516	0.565	-0.209	1.866	0.044	0.362	5.916	0.131
Femur width	0.305	2.227	0.093	-0.619	16.396	0.383	0.659	19.611	0.435

Logistic regression

Logistic regressions revealed that five of the ten analysed traits had a significant effect on the ability to fly: wing length, sternum length, sternum depth, pelvis width and femur width (Table 3). The regression coefficients were positive for the wing length, the sternum length and the sternum depth but negative for pelvis width and femur width.

This means that the possibility of being volant increases when the wing length and the sternum size increase but decreases when the pelvis and femur width are large.

Table 3: Logistic regression performed on a 52 species dataset showing the relationship between 10 morphological traits and the ability to fly. Asterisks show significance of P-values; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. A p-value under 0.05 for the normality test (Shapiro-Wilk) indicates a rejection of the null hypothesis that the sample is normally distributed. Blomberg's K measures the phylogenetic signal, if it is <0.5 the variable is phylogenetically independent.

	Number of species		Logistic regression			Phylo. signal
Trait	Volant	Flightless	Coefficient	Statistic	Significance	Blomberg's K
Wing length	38	14	27.696	3.194	**0.001*** **	0.322
Body Mass (log)	38	11	-1.369	-1.581	0.114	0.493
Cranial length	33	9	-21.834	-0.870	0.384	0.348
Cranial depth	33	9	-16.101	-0.392	0.695	0.384
Cranial width	33	9	-33.426	-0.958	0.338	0.388
Sternum length	35	10	126.575	2.958	**0.003** **	0.279
Sternum depth	35	10	530.465	2.383	**0.017***	0.290
Pelvis width	35	10	-179.97	-2.469	**0.013***	0.310
Femur length	35	10	-17.808	-1.123	0.261	0.279
Femur width	35	10	-255.919	-2.589	**0.009** **	0.276

Ratio comparison

The volant group showed significantly higher ratio values in all the comparisons except two, body length divided by body mass and wing length divided by body mass (Fig. 3). This was expected as traits associated with flight should be higher in volant rails. We note that the ratio between the depth and the length of the sternum showed a significant group difference. This suggests that a single bone may give an indication regarding the flight capacity of a bird, although the ratio values between flightless and volant groups overlap. The flightless group always had a lower ratio value when a trait associated with walking was involved.

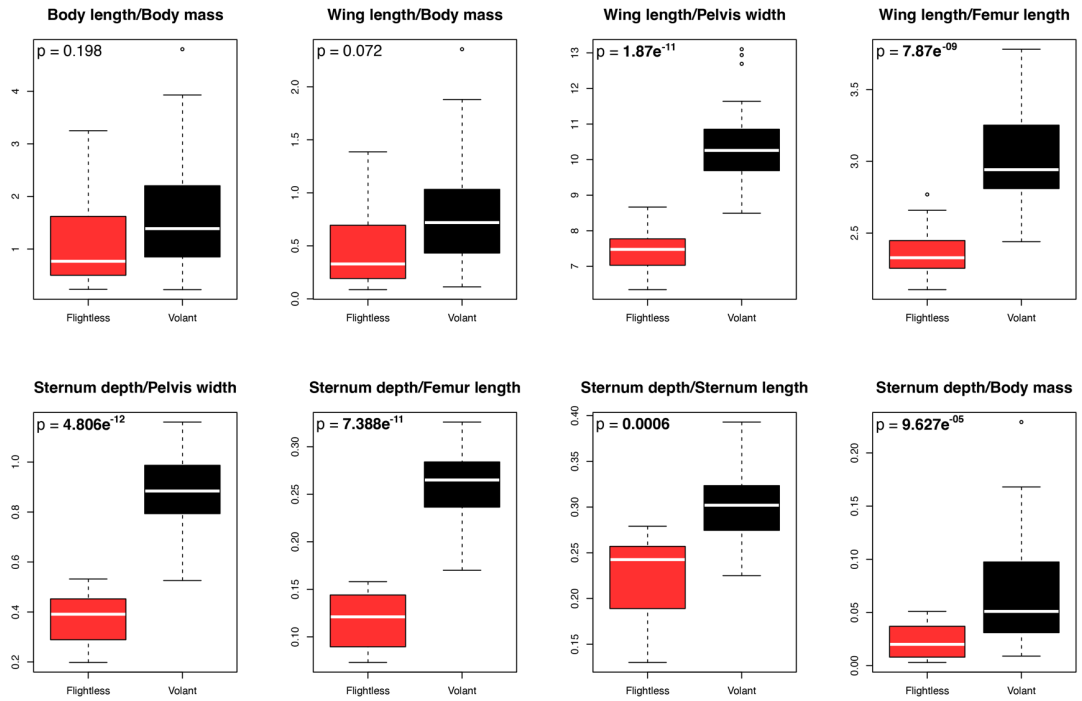


Figure 3: Comparison of ratio values for volant (black) and flightless (red) groups. Ratios were created by dividing the trait values from the upper part of the body (sternum depth and wing length) by the trait values from the lower part of the body (pelvis width and femur length) and body mass. Body length divided by body mass, wing length divided by body mass, sternum depth divided by sternum length and sternum depth divided by body mass are shown as well. P-values below 0.05 indicate a significant group difference (T-test).

Phylogenetic tree

A maximum likelihood phylogeny was generated using 5 genes and 95 birds species (88 rails and 7 birds from other families as an outgroup) (Appendix C2 Figure 1). Maximum Likelihood bootstrap support was largely consistent with the phylogeny of Garcia-R et. al (2014a).

A subset of the phylogenetic tree was obtained comprising only the species for which we had morphological data (Fig. 4). Flying ability and geographic distribution were mapped on this tree. The majority of available species (38) in the analysis were classified as volant and, of the flightless ones (14 species), many were in the *Rallus* group although *Fulica, Gallicrex* and *Porphyrio* each had one flightless species.

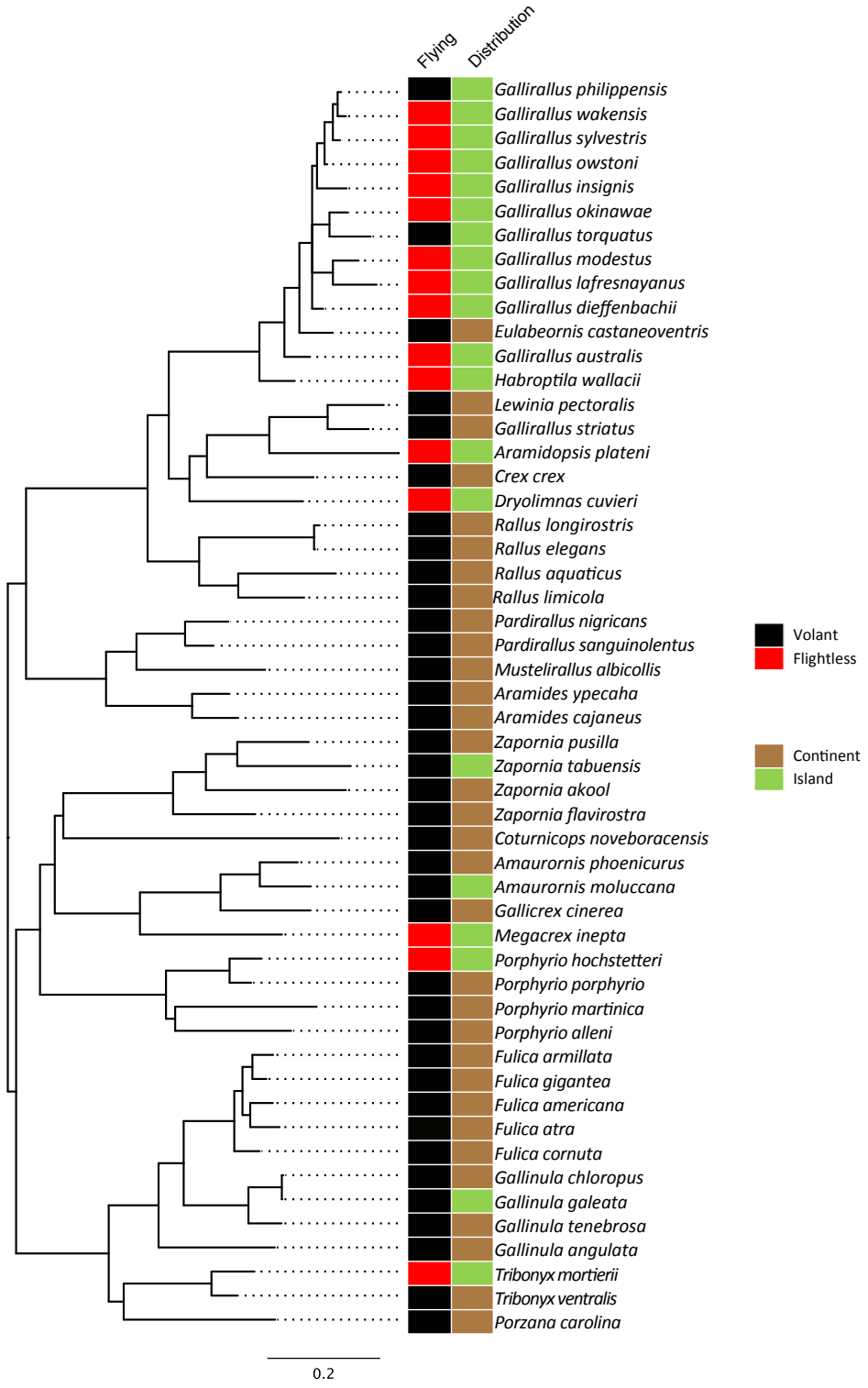


Figure 4: Flying ability and distribution (island or continent) mapped on a 52 rail species Maximum Likelihood phylogenetic tree.

When the ability to fly was compared to the habitat, a clear relationship was observed between the flightless trait and the island habitat (Fig 4). Indeed, all 14 flightless species represented in the phylogenetic tree live on islands, although islands differ in terms of the habitat they provide. For instance, *Gallirallus modestus* is endemic to the small Chatham Islands, while other species including *Porphyrio hochstetteri* and *Dryolimna cuiveri* inhabit larger continental islands, New Zealand and Madagascar respectively.

The 52 species phylogenetic tree was used to quantify the phylogenetic signal of each morphological trait using Blomberg's K (Table 3). All ten traits tested showed a K value lower than 1 suggesting phylogenetic relatives resemble each other less than expected under Brownian motion evolution along the candidate tree (Blomberg et al. 2003). These K values imply that the evolution of the morphological traits is uncorrelated with phylogeny. Data from the principal component analysis and phylogenetical analysis for 52 species were then combined to produce a phylomorphospace graph (Fig. 5). This graph suggests that the clustering observed in the morphospace (PCA result, Fig. 2) was not correlated with the phylogenetic tree as multiple branches extend between the volant and the flightless group.

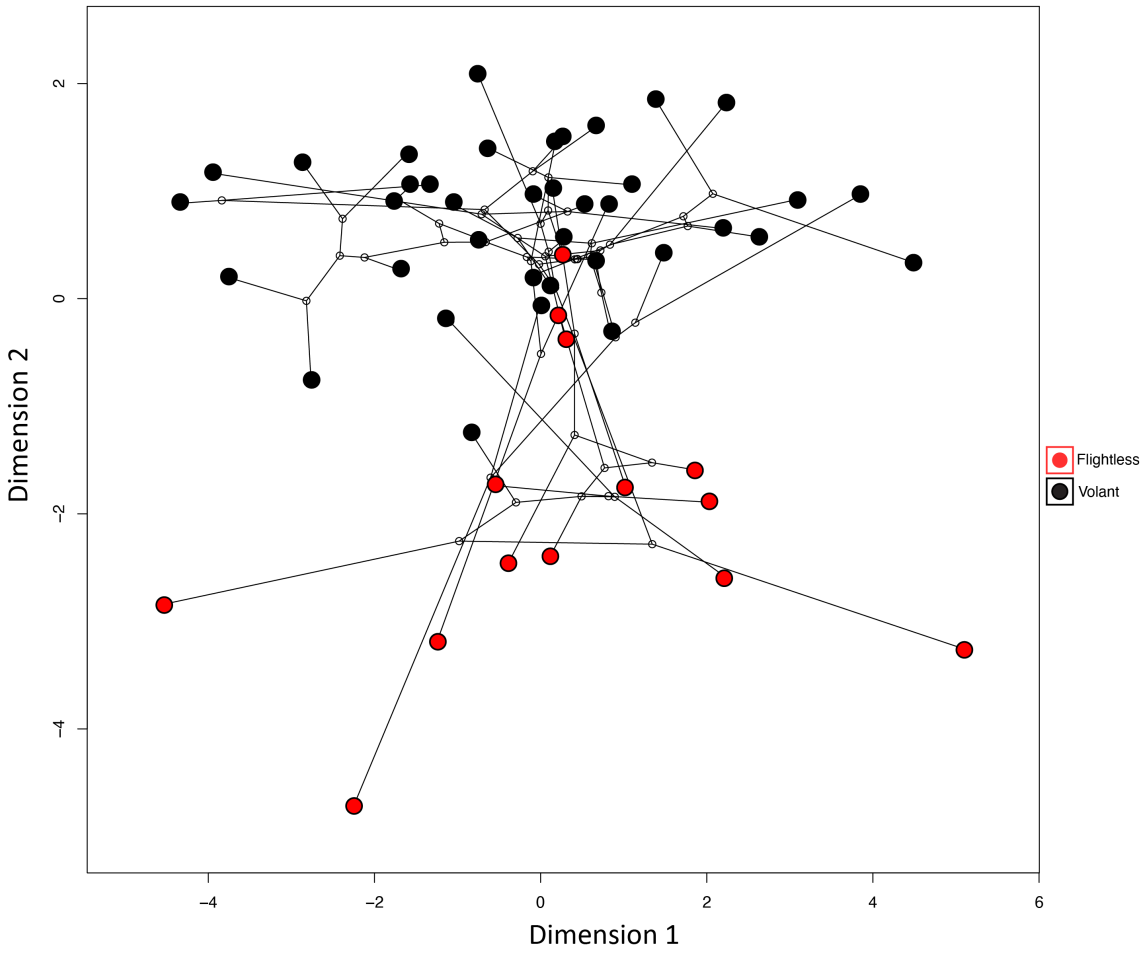


Figure 5: Phylomorphospace. Projection of the 52 rail species Maximum Likelihood phylogenetic tree within the two first dimensions of a PCA performed on 10 morphological traits. Black dots indicate volant species and red dots indicate flightless species. Black lines indicate phylogenetic relationships between species.

Discussion

Morphological differences between volant and flightless rails

The trait correlation analysis showed multiple trend differences between volant and flightless groups. Such differences were also observed in principal component analysis with clear group clustering, confirming the existence of a strong link between flight ability and phenotypic evolution in rails (Livezey, 2003). This phenomenon has now been observed in several bird families (Diamond 1991, Gussekloo and Cubo 2013, Baker et al. 2014).

Results from logistic regression suggest that the transition from volant to flightless involves a reduction of the sternum depth and length together with a shortening of the wing length. Flightless rails also exhibit wider pelves and femora which is consistent with the informal description of flightless birds as species with bigger feet, legs and leg muscles to support a heavier body (Roots 2006). The patterns revealed by the logistic regression are also observed on the PCA graph (Fig. 2) where the distinction between groups is mostly present on the second dimension (covering 23.4% of the variance). The variables influencing the variance within that component are, by order of importance, sternum depth, sternum length, femur width, wing length, body mass and pelvis width. This is consistent with the inference that, in rails, the transition to flightlessness involves an increase in body size, body mass, pelvis and cranium size as well as a reduction in lengths of wings and sternum size (Livezey, 2003). The fact that body mass appears to be associated with flightlessness in the PCA but is not statistically significant in the logistic regression is probably due to the high variance of that trait. Indeed, some flightless rails exhibit relatively high body mass (McNab 1994, Roots 2006) but others are much smaller. In New Zealand for example, the flightless takahe *Porphyrio hochsteterri* (2.7kg) is about 450 times heavier than the flightless Chatham rail *Gallirallus modestus* (60g).

No evidence of an association between the cranial traits and the ability to fly was found in the logistic regression and the PCA results. This confirms our expectation that these characters are not directly linked to flight or walking efficiency. The contrary inference of Livezey (2003) likely reflects skull size being confounded with body size and not corrected for as in our analysis. Although we note that, at an ordinal level, skull size and the ability to fly appear to be linked (Cubo and Arthur 2001, Gussekloo and Cubo 2013). Femur length also does not significantly correlate with flight capacity while femur width does which expresses the link between femur cross-sectional area and body mass (Trewick, 1996).

We find that three putatively flightless rails have morphological traits that appear closer to the volant rails (the three red dots that fell within the 95% confidence ellipse of the volant rails in Figure 2). There is limited information on the ecology of *Aramidopsis plateni, Megacrex inepta* and *Gallirallus insignis* (Fig. 2). *Gallirallus insignis* has been described as both almost flightless (del Hoyo et al. 2015) and not flightless (Gillard

1967). Based on the PCA results, and given that uncertainty and the lack of information, one can assume that they have been assigned to the wrong group. They might also be considered as part of an intermediate group including the “almost flightless”. Such a group would also include *Eulabeornis castaneoventris* (the only “volant” species that falls outside the confidence ellipse and closer to the flightless group in Figure 2) and is described as a weak flyer (Taylor 1998). We find that dividing the rail data into three categories (flying, flightless and almost flightless), or removing *Aramidopsis plateni, Megacrex inepta* and *Gallirallus insignis* do not significantly change the result (data not shown).

Similarly to other bird families (Cubo and Arthur 2001), volant rails develop their forelimbs more than their hindlimbs as they always exhibit a larger ratio when traits associated with flight (wing length, sternum depth and length) are divided by traits associated with walking (pelvis width and femur length) (Fig. 3). This reflects the different selective pressures acting on volant and flightless rails. Flight requires powerful pectoral muscles which explains the bigger wings and sterna in volant species. Regarding the legs, we hypothesize that the weight constraint associated with flight maintains the hindlimb and the pelvic girdle to a reduced size in volant species but once the flight constraint is removed, rails evolve bigger legs. In other words, the differences in leg structures between volant and flightless rails are most likely due to the release of light body mass constraint in flightless rails and not to the fact that volant species walk less.

A clear difference is observed between groups in the sternum depth/sternum length ratio (Fig. 3). Indeed, volant rails have a deeper and longer sternum allowing the insertion of powerful pectoral muscles involved in flight. The flightless rails possess a shallower sternum relative to their length. This is not surprising because sternum reduction is observed in many flightless birds as this bone is associated with flight muscles (Lambertz and Perry 2015). This means that two measurements of the same bone can give an indication regarding the flying ability of a species (Bickley and Logan 2014). This might be useful in the context of paleontological research where it is not possible to directly assess pectoral muscle mass. It has been used in the past to investigate the ability to fly of ancient species based on fossils (Howard 1964, Trewick 1997b).

The transition to flightlessness in rails shows similarities with the transition to flightlessness in other bird families which could indicate a convergent evolution on a much broader phylogenetic scale. Examination of shape differences in volant island bird populations suggests a tendency to change shape in a way that converges subtly toward the flightless form (Wright and Steadman 2012). An energy allocation from the forelimbs to the hindlimbs is observed in most islands birds (Wright et al. 2016) and is associated with smaller flight muscles (Wright and Steadman 2012). Other phenotypic changes associated with the loss of flight reflect a reduction of selective pressure, such as the tendency for flightless birds to exhibit larger pelves and skulls (Cubo and Arthur 2001, Gussekloo and Cubo 2013). Penguins and many flightless ducks do not, however, strictly follow the same morphology trends as their pectoral muscles are not significantly reduced compared to flying birds because of adaptation for aquatic 'flying' (McNab 1994).

Phylogeny and evolution

Flightlessness has evolved multiple times in rails (Fig. 4 and 5) and as most of the flightless species are endemic to islands, it means they all had volant ancestors who dispersed to these islands (Kirchman 2009, Garcia–R et al. 2014). Therefore, the loss of flight probably occurred as many times as there are flightless species in the phylogenetic tree (Fig. 4).

The *Rallus* clade includes a majority of flightless species which might be the result of the sampling (we may have more data on flightless birds than on volant ones) or the consequence of the extinction of most of the volant ancestors. It is also possible that a few flying ancestors colonised many different islands resulting in a phylogenetic cluster of several flightless island endemic species and only a few volant ones (Diamond 1991). Finally, volant lineages may not speciate so readily as the flightless ones because they are not as geographically subdivided (Trewick, 1997; Garcia-R et al., 2017). The large number of flightless species within the *Rallus* clade might potentially introduce a bias in the morphological analysis. Indeed the phenotypic trends observed in flightless species could reflect the overall body shape of the *Rallus* clade rather than convergent evolution within flightless rails. However, the phylogenetic signal test showed this is not the case.

The phylomorphospace (Fig. 5) shows that the volant-flightless pairs of closely related species can be morphologically very distant. This phenomenon emphasizes that the morphology of rails (or at least the morphological characters selected in this analysis) is more affected by the ability to fly than by the phylogeny. This is confirmed by the phylogenetic signal analysis (Table 3). Indeed, none of the morphological traits involved in this study shows a significant phylogenetic signal (Blomberg's K was always <1) which means the phenotype can be correlated with the ability to fly but not with the phylogeny. This result is in concordance with Olson (1973) who described the flightless condition as a rapid evolution that involves little genetic modifications and is without major phylogenetic significance.

Flight involves strict physical constraints in terms of body size, shape and weight (Vizcaíno and Fariña 1999, Hone et al. 2008), which implies that most volant birds in this analysis have similar shapes (though size may differ). Flightless rails on the other hand show wider overall variation (Fig. 2). This may be linked to the genetic isolation associated with the island habitat in flightless rails. Their evolution can be intensely constrained by local selective pressure leading to rapid morphogenesis in response to the environmental conditions (Garcia–R et al. 2019). For example, closely related species of *Gallirallus* on the Chatham Islands share reduction in sternum size but show very different responses in body size and relative beak length (Trewick 1997b). The variation within the flightless group also indicates that the loss of flight results in changes that can result in many different evolutionary trajectories. Freed from the constraint of flight, several different viable ecological opportunities for trait evolution may emerge. Without the physiological demands of flying a population can evolve as a function of the ecological opportunities available to them resulting in a wider range of phenotypic outcomes among species (Trewick, 1997b). At the species level, directional evolution is still involved, but when all flightless species are considered, the range of ecological outcomes results in a wide phenotypic variance. In other words, during the transition to flightlessness morphological traits can diverge in many different ways because there are more viable options as a result of less strict morphological constraints.

In conclusion, this study confirms the convergent evolution of multiple morphological traits in flightless rails. These birds exhibit short wings and small sterna as well as wide pelves and femora whereas volant species have long wings, deep sterna and smaller

femora and pelves. We find no evidence that cranial traits are correlated with the ability to fly (c.f. Livezey 2003), and this likely reflects the correlation between cranium size and the overall size of the birds which we corrected for. Using independent phylogenetic data, we show that traits of flightless rails are not subject to phylogenetic constraint but rather reflect a morphological convergent adaptation to the loss of flight.

Data availability

Supplementary material can be found on dryad https://doi.org/10.5061/dryad.dz08kprsz and in Appendices.

Acknowledgements

This study was supported by the New Zealand Marsden Fund Council from Government funding, managed by Royal Society Te Apārangi, grant MAU1601 to GCG.

References

Alexander, R. M. 1998. All-time giants: the largest animals and their problems. – Palaeontology Volume 41, Part 6: 1231–1245.

Baker, A. J., Haddrath, O., McPherson, J. D. and Cloutier, A. 2014. Genomic Support for a Moa–Tinamou Clade and Adaptive Morphological Convergence in Flightless Ratites. – Mol Biol Evol 31: 1686–1696.

Bickley, S. R. B. and Logan, M. P. O. 2014. Regulatory modulation of the T-box gene Tbx5 links development, evolution, and adaptation of the sternum. – PNAS 111: 17917–17922.

Blomberg, S. P., Garland, T. and Ives, A. R. 2003. Testing for Phylogenetic Signal in Comparative Data: Behavioral Traits Are More Labile. – Evolution 57: 717–745.

Brusatte, S. L., O'Connor, J. K. and Jarvis, E. D. 2015. The Origin and Diversification of Birds. – Current Biology 25: R888–R898.

Clements, J. F., Schulenberg, M. J. I., Roberson, D., Fredericks, T. A., Sullivan, B. L. and Wood, C. L. 2018. The eBird/Clements checklist of birds of the world: v2018.

Cubo, J. and Arthur, W. 2001. Patterns of correlated character evolution in flightless birds: a phylogenetic approach. – Evolutionary Ecology 14: 693–702.

del Hoyo, J., Elliott, A., Sargatal, J., Christie, D. A. and de Juana, E. 2015. Handbook of the Birds of the World Alive. – Lynx Edicions.

Diamond, J. 1991. A New Species of Rail from the Solomon Islands and Convergent Evolution of Insular Flightlessness. – Auk 108: 461–470.

Elliott, K. H., Ricklefs, R. E., Gaston, A. J., Hatch, S. A., Speakman, J. R. and Davoren, G. K. 2013. High flight costs, but low dive costs, in auks support the biomechanical hypothesis for flightlessness in penguins. – PNAS 110: 9380–9384.

Fox, J. and Weisberg, S. 2018. Package 'car.'

Garcia-R., J. C. and Trewick, S. A. 2014. Dispersal and speciation in purple swamphens (Rallidae: Porphyrio). – The Auk 132: 140–155.

Garcia–R, J. C., Gibb, G. C. and Trewick, S. A. 2014. Eocene Diversification of Crown Group Rails (Aves: Gruiformes: Rallidae). – PLOS ONE 9: e109635.

Garcia-R, J. C., Gibb, G. C. and Trewick, S. A. 2014. Deep global evolutionary radiation in birds: Diversification and trait evolution in the cosmopolitan bird family Rallidae. – Molecular Phylogenetics and Evolution 81: 96–108.

Garcia–R, J. C., Gonzalez-Orozco, C. E. and Trewick, S. A. 2019. Contrasting patterns of diversification in a bird family (Aves: Gruiformes: Rallidae) are revealed by

analysis of geospatial distribution of species and phylogenetic diversity. – Ecography 42: 500–510.

Gilliard, E. T. (Ernest T., Lecroy, M., Gilliard, M. and Expedition (1958-1959), G. N. B. 1967. Results of the 1958-1959 Gilliard New Britain Expedition. 4, Annotated list of birds of the Whiteman Mountains, New Britain. Bulletin of the AMNH ; v. 135, article 4. – Annotated list of birds of the Whiteman Mountains, New Britain.

Gussekloo, S. W. S. and Cubo, J. 2013. Flightlessness affects cranial morphology in birds. – Zoology 116: 75–84.

Harshman, J., Braun, E. L., Braun, M. J., Huddleston, C. J., Bowie, R. C. K., Chojnowski, J. L., Hackett, S. J., Han, K.-L., Kimball, R. T., Marks, B. D., Miglia, K. J., Moore, W. S., Reddy, S., Sheldon, F. H., Steadman, D. W., Steppan, S. J., Witt, C. C. and Yuri, T. 2008. Phylogenomic evidence for multiple losses of flight in ratite birds. – PNAS 105: 13462–13467.

Hone, D. W. E., Dyke, G. J., Haden, M. and Benton, M. J. 2008. Body size evolution in Mesozoic birds. – Journal of Evolutionary Biology 21: 618–624.

Howard, H. 1964. Further Discoveries concerning the Flightless “Diving Geese” of the Genus Chendytes. – The Condor 66: 372–376.

Keck, F., Rimet, F., Bouchez, A. and Franc, A. 2016. phylosignal: an R package to measure, test, and explore the phylogenetic signal. – Ecology and Evolution 6: 2774–2780.

Kirchman, J. J. 2009. Genetic tests of rapid parallel speciation of flightless birds from an extant volant ancestor. – Biol J Linn Soc 96: 601–616.

Kirchman, J. J. 2012. Speciation of Flightless Rails on Islands: A DNA-Based Phylogeny of the Typical Rails of the Pacific. – The Auk 129: 56–69.

Lambertz, M. and Perry, S. F. 2015. Remarks on the evolution of the avian sternum, dinosaur gastralia, and their functional significance for the respiratory apparatus. – Zoologischer Anzeiger - A Journal of Comparative Zoology 255: 80–84.

Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T. and 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.

Lê S, Husson. 2008. FactoMineR: An R Package for Multivariate Analysis. – Journal of Statistical Software 25: 1–18.

Lighthill, J. 1975. Aerodynamic Aspects of Animal Flight. – In: Wu, T. Y.-T. et al. (eds), Swimming and Flying in Nature: Volume 2. Springer US, pp. 423–491.

Livezey, B. C. 1992. Flightlessness in the Galápagos cormorant (Compsohalieus [Nannopterum] harrisi): heterochrony, giantism and specialization. – Zool J Linn Soc 105: 155–224.

Livezey, B. C. 2003. Evolution of Flightlessness in Rails. – American Ornithologists' Union.

Maina, J. N. 2006. Development, structure, and function of a novel respiratory organ, the lung-air sac system of birds: to go where no other vertebrate has gone. – Biological Reviews 81: 545–579.

McNab, B. K. 1994. Energy Conservation and the Evolution of Flightlessness in Birds. – The American Naturalist 144: 628–642.

Miller, M. A., Pfeiffer, W. and Schwartz, T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. – 2010 Gateway Computing Environments Workshop (GCE). : 1–8.

Møller, A. P. 2009. Basal metabolic rate and risk-taking behaviour in birds. – J Evol Biol 22: 2420–2429.

Olson, S. L. 1973. Evolution of the rails of the South Atlantic islands (Aves: Rallidae). – Smithsonian Contributions to Zoology.

Phillips, M. J., Gibb, G. C., Crimp, E. A. and Penny, D. 2010. Tinamous and Moa Flock Together: Mitochondrial Genome Sequence Analysis Reveals Independent Losses of Flight among Ratites. – Syst Biol 59: 90–107.

R Core Team. 2014. R: A language and environment for statistical computing. – R Foundation for Statistical Computing.

Rayner, J. M. V. 1988. Form and Function in Avian Flight. – In: Johnston, R. F. (ed), Current Ornithology, Current Ornithology. Springer US, pp. 1–66.

Revell, L. J. 2012. phytools: an R package for phylogenetic comparative biology (and other things). – Methods in Ecology and Evolution 217–223.

Ricklefs, R. E. 1973. Patterns of Growth in Birds. Ii. Growth Rate and Mode of Development. – Ibis 115: 177–201.

Ripley, S. D., Lansdowne, J. F. and Olson, S. L. 1977. Rails of the World: A Monograph of the Family Rallidae. – M. F. Feheley Publishers.

Roff, D. A. 1994. The evolution of flightlessness: Is history important? – Evol Ecol 8: 639–657.

Roots, C. 2006. Flightless Birds. – Greenwood Publishing Group.

Slikas, B., Olson, S. L. and Fleischer, R. C. 2002. Rapid, independent evolution of flightlessness in four species of Pacific Island rails (Rallidae): an analysis based on mitochondrial sequence data. – Journal of Avian Biology 33: 5–14.

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

Steadman, D. W. 1995. Prehistoric Extinctions of Pacific Island Birds: Biodiversity Meets Zooarchaeology. – Science 267: 1123–1131.

Steadman, D. W. 2006. Extinction and Biogeography of Tropical Pacific Birds. – University of Chicago Press.

Taylor, B. 1998. Rails: A Guide to the Rails, Crakes, Gallinules and Coots of the World. – Bloomsbury Publishing.

Trewick, S. A. 1996. Morphology and evolution of two takahe: flightless rails of New Zealand. – Journal of Zoology 238: 221–237.

Trewick, S. A. 1997a. Flightlessness and phylogeny amongst endemic rails (Aves: Rallidae) of the New Zealand region. – Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 352: 429–446.

Trewick, S. A. 1997b. Sympatric flightless rails Gallirallus dieffenbachii and G. modestus on the Chatham Islands, New Zealand; morphometrics and alternative evolutionary scenarios. – Journal of the Royal Society of New Zealand 27: 451–464.

Turner, A. H., Pol, D., Clarke, J. A., Erickson, G. M. and Norell, M. A. 2007. A Basal Dromaeosaurid and Size Evolution Preceding Avian Flight. – Science 317: 1378–1381.

Vizcaíno, S. F. and Fariña, R. A. 1999. On the ftight capabilities and distribution of the giant Miocene bird Argentavis magnificens (Teratornithidae). – Lethaia 32: 271–278.

Wickham, H. 2011. ggplot2. – Wiley Interdisciplinary Reviews: Computational Statistics 3: 180–185.

Wright, N. A. and Steadman, D. W. 2012. Insular avian adaptations on two Neotropical continental islands. – Journal of Biogeography 10: 1891–1899.

Wright, N. A., Steadman, D. W. and Witt, C. C. 2016. Predictable evolution toward flightlessness in volant island birds. – PNAS 113: 4765–4770.

Xu, X., Zhou, Z., Dudley, R., Mackem, S., Chuong, C.-M., Erickson, G. M. and Varricchio, D. J. 2014. An integrative approach to understanding bird origins. – Science 346: 1253293.

MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION
DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Julien Gaspar
Name/title of Primary Supervisor:	Gillian Gibb
Name of Research Output and full reference:	
Gaspar, J., Gibb, G. C. and Trewick, S. A. 2020. Convergent morphological response	
In which Chapter is the Manuscript /Published work:	Chapter Two
Please indicate:	
• The percentage of the manuscript/Published Work that was contributed by the candidate:	80%
and	
• Describe the contribution that the candidate has made to the Manuscript/Published Work:	
Julien Gaspar carried out the data analysis and writing under the supervision of Gillian C. Gibb and Steve A. Trewick	
For manuscripts intended for publication please indicate target journal:	
Candidate's Signature:	Gaspar, Julien Digitally signed by Gaspar, Julien Date: 2022.02.23 13:00:56 +13'00'
Date:	23 February 2022
Primary Supervisor's Signature:	Gillian Gibb Digitally signed by Gillian Gibb Date: 2022.02.23 12:26:18 +13'00'
Date:	23 February 2022

(This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis)

Chapter Three

De-novo genome assembly of four rails (Aves: Rallidae): a resource for comparative genomics

Abstract

The rails (Aves: Rallidae) are a phenotypically diverse family of birds that includes around 130 species and displays a wide distribution around the world. Prior to this study, the genomic data in this clade was limited and here we present annotated genome assemblies for four rails of Aotearoa New Zealand: two native volant species, pūkeko *Porphyrio melanotus* and mioweka *Gallirallus philippensis,* and two endemic flightless species takahē *Porphyrio hochstetteri* and weka *Gallirallus australis*. The quality checks and comparison with existing rallid genomes showed that the new assemblies were of high quality and that the annotations could be trusted. Using the sequence read data, heterozygosity was found to be lowest in the endemic flightless species and this probably reflects their relatively small populations. This study doubles the number of available rallid genomes and will enable future genomic studies on the evolution of this family.

Keywords: Genome assemblies, Rallidae, Flightlessness, Heterozygosity, *Porphyrio melanotus, Porphyrio hochstetteri, Gallirallus philippensis, Gallirallus australis*

Introduction

The bird family known as rails (Aves: Rallidae) has its origin during the Eocene around 40 million years ago (Garcia–R et al. 2014) and has diversified into over 130 extant species (Steadman 1995, Kirchman 2012, Garcia–R et al. 2014). They are a phenotypically diverse family of primarily terrestrial birds with relatively short wings and strong, variably elongated bills (Ripley et al. 1977, Taylor 1998, Livezey 2003). Despite the terrestrial lifestyle of the majority of the species (Taylor 1998), this bird family displays remarkable dispersal capacity resulting in broad distribution and the colonization of numerous oceanic islands (Olson 1973, Ripley et al. 1977, Garcia-R et al. 2017). At the same time, more than 30 flightless rail species are known (Steadman 1995, Kirchman 2012) and a large proportion of them are endemic to single oceanic islands, demonstrating that their ancestors had been volant (Trewick 1997a, b).

In birds, transitions to flightlessness appear to be rapid and irreversible (McNab 1994, Slikas et al. 2002) and this phenomenon applies to rails too. For instance, the flightless Aldabra rail, *Dryolimnas cuvieri aldabranus,* is found on an island that was subject to marine inundation around 127,000 YBP (Hume and Martill 2019), implying that flightlessness has evolved in this lineage within a small number of generations.

Rails are part of the order Gruiformes that includes two suborders; the Gruoidea containing, among others, the cranes (family Gruidae) and the Ralloidea that is dominated by the rails (family Rallidae) (Fain et al. 2007, Boast et al. 2019). Prior to this study, four rail genome assemblies were available on GenBank, three representing flightless species: the Henderson crake *Zapornia atra* (North, 1908), the Inaccessible Island rail *Atlantisia rogersi* (Lowe, 1923), the Okinawa rail *Gallirallus okinawae* (Yamashina & Mano, 1981) and one volant species the Eurasian coot *Fulica atra* (Linnaeus, 1758) (Fig.1).

Phylogenetic analyses of mitochondrial and nuclear genes show that rails comprise eight clades *Fulica*, *Aramides*, *Porphyrio*, *Rallina*, *Porzana*, *Laterallus*, *Gallicrex*, and *Rallus* (Garcia-R et al. 2014). The large number of flightless species as well as the fact that flightlessness evolved many times amongst extant rails provides a suitable system with which to study genomic changes associated with maintenance and loss of flight in birds.

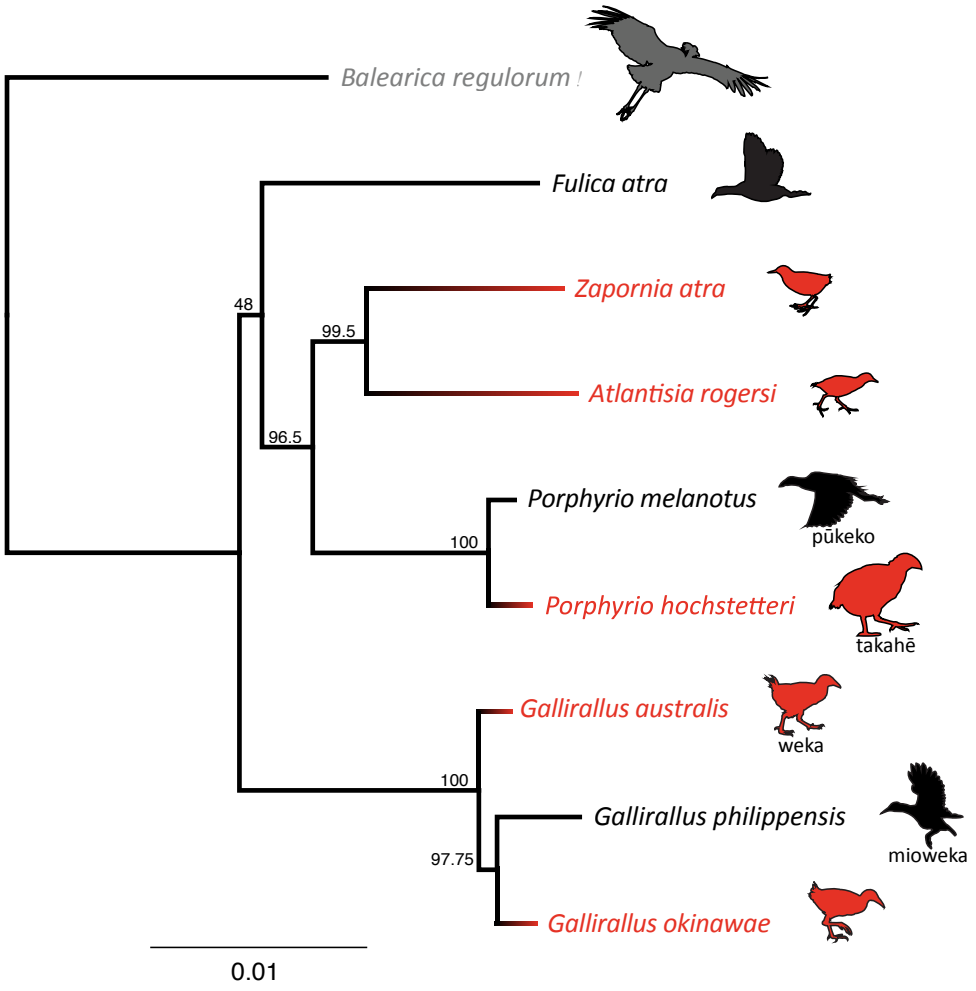


Figure 1: Maximum likelihood (RAxML V.8) phylogeny of three volant (black) and five flightless (red) rail lineages (Aves: Rallidae) based on 10 concatenated nuclear genes (*ADNP, BEGAIN, INO80D, KBTBD8, NCOA6, RHOBTB1, S1PR3, SPECC1L, ZNF618, ZNF654*) analysed with *Balearica regulorum* crane (Aves: Gruidae) (grey) as outgroup; the genes were selected from a set of universal nuclear protein-coding loci markers for avian phylogenetics (Liu et al. 2018); bootstrap supports are indicated for each node.

Despite their phylogenetic diversity, flightless rails typically exhibit smaller sterna and wings than volant taxa along with wider pelves and more robust femora (Livezey 2003, Gaspar et al. 2020). Moreover, it has been shown that these differences are independent of phylogeny and instead demonstrate convergent evolution associated with a walking ecology (Gaspar et al. 2020). Despite some research using short markers at the population level (Garcia-R. and Trewick 2014, Garcia-R et al. 2017, Trewick et al. 2017), the molecular basis underlying the convergent evolution of flightless rails remain unknown. To investigate that question, more genomic data are needed. Here we present new, high quality, annotated rail genome assemblies of four rail species; two volant, Purple

swamphen (called pūkeko in Aotearoa New Zealand) *Porphyrio melanotus* (Temminck, 1820) and buff-banded rail (also called mioweka and moho pererū) *Gallirallus philippensis* (Linnaeus, 1766), and two flightless species, takahē *Porphyrio hochstetteri* (Meyer, 1883) and weka *Gallirallus australis* (Sparrman, 1786). These four genome assemblies were generated to provide two volant-flightless pairs of closely related living species, that will enable future genomic comparisons to highlight the differences and similarities in evolutionary trends between rails with and without the ability to fly.

Methods

DNA extraction and sequencing

DNA was extracted from muscle samples of four rails sampled in Aotearoa New Zealand: *Porphyrio melanotus, Gallirallus philippensis*, *Porphyrio hochstetteri* and *Gallirallus australis*. Extraction used the Geneaid© Genomic DNA Mini Kit following the kit instructions and eluted in 100 µl. DNA quality was then verified by gel electrophoresis and quantified using Qubit 2.0 (Table. 1). Library preparation using the TruSeq Nano DNA kit and quality check were performed by The Massey Genome Service, Massey University (New Zealand) with sequencing by Novogene (Hong Kong). Libraries were sequenced on the Illumina HiSeq™ X platform generating non-overlapping 150 bp paired-end reads with an insert size of 550 bp. Fastp V0.19.4 (Chen et al. 2018) was used with default settings for paired-end data to trim the adapters as well as filter and assess the read quality.

Table 1: DNA concentration and sampling information including the location and time of collection as well as museum ID (when known)

Species	DNA concentration	Sampling	Sex	ID
Porphyrio melanotus	35.9 ng/μl	Roadkill, Turitea Valley near Palmerston North, North Island, New Zealand, within the rohe of Rangitāne o Manawatū. October 2018	Male	MUNZ12900
Porphyrio hochstetteri	5.0 ng/μl	Provided by the Department of Conservation via Massey University Veterinary Pathology. A translocated individual on Maud Island, Marlborough Sounds, New Zealand	Male	NA
Gallirallus philippensis	41.4 ng/μl	Roadkill, Whananāki estuary, Northland, North Island, New Zealand, within the rohe of Ngatiwai. Retrieved March 2011	Male	MUNZ12901
Gallirallus australis	40.0 ng/μl	Roadkill Granity, West Coast, South Island, New Zealand, within the rohe of Ngāi Tahu. Retrieved July 2012.	Male	MUNZ12767

Genome Assembly

De novo assembly was performed for each of the genomes using Meraculous (Chapman et al. 2011). Average insert size, standard deviation and average read lengths were estimated using sequence reads mapped to a nuclear gene of a close species. Following the Meraculous manual instructions, a range of k-mer sizes was analysed using KmerGenie V1.7051 (Chikhi and Medvedev 2014). The k-mer frequency histograms were reviewed and k for which the main haploid peak had a coverage of at least 30X and a distinct trough to its left that was at most 1/10 of the peak height was chosen. These were 61, 87, 61, 57 for respectively *Porphyrio melanotus, Porphyrio hochstetteri, Gallirallus philippensis* and *Gallirallus australis* (Figure 2). High heterozygosity for *G. philippensis* meant that completely optimal peak height/trough specs could not be met but the assembly was still successful. See supplementary material for full details of settings used in all Meraculous runs.

Meraculous (Chapman et al. 2011) was implemented using a docker container we created, which is publicly available at both Github and docker (https://github.com/GenomicsForAotearoaNewZealand/genomics-tools, https://hub.docker.com/r/gfanz/meraculous). The assembly was run through the Catalyst

Cloud server (https://catalystcloud.nz) using a cloud instance with 32 vCPU and 256 GB RAM. Runs took between 1 and 3 days per assembly.

Additional Genomes

We incorporated data for four additional rail genomes in this study, the Okinawa rail *Gallirallus okinawae*, GenBank assembly accession: GCA_002003005 (Ozaki et al. 2010), Henderson Crake *Zapornia atra* (formerly *Porzana atra*) GCA_013400835, Eurasian coot *Fulica atra* GCA_013372525, and Inaccessible Island rail *Atlantisia rogersi* GCA_013401215. The genome of a Grey crowned crane *Balearica regulorum* (order Gruiformes, family Gruidae; Bennett, 1834) GCA_011004875 was used as a reference for the gene annotations.

Quality assessment

Meraculous outputs were used to compare the sequence length of the shortest scaffold at 50% of the total genome length (N50) and the smallest number of scaffolds whose total length makes up half of the genome size (L50) values as well as the assembly length and the number of contigs and scaffolds. Busco v4 (Seppey et al. 2019) was implemented using a Docker (Merkel 2014) container (default parameters, mode: genome) on the genomes using the aves_odb10 dataset to assess the assembly completeness.

Genome annotation

Geneious R.11 (https://www.geneious.com) was used to extract the coding sequences (CDS) from *B. regulorum* genome (GCA_000709895) and these were filtered to retain only the longest CDS per gene where multiple annotations existed. Gmap (version 2019-09-12) (Wu and Watanabe 2005) was used to annotate the newly assembled genomes. Each assembly was first indexed using the gmap_buil function, then *B. regulorum* CDS were mapped to it with the setting -f 2 to obtain a GFF3 formatted annotation.

Extracting coding regions

During the assembly process, exons from the same gene are sometimes assembled into different scaffolds. To obtain a sequence list containing the entire coding region for each gene, the exons were extracted using Geneious R.11 and remapped to the *B. regulorum* CDS with BWA (0.7.17-r1188) using BWA-mem with the default settings (Li 2013).

To assess the size and quality of the extracted CDS for each genome they were compared to the *B. regulorum* reference. The quality (complete or partial) of coding regions retrieved was assessed using the samtools V.1.9 (Li et al. 2009) faidx tool (to obtain the length of each sequence) and a custom R script to compare the CDS sequences with the reference (see supplementary data).

Heterozygosity

Read depth, coverage and heterozygosity of the newly assembled genomes were estimated using twenty genes randomly selected (ID: *ADA*, *DHX40*, *ENPEP*, *EXOG*, *FAM196B*, *FUBP3*, *GOLGA7B*, *GRHL3*, *KCNK5*, *LEMD3*, *LOC104630315*, *LOC104633950*, *LOC104643156*, *MLNR*, *MMS19*, *PIANP*, *THOC3*, *ZCCHC2*, *ZNF410*, and *ZRANB1*) for a total length of 266,456 bp and the paired reads for each genome mapped to them in Geneious R.11 with low sensitivity/fast mapping settings. The Geneious 'Find variations/SNPs' tool in the 'Annotate & Predict' section was used with the following settings: minimum coverage of 50 and minimum variant frequency of 0.3 to locate the heterozygous sites. Heterozygosity was then estimated by dividing the number of heterozygous sites by the total length of the concatenated gene sequences. This method, despite not using the whole genome to assess the heterozygosity level of each species, generates reliable estimates that can be compared between lineages.

Results

DNA extraction and sequencing

The raw data comprised between 780 million (*G. philippensis*) and 936 million (*G. australis*) paired reads per species. Most of these were retained after the filtering and cleaning step (Table 2). Fastp generates a Phred quality score (Q score) for each of the

species that represents the ratio of bases with a probability of containing no more than 1/100 (Q20) or in 1/1000 (Q30) errors (Ewing and Green 1998, Ewing et al. 1998, Richterich 1998). These scores range between 97.37% and 98.5% for Q20 and between 93.74% and 95.23% for Q30 implying high sequencing quality for all four species.

Table 2: Fastp outputs after the sequencing of four rail species indicating the number of reads before and after filtering as well as the quality assessment

Species	Before fastp filtering	After fastp filtering			
	Total reads	Total reads	% reads conserved	Q20 bases	Q30 bases
P. melanotus	894.570034 M	881.624382 M	98.55%	97.73%	94.46%
P. hochstetteri	845.999032 M	817.395666 M	96.62%	97.37%	93.84%
G. philippensis	781.084610 M	760.059606 M	97.31%	97.39%	93.74%
G. australis	936.861886 M	917.214824 M	97.90%	98.05%	95.23%

K-mer frequency plots can be used to estimate the level of heterozygosity for each individual and by proxy each species. Indeed, k-mers from the heterozygous regions (left peak on Fig. 2) will have half the sequencing coverage (i.e., K-mer frequency) compared to the homozygous regions (right peak). The higher the left peak the higher the heterozygosity. The two volant species *G. philippensis* and *P. melanotus* exhibited high heterozygosity with the left peak being higher than the right for *G. philippensis*. A very low left peak was found for the *G. australis* data and only one peak was observed for *P. hochstetteri.* This implies a lower level of heterozygosity for both of the endemic, flightless species that have limited populations.

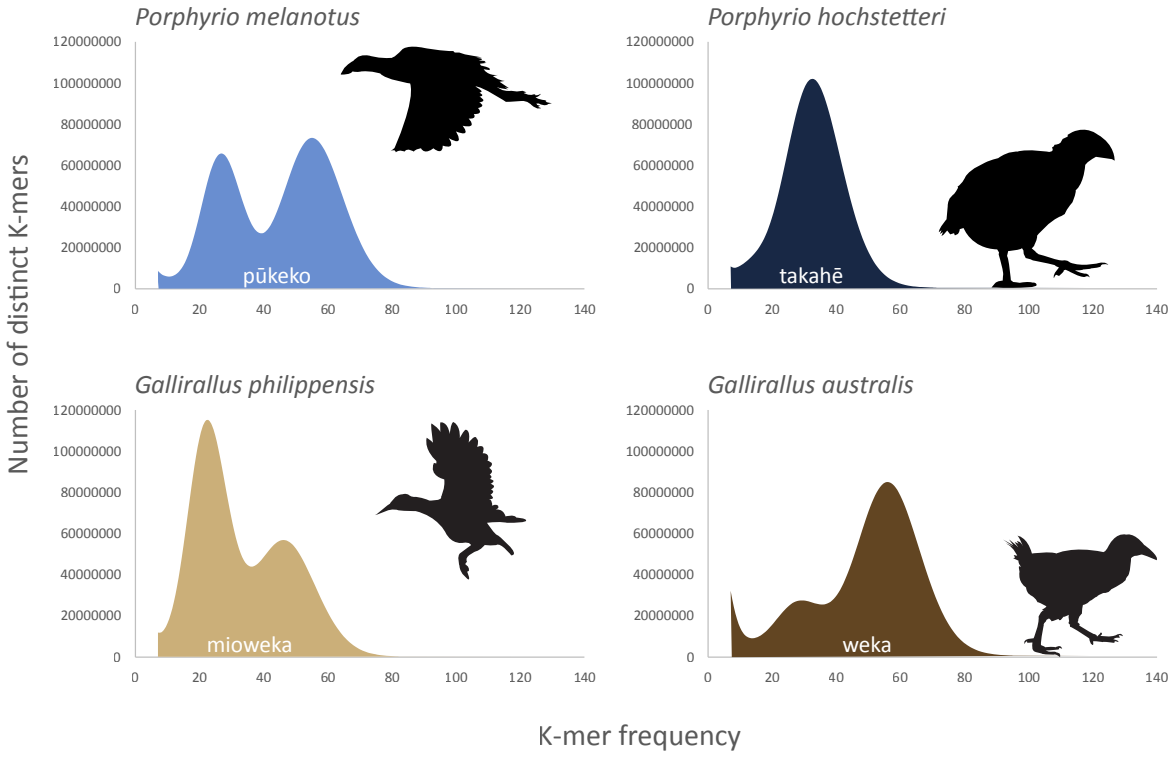


Figure 2: K-mer frequency in four rails from Aotearoa New Zealand. K-mer (nucleotide sequence of a certain length) were 57, 61, 61, 87 for *G. australis*, *G. philippensis*, *P. melanotus* and *P. hochstetteri* respectively. In each distribution, two main peaks correspond to the genomic K-mers for the heterozygous (left) and homozygous (right) parts of the genome. The single main peak of *P. hochstetteri* indicates high homozygosity. Low depth peaks corresponding to erroneous K-mer populations have been masked for clarity. Icons indicate flightless and volant species.

To compare heterozygosity between the newly assembled genomes, paired reads were mapped to a set of 20 genes for each species and the ratio of heterozygous sites divided by the total sequence length was calculated (Fig. 3). The two volant species showed a higher heterozygosity level than the two flightless species. Based on the paired reads mapping, the mean depth of coverage was calculated for each species (Table. 3) with the overall average being 96.4.

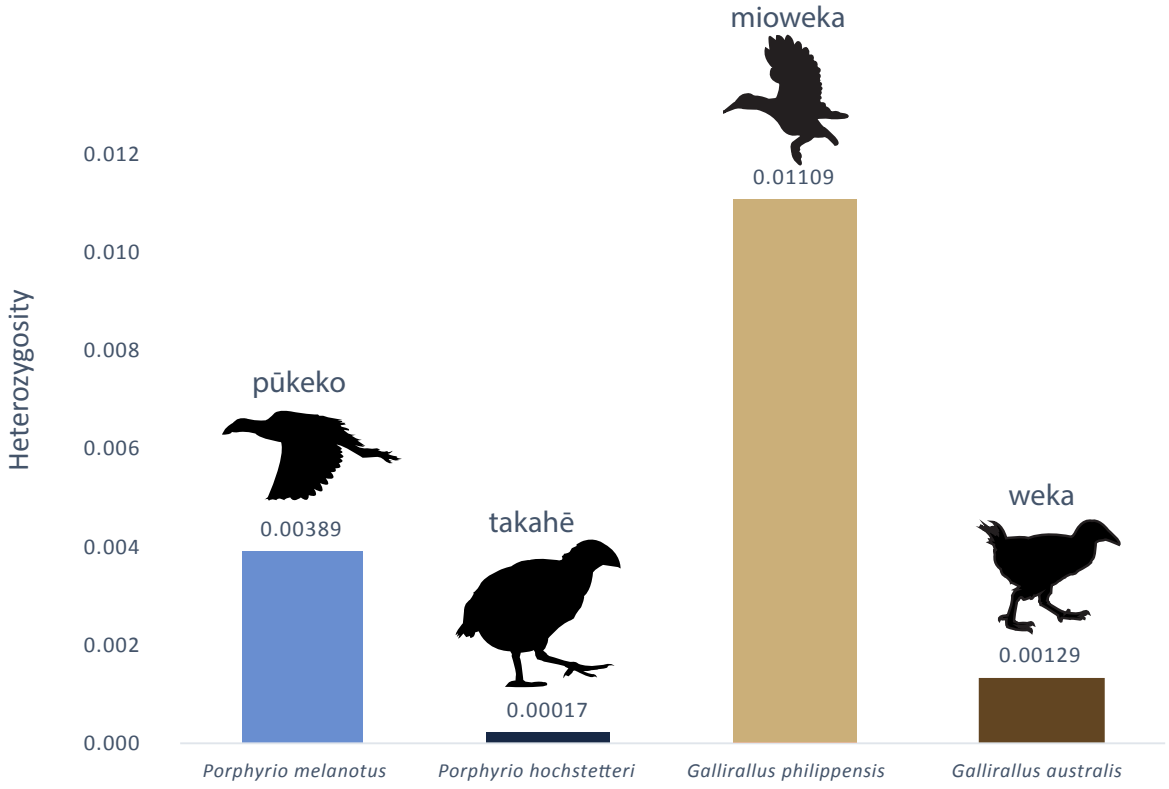


Figure 3: Average heterozygosity at 20 randomly selected genes from four newly assembled and annotated rail genomes (average total length 266,456 bp). Heterozygosity is the proportion of total nucleotide sites per individual site having two bases. Icons indicate flightless and volant species.

Genome assembly

Meraculous de novo assemblies yielded scaffold N50 between 126kb (*G. australis*) and 30kb (*P. hochstetteri*) and scaffold L50 between 2,365 (*G. australis*) and 6,047 (*G. philippensis*) (Table.2). The total genome assembly size of the four newly assembled rails differed little with a range from 1.07 Gb (*G. philippensis*) to 1.16 Gb (*G. australis*). This was similar to the previously assembled rails (between 1.11 and 1.17 Gb, see Table 2) and slightly shorter than the crane *B. regulorum* (1.22 Gb). Total gap length was bigger in volant species (24 Mb and 26 Mb for *P. melanotus* and *G. philippensis* respectively) than in flightless ones (9 Mb for *P. hochstetteri* and 13.6 Mb for *G. australis*).

BUSCO scores were similar for *P. melanotus*, *P. hochstetteri* and *G. australis* with close to 80% of single copy genes were found complete. In contrast, *G. philippensis* comprised 69% of "Complete single copy" and had a higher proportion (17%) of missing genes (Fig. 4).

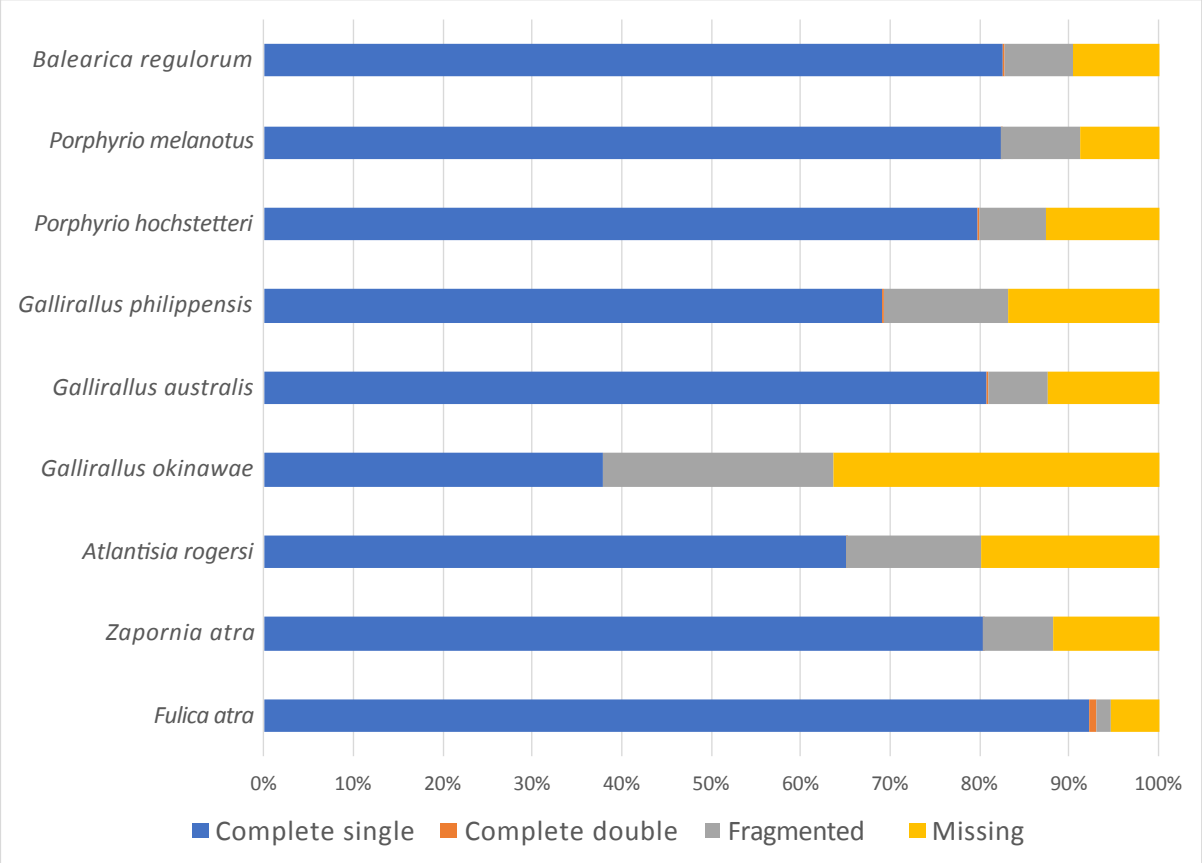


Figure 4: BUSCO V4 results (mode: genome) using the aves_odb10 dataset. Total number of genes (Benchmarking Universal Single-Copy Orthologs): 8338. Complete: ortholog is present in the genome in a single copy (Complete single) or in two copies (Complete double), Fragmented: gene is only partially present, Missing: no significant match in the genome.

Table 3: De novo genome assembly metrics among 8 rail species and one crane (*B. regulorum*)

Species	Genome Assembly size (Gb)	Largest scaffold	Number of scaffolds	Scaffold N50	Scaffold L50	Number of contigs	Contig N50 (KB)	Contig L50	Depth of coverage	# Ns in assembly (KB)
Balearica regulorum	1.221	219,267,915	105	82,577,926	5	249	23.3	14	na	0
Porphyrio melanotus	1.114	1,068,914	34,563	82,204	3,707	159,218	16.5	16,605	102.26	24.745
Porphyrio hochstetteri	1.116	1,015,032	30,278	30,278	3,641	76,213	40.8	7,446	94.53	8.974
Gallirallus philippensis	1.07	722,688	55,205	46,737	6,047	209,712	9.6	28,118	103.04	26.818
Gallirallus australis	1.158	1,692,012	36,524	126,032	2,365	96,978	41.2	7,224	85.59	13.634
Galirallus okinawae	1.114	97,103	768,680	6,245	48,754	768,680	6.2	48,754	na	na
Atlantisia rogersi	1,168	1,015,111	159,311	36,139	8,295	160,845	36,1	8,295	na	na
Zapornia atra	1.119	1,795,565	58,849	134,191	2,049	113,655	44,130	6,609	na	na
Fulica atra	1,167	27,139,163	17,827	6,390,841	46	31,348	246,1	1,314	na	na

Extracting coding regions

For each species, the coding regions of each gene (the CDS) were extracted based on the annotations and compared with the respective *B. regulorum* CDS. Over 9000 gene CDSs were retrieved near-complete (above 95% of the reference CDS nucleotide sequence length) for *Zapornia atra*, the two *Porphyrio* species and *Gallirallus australis* (Fig.5). *G. philippensis* exhibited a slightly lower proportion (8,259 CDS over 95%) which was consistent with the BUSCO results and *G. okinawae* included a large number of fragmented CDS. The CDSs present in the reference genome but not in the Rail data ("Not found") represent less than 8% of the CDS for all species.

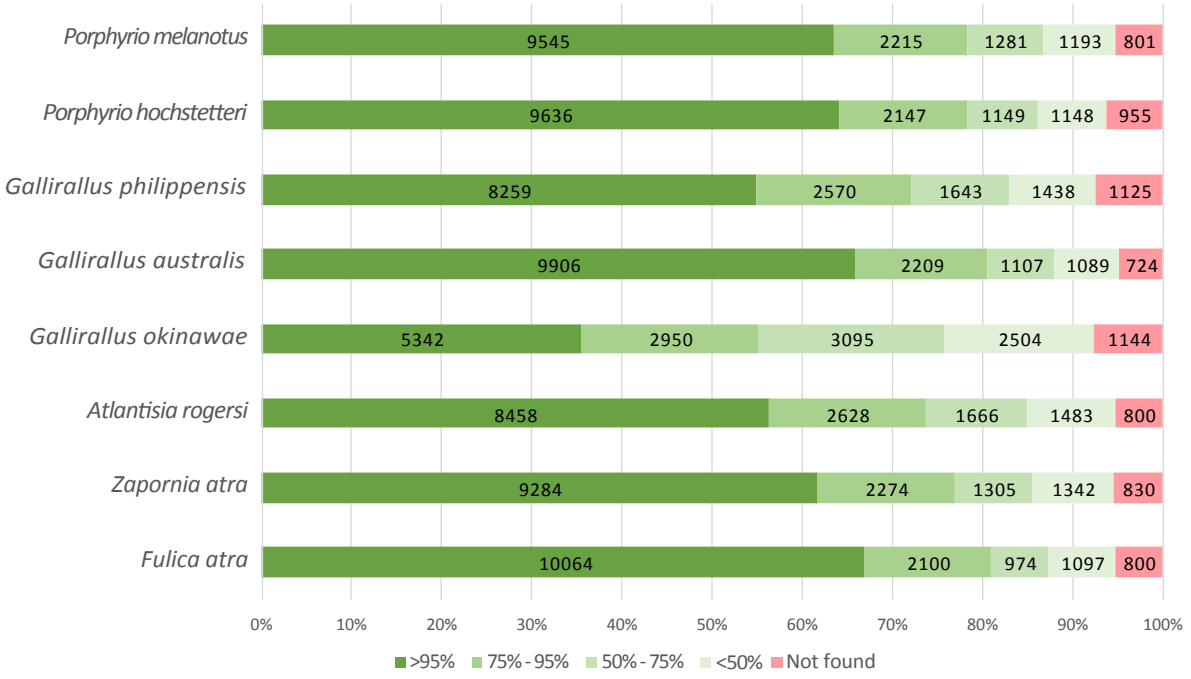


Figure 5: Completeness of CDSs retrieved from eight rail genomes compared to the reference crane *B. regulorum* genome that has a total of 15,035 annotated CDSs. Colours indicate the proportion of genes retrieved from a sample at various scales of completeness.

Discussion

Considerable variation was observed between species heterozygosity (Fig. 2 and 3). Indeed, the two volant species were more heterozygous than the flightless ones (Fig. 3) with big differences observed between the most heterozygous species, *G. philippensis* (frequency of heterozygous site of 0.01) and the least heterozygous species *P. hochstetteri* (0.0002). The low

level of heterozygosity in flightless species probably reflects the relative isolation and reduced size of the habitat as well as population collapse (Baker et al. 1995, Burga et al. 2017, White et al. 2018). The takahē *P. hoschetteri* is a critically endangered flightless species with a population of only 445 in 2020 (www.doc.govt.nz), all derived from a remnant discovered in the 1950s that may have numbered as low as two individuals (Wallace 2002). The resulting inbreeding depression likely explains its extremely low level of heterozygosity (Grueber et al. 2010). *Gallirallus philippensis* on the other hand is a relatively abundant species with a geographic range that covers the islands of Aotearoa New Zealand and the western Pacific (Trewick 1997b, Garcia-R et al. 2017) which is likely to maintain high heterozygosity at the species level.

The four newly assembled genomes have similar or better characteristics than the previously assembled rail genomes (Fig. 4 and 5, Table 3). The N50 and L50 scaffolds were within the same range as the other rails with the exception of *G. okinawae* which is the oldest assembly and presents a lower quality. *Fulica atra* on the other hand had a high N50 (6,390,841) as well as an extremely low L50 (46) indicating a good assembly quality.

The BUSCO results (Fig. 4) and CDS extractions (Fig. 5) showed similar trends with numerous fragmented or missing genes in the *G. okinawae* genome contrasting with a large majority of the genes being retrieved for *F. atra*. The latter genome, however, appears to contain a considerable number of genes represented twice (Complete double in Fig. 4) suggesting that some parts of the genome may have been duplicated during bioinformatic assembly.

Among the four newly assembled rallid genomes, *G. philippensis* had the lowest proportion of complete genes according to BUSCO (Fig. 4) and extracted CDS comparison (Fig. 5). This can be attributed to the high heterozygosity level which generally makes the assembly process more challenging due to the increased complexity of the de Bruijn graph structure (Kajitani et al. 2014). Nonetheless, *G. philippensis* genome is a high-quality assembly that can without a doubt be used to investigate evolutionary processes along with the three other assembled genomes.

With the exception of *G. okinawae*, where the assembly quality was much lower than for other species, the annotation process was stable, and a large majority of the genes were retrieved. In each species, over 70% of the genes were identified with greater than 75% completeness.

To conclude, we provide here four high-quality assemblies which represent valuable genomic resources to investigate evolutionary processes within the rail family. The quality checks that were performed showed that the generated assemblies are reliable and that the annotations can be trusted. Comparing the newly assembled genomes showed lower levels of heterozygosity in flightless species which probably reflects their relatively small populations. This study doubles the number of available rallid genomes bringing the total to eight assemblies. Five of them belong to species that have independently lost the ability to fly and three are volant relatives; this creates new opportunities to investigate the evolution of avian flightlessness.

Data availability

The genomes and annotations are available on NCBI, BioProject PRJNA782688. The configuration files, command lines used, CDS lists, and R scripts are available in the supplementary data. URL: https://figshare.com/s/3a89eea20c4607abbefe.

Acknowledgements

This study was supported by the New Zealand Marsden Fund Council from Government funding, managed by Royal Society Te Apārangi, grant MAU1601 to GCG. Thanks to Roger Moraga for initial bioinformatic discussions and assistance using the software Meraculous. The genome assemblies were generated with the help of Genomics for Aotearoa New Zealand (GFANZ, genomics.nz) thanks to Rob Elshire. The authors would like to thank Richard Witehira who provided the helpful local knowledge about the mioweka name. Thanks also to Jonathan Proctor (Rangitāne o Manawatū) for ongoing consultation around the role of Rangitāne o Manawatū as kaitiaki (guardians) of the museum collection samples held by Massey University Palmerston North.

References

Baker, A. J., Daugherty, C. H., Colbourne, R. and McLennan, J. L. 1995. Flightless brown kiwis of New Zealand possess extremely subdivided population structure and cryptic species like small mammals. – PNAS 92: 8254–8258.

Boast, A. P., Chapman, B., Herrera, M. B., Worthy, T. H., Scofield, R. P., Tennyson, A. J. D., Houde, P., Bunce, M., Cooper, A. and Mitchell, K. J. 2019. Mitochondrial Genomes from New Zealand's Extinct Adzebills (Aves: Aptornithidae: Aptornis) Support a Sister-Taxon Relationship with the Afro-Madagascan Sarothruridae. – Diversity 11: 24.

Burga, A., Wang, W., Ben-David, E., Wolf, P. C., Ramey, A. M., Verdugo, C., Lyons, K., Parker, P. G. and Kruglyak, L. 2017. A genetic signature of the evolution of loss of flight in the Galapagos cormorant. – Science 356: eaal3345.

Chen, S., Zhou, Y., Chen, Y. and Gu, J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. – Bioinformatics 34: i884–i890.

Chikhi, R. and Medvedev, P. 2014. Informed and automated k-mer size selection for genome assembly. – Bioinformatics 30: 31–37.

Ewing, B. and Green, P. 1998. Base-Calling of Automated Sequencer Traces Using Phred. II. Error Probabilities. – Genome Res. 8: 186–194.

Ewing, B., Hillier, L., Wendl, M. C. and Green, P. 1998. Base-Calling of Automated Sequencer Traces UsingPhred. I. Accuracy Assessment. – Genome Res. 8: 175–185.

Fain, M. G., Krajewski, C. and Houde, P. 2007. Phylogeny of "core Gruiformes" (Aves: Grues) and resolution of the Limpkin–Sungrebe problem. – Molecular Phylogenetics and Evolution 43: 515–529.

Garcia-R., J. C. and Trewick, S. A. 2014. Dispersal and speciation in purple swamphens (Rallidae: Porphyrio). – The Auk 132: 140–155.

Garcia–R, J. C., Gibb, G. C. and Trewick, S. A. 2014. Eocene Diversification of Crown Group Rails (Aves: Gruiformes: Rallidae). – PLOS ONE 9: e109635.

Garcia-R, J. C., Gibb, G. C. and Trewick, S. A. 2014. Deep global evolutionary radiation in birds: Diversification and trait evolution in the cosmopolitan bird family Rallidae. – Molecular Phylogenetics and Evolution 81: 96–108.

Garcia-R, J. C., Joseph, L., Adcock, G., Reid, J. and Trewick, S. A. 2017. Interisland gene flow among populations of the buff-banded rail (Aves: Rallidae) and its implications for insular endemism in Oceania. – Journal of Avian Biology 48: 679–690.

Gaspar, J., Gibb, G. C. and Trewick, S. A. 2020. Convergent morphological responses to loss of flight in rails (Aves: Rallidae). – Ecology and Evolution 10: 6186–6207.

Grueber, C. E., Laws, R. J., Nakagawa, S. and Jamieson, I. G. 2010. Inbreeding Depression Accumulation across Life-History Stages of the Endangered Takahe. – Conservation Biology 24: 1617–1625.

Hume, J. P. and Martill, D. 2019. Repeated evolution of flightlessness in Dryolimnas rails (Aves: Rallidae) after extinction and recolonization on Aldabra. – Zool J Linn Soc 186: 666–672.

Kajitani, R., Toshimoto, K., Noguchi, H., Toyoda, A., Ogura, Y., Okuno, M., Yabana, M., Harada, M., Nagayasu, E., Maruyama, H., Kohara, Y., Fujiyama, A., Hayashi, T. and Itoh, T. 2014. Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. – Genome Res. 24: 1384–1395.

Kirchman, J. J. 2012. Speciation of Flightless Rails on Islands: A DNA-Based Phylogeny of the Typical Rails of the Pacific. – The Auk 129: 56–69.

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

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

Liu, Y., Liu, S., Yeh, C.-F., Zhang, N., Chen, G., Que, P., Dong, L. and Li, S. 2018. The first set of universal nuclear protein-coding loci markers for avian phylogenetic and population genetic studies. – Scientific Reports 8: 15723.

Livezey, B. C. 2003. Evolution of Flightlessness in Rails. – American Ornithologists' Union.

McNab, B. K. 1994. Energy Conservation and the Evolution of Flightlessness in Birds. – The American Naturalist 144: 628–642.

Merkel, D. 2014. Docker: lightweight Linux containers for consistent development and deployment. – Linux J. 2014: 2:2.

Olson, S. L. 1973. Evolution of the rails of the South Atlantic islands (Aves: Rallidae). – Smithsonian Contributions to Zoology.

Ozaki, K., Yamamoto, Y. and Yamagishi, S. 2010. Genetic diversity and phylogeny of the endangered Okinawa Rail, Gallirallus okinawae. – Genes Genet. Syst. 85: 55–63.

Richterich, P. 1998. Estimation of Errors in "Raw" DNA Sequences: A Validation Study. – Genome Res. 8: 251–259.

Ripley, S. D., Lansdowne, J. F. and Olson, S. L. 1977. Rails of the World: A Monograph of the Family Rallidae. – M. F. Feheley Publishers.

Seppey, M., Manni, M. and Zdobnov, E. M. 2019. BUSCO: Assessing Genome Assembly and Annotation Completeness. – In: Kollmar, M. (ed), Gene Prediction: Methods and Protocols, Methods in Molecular Biology. Springer, pp. 227–245.

Slikas, B., Olson, S. L. and Fleischer, R. C. 2002. Rapid, independent evolution of flightlessness in four species of Pacific Island rails (Rallidae): an analysis based on mitochondrial sequence data. – Journal of Avian Biology 33: 5–14.

Steadman, D. W. 1995. Prehistoric Extinctions of Pacific Island Birds: Biodiversity Meets Zooarchaeology. – Science 267: 1123–1131.

Taylor, B. 1998. Rails: A Guide to the Rails, Crakes, Gallinules and Coots of the World. – Bloomsbury Publishing.

Trewick, S. A. 1997a. Flightlessness and phylogeny amongst endemic rails (Aves: Rallidae) of the New Zealand region. – Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 352: 429–446.

Trewick, S. A. 1997b. Sympatric flightless rails Gallirallus dieffenbachii and G. modestus on the Chatham Islands, New Zealand; morphometrics and alternative evolutionary scenarios. – Journal of the Royal Society of New Zealand 27: 451–464.

Trewick, S. A., Pilkington, S., Shepherd, L. D., Gibb, G. C. and Morgan-Richards, M. 2017. Closing the gap: Avian lineage splits at a young, narrow seaway imply a protracted history of mixed population response. – Molecular Ecology 26: 5752–5772.

Wallace, G. E. 2002. The Takahe: Fifty Years of Conservation Management and Research. – The Auk 119: 291–293.

White, D. J., Ramón-Laca, A., Amey, J. and Robertson, H. A. 2018. Novel genetic variation in an isolated population of the nationally critical Haast tokoeka (Apteryx australis ‘Haast’) reveals extreme short-range structure within this cryptic and flightless bird. – Conserv Genet 19: 1401–1410.

Wu, T. D. and Watanabe, C. K. 2005. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. – Bioinformatics 21: 1859–1875.

Chapter Four

Genetic pathway to flightlessness: evidence of convergent evolution in rails (Aves: Rallidae)

Abstract

Environmental changes can reduce or eliminate a source of selection that was formerly involved in the maintenance of a phenotypic trait. This phenomenon is called relaxed selection and an example of it is the loss of flight in island birds. When organisms evolve flightlessness it is expected that the genomic regions involved in flight capacity will undergo relaxed selection. Several species from the rail family (Aves: Rallidae) independently lost the ability to fly which makes this clade a good model to study the evolution of avian flightlessness. Here, we used branch-site models to investigate the convergent evolution of that trait. To do so, the coding regions from seven rails: three volant, *Porphyrio melanotus, Gallirallus philippensis* and *Fulica atra* and four flightless *Porphyrio hochstetteri, Gallirallus australis, Atlantisia rogersi,* and *Zapornia atra* were compared to identify genes under different selective pressure associated with flightlessness. Evidence of selection was detected in hundreds of genes. Moreover, genes under relaxed selection were enriched in GO functional categories that were either directly associated with flight like muscle development and regulation of the circulatory system or associated with the ecological changes consequent to the loss of flight such as immune response and cognition.

Keywords: Rallidae, Flightlessness, Relaxed selection, Convergent evolution

Introduction

Different species often evolve phenotypic similarities in response to ecological challenges. That convergence shows that environmental circumstances can apply selective pressure which results in similar evolutionary solutions (Stern 2013). The molecular basis associated with these adaptative traits can be similar mutations in different species (Besnard et al. 2009, Zhen et al. 2012) or "lineage specific" variations that result in the same phenotype (Cooper et al. 2014). A remarkable example of convergent evolution is the loss of flight in birds which repeatedly occurs in avian evolution (Roff 1994). Flightlessness has been observed in many island species and is interpreted as an effect of the insular conditions which often provide a habitat with few or no predators and limited competition for resources (McNab 1994).

Rallidae

The bird family known as rails (Aves: Rallidae) originated during the Eocene around 40 million years ago (Garcia–R et al. 2014) and diversified into over 130 extant species (Steadman 1995, Kirchman 2012, Garcia–R et al. 2014). Despite the terrestrial lifestyle of the majority of the species (Taylor 1998), this bird family displays remarkable dispersal capacity resulting in broad distribution and the colonization of numerous oceanic islands (Olson 1973, Ripley et al. 1977, Garcia–R et al. 2014). The extant diversity includes (or recently included) more than 30 flightless species, the majority of which are endemic to single islands (Steadman 1995, Kirchman 2012). This demonstrates their ancestors had been volant (Trewick 1997a, b) and therefore that flightlessness evolved multiple times. As a consequence, the rail family is an excellent model to study the evolutionary process behind the transition to flightlessness.

Flightless rails share morphological characters that distinguish them from their volant relatives, they exhibit smaller sterna and wings than volant taxa along with wider pelves and more robust femora (Livezey 2003, Gaspar et al. 2020). Moreover, it has been shown that these differences are independent of phylogeny and instead demonstrate convergent morphological adaptation associated with a walking ecology (Gaspar et al. 2020).

Measuring selection

While much is known about the morphological characteristics of flightlessness, we are only beginning to understand the genetic signals associated with this phenotypic evolution. One approach to studying the molecular basis of flightlessness is to measure selection on coding regions. Amino acids are specified by more than one mRNA codon, as a consequence, not all nucleotide substitutions result in amino acid replacement. Synonymous nucleotide substitutions have no effect on the amino acid sequence whereas nonsynonymous substitutions cause amino acid change. As natural selection operates mainly at the protein level, relative rates of nonsynonymous (d_N) and synonymous (d_S) substitution are useful to infer selective pressures (Kimura 1977). The ratio $\omega=d_N/d_S$ can be used to measure the balance between positive, neutral, and purifying selection. It can also be combined with phylogenetic information and calculated for any codon site, these branch-site models aim to detect relaxed or intensified selection at the lineage level (Yang 2007). Relaxed selection refers to release from selective constraint, in our case the constraint is flight. This evolutionary phenomenon is characterized by a reduction in the intensity of purifying and positive selection (Lahti et al. 2009, Wertheim et al. 2015). To investigate signs of convergent relaxed selection, volant species (background lineages) can be compared with flightless species (foreground lineages) in a phylogenetic context. At a molecular level, the selection is considered relaxed if the sites with smaller ω values (purifying selection, blue in Fig. 1) in the background lineages increase toward one (neutral) in the foreground lineages whereas the ω values greater than one in the background lineages (positive selection, red in Fig. 1) decrease in the foreground lineages. On the other hand, if the smaller ω values are even smaller in the foreground lineages and the ω values above one increase in the foreground lineages the selection is considered intensified (Fig.1).

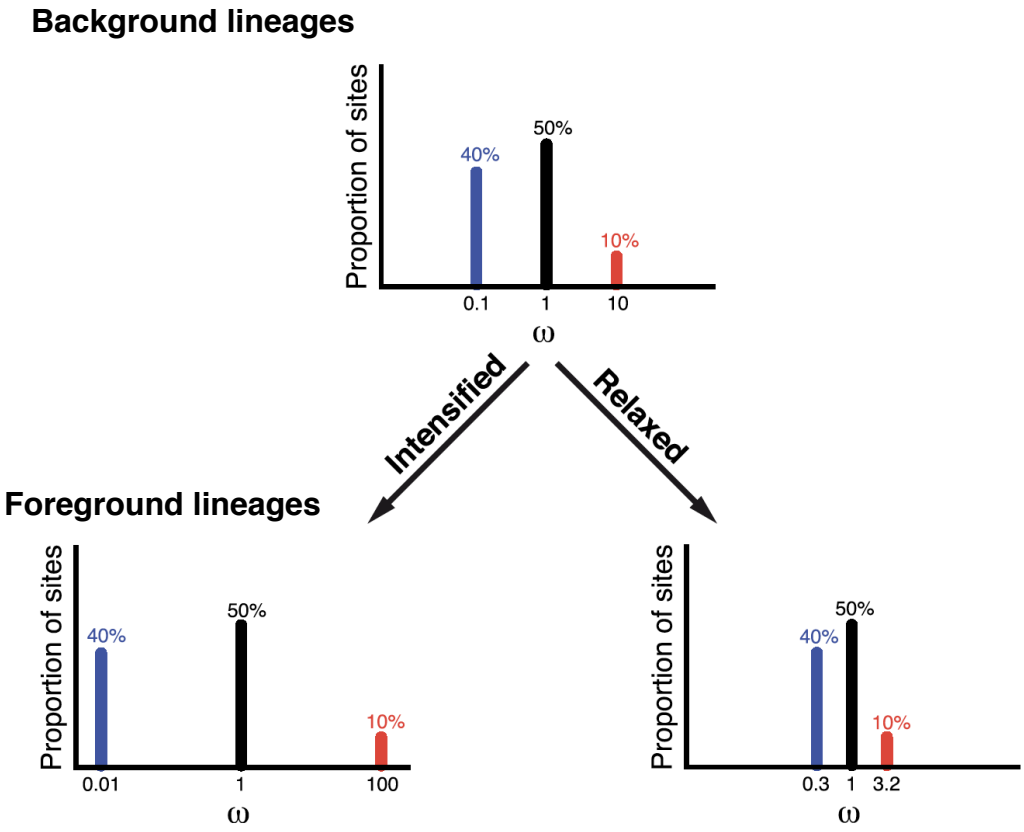


Figure 1: Relaxed and intensified selection. Sites under positive selection are shown in red and sites under purifying selection are shown in blue. Intensified selection results in all ω categories being pushed away from neutrality (ω=1) and relaxed selection results in all ω categories being pushed toward neutrality. (Redrawn after Wertheim et al. 2015).

Some studies have investigated the molecular processes associated with the ability to fly in birds (Bickley and Logan 2014, Machado et al. 2016, Burga et al. 2017, Farlie et al. 2017, Campagna et al. 2019, Sackton et al. 2019) however, none of them studied rails. The present research uses the branch-site model to investigate the convergent evolution of flightlessness in that bird family. To do so, the coding regions from seven rails: three volant, *Porphyrio melanotus, Gallirallus philippensis* and *Fulica atra* and four flightless *Porphyrio hochstetteri, Gallirallus australis, Atlantisia rogersi,* and *Zapornia atra* were compared to look for evidence of selection in flightless rails. Our hypothesis is that most genes associated with the loss of flight in bird lineages are subject to relaxed selection in flightless species. After identifying the genomic regions of interest, gene ontology enrichment analyses were carried out to determine what functional categories were affected by the transition to flightlessness.

Methods

Gene alignments and filtering

The genomes of four New Zealand rail species: two volant- Purple swamphen (called pūkeko in Aotearoa New Zealand) *Porphyrio melanotus* and buff-banded rail (also called mioweka and moho pererū) *Gallirallus philippensis*; and two flightless- takahē *Porphyrio hochstetteri* and weka *Gallirallus australis* were assembled and annotated (See Chapter 3 for details). Data from a further three rail species was included in the study as their genomes were available via GenBank: Henderson crake *Zapornia atra* (flightless) GenBank assembly accession GCA_013400835.1, Eurasian coot *Fulica atra* (volant) GCA_013372525.1, Inaccessible Island rail *Atlantisia rogersi* (flightless) GCA_013401215.1.

Coding sequences (CDS) were retrieved from these seven rail genomes following the methods described in Chapter 3. The CDS list of each species was organised by gene using Geneious 11.1.4 (https://www.geneious.com) then aligned by codon sequence with MACSE v2.03 (Ranwez et al. 2018) using the default settings. Ambiguous alignment positions and gaps were eliminated using GBlocks v0.91b (Castresana 2000) with settings: minimum length of a block =5 and default settings for the rest. Alignments that were shorter than 100 base pairs and had a pairwise identity (visualized on Geneious) lower than 85% were filtered out. The dataset comprised 12,463 CDS alignments before filtering and 11,577 (7.1% filtered) after filtering. Comparison using NCBI nucleotide BLAST (blast.ncbi.nlm.nih.gov, Chen et al. 2015) were used to assign putative names to unidentified genes from incomplete genome annotations.

Phylogeny

Phylogenetic inference for the seven rails with *B. regulorum* as an outgroup was performed using data for 45 genes selected from a set of markers recommended for avian phylogenetic reconstruction (Liu et al. 2018). The 45 CDS alignments were concatenated into a 100,416 bp alignment using Phyluce 1.7.1 (Faircloth 2016) with the default settings. The best-fit partitioning scheme was then determined with PartitionFinder2 (Lanfear et al. 2017) via the CIPRES Science Gateway (Miller et al. 2010, see partitions and alignment in supplementary data) using the default settings. Maximum Likelihood (ML) analyses were implemented in RAxML v8.2.10 (Stamatakis 2014) via the CIPRES Science Gateway with bootstrapping

automatically stopped employing the majority rule criterion, and a consensus tree generated in Geneious (Fig. 2). This tree is required as an input in many of the subsequent tests.

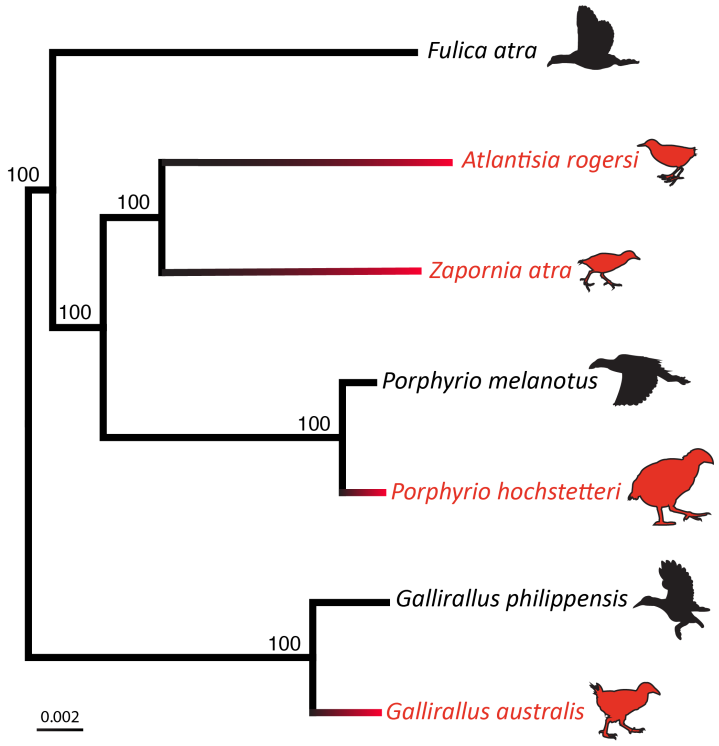


Figure 2: Maximum likelihood (RAxML V.8) phylogeny of three volant (black) and four flightless (red) rails (Aves: Rallidae) based on 45 concatenated nuclear gene sequences selected from a set of universal nuclear protein coding loci markers for avian phylogenetics (Liu et al. 2018). Bootstrap support is indicated for each node. The grey crowned crane *Balearica regulorum* root species is not shown on this tree. Red terminals branches symbolise the independent emergence of flightlessness.

Analyses of protein-coding regions

To investigate the convergent evolution of flightlessness at the gene level several phylogeny-based tests were run. As the four flightless species are endemic to islands, we know that their ancestors were volant and that flightlessness evolved independently in each of them (Trewick 1997a). Therefore, we expect the flightless traits evolved closer to the tips of the branches, with flight considered the ancestral state in each lineage (Fig. 2). To determine which genes might be associated with the flight/flightlessness transition the following analyses compare a set of foreground lineages (the flightless species) to a set of background lineages (the volant species).

Significant differences between these two groups can indicate evidence of convergent evolution of flightlessness.

Testing for bias due to incomplete lineage sorting

To avoid bias caused by incomplete lineage sorting (ILS) (Mendes and Hahn 2016), Approximately Unbiased (AU) tests (Shimodaira 2002) were applied to all CDS alignments. This test detects gene trees that significantly deviate from the species tree by assessing the confidence of a tree selection given a nucleotide alignment. The test was implemented using IQ-TREE v2.0.3 (Minh et al. 2020) on 11,577 rail CDS alignments using the species tree (Fig. 2) as an input. The following settings were used: search iteration set to 0 and 1000 resampling estimated log-likelihood (RELL) replicates. At $p<0.05$, genes were considered to significantly deviate from the species tree.

Branch-site models

Once all the CDS alignments were generated and quality checked, each of the 11,577 alignments was screened to detect signal of selection in lineages leading to flightless species using two different branch-site model tools: CodeML (Yang 2007) and HyPhy RELAX (Wertheim et al. 2015). In these analyses, the objective was to identify whether any genes were subject to the same type of selection in all the four flightless rail species. To do this, the four flightless rails were together considered 'foreground' lineages, and the three volant species as 'background' lineages (Fig. 2).

Likelihood comparison tests

CodeML is a part of the Phylogenetic Analysis by Maximum Likelihood v4.9 (PAML) package (Yang 2007) which performs phylogenetic analysis of DNA and protein sequences using maximum likelihood. It was used to apply two sets of model likelihood comparison tests on the CDS alignments in the context of the species tree topology.

To assess the potential difference of selective pressures between lineages, we compared likelihood scores from M0 model (one ratio) that supposes a single ω for all lineages, and M2 that allows different ω between foreground (flightless species) and the background (volant species) lineages (the CodeML control files are available in supplementary data). Likelihood

ratio tests (LRT) (Felsenstein 1981) were used to compare the fit of different models applied to each alignment with the likelihood ratio LR = 2*(likelihood model 1 – likelihood model 2). Significant model likelihood differences between M0 and M2 would suggest that volant and flightless rails experience different selective pressures.

To assess evidence of significant relaxed or positive selection in the foreground flightless lineages the results from three models were compared (Chikina et al. 2016); a neutral site model (M1), the branch-site neutral model (BS1), and the branch-site selection model (BS2) (Yang 2007). Rejection of M1 in favour of BS1 indicates relaxed selection in flightless lineages, and further rejection of BS1 in favour of BS2 indicates positive selection in flightless lineages. In other words, relaxed selection is inferred if the BS1 model is significantly different to M1 but that BS2 and BS1 are not significantly different. Positive selection is inferred when BS1 is significantly different from M1 and that BS2 is significantly different from BS1 (Zhang et al. 2005, Yang 2007). CodeML was implemented using a custom python script (see supplementary data).

An additional set of tests was implemented using RELAX v3.1 (Wertheim et al. 2015) in the HyPhy platform (Pond et al. 2005), to detect signal of relaxed or intensified selection among the foreground flightless species. RELAX allows for a free relaxation parameter K to be inferred on each branch. K is an exponent for selection ω between the foreground and the background lineages, where $\omega_{background}^{K} = \omega_{foreground}$. Therefore, K greater than 1 implies intensified selection in flightless rails and a K smaller than 1 suggests relaxed selection. A null model in which K is constrained to 1 was compared, using LRT, to the alternative model where K is a free parameter (Wertheim et al. 2015). When K significantly improved the model the flightless lineages are considered to be under relaxed (K<1) or intensified (K>1) selection.

To distinguish results from CodeML and RELAX, genes inferred to be under relaxed selection according to CodeML are indicated as “CodeML relaxed” and those inferred to be under relaxed selection according to RELAX are indicated as “HyPhy relaxed”.

Testing for nonsynonymous multinucleotide mutations

Branch-site tests (CodeML and HyPhy RELAX) assume that nucleotide substitutions happen independently, however, the likelihood of DNA replication errors at adjacent sites is higher

than the background rate. This can result in nonsynonymous multinucleotide mutations (MNMs) (Schrider et al. 2011, Saribasak et al. 2012) that bias branch-site tests to return false support for positive selection (Venkat 2018). Although the present research focuses on relaxed selection rather than positive selection, MNMs could affect ω (the d_N/d_S ratio) used to evaluate relaxed selection so this issue needs to be addressed (Cui et al. 2019). Therefore, genes inferred to be under selection by previous analyses were tested for Codons with Multiple Differences (CMD). Codon sites that contained more than one substitution in at least one of the species were removed from the alignments and the analyses were rerun on the modified alignments. This resulted in a 7% reduction in the number of genes interpreted as displaying significant selection.

Gene ontology enrichment

A conservative set of relaxed genes was generated by including all the genes significant for both CodeML relaxed and HyPhy relaxed. The genes with ILS (Incomplete Lineage Sorting) were excluded (resulting in a list of 230 relaxed genes) and an overrepresentation test for Gene Ontology (GO) biological process was implemented online using PANTHER (Protein Analysis Through Evolutionary Relationships, http://pantherdb.org) (Mi et al. 2019) with the reference list from *Gallus gallus* (Reference Proteome 2020_04, http://pantherdb.org/data/). The overrepresentation test results showed the observed and expected number of genes corresponding with each GO term as well as the fold enrichment, the p-value as determined by Fisher's exact test, and False Discovery Rate (FDR) as calculated by the Benjamini-Hochberg procedure. Only GO terms with an enrichment greater than two and more than four identified relaxed genes were retained. A non-redundant functionally-grouped network was generated using the ClueGO v2.5.8 (Bindea et al. 2009) plug-in of Cytoscape v3.8.2 (Shannon et al. 2003) for the significant GO terms (Appendix C4 Table 1). The preselected function tool was used with *Gallus gallus* as a reference organism and medium network specificity. Overrepresentation tests were also applied for a set of intensified genes (the genes under intensified selection, significant for M0-M2 and without ILS; 374 genes) and all the genes significant for M0-M2 test (not including ILS; 731 genes).

The pipeline for the methods described here (from genome assembly to overrepresentation test) is shown in Figure 3.

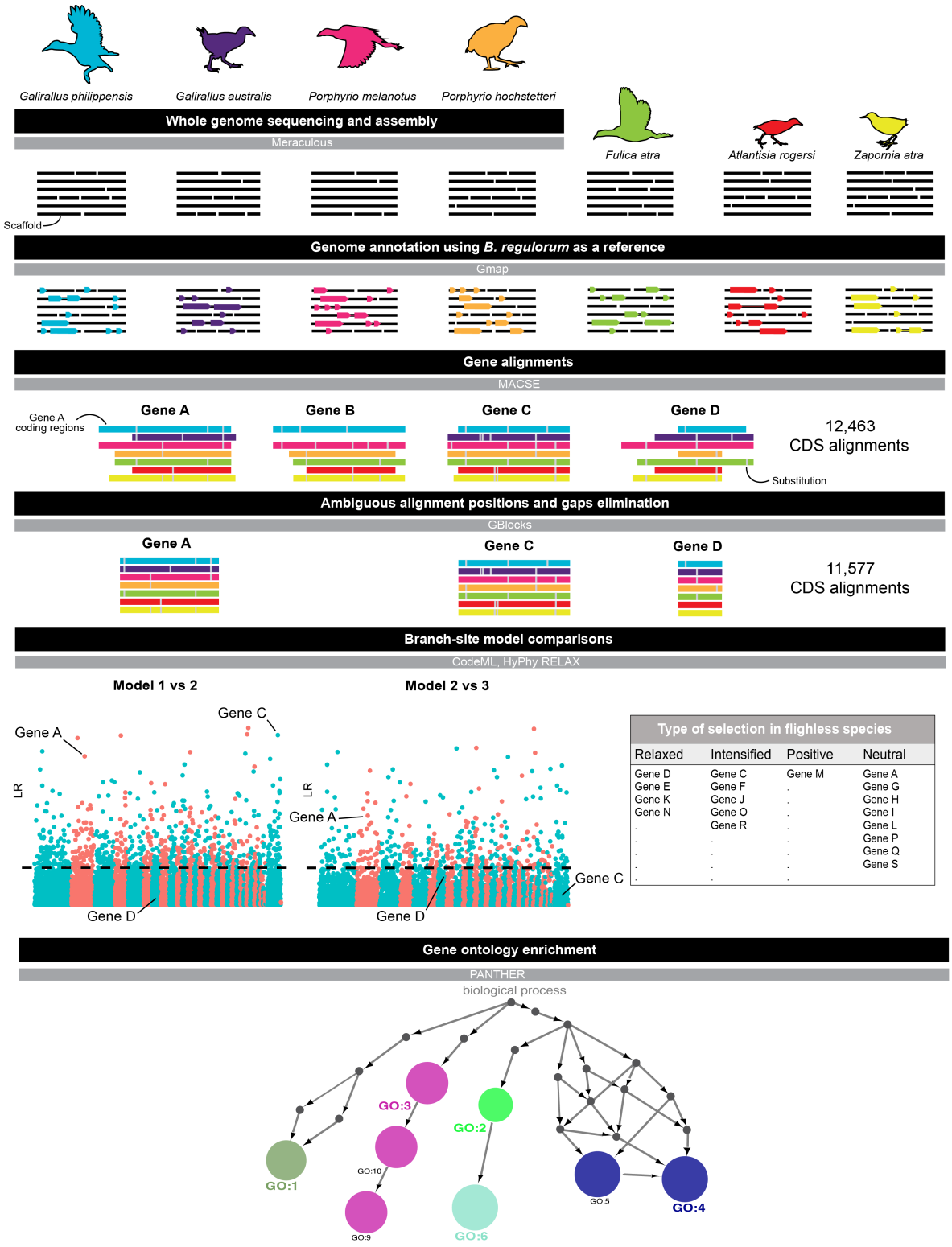


Figure 3: Pipeline of methods used to generate and screen thousands of coding sequence alignments for selection signal in flightless species.

Results

Branch site model comparisons

The seven species maximum likelihood phylogeny (Fig. 2) was consistent with recent literature (Garcia–R et al. 2014, Kirchman et al. 2021) and was used for the subsequent analyses.

Selection on flightless rail lineages was tested for 11,577 CDS alignments using four different branch site model comparisons. Out of the 11,577 CDS alignments, 791 (including the alignments with ILS) showed a significant result for the M0-M2 model comparisons test (Fig. 4a) that detects differences in selective pressure between foreground flightless and background volant rails. One gene, Ankyrin Repeat domain 52 (*ANKRD52*), exhibited an LR value much higher than the others (28.3).

Comparisons between the three likelihood models M1, BS1, and BS2 were used to detect evidence of relaxed or positive selection. Rejection of M1 in favour of BS1 (without being further rejected in favour of BS2) was observed in 388 genes (Fig. 4b) indicating evidence of relaxed selection in flightless rails. The gene Protein Kinase, DNA-Activated, Catalytic Subunit (*PRKDC*) exhibited the highest value (24.5). The BS1 model was rejected in favour of BS2 in 272 CDS alignments (Fig. 4c), of which 24 genes exhibited rejection of both M1 and BS1 in favour of BS2 indicating evidence of positive selection in flightless species.

Genes that exhibited high LR values in more than one model comparisons included *ANKR252* and the gene Mitogen-Activated Protein Kinase Kinase Kinase 14 (*MAP3K14*) that showed evidence of positive selection in flightless species and *PRKDC*, Sorilin Related Receptor 1 (*SORL1*), and Protein Disulfide Isomerase Family A Member 4 (*PDIA4*) that exhibited evidence of relaxed selection in flightless rails.

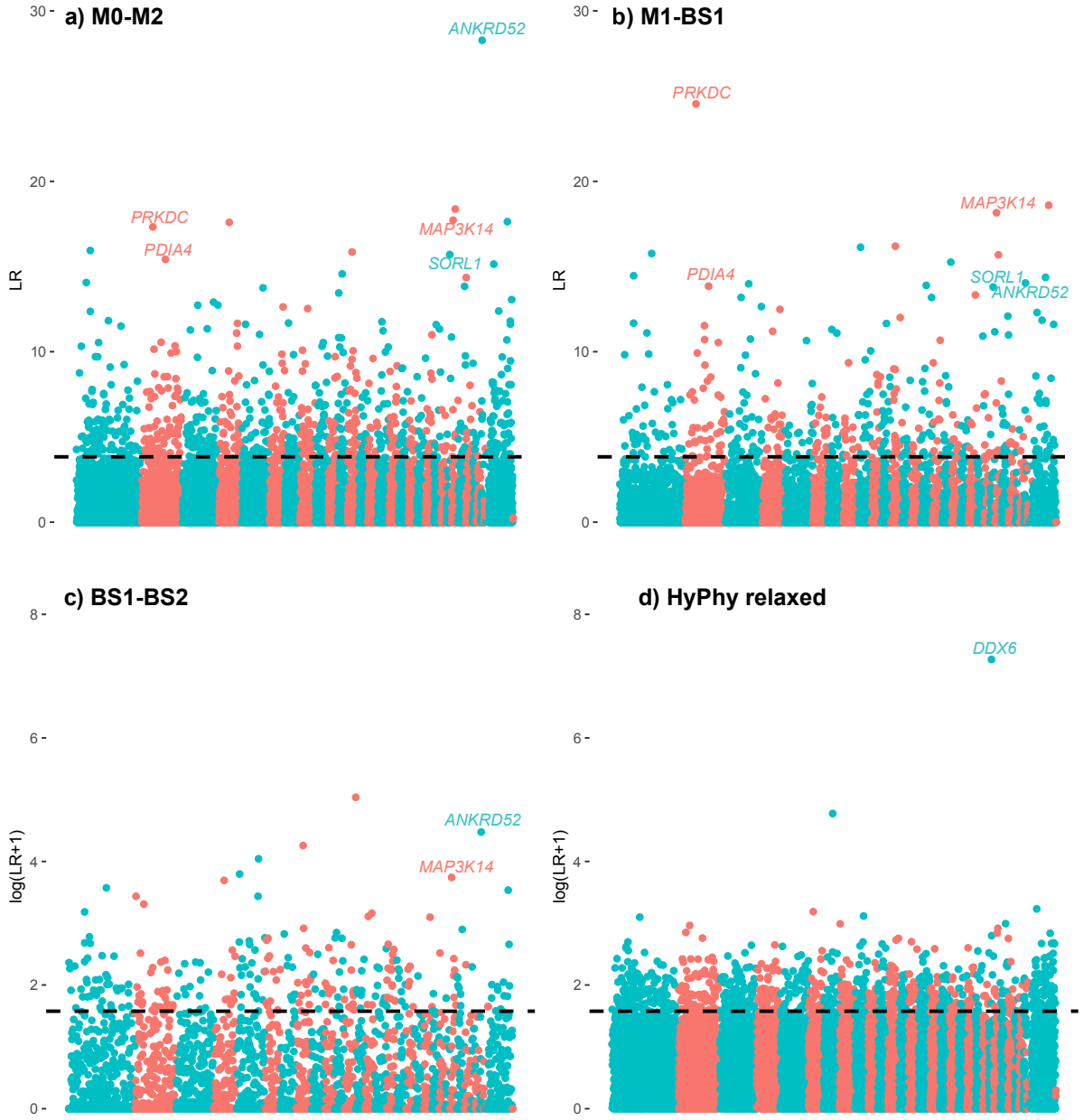


Figure 4: Likelihood ratio test (LRT) results for 11,577 CDS alignments. Four branch-site model comparisons were tested with four flightless 'foreground' rails and three volant 'background' rails. Alternating colours indicate putative chromosomal positions of the genes estimated from *Balearica regulorum* chromosome level genome assembly (GCA_011004875.1) arranged with chromosome 1 (left) to Z (no W Chromosomes were found as most specimens were males) and non-assigned genes on the right. LR = 2*(likelihood model 1 - likelihood model 2). a) LRT results between M0 and M2 models, genes above the dashed line present significant differences between the models indicating evidence of selective pressure in flightless rails for that gene. b) LRT results between M1 and BS1 models, if not further rejected in favour of BS2, all the genes above the dashed line are inferred to be under relaxed selection in flightless rails. c) LRT results between BS1 and BS2 models. Genes above the dashed line in both M1-BS1 and BS1-BS2 show evidence of positive selection in flightless rails; LR values were log transformed for clarity. d) LRT results for RELAX v3.1. Genes above the dashed line show evidence of relaxed or intensified selection in flightless species; LRT values were log transformed for clarity. The genes named on the plots show high LR values in multiple likelihood ratio tests and/or are further discussed in the text.

Assessing relaxed selection

Comparing a null model in which the free relaxation parameter K was constrained to 1 to the alternative model where K was a free parameter allows assessment of relaxed selection (using HyPhy RELAX). K significantly improved the model in 886 CDS alignments (Fig. 4d) implying these genes were under relaxed or intensified selection. For 479 genes, K was lower than 1 which means they showed evidence of relaxed selection in flightless species, and for 407 genes K was greater than 1 indicating evidence of intensified selection. The gene DEAD-Box Helicase 6 (*DDX6*) presented a much higher value (1436.3) than the other genes; as K=6.82, that gene was under intensified selection. As in the CodeML, tests *PRKDC* exhibited a high LR value; furthermore, K was lower than 1 (0.46) indicating a relaxed selection.

Genes under selective pressure (those with high LRT scores) appear to be evenly distributed between all the chromosomes (Fig. 4).

Interrelations between test results

Branch site test results for each gene were compared to one another using the Phi-coefficient (or mean square contingency coefficient) that measures the association between two variables (Ekström 2011) (Fig. 5). As expected, the two tests for relaxed selection showed the highest association with a Phi-coefficient of 0.57; a gene inferred to be under relaxed selection according to the CodeML test is likely to also be found to be under relaxed selection according to the HyPhy test. As CodeML M0-M2 is a test that detects genes under selective pressure it is not surprising to see that it was associated with all the other tests. Positive and intensified selection did not show a high level of association implying that what is considered positive and intensified selection are different selective pressures. Finally, as a gene can be either under positive or relaxed selection for the CodeML tests and either intensified or relaxed for the HyPhy tests these two pairs showed a Phi-coefficient close to zero.

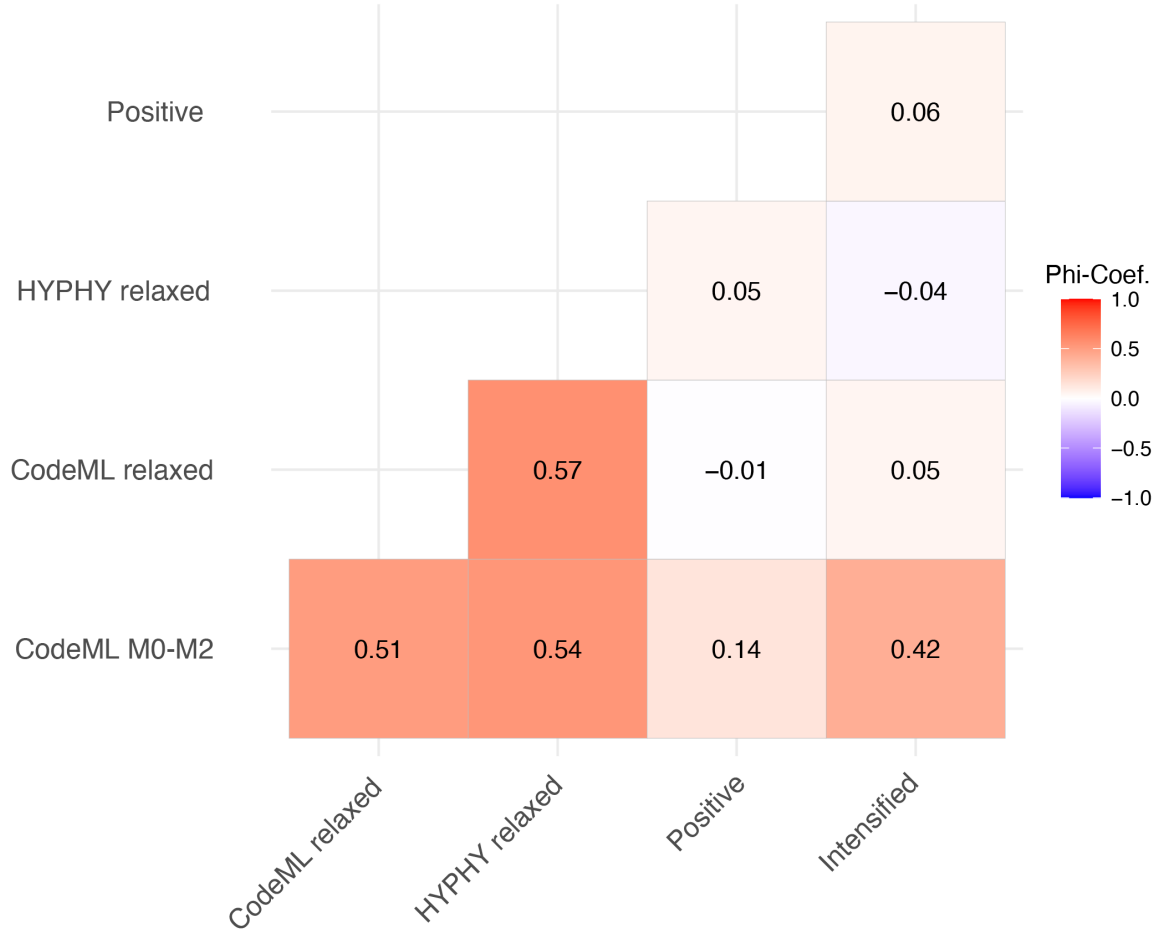


Figure 5: Variable association matrix between CodeML and HyPhy tests results. A positive Phi-coefficient (red) indicates an association between two test results and a Phi-coefficient close to zero indicates an absence of association.

The majority of genes that showed evidence of selective pressure according to the CodeML M0-M2 test were under relaxed (Fig. 6a), intensified, or positive selection (Fig. 6b). CodeML and HyPhy tests reached the same conclusions on relaxed selection for roughly half the genes.

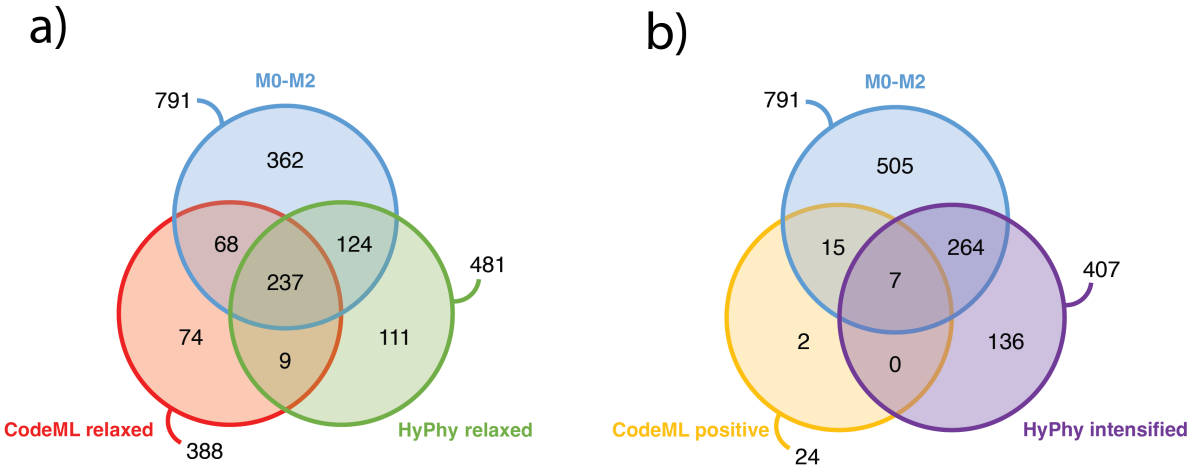


Figure 6: Venn diagrams showing the overlapping of CodeML and HyPhy test results; a) comparison between M0-M2 (different selective pressure) and genes under relaxed selection; b) comparison between M0-M2, genes under positive and intensified selection. Genes with Incomplete Lineage Sorting (ILS) are included in these diagrams.

Testing for bias due to incomplete lineage sorting and gene length

Approximately Unbiased (AU) tests were run on all CDS alignments to detect bias caused by incomplete lineage sorting (ILS). Out of the 11,557 genes, 802 significantly deviated from the species tree (Table 1). Between 5% and 13% of the genes that appeared to be under selection according to the different tests exhibited signs of incomplete lineage sorting.

Table 1: Number of genes under different types of selective pressures according to CodeML, HyPhy and Approximately Unbiased (AU) test results. The average length is the CDS alignment length after quality filtering.

	Number of genes	ILS (percentage)	Average length (bp)
M0-M2	791	61 (7.77%)	1145
CodeML relaxed	388	28 (7.21%)	1290
Positive	24	3 (12.50%)	980
HyPhy relaxed	481	28 (5.85%)	1284
Intensified	407	33 (8.11%)	1171
AU test	802	-	487
All genes	11557	802 (6.94%)	1080

To determine if alignment length influenced these results, a T-test was used for each comparison between the genes that showed significant results and those that did not. HyPhy

relaxed and CodeML relaxed both showed significant group differences with relaxed genes having, on average, longer CDS alignments than non-relaxed genes (Fig. 7). M0-M2, positive and intensified showed no significant differences. Genes that were significant for the approximately unbiased (AU) test were overwhelmingly shorter than the genes that were not significant.

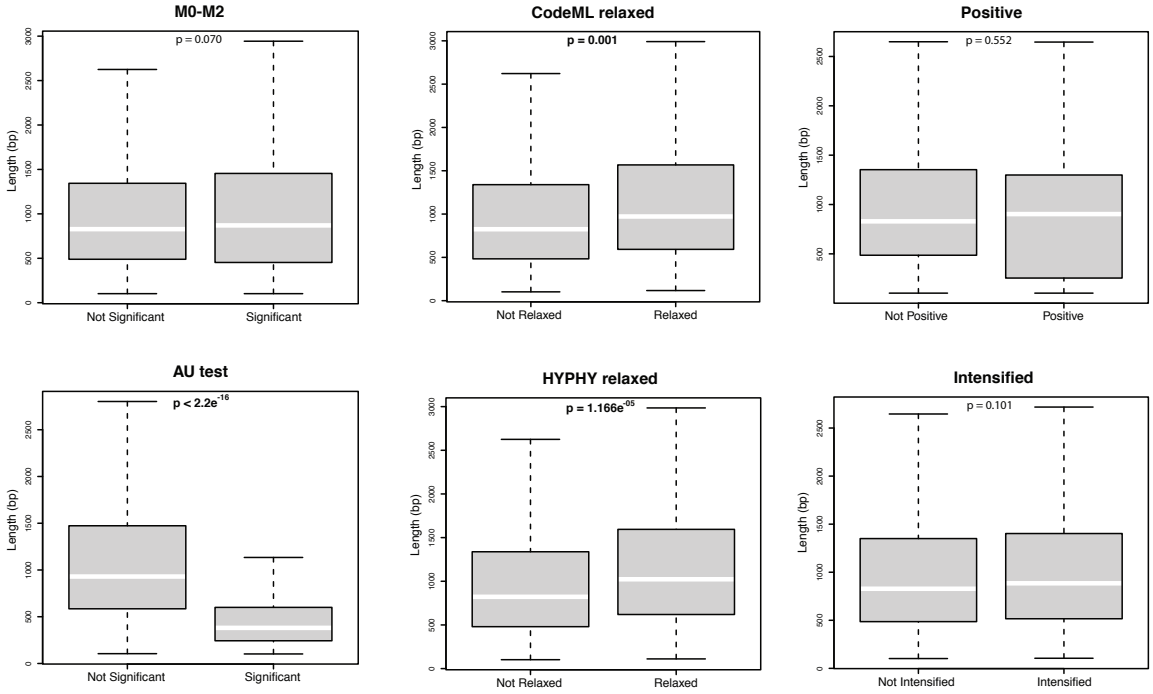


Figure 7: Comparison of gene length difference in relation to CodeML, HyPhy and AU test results.

Functions of genes under selection

The overrepresentation test results for relaxed genes, intensified genes and genes under selection (M0-M2) showed compelling evidence of enrichment fold and observed-expected values although almost none of the GO terms exhibited significant results for the False Discovery Rate (FDR) (Appendices C4 Tables 1, 2, and 3). The gene ontology overrepresentation test showed that genes under relaxed selection contained enrichment for functional categories that align well with the expected trait evolution associated with flightlessness such as muscle development function and circulatory system (Fig. 8). GO terms associated with the immune system, memory, visual perception and inner ear development (Appendix C4 Table 1) were also enriched. Genes under intensified selection in flightless rails were enriched for GO terms that did not seem to be directly associated with the evolution of

flightlessness (Appendix C4 Table 2). The number of genes under positive selection (24) was too low for robust overrepresentation testing. As expected, genes significant for the M0-M2 test were enriched for similar GO terms as either the relaxed or the intensified set of genes (Appendix C4 Table 3).

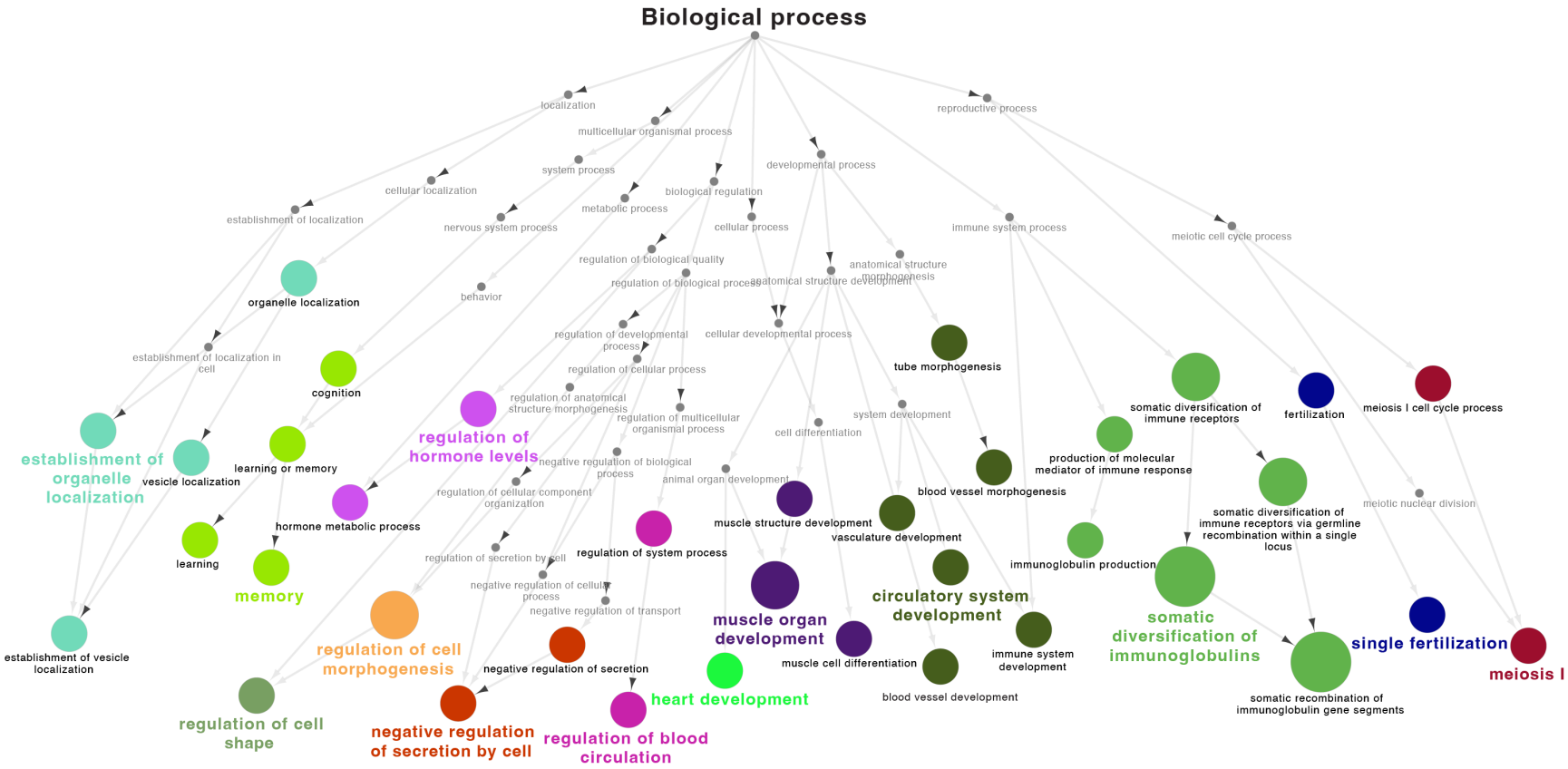


Figure 8: Hierarchical tree graph of overrepresented GO terms for relaxed genes on flightless rail lineages. Coloured nodes represent significantly enriched categories with node diameter inversely proportional to the Fisher's exact test p-value. Categories displayed were restricted to those with enrichment greater than two fold and associated with at least 4 genes. Node colours correspond to GO categorical function groups. Small grey nodes are not part of the enriched GO terms but show the GO pathways to enriched terms.

Discussion

After screening thousands of CDS alignments in seven rail species, hundreds of genes are found to be subject to selection of one sort or another in flightless species. Genes with high LR values for both CodeML (Figs. 4a, b and c) and HyPhy (Fig. 4d) tests appear to be evenly distributed across the genome.

We describe what is known about the functions of the six genes that show the strongest evidence of selection in flightless rails. The genes *ANKRD52* and *MAP3K14* show high LR values when comparing M1 and BS1 and when comparing BS1 and BS2 models which is clear evidence of positive selection (Fig. 4b and c). *ANKRD52*, according to Uniprot, is involved in the regulation of cardiac muscle contraction (The UniProt Consortium 2021). *MAP3K14* is part of the MAPK cascade that is involved in heart development (Saydmohammed et al. 2018, Romero-Becerra et al. 2020) and encodes mitogen-activated protein kinase kinase kinase 14. *PRKDC* has high LR values for M1-BS1 but not for BS1-BS2 tests and is therefore inferred as being subject to relaxed selection. It encodes a kinase that acts as a molecular sensor for DNA damage in chicken (The UniProt Consortium 2021) and has been identified as being responsible for severe combined immunodeficiency in mice (Fujimori et al. 2002). *SORL1* and *PDIA4* also show high LR values (Fig. 4a and b) and are under relaxed selection. *SORL1* is associated with Alzheimer's disease in humans (Rogaeva et al. 2007, Shafiei et al. 2019) and *PDIA4* GO terms are related to response to endoplasmic reticulum stress and protein folding (The UniProt Consortium 2021). Finally, *DDX6* shows extremely high LR and K values (HyPhy test, 1436 and 6.81 respectively) implying a strong intensified selection. This gene's GO terms contain, among others, negative regulation of neuron differentiation, P-body assembly, and stem cell population maintenance. Of the six genes listed above, the biological functions of *ANKRD52* and *MAP3K14* can be directly linked to the loss of flight. Indeed, cardiac muscle contraction is associated with the energy needed to fly, as we discuss later. Although these six genes are promising targets to investigate the molecular basis of flightlessness, it is likely that they are not the only genomic regions responsible for phenotypic changes. We rather expect a cluster of genes or small effects of hundreds of genes to be involved in the transition to flightlessness in birds (Machado et al. 2016, Burga et al. 2017, Sackton et al. 2019).

The correspondence of M0-M2 test results with the range of more specific tests (Fig. 6) indicates that this model comparison can effectively infer deviation from neutrality in volant and flightless species. CodeML and HyPhy relaxed tests agree on roughly 50% of the genes under selection, but despite their high Phi-coefficient (Fig. 5), substantial differences are observed between these two model likelihood comparison analyses. CodeML appears to be more conservative with 20% fewer genes (388) inferred as relaxed than HyPhy (481).

Tests for intensified and positive selection show very different results (Figs. 5 and 6) indicating that these two types of tests respond to different signals of non-neutral nucleotide substitution rates. A gene is considered to be under positive selection when $\omega>1$ in the flightless (foreground) species; nonsynonymous substitutions are fixed at a higher rate than synonymous substitution (Yang and Bielawski 2000). Intensified selection is the opposite of relaxed selection and is characterized by an increase of the intensity of both purifying and positive selection in the flightless (foreground) lineages compared to the volant (background) lineages (Fig. 1).

CDS alignment length differences are observed between gene groups; for instance, relaxed genes (according to both CodeML and HyPhy) are significantly longer than genes not under relaxed selection (Fig. 7). This result could either indicate that the branch-site models tests for relaxed selection are biased in favour of longer CDS alignment or that the d_N/d_S ratio is influenced by the gene length (Eyre-Walker 1996, Stoletzki and Eyre-Walker 2007). Genes that exhibited incomplete lineage sorting according to the AU tests were significantly shorter than genes that did not. This plausibly reflects the fact that short gene alignments are likely to have less phylogenetic signal on average, leading to selection of 'optimal' gene trees that do not reflect the true species tree (Shimodaira 2002). As a consequence, the AU test will tend to reject gene trees from short CDS alignments.

Several GO terms show evidence of overrepresentation in relaxed genes according to Fisher's exact test (p-values in Appendix C4 Table 1) but are not significant for FDR which assesses the rate of false positives when conducting multiple comparisons. The goal of this analysis is to detect overall functional trends in genes under relaxed selection.

Given the compelling ratios of observed/expected genes for significant GO terms, we are confident that this analysis is relevant and reliable for descriptive purposes.

Genes under relaxed selection are enriched for functional categories that are associated with the circulatory system (Fig. 8), suggesting that selective pressure on genes associated with the circulatory system in volant rails is significantly reduced or absent in flightless rails. Powered flight involves intense aerobic activity and an efficient cardiovascular system to deliver oxygen to the pectoral muscles during flight (Grubb 1983, Brusatte et al. 2015). One can assume that this physiological constraint is reduced in flightless species (McNab 1994). In birds, heart size is associated with flight capacity (Berger et al. 1970, Nespolo et al. 2018). A study on tinamous, the living birds with the smallest hearts in relation to their body size, showed that the aerobic metabolism required for flight is limited by heart size (Altimiras et al. 2017). This study also highlights the correlation between heart development and differential gene expression in MAPK kinases. Flightless birds do not need to regulate their body temperature as much as their volant relatives (Scott 2011) which could also explain why genes associated with the circulatory system are relaxed. Flightless birds have evolved smaller flight muscles than volant birds (Wright et al. 2016, Gaspar et al. 2020) which can explain why we observe the relaxation of genes associated with muscle development.

Of the nuclear genes previously identified as associated with the ability to fly (Bickley and Logan 2014, Machado et al. 2016, Burga et al. 2017, Farlie et al. 2017), only *AHSG* (Alpha-2-HS-glycoprotein) exhibits signs of relaxed selection in flightless rails (according to CodeML but not HyPhy). That gene is a serum and bone resident glycoprotein that is involved in bone geometry and mineral density (Jiang et al. 2007, Yang et al. 2007). As birds developed a lightweight skeleton (Dumont 2010), genes involved in bone development are known to be associated with flight (Machado et al. 2016). It is therefore not surprising to observe a relaxation of the selective pressure applying on *AHSG* in flightless rails.

Genes associated with the immune response (GO “somatic diversification of immunoglobulins”) show evidence of relaxed selection (Fig. 8). However, genes under intensified selection were also enriched in immune functions such as “regulation of

leukocyte migration" and "inflammatory response". It, therefore, appears that volant and flightless rails took different evolutionary trajectories in terms of physiological defence against infections. Based on studies made on migratory birds, immune defences increase with the number of different habitats the birds are exposed to (Møller and Erritzøe 1998, Eikenaar et al. 2020). The loss of flight undeniably reduces the number and diversity of those habitats. As maintaining a strong immune system can be energy consuming (Norris and Evans 2000), one can suppose that physiological defence against infections would be different between flightless and volant birds.

Flightless rails limited habitat compared to their volant relatives may also explain the overrepresentation of relaxed genes associated with memory and visual perception. Indeed, one can assume that these traits are under less selective pressure in a smaller less variable environment.

P. hochstetteri and *G. australis* are endemic flightless birds from New Zealand, and *A. rogersi* and *Z. atra* are found on very small islands, respectively Inaccessible Island and Henderson Island. Due to this sampling, it is likely that the genes that appear to evolve in a similar way in flightless rails not only reflect an adaptation to flightlessness but also result from selective pressures associated with the island habitat.

It is important to note that coding regions represent a small part of the genome and a significant proportion of the genetic basis of flightlessness could also be nonexonic. For instance, a recent study on the origin of flightlessness in ratites shows that evolutionary changes associated with the loss of flight are primarily regulatory. It also suggests that protein-coding evolution, in contrast to regulatory evolution, is mainly lineage-specific (Sackton et al. 2019) and therefore is less appropriate to study convergent evolution. Burga et. al (2017), on top of finding a function-altering variant in the *CUX1* (Cut like Homebox 1) gene to be a strong candidate to contribute to loss of flight, found 11 ultraconserved noncoding regions that show accelerated evolution in the flightless cormorant.

Based on our robust and consistent results we think that investigating the coding regions is a necessary first step to take to resolve the molecular basis of flightlessness. Of course,

more work on non-coding regions and potentially on gene expression is also needed to reveal more of the genetic complexity behind this evolutionary phenomenon.

Furthermore, as flightlessness independently evolved multiple times in rails, it is likely that part of the evolutionary processes that led to loss of flight is specific to each flightless species. We however decided to take a wider approach in the present study to assess the genetic regions that evolve in similar ways in all flightless rails.

In conclusion, out of the 11,577 investigated genes, 6 show obvious evidence of strong selection in flightless rails: *ANKRD52*, *MAP3K14*, *PRKDC*, *PDIA4*, *SORL1*, and *DDX6*. We observe that genes under relaxed selection are enriched in GO functional categories that are either directly associated with flight (regulation of blood circulation and muscle development) or with the ecological change consequent to the loss of flight (defence against infections and memory). This study presents a reliable method to assess the selective pressures involved in avian flightlessness at the genome level.

Data availability

The phylogenetic data, control files and script to run CodeML are available in supplementary data on the following website:

https://figshare.com/s/aa0ffc1312d95d566ae2

References

Altimiras, J., Lindgren, I., Giraldo-Deck, L. M., Matthei, A. and Garitano-Zavala, Á. 2017. Aerobic performance in tinamous is limited by their small heart. A novel hypothesis in the evolution of avian flight. – Sci Rep 7: 15964.

Berger, M., Hart, J. S. and Roy, O. Z. 1970. Respiration, oxygen consumption and heart rate in some birds during rest and flight. – Z. Vergl. Physiol. 66: 201–214.

Besnard, G., Muasya, A. M., Russier, F., Roalson, E. H., Salamin, N. and Christin, P.-A. 2009. Phylogenomics of C(4) photosynthesis in sedges (Cyperaceae): multiple appearances and genetic convergence. – Mol Biol Evol 26: 1909–1919.

Bickley, S. R. B. and Logan, M. P. O. 2014. Regulatory modulation of the T-box gene Tbx5 links development, evolution, and adaptation of the sternum. – PNAS 111: 17917–17922.

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W.-H., Pagès, F., Trajanoski, Z. and Galon, J. 2009. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. – Bioinformatics 25: 1091–1093.

Brusatte, S. L., O'Connor, J. K. and Jarvis, E. D. 2015. The Origin and Diversification of Birds. – Current Biology 25: R888–R898.

Burga, A., Wang, W., Ben-David, E., Wolf, P. C., Ramey, A. M., Verdugo, C., Lyons, K., Parker, P. G. and Kruglyak, L. 2017. A genetic signature of the evolution of loss of flight in the Galapagos cormorant. – Science 356: eaal3345.

Campagna, L., McCracken, K. G. and Lovette, I. J. 2019. Gradual evolution towards flightlessness in steamer ducks*. – Evolution 73: 1916–1926.

Castresana, J. 2000. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. – Mol Biol Evol 17: 540–552.

Chen, Y., Ye, W., Zhang, Y. and Xu, Y. 2015. High speed BLASTN: an accelerated MegaBLAST search tool. – Nucleic Acids Research 43: 7762–7768.

Chikina, M., Robinson, J. D. and Clark, N. L. 2016. Hundreds of Genes Experienced Convergent Shifts in Selective Pressure in Marine Mammals. – Mol Biol Evol 33: 2182–2192.

Cooper, K. L., Sears, K. E., Uygur, A., Maier, J., Baczkowski, K.-S., Brosnahan, M., Antczak, D., Skidmore, J. A. and Tabin, C. J. 2014. Patterning and post-patterning modes of evolutionary digit loss in mammals. – Nature 511: 41–45.

Cui, R., Medeiros, T., Willemsen, D., Iasi, L. N. M., Collier, G. E., Graef, M., Reichard, M. and Valenzano, D. R. 2019. Relaxed Selection Limits Lifespan by Increasing Mutation Load. – Cell 178: 385–399.

Dumont, E. R. 2010. Bone density and the lightweight skeletons of birds. – Proceedings of the Royal Society B: Biological Sciences 277: 2193–2198.

Eikenaar, C., Hessler, S. and Hegemann, A. 2020. Migrating birds rapidly increase constitutive immune function during stopover. – Royal Society Open Science 7: 192031.

Ekström, J. 2011. The Phi-coefficient, the Tetrachoric Correlation Coefficient, and the Pearson-Yule Debate.

Eyre-Walker, A. 1996. Synonymous codon bias is related to gene length in Escherichia coli: selection for translational accuracy? – Molecular Biology and Evolution 13: 864–872.

Faircloth, B. C. 2016. PHYLUCE is a software package for the analysis of conserved genomic loci. – Bioinformatics 32: 786–788.

Farlie, P. G., Davidson, N. M., Baker, N. L., Raabus, M., Roeszler, K. N., Hirst, C., Major, A., Mariette, M. M., Lambert, D. M., Oshlack, A. and Smith, C. A. 2017. Co-

option of the cardiac transcription factor Nkx2.5 during development of the emu wing. – Nature Communications 8: 132.

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. – J Mol Evol 17: 368–376.

Fujimori, A., Hashimoto, H., Araki, R., Saito, T., Sato, S., Kasama, Y., Tsutsumi, Y., Mori, M., Fukumura, R., Ohhata, T., Tatsumi, K. and Abe, M. 2002. Sequence Analysis of 193.4 and 83.9 kbp of Mouse and Chicken Genomic DNAs Containing the Entire Prkdc (DNA-PKcs) Gene1. – Radiation Research 157: 298–305.

Garcia–R, J. C., Gibb, G. C. and Trewick, S. A. 2014. Eocene Diversification of Crown Group Rails (Aves: Gruiformes: Rallidae). – PLOS ONE 9: e109635.

Gaspar, J., Gibb, G. C. and Trewick, S. A. 2020. Convergent morphological responses to loss of flight in rails (Aves: Rallidae). – Ecology and Evolution 10: 6186–6207.

Grubb, B. R. 1983. Allometric relations of cardiovascular function in birds. – American Journal of Physiology-Heart and Circulatory Physiology 245: H567–H572.

Jiang, H., Lei, S., Xiao, S., Chen, Y., Sun, X., Yang, F., Li, L., Wu, S. and Deng, H. 2007. Association and linkage analysis of COL1A1 and AHSG gene polymorphisms with femoral neck bone geometric parameters in both Caucasian and Chinese nuclear families5. – Acta Pharmacologica Sinica 28: 375–381.

Kimura, M. 1977. Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. – Nature 267: 275–276.

Kirchman, J. J. 2012. Speciation of Flightless Rails on Islands: A DNA-Based Phylogeny of the Typical Rails of the Pacific. – The Auk 129: 56–69.

Kirchman, J. J., Rotzel McInerney, N., Giarla, T. C., Olson, S. L., Slikas, E. and Fleischer, R. C. 2021. Phylogeny based on ultra-conserved elements clarifies the evolution

of rails and allies (Ralloidea) and is the basis for a revised classification. – Ornithology 138: ukab042.

Lahti, D. C., Johnson, N. A., Ajie, B. C., Otto, S. P., Hendry, A. P., Blumstein, D. T., Coss, R. G., Donohue, K. and Foster, S. A. 2009. Relaxed selection in the wild. – Trends in Ecology & Evolution 24: 487–496.

Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T. and 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.

Liu, Y., Liu, S., Yeh, C.-F., Zhang, N., Chen, G., Que, P., Dong, L. and Li, S. 2018. The first set of universal nuclear protein-coding loci markers for avian phylogenetic and population genetic studies. – Scientific Reports 8: 15723.

Livezey, B. C. 2003. Evolution of Flightlessness in Rails. – American Ornithologists' Union.

Machado, J. P., Johnson, W. E., Gilbert, M. T. P., Zhang, G., Jarvis, E. D., O'Brien, S. J. and Antunes, A. 2016. Bone-associated gene evolution and the origin of flight in birds. – BMC Genomics 17: 371.

McNab, B. K. 1994. Energy Conservation and the Evolution of Flightlessness in Birds. – The American Naturalist 144: 628–642.

Mendes, F. K. and Hahn, M. W. 2016. Gene Tree Discordance Causes Apparent Substitution Rate Variation. – Systematic Biology 65: 711–721.

Mi, H., Muruganujan, A., Ebert, D., Huang, X. and Thomas, P. D. 2019. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. – Nucleic Acids Research 47: D419–D426.

Miller, M. A., Pfeiffer, W. and Schwartz, T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. – 2010 Gateway Computing Environments Workshop (GCE): 1–8.

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A. and Lanfear, R. 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. – Molecular Biology and Evolution 37: 1530–1534.

Møller, A. P. and Erritzøe, J. 1998. Host immune defence and migration in birds. – Evolutionary Ecology 12: 945–953.

Nespolo, R. F., González-Lagos, C., Solano-Iguaran, J. J., Elfwing, M., Garitano-Zavala, A., Mañosa, S., Alonso, J. C. and Altimiras, J. 2018. Aerobic power and flight capacity in birds: a phylogenetic test of the heart-size hypothesis. – Journal of Experimental Biology 221: jeb.162693

Norris, K. and Evans, M. R. 2000. Ecological immunology: life history trade-offs and immune defense in birds. – Behavioral Ecology 11: 19–26.

Olson, S. L. 1973. Evolution of the rails of the South Atlantic islands (Aves: Rallidae). – Smithsonian Contributions to Zoology.

Pond, S. L. K., Frost, S. D. W. and Muse, S. V. 2005. HyPhy: hypothesis testing using phylogenies. – Bioinformatics (Oxford, England) 21: 676–679.

Ranwez, V., Douzery, E. J. P., Cambon, C., Chantret, N. and Delsuc, F. 2018. MACSE v2: Toolkit for the Alignment of Coding Sequences Accounting for Frameshifts and Stop Codons. – Mol Biol Evol 35: 2582–2584.

Ripley, S. D., Lansdowne, J. F. and Olson, S. L. 1977. Rails of the World: A Monograph of the Family Rallidae. – M. F. Feheley Publishers.

Roff, D. A. 1994. The evolution of flightlessness: Is history important? – Evol Ecol 8: 639–657.

Rogaeva, E., Meng, Y., Lee, J. H., Gu, Y., Kawarai, T., Zou, F., Katayama, T., Baldwin, C. T., Cheng, R., Hasegawa, H., Chen, F., Shibata, N., Lunetta, K. L., Pardossi-Piquard, R., Bohm, C., Wakutani, Y., Cupples, L. A., Cuenco, K. T., Green, R. C., Pinessi, L., Rainero, I., Sorbi, S., Bruni, A., Duara, R., Friedland, R. P., Inzelberg, R., Hampe, W., Bujo, H., Song, Y.-Q., Andersen, O. M., Willnow, T. E., Graff-Radford, N., Petersen, R. C., Dickson, D., Der, S. D., Fraser, P. E., Schmitt-Ulms, G., Younkin, S., Mayeux, R., Farrer, L. A. and St George-Hyslop, P. 2007. The neuronal sortilin-related receptor SORL1 is genetically associated with Alzheimer disease. – Nat Genet 39: 168–177.

Romero-Becerra, R., Santamans, A. M., Folgueira, C. and Sabio, G. 2020. p38 MAPK Pathway in the Heart: New Insights in Health and Disease. – International Journal of Molecular Sciences 21: 7412.

Sackton, T. B., Grayson, P., Cloutier, A., Hu, Z., Liu, J. S., Wheeler, N. E., Gardner, P. P., Clarke, J. A., Baker, A. J., Clamp, M. and Edwards, S. V. 2019. Convergent regulatory evolution and the origin of flightlessness in palaeognathous birds. – Science 364: 74–78.

Saribasak, H., Maul, R. W., Cao, Z., Yang, W. W., Schenten, D., Kracker, S. and Gearhart, P. J. 2012. DNA polymerase ζ generates tandem mutations in immunoglobulin variable regions. – Journal of Experimental Medicine 209: 1075–1081.

Saydmohammed, M., Vollmer, L. L., Onuoha, E. O., Maskrey, T. S., Gibson, G., Watkins, S. C., Wipf, P., Vogt, A. and Tsang, M. 2018. A High-Content Screen Reveals New Small-Molecule Enhancers of Ras/Mapk Signaling as Probes for Zebrafish Heart Development. – Molecules 23: 1691.

Schrider, D. R., Hourmozdi, J. N. and Hahn, M. W. 2011. Pervasive Multinucleotide Mutational Events in Eukaryotes. – Current Biology 21: 1051–1054.

Scott, G. R. 2011. Elevated performance: the unique physiology of birds that fly at high altitudes. – Journal of Experimental Biology 214: 2455–2462.

Shafiei, H., Bakhtiarizadeh, M. R. and Salehi, A. 2019. Large-scale potential RNA editing profiling in different adult chicken tissues. – Animal Genetics 50: 460–474.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. – Genome Res 13: 2498–2504.

Shimodaira, H. 2002. An Approximately Unbiased Test of Phylogenetic Tree Selection. – Systematic Biology 51: 492–508.

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

Steadman, D. W. 1995. Prehistoric Extinctions of Pacific Island Birds: Biodiversity Meets Zooarchaeology. – Science 267: 1123–1131.

Stern, D. L. 2013. The genetic causes of convergent evolution. – Nat Rev Genet 14: 751–764.

Stoletzki, N. and Eyre-Walker, A. 2007. Synonymous Codon Usage in Escherichia coli: Selection for Translational Accuracy. – Molecular Biology and Evolution 24: 374–381.

Taylor, B. 1998. Rails: A Guide to the Rails, Crakes, Gallinules and Coots of the World. – Bloomsbury Publishing.

The UniProt Consortium. 2021. UniProt: the universal protein knowledgebase in 2021. – Nucleic Acids Research 49: D480–D489.

Trewick, S. A. 1997a. Flightlessness and phylogeny amongst endemic rails (Aves: Rallidae) of the New Zealand region. – Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 352: 429–446.

Trewick, S. A. 1997b. Sympatric flightless rails Gallirallus dieffenbachii and G. modestus on the Chatham Islands, New Zealand; morphometrics and alternative evolutionary scenarios. – Journal of the Royal Society of New Zealand 27: 451–464.

Venkat, A., Hahn, M. W. and Thornton, J. W. 2018. Multinucleotide mutations cause false inferences of lineage-specific positive selection. – Nature Ecology & Evolution 2: 1280–1288.

Wertheim, J. O., Murrell, B., Smith, M. D., Kosakovsky Pond, S. L. and Scheffler, K. 2015. RELAX: Detecting Relaxed Selection in a Phylogenetic Framework. – Mol Biol Evol 32: 820–832.

Wright, N. A., Steadman, D. W. and Witt, C. C. 2016. Predictable evolution toward flightlessness in volant island birds. – PNAS 113: 4765–4770.

Yang, Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. – Mol Biol Evol 24: 1586–1591.

Yang, Z. and Bielawski, J. P. 2000. Statistical methods for detecting molecular adaptation. – Trends in Ecology & Evolution 15: 496–503.

Yang, Y.-J., Wang, Y.-B., Lei, S.-F., Long, J.-R., Shen, H., Zhao, L.-J., Jiang, D.-K., Xiao, S.-M., Chen, X.-D., Chen, Y. and Deng, H.-W. 2007. AHSG gene polymorphisms are associated with bone mineral density in Caucasian nuclear families. – Eur J Epidemiol 22: 527–532.

Zhang, J., Nielsen, R. and Yang, Z. 2005. Evaluation of an Improved Branch-Site Likelihood Method for Detecting Positive Selection at the Molecular Level. – Mol Biol Evol 22: 2472–2479.

Zhen, Y., Aardema, M. L., Medina, E. M., Schumer, M. and Andolfatto, P. 2012. Parallel Molecular Evolution in an Herbivore Community. – Science 337: 1634.

Chapter Five

Convergent accelerated rate of molecular evolution and functional impact of amino acid substitutions in flightless rails (Aves: Rallidae)

Abstract

Distantly related species can independently evolve similar phenotypic traits in response to similar environmental challenges. A classic example of that convergent evolution is the loss of flight in island birds. Several species from the rail family (Aves: Rallidae) independently lost the ability to fly which makes this clade a good model to study the evolution of avian flightlessness. Here we used genomic data from rails and their distant relatives the cranes (Aves: Gruidae) to investigate the molecular basis of the loss of flight. Hundreds of genes exhibiting an accelerated evolutionary rate in flightless species were identified. These genes were enriched in biological functions associated with gene expression regulation or in functions that align well with the expected adaptation to flightlessness like the development of bones, limbs, muscles, and heart. Moreover, several proteins carrying function-altering amino acid substitutions associated with flightlessness were found. These proteins were enriched in biological GO functional categories such as immune response, DNA conformation, cilium, muscle and circulatory system development and lipid metabolic process. Regarding the genetic pathway to flightlessness, we observed signs of both convergent whole-gene evolution and lineage-specific pathways. This suggests that to converge toward the flightless phenotype, some genomic regions undergo similar selective pressures in all flightless species when other regions go through distinct evolutionary mechanism in each flightless rail.

Keywords: Rallidae, Flightlessness, Convergent evolution, Function-altering substitution, Accelerated evolutionary rate

Note for the reader

The content and methods presented in this section may appear similar to Chapter 4 as both chapters compare the genetics of volant and flightless rails to identify the genomic regions involved of flightlessness. The difference resides in the fact that the current chapter only includes the analyses with amino acid sequences while nucleotide sequences are used in Chapter 4. Five crane species are also added in some of the analysis to make evolutionary rate estimation for the background (volant) lineages more reliable (Partha et al. 2019). The cranes were not added to Chapter 4 as nucleotide alignments exhibit more phylogenetic signal than amino acid alignments. I decided to separate nucleotide and amino acid analyses in two chapters for ease of reading.

Introduction

Rallidae

The bird family known as rails (Aves: Rallidae) originated during the Eocene around 40 million years ago (Garcia–R et al. 2014) and diversified into over 130 extant species (Steadman 1995, Kirchman 2012, Garcia–R et al. 2014). Despite the terrestrial lifestyle of the majority of the species (Taylor 1998), this bird family displays remarkable dispersal capacity resulting in broad distribution and the colonization of numerous oceanic islands (Olson 1973, Ripley et al. 1977, Garcia–R et al. 2014). The extant diversity includes (or recently included) more than 30 flightless species the majority of which are endemic to single islands (Steadman 1995, Kirchman 2012). This demonstrates their ancestors had been volant (Trewick 1997a, b) and that loss of flight independently evolved multiple times which makes the rail family a good model to study the evolutionary process behind the transition to flightlessness.

Evolution of avian flightlessness

Flightless species share morphological characters that distinguish them from their volant relatives. Flightless rails typically exhibit smaller sterna and wings than volant taxa along with wider pelves and more robust femora (Livezey 2003, Gaspar et al. 2020). Moreover, it has been shown that these differences are independent of phylogeny and instead

demonstrate convergent morphological adaptation associated with a walking ecology (Gaspar et al. 2020). Working with genomes of representative volant and flightless rails present the opportunity to identify the genetic basis for this convergent evolution.

We hypothesize that loss of flight is likely to have a polygenic basis (Stern 2000, Machado et al. 2016) and that the underlying genomic regions would be enriched in biological pathways involved in functions associated with the ability to fly (Burga et al. 2017). As natural selection operates mainly at the protein level (Kimura 1977), we screened whole genomes from seven rails and compared amino acid sequences between volant and flightless species. To study the genetics behind the transition to flightlessness two approaches were used; one based on the convergent changes in the evolutionary rates of genes and the other based on the functional impact of protein amino acid changes in flightless species.

If genes associated with the evolution of flightlessness in independent lineages were subject to parallel changes in natural selection this would result in convergent shift in their evolutionary rates (Chikina et al. 2016). To test this idea we looked at genomic regions of flightless taxa that showed similar rate shift in comparison to volant taxa.

It is also likely that genes closely linked to functional flight are less constrained after a transition to flightlessness. Therefore, the proteins encoded by these genes would retain altered structure more frequently in flightless rails. Our strategy used genome sequences from four flightless rails and their closest volant relatives to evaluate the functional impact of genetic variants.

An alternative, but not necessarily conflicting, hypothesis to the evolution of loss of flight through similar molecular mechanisms in all flightless species would involve lineage-specific genetic pathways to flightlessness. Indeed, as flightlessness evolved independently in most flightless rails, the molecular basis underlying that transition may be different in each species. To investigate that we also examined function altering amino acid changes in each flightless species rather than considering all the flightless birds as a group.

In this research, we identified hundreds of genes that have an accelerated evolutionary rate in flightless species and found several proteins with altered function that are robust candidates to explain the molecular basis of flightlessness. These proteins are enriched in biological functions that can be associated with the ability to fly.

Methods

Gene alignments and filtering

Coding sequences (CDS) were retrieved from seven rail (family Rallidae) genomes, *Porphyrio melanotus* (Temminck, 1820), *Porphyrio hochstetteri* (Meyer, 1883), *Gallirallus philippensis* (Linnaeus, 1766), *Gallurallus australis* (Sparrman, 1786), *Atlantisia rogersi* (Lowe, 1923), *Fulica atra* (Linnaeus, 1758) and *Zapornia atra* (North, 1908), and five crane (family Gruidae) genomes *Balearica regulorum* (Bennett, 1834), *Antigone vipio* (Pallas, 1811), *Grus americana* (Linnaeus, 1758), *Grus monachal* (Temminck, 1835), and *Grus nigricollis* (Przhevalsky, 1876) (see Chapter 3 for more information about genome assembly, annotations and CDS retrieval).

Cranes (family Gruidae), like rails (family Rallidea) are part of the Gruiformes order and were included to increase the number of species in the analysis. As data for only three volant rails were available, we added five volant crane species, which, although more distantly related, make evolutionary rate estimation for the background (volant) lineages more reliable (see "Relative Evolution Rate" section below) (Partha et al. 2019).

The CDSs of each species were grouped by gene using Geneious v11.1.4 (https://www.geneious.com) and aligned based on amino acid codons using MACSE v2.03 (Ranwez et al. 2018) with default settings. MACSE outputs nucleotide alignments and the corresponding amino acid alignments, both of which were used in the following analyses. Ambiguous alignment positions and gaps were eliminated using GBlocks v0.91b (Castresana 2000) (settings: minimum length of a block =5 and default settings for the rest). The alignments that were shorter than 100 base pairs or 33 amino acids and the alignments that had a nucleotide pairwise identity (visualized on Geneious) lower than 85% were excluded from further analysis. The resulting dataset comprised 12,316 protein

alignments before filtering and 11,192 after (9.1% filtered). Of these 11,192 alignments, 10,243 contained at least one amino acid substitution.

Phylogeny

A species tree topology was required to evaluate difference in evolutionary rate between volant and flightless species. Therefore, phylogenetic inference for 12 species (7 rails and 5 cranes) was performed using 20 genes selected as suitable markers for avian phylogenetic reconstruction (Liu et al. 2018). The 20 CDS alignments were concatenated into a 46,972 bp alignment using Phyluce v1.7.1 (Faircloth 2016) with the default settings and the best-fit partitioning scheme was determined using PartitionFinder2 (Lanfear et al. 2017) via the CIPRES Science Gateway (Miller et al. 2010)(see list of genes and partitions in supplementary data). Maximum Likelihood (ML) analyses were implemented in RaxML v8.2.10 (Stamatakis 2014) via the CIPRES Science Gateway with bootstrapping automatically stopped employing the majority rule criterion. The consensus tree was then visualized in Geneious (Fig. 1).

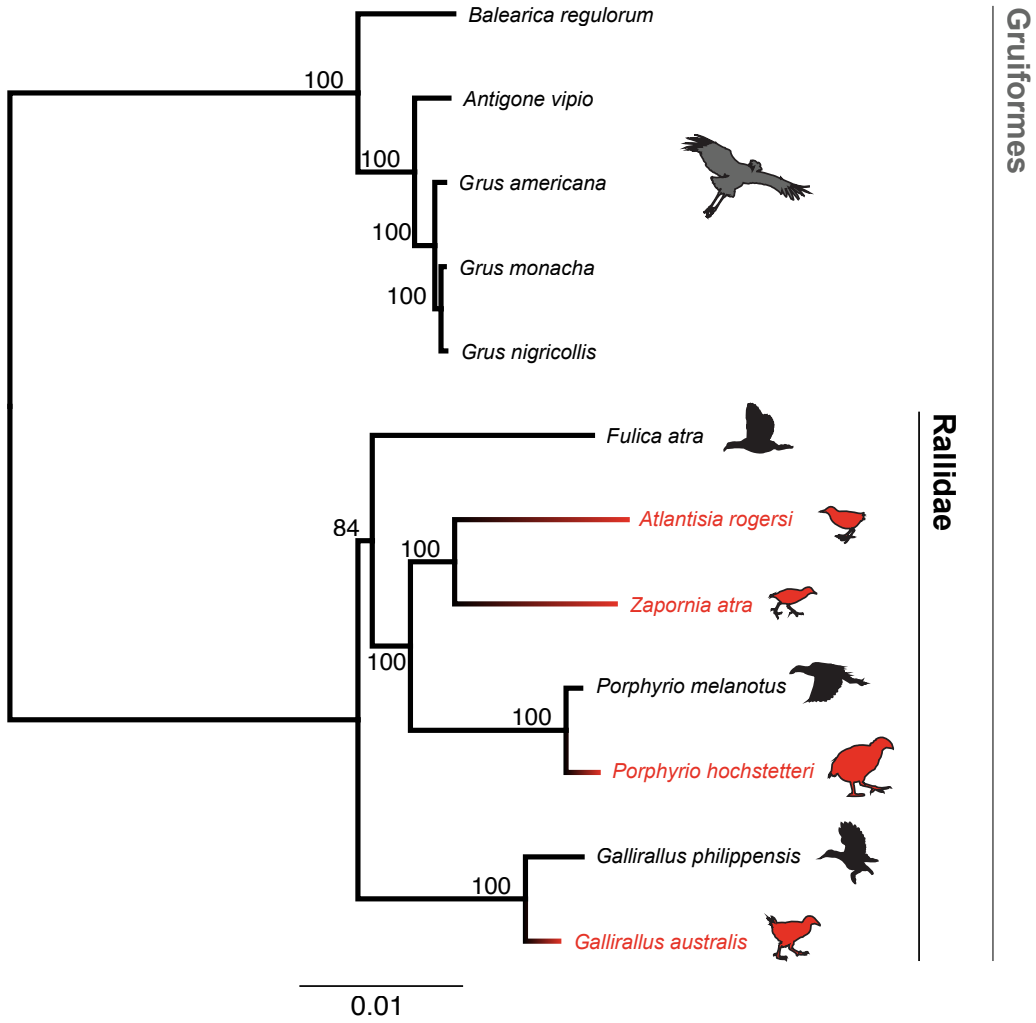


Figure 1: Maximum likelihood (RaxML V8) phylogeny of three volant (black) and four flightless (red) rails (Aves: Rallidae) and five cranes (all volant), inferred from a 46,972bp alignment of 20 nuclear genes (Liu et al. 2018); bootstrap supports are indicated for each node. Red terminals branches symbolise the independent emergence of flightlessness.

Relative Evolution Rate

We tested for convergent shifts in rates of molecular evolution in flightless species using the R package *RERconverge* (Kowalczyk et al. 2019). *RERconverge* takes a list of protein trees as an input to calculate the relative evolution rate (RER) of each branch and determines if, for each protein tree, the foreground branches have a significantly different evolution rate than the background branches. In this context the foreground lineages were the flightless species (*A. rogersi*, *G. australis*, *P. hochstetteri* and *Z. atra*) and the background lineages were the eight volant species (Fig. 1). The PAML (Yang 2007)

program *aaml* was applied to the amino acid alignments of the 12 species. For each gene amino acid alignment, branch lengths were estimated from phylogenetic analysis under the empirical model of amino acid substitution rate (aaml model "Empirical + F", see control file in supplementary data). *RERconverge* was then implemented using a Docker image following the instructions in https://github.com/nclark-lab/RERconverge. The protein trees were extracted from the *aaml* outputs and raw branch lengths were transformed (transform = "sqrt", weighted= T, and scale =T) into relative rates using a projection operator method (Sato et al. 2005). The flightless lineages were defined as foreground branches and the flightlessness character was set as terminal. Proteins that exhibited significantly different evolution rates between flightless and volant species ($p<0.05$) were assigned to either of two groups; one where the evolution rate was higher in the volant species (Rho <0) and the other where the evolution rate was higher in the flightless species (Rho >0).

Protein variants with altered functions

The phylogeny-corrected variant effect predictor PROVEAN (Choi et al. 2012) was applied at a genome-scale to evaluate the impact of amino acid substitutions on protein functions. To do so, differences between the protein sequences of volant and flightless species were investigated. As all flightless rails have volant ancestors, we predict that proteins involved in flight may carry function-altering mutations in flightless rails when compared to their volant relatives (which is the closest we have to their volant ancestor). PROVEAN requires two inputs, a reference protein sequence (in this case a volant species sequence) and a mutation file that indicates the differences between the reference sequence and a test sequence (in this case a flightless species). After confirming that the amino acid alignments did not contain any gaps or ambiguous sites, pairs of volant-flightless species were assembled based on phylogenetic inference. *Gallirallus philippensis* was used as a reference for *Gallirallus australis* and *Porphyrio melanotus* was used as a reference for *Porphyrio hochstetteri* on the basis of their respective relatedness. However, the flightless species *Atlantisia rogersi* and *Zapornia atra* did not have closely related volant species for which the whole genome data were available (Fig. 1). Therefore, for these two species PROVEAN analyses were run twice, once using *P. melanotus* as a reference and once using *F. atra* as a reference (these are the two closest

volant species). As results using these two references were highly correlated (see results section) we subsequently used *P. melanotus* as the reference for both *A. rogersi* and *Z. atra.*

In reporting results, reference to a flightless species containing function-altering substitutions indicates that the protein function is altered when compared to the reference branch (Table 1). All pairwise permutations of volant-volant species, considering the volant species in each case as both reference and test sequence, were also investigated to serve as a negative control. Each control has a code that starts with the first letters of the reference species and finished with the first letters of the test species; for example, *F. atra* - *G. philippensis* code is Fa_Gp. The genetic distances were estimated based on the phylogenetic inference presented before (Fig. 1).

Table 1: PROVEAN pairs: 4 pairs of volant-flightless rails and 6 pairs of volant-volant rails (as control). # proteins is the number of AA alignments that contains at least one amino acid substitution. Genetic distance is given as substitutions per site.

	Reference branch	Test branch	Genetic distance	# Proteins	Code
Volant-Flightless	*G. philippensis*	*G. australis*	0.0068	5,280	*G. australis*
	P. melanotus	*A. rogersi*	0.0309	8,361	*A. rogersi*
	P. melanotus	*P. hochstetteri*	0.0037	4,801	*P. hochstetteri*
	P. melanotus	*Z. atra*	0.0294	8,336	*Z. atra*
Volant-Volant (Control)	*F. atra*	*G. philippensis*	0.0371	8,486	Fa_Gp
	F. atra	*P. melanotus*	0.0341	8,446	Fa_Pm
	G. philippensis	*F. atra*	0.0371	8,486	Gp_Fa
	G. philippensis	*P. melanotus*	0.0351	8,465	Gp_Pm
	P. melanotus	*F. atra*	0.0341	8,446	Pm_Fa
	P. melanotus	*G. philippensis*	0.0351	8,465	Pm_Gp

Of the 11,192 amino acids alignments that were examined, ten sets of pairwise alignments were created: one for each bird pair (Table 1). Only alignments that included at least one substitution were retained for the PROVEAN analysis resulting in sets containing between 4,801 (*P. melanotus* - *P. hochstetteri*) and 8,486 (*F. atra* - *G. philippensis*) gene alignments.

Based on these ten sets of two species alignments, mutation files were generated using a custom R script (see supplementary data). PROVEAN v1.1.5 was implemented on the

high-performance computing facility provided by the New Zealand eScience Infrastructure (NeSI, www.nesi.org.nz) using a slurm script (supplementary material) with the following prerequisites: NCBI BLAST 2.2.30, CD-HIT 4.6, NCBI nr (non-redundant) protein database 2020-01. A PROVEAN score was generated for each substitution; the more negative the score the more likely the substitution impacts protein function. The score threshold was set at −2.5 following the software instructions. For practical reasons and due to the high number of proteins to study, only the substitutions and not the indels were investigated.

Overrepresentation tests

For each flightless species, a list of proteins with altered functions was generated. These proteins were defined as amino acid sequences that contained at least two function-altering substitutions (receiving a PROVEAN score < −2.5) and a proportion of function-altering substitutions over 25%. The same approach was used for the six control pairs. The proteins found to be altered in both a flightless species and at least one of the controls were excluded. For the list of protein with altered functions of each flightless species an overrepresentation test for Gene Ontology (GO) biological process was applied online using PANTHER (Protein Analysis Through Evolutionary Relationships, http://pantherdb.org) (Mi et al. 2019) with the *Gallus gallus* chicken genome annotations for reference. The overrepresentation test results showed the observed and expected number of genes corresponding with each GO term as well as the fold enrichment, the p-value as determined by Fisher's exact test, and a False Discovery Rate (FDR) as calculated by the Benjamini-Hochberg procedure (Benjamini and Hochberg 1995). GO terms were filtered to retain only those with a p-value below 0.05, an enrichment greater than two, and more than four genes found. These GO terms were pooled in clusters of similar functions with Revigo online (Supek et al. 2011) using the chicken genome and setting the resulting list to be "tiny" (0.4). We then looked for similarities in enriched biological function between the four lists. An overrepresentation test was performed on the proteins showing evidence of accelerated and decelerated RER in flightless rails following the same protocol.

Results

Relative evolution rate

Out of 11,192 protein alignments analysed, 861 showed significant relative evolution rate differences between flightless and volant birds. Of these, 247 protein alignments had a higher rate in flightless species (Rho>0) and 614 had a higher rate in volant species (Rho<0) (Fig. 2).

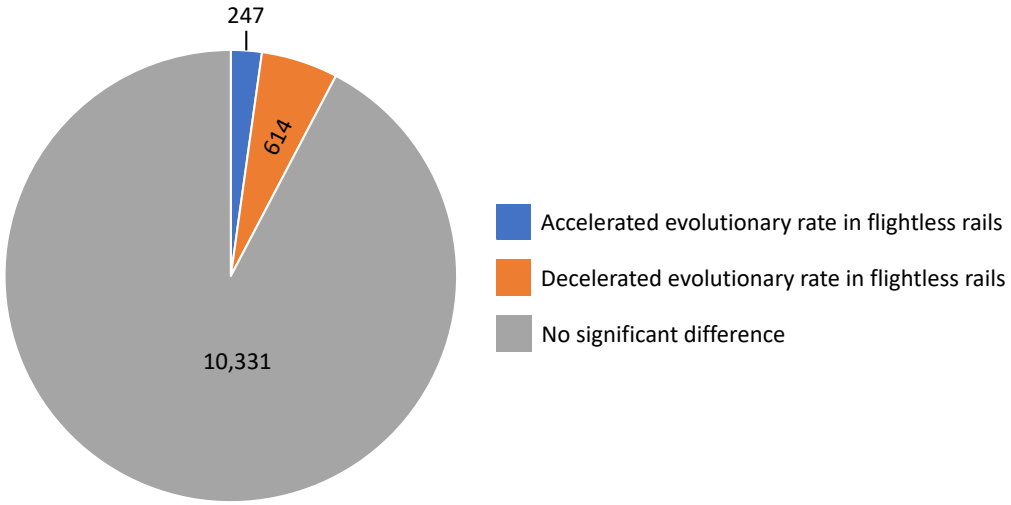


Figure 2: Genome-wide evolutionary rate convergent shifting in four rail species that evolved flightless independently. Eight volant Gruiformes species were used as background. Significant accelerated or decelerated evolutionary rate in flightless rails was assessed using Wilcoxon rank-sum test via the R package *RERconverge*.

To find out if proteins with accelerated evolutionary rates in flightless species were enriched in certain biological functions, an overrepresentation test was carried out (Appendix C5 Table 1). That test showed an overrepresentation of genes implicated in histone and chromatin remodelling as well as the biological processes associated with protein modifications: methylation, demethylation, dealkylation, hydroxylation, lipidation. Functions implicated in the development of bones, limbs, blood vessels, heart, kidney, and muscles were enriched as well. The proteins with a decelerated evolutionary rate in flightless rails do not exhibit functional enrichments that are obviously associated with loss of flight (Appendix C5 Table 2).

Proteins with function altering amino acid substitutions

For the PROVEAN analysis, the volant species *P. melanotus* and *F.atra* were each used as a reference for *Z. atra* and *A. rogersi* and the results were compared. Using a function-altering substitutions presence/absence dataset, the percentage of similar results, as well as a Pearson correlation coefficient r, were calculated. Results for the pairs *P. melanotus – A. rogersi* and *F. atra – A. rogersi* were similar at 91.80% and highly correlated (r = 0.76) and results for the pairs *P. melanotus – Z. atra* and *F. atra – Z. atra* were similar at 91.86% with a Pearson correlation coefficient equal to 0.74. Based on these high correlations, *P. melanotus* was preferred to be the reference branch for both *Z. atra* and *A. rogersi* as it is the most closely related volant species for which the genome is available.

In flightless rails, between 4801 (*P. hochstetteri*) and 8361 (*A. rogersi*) proteins contained at least one substitution for a total of between 11,135 (*P. hochstetteri*) and 69,623 (*A. rogersi*) substitutions. 21.5% of the substitutions were considered to impact the protein function (PROVEAN score < −2.5) in *P. hochstetteri* (Fig. 3). This proportion was higher than the other flightless species which exhibited between 13.4% (*Z. atra*) and 14.1% (*G. australis*) of function-altering substitutions. Volant pairs used as a negative control showed fewer variations in the number and proportions of function-altering substitutions. Indeed, the number of proteins containing a least one substitution was between 8446 (Pp_Fa) and 8486 (Fa_Gp) and the proportions of function-altering substitutions were between 11.8 and 13.2%.

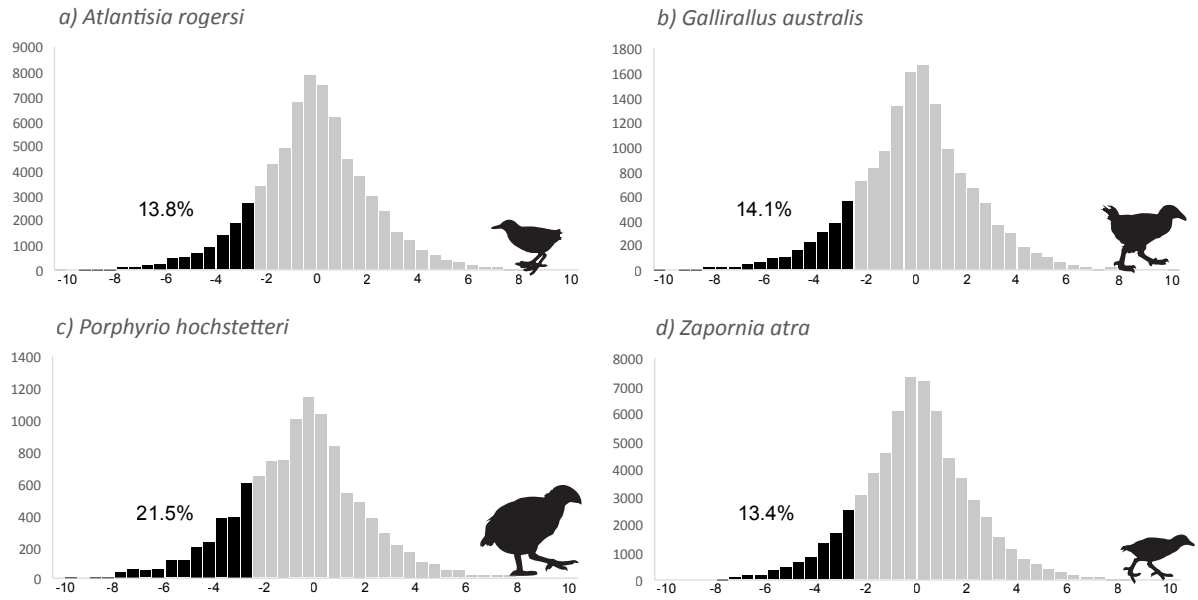


Figure 3: Distribution of the effect of variant (PROVEAN score) between four flightless rails and their closest volant relatives for which the whole genome is available. PROVEAN was used genome-wide to predict the effect on protein function of all substitutions. The more negative the score the more likely the variant impacts the protein function. The threshold was set at –2.5 and the percentage of function-altering substitutions (black boxes) is indicated for each flightless rail. The reference species for *G. australis* was *G. philippensis* and *P. melanotus* was used as a reference for the three other flightless rails.

Single substitutions

As flight is the ancestral state, we expect the molecular basis for that condition to be conserved in cranes and volant rails and to evolve differently in flightless rails. At the single amino acid level, the four flightless species carried a function-altering difference at the same amino acid site (95) in the protein "NHL repeat-containing protein 3" (*NHLRC3*). However, that AA was also present in the volant crane *B. regulorum*. A further 16 proteins contained single function-altering amino acid substitutions in three out of the four flightless species (Fig. 4), but some differences appeared to be specific to the rail clade (*FBH1* and *MFAT4D*).

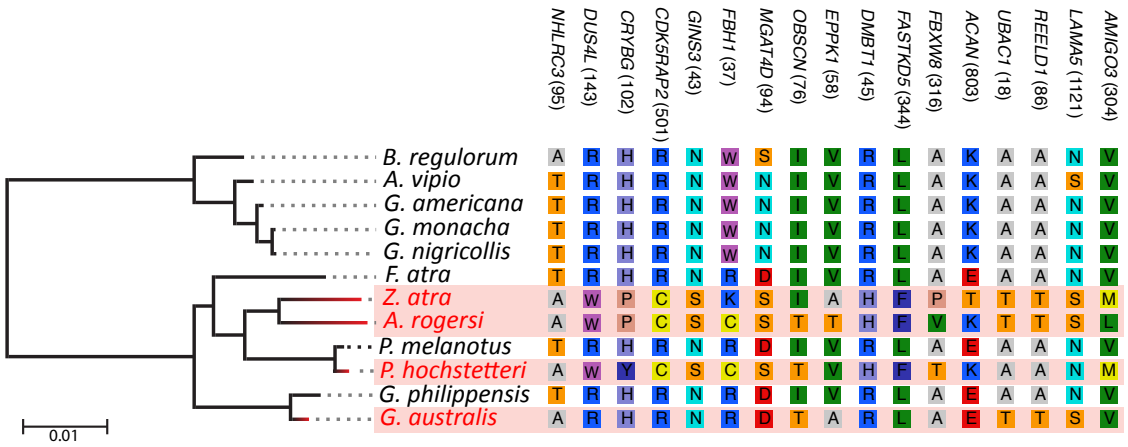


Figure 4. Single amino acid substitutions correlated with the transition to flightlessness in 17 protein sites. The gene encoding each protein as well as the amino acid site (in brackets) is indicated on top of the graph. The red colour in the phylogenetic tree indicates the independent evolution of flightlessness in the four species. All substitutions present in flightless birds are function-altering (PROVEAN score <–2.5).

Protein level

For each protein, the number of function-altering substitutions and the total number of substitutions were compared between the flightless species and the 6 negative controls. Proteins were considered likely to be associated with loss of flight if they contained more function-altering substitutions in flightless rails than in the controls (Table 2). Three proteins had function-altering substitutions in all four flightless rails and none in the volant rails, *MAP7*, *SORL1*, and *USP31*.

Table 2: Function-altering substitutions/Total number of substitutions. Proportions of function-altering substitutions in four flightless rails when compared to their closest volant relatives. Six pairs of volant-volant rails were used as negative controls to identify proteins carrying function-altering variants in flightless species. The length of each protein alignment is given in amino acids. "Acc." in the RER (Relative Evolutionary Rate) column indicates a significant accelerated convergent evolutionary rate in all flightless rails compared to the volant species for that protein.

Gene	Length	Flightless				Volant (control)						RER
		A. rogersi	*G. australis*	*P. hochstetteri*	*Z. atra*	Fa_Pm	Fa_Gp	Gp_Fa	Gp_Pm	Pm_Fa	Pm_Gp	
MAP7	579	3/10	1/4	2/8	1/14	0/10	0/10	0/10	0/13	0/10	0/13	
SORL1	823	1/11	1/3	1/2	3/10	0/5	0/1	0/1	0/4	0/5	0/4	
USP31	1026	1/6	1/1	2/4	2/5	0/6	0/8	0/8	0/3	0/6	0/3	
MTRR	335	2/21	2/4	1/3	3/20	0/13	0/13	0/13	0/14	1/13	0/14	
DUS4L	316	1/13	1/1	1/2	1/7	1/12	0/5	0/5	0/7	0/12	0/7	
NR1H4	422	3/10	1/5	0/1	2/5	0/4	0/4	0/4	0/4	0/4	0/4	
GRIFIN	134	1/5	0/0	2/4	3/6	0/5	0/4	0/4	0/6	0/5	0/6	
PRDM16	1048	0/2	3/3	2/2	1/3	0/2	0/4	0/4	0/6	0/2	0/6	Acc.
BIRC6	1071	2/6	1/1	2/4	0/1	0/2	0/1	0/1	0/1	0/2	0/1	
TSPAN32	88	3/5	1/1	0/0	1/2	0/3	0/2	0/2	0/1	0/3	0/1	
MED13	2149	1/5	0/0	2/2	2/5	0/5	0/5	0/5	0/4	0/5	0/4	
SLC20A1	301	3/13	0/0	1/1	1/11	0/8	0/10	0/10	0/10	0/8	0/10	
TMEM144	309	2/8	0/0	2/5	1/6	0/8	0/7	0/7	0/5	0/8	0/5	
USP13	576	1/5	1/1	0/1	3/6	0/2	0/2	0/2	0/2	0/2	0/2	Acc.
ANKRD66	80	2/3	1/2	0/0	1/2	0/3	0/2	0/2	0/3	0/3	0/3	
ELF2	462	2/4	0/2	1/2	1/5	0/5	0/8	0/8	0/5	0/5	0/5	
EMC1	951	1/6	1/3	0/0	2/7	0/4	0/9	0/9	0/5	0/4	0/5	
FAM76A	281	1/2	1/2	0/0	2/3	0/2	0/3	0/3	0/1	0/2	0/1	
PCSK6	213	2/5	0/0	1/3	1/2	0/1	0/1	0/1	0/2	0/1	0/2	
CDH18	247	1/1	2/2	0/0	1/1	0/0	0/0	0/0	0/0	0/0	0/0	Acc.
LRRC40	267	1/8	0/2	1/3	2/9	0/4	0/5	0/5	0/5	0/4	0/5	
METTL4	98	2/11	1/2	0/0	1/6	0/9	0/12	0/12	0/8	0/9	0/8	
PPFIBP1	313	0/3	1/1	2/2	1/4	0/2	0/0	0/0	0/2	0/2	0/2	Acc.
RNF223	258	2/5	0/0	1/2	1/5	0/2	0/1	0/1	0/1	0/2	0/1	
CNKSR3	377	1/1	2/2	4/4	2/3	0/1	0/2	0/2	0/3	1/1	1/3	
ZK1290.5	260	1/8	2/2	2/2	2/8	1/6	0/5	0/5	1/3	0/6	0/3	
RREB1	968	2/6	1/4	1/3	3/9	1/8	0/11	0/11	1/9	0/8	0/9	
PRKDC	1354	4/12	3/7	1/2	3/8	2/5	0/5	0/5	2/6	0/5	0/6	
ITGB8	694	2/9	1/5	1/1	2/11	0/8	0/9	1/9	0/9	1/8	0/9	
KCNK17	253	3/11	1/2	1/1	1/8	0/13	0/11	1/11	0/6	1/13	0/6	
PTPN14	1149	2/5	1/1	1/2	2/4	0/7	0/7	1/7	1/4	0/7	0/4	
TTC27	433	2/5	2/3	1/2	1/4	0/5	0/7	1/7	1/4	0/5	0/4	Acc.
FEM1C	380	0/1	0/0	1/1	11/19	0/6	0/7	0/7	0/1	0/6	0/1	
WIZ	516	4/20	0/0	0/0	8/20	0/12	0/9	0/9	0/7	0/12	0/7	
ZFAND4	600	19/35	0/4	2/4	46/62	0/9	3/15	1/15	0/14	1/9	3/14	Acc.

Lineage specific

In order to test for a potential lineage species genetic pathway to flightlessness, each flightless species was considered independently. A list of all the proteins with altered function was generated for each species and an overrepresentation test was performed to

find out if these proteins lists were enriched in certain biological functions that could be linked with the ability to fly. Between 138 (*G. australis*) and 362 (*A. rogersi*) proteins with altered function were found (Table 3). For the four flightless species, these proteins were enriched in GO terms associated with immune response and DNA conformation (Table 4 and Appendix C5 Table 3). The proteins with altered function of three out of four flightless rails showed evidence of overrepresentation of biological functions associated with lipids metabolic processing and cilium formation and function (including cytoskeleton and microtubule development). Two out of four flightless species were enriched for GO terms associated with the circulatory system, muscle development and contraction, methylation and neuron development. Finally, it is worth mentioning that *G. australis* proteins with altered functions were enriched in GO terms associated with respiratory system and kidney development.

Table 3. Lineage specific proteins with altered function in four flightless rails. The number of proteins with at least two and more than 25% of function-altering substitutions (PROVEAN score < –2.5) is shown for each flightless species. Tests for functional enrichment were performed on the four protein lists. Enriched GO terms containing at least 4 genes, exhibiting an enrichment fold above 2, and a p-value < 0.05 were kept.

	A. rogersi	***G. australis***	***P. hochstetteri***	***Z. atra***
Proteins with altered function	362	138	198	334
Overrepresented GO terms	64	50	69	75

Discussion

We expect genes affected by the convergent evolution toward the flightless phenotype to experience similar changes in selective pressure which result in convergent shifts in whole-gene evolutionary rate. A set of 247 genes showed evidence of accelerated evolutionary rate in flightless rails and this set was enriched in biological functions that can be directly linked to the ability to fly like the development of bones, limbs, muscles, and heart (Appendix C5 Table 1). Indeed, powered flight requires a demanding aerobic activity and an efficient cardiovascular system to deliver oxygen to the pectoral muscles during flight. In birds, heart size is associated with flight capacity (Berger et al. 1970, Nespolo et al. 2018). Flightless birds have evolved smaller wings and flight muscles than volant birds (Wright et al. 2016, Gaspar et al. 2020) and genes involved in the ability to

fly are associated with bones, muscles, and heart development (Machado et. al 2016). This is consistent with the accelerated evolutionary rate observed here in genes associated with muscle and limb development in flightless rails.

Table 4. Biological function enriched in proteins with altered function for each flightless species. Numbers indicate how many enriched GO terms were present for each flightless rail.

Biological function	*Atlantisia rogersi*	*Gallirallus australis*	*Porphyrio hochstetteri*	*Zapornia atra*
Immune response	3	3	2	8
DNA conformation, chromosome, chromatin and histone	3	6	13	4
Lipid metabolic process	3	1		14
Cilium		9	2	1
Circulatory system	7		1	
Muscle development and contraction	4	2		
Methylation	3		1	
Neuron development	2			2
Respiratory system development		1		
Kidney development		1		
Others	39	27	50	47

The enrichment of other biological functions such as histone and chromatin remodelling and various protein modifications suggests that evolving flightlessness involves a change in gene expression and in post-translational mechanisms. A similar inference has been made from analysis of flightlessness in ratites (kiwis, emus, ostriches etc.) (Sackton et al. 2019). As flight is the ancestral state, it is unlikely that decelerated (in other words, more conserved) genes in flightless rails are involved in flightlessness. The overrepresentation test on decelerated genes shows that this set encompasses more genes but their association with the loss of flight is not obvious (Appendix C5 Table 2).

GO terms are ordered as a complex hierarchical system which means if a term is enriched, chances are that its parents' terms will be too. As a consequence, in a case were an enriched term has numerous parents terms, the importance of the biological function it belongs to may appear biased. It is useful to visualise the results in terms of enriched

functions rather than a precise quantification of overrepresentation in a set of genes (Table 4 and Appendices C5).

We find 21.5% of *P. hochstetteri* substitutions are likely to alter the protein function, which is much higher than all the other rails (Fig. 3). The takahē *P. hoschetteri* is a critically endangered flightless species recently subject to an extreme bottleneck and currently with a small recovering population (445 in 2020 -www.doc.govt.nz). All individuals are derived from as few as two individuals discovered in the 1950s (Wallace 2002). The small population and the resulting inbreeding depression has resulted in a high level of homozygosity (See Chapter 3) that likely explains why numerous function-altering mutations have not been filtered out through selective processes.

Only three proteins contain function-altering substitutions in all flightless rails and none in volant species. These are the proteins encoded by Microtubule Associated Protein 7 (*MAP7*), Sortilin Related Receptor 1 (*SORL1*), and Ubiquitin Specific Peptidase 31 (*USP31*). Ensconsin/MAP7 is a microtubule-stabilizing protein encoded by the *MAP7* gene that promotes microtubule growth in neural stem of *Drosophila* cells and is associated with muscle dysfunction in humans (Metzger et al. 2012, Gallaud et al. 2014). *SORL1* is mostly expressed in the central nervous system and its genetic variants have been associated with Alzheimer disease (Holstege et al. 2017). *USP31* gene is involved in protein deubiquitination in humans (Lockhart et al. 2004). None of these three genes show signs of accelerated evolutionary rate in flightless rails. Moreover, evidence of evolutionary rate acceleration is limited in proteins with altered function in flightless birds (Table 2). This is likely caused by the fact that one analysis takes into account the number of substitutions when the other considers the functional impact of these substitutions. In other words, the evolutionary rate test calculates the amount of change observed between volant and flightless species when the variant effect predictor calculates the selective strength associated with these changes.

We hypothesize that the transition to flightlessness results in a reduced purifying selection acting on genes associated with flight which are therefore more likely to contain function-altering mutations. The set of genes affected by these changes may be different in each flightless rail. We observe that the proteins with altered function in each flightless species

are enriched in biological function associated the immune response. Moreover, immune functions are enriched in genes with both accelerated and decelerated evolutionary rate (Appendices C5 Tables 1 and 2). Based on studies made on migratory birds, immune defences increase with the number of different habitats the birds are exposed to (Møller and Erritzøe 1998). The loss of flight undeniably reduces the number and diversity of those habitats. As maintaining a strong immune system can be energy consuming (Norris and Evans 2000) one can suppose that physiological defence against infections would be different between flightless and volant birds.

We observe that the GO terms associated with DNA conformation (chromosome, chromatin and histone) as well as those associated with protein methylation are overrepresented in proteins carrying function-altering substitutions in flightless rails. This suggests once more that the transition to flightlessness involves a change in gene expression and post-translational mechanisms. All flightless species except *P. hochstetteri* exhibit enrichment of biological functions associated with lipid metabolic processing. As energy expenditure is extremely important in the ability to fly, it is not surprising to observe physiological differences between volant and flightless species. A large scale comparative avian genomic study identified convergence signature occurring on enzymes associated with lipid metabolism in flight-degenerate birds (Pan et al. 2019).

Proteins carrying altered functions in *G. australis*, *P. hochstetteri and Z. atra* are enriched in functions associated with cilium development. A similar response was found in the flightless Galapagos cormorant, *Phalacrocorax harrisi* (Burga et al. 2017), where proteins carrying function-altering variants in the flightless cormorant were enriched in genes that affect limb development and are associated with human disorders known as skeletal ciliopathies (Waters and Beales 2011). These disorders result from defects in the formation and function of cilia which are hairlike microtubule-based structures that project from the surface of the cell. In vertebrates, this organelle has crucial roles in embryonic development (Goetz and Anderson 2010).

Consistently with the proteins that show accelerated evolutionary rates in flightless rails, the proteins carrying function-altering substitutions were also enriched in functions that can be readily linked with the ability to fly including the circulatory and respiratory

systems and muscle and neuron development. *G. australis* exhibits enrichment of biological function associated with renal system development. The high level of energy expenditure associated with flight results in an important evaporative water loss which is why birds evolved physiological adaptations to minimize their excretion of water (Gerson and Guglielmo 2013). As in most vertebrates, one of these is to concentrate urine by initial filtration of the plasma and water reabsorption in the kidneys. In birds, urine can also be processed in the cloaca and the lower parts of the gastrointestinal tract (Giladi et al. 1997). It appears that the selective pressure to reduce water loss decreases in flightless birds as, for instance, the emu has a low ability to concentrate its urine (Dawson et al. 1991). It is therefore not surprising to find function-altering substitutions in genes associated with kidney development in a flightless rail.

In order to streamline the analysis, the functional effects of amino acid insertions and deletions are not investigated in the present research, only amino acid replacements are considered. Despite the relatively low number of variants caused by insertion and deletion (3.2% of all variations in the Galapagos flightless cormorant), these changes can have a significant impact on protein function (Burga et al. 2017). For that reason, it would be interesting in future research to investigate this aspect of protein variants in flightless rails.

In the variant effect predictor tests (PROVEAN), the fact that *P. melanotus* was used as a reference species for three out of the four flightless species may introduce a bias. Indeed, a sequencing error or a substitution present only in *P. melanotus* can potentially impact the results for three flightless species. However, the way PROVEAN collects a set of homologous proteins related to the reference sequence, aligns them, and assigns a score to each variation based on the alignment prevents rare variations in the reference sequence from having a big influence on the predicted effect (Choi et al. 2012). Or in other words, a rare mutation in the reference sequence (the volant species) will most likely not result in a PROVEAN score < -2.5 whereas a rare mutation in the test sequence (the flightless species) will.

In conclusion, after screening 11,192 amino acid sequences in 12 bird species, we identified hundreds of genes exhibiting a convergent accelerated rate of molecular evolution in flightless species. These genes are enriched in biological functions associated

with gene expression regulation or in functions that align well with the expected adaptation to flightlessness like the development of bones, limbs, muscles, and heart. A variant effect assessment of all the flightless species substitutions highlights proteins of interest that can be associated with the transition to flightlessness, especially those encoded by the genes *MAP7*, *SORL1*, and *USP31*. The functional impact of amino acid substitutions provides limited evidence of similar molecular mechanisms being involved, at the protein level, in separate lineages of rails evolving flightlessness. At the level of biological processes however, similar functions were enriched in proteins carrying function-altering substitutions in different flightless rails, namely the immune response, the DNA conformation, cilium, muscle and circulatory system development as well as lipid metabolic processing. Regarding the genetic pathway to flightlessness, we observe signs of both convergent whole-gene evolution and lineage-specific evolution. This suggests that to converge toward the flightless phenotype, some genomic regions undergo similar selective pressures in all flightless species when other regions go through distinct evolutionary mechanism in each flightless rail.

Data availability

Supplementary data can be accessed on https://figshare.com/s/44475dcf0f6e3a6b55d9.

Acknowledgments

The authors wish to acknowledge the use of New Zealand eScience Infrastructure (NeSI) high-performance computing facilities, consulting support, and/or training services as part of this research. New Zealand's national facilities are provided by NeSI and funded jointly by NeSI's collaborator institutions and through the Ministry of Business, Innovation & Employment's Research Infrastructure programme. URL https://www.nesi.org.nz.

References

Benjamini, Y. and Hochberg, Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. – Journal of the Royal Statistical Society: Series B (Methodological) 57: 289–300.

Berger, M., Hart, J. S. and Roy, O. Z. 1970. Respiration, oxygen consumption and heart rate in some birds during rest and flight. – Z. Vergl. Physiol. 66: 201–214.

Burga, A., Wang, W., Ben-David, E., Wolf, P. C., Ramey, A. M., Verdugo, C., Lyons, K., Parker, P. G. and Kruglyak, L. 2017. A genetic signature of the evolution of loss of flight in the Galapagos cormorant. – Science 356: eaal3345.

Castresana, J. 2000. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. – Mol Biol Evol 17: 540–552.

Chikina, M., Robinson, J. D. and Clark, N. L. 2016. Hundreds of Genes Experienced Convergent Shifts in Selective Pressure in Marine Mammals. – Mol Biol Evol 33: 2182–2192.

Choi, Y., Sims, G. E., Murphy, S., Miller, J. R. and Chan, A. P. 2012. Predicting the functional effect of amino acid substitutions and indels. – PLoS One 7: e46688.

Dawson, T. J., Maloney, S. K. and Skadhauge, E. 1991. The role of the kidney in electrolyte and nitrogen excretion in a large flightless bird, the emu, during different osmotic regimes, including dehydration and nesting. – J Comp Physiol B 161: 165–171.

Faircloth, B. C. 2016. PHYLUCE is a software package for the analysis of conserved genomic loci. – Bioinformatics 32: 786–788.

Gallaud, E., Caous, R., Pascal, A., Bazile, F., Gagné, J.-P., Huet, S., Poirier, G. G., Chrétien, D., Richard-Parpaillon, L. and Giet, R. 2014. Ensconsin/Map7 promotes

microtubule growth and centrosome separation in Drosophila neural stem cells. – J Cell Biol 204: 1111–1121.

Garcia–R, J. C., Gibb, G. C. and Trewick, S. A. 2014. Eocene Diversification of Crown Group Rails (Aves: Gruiformes: Rallidae). – PLOS ONE 9: e109635.

Gaspar, J., Gibb, G. C. and Trewick, S. A. 2020. Convergent morphological responses to loss of flight in rails (Aves: Rallidae). – Ecology and Evolution 10: 6186–6207.

Gerson, A. R. and Guglielmo, C. G. 2013. Measurement of glomerular filtration rate during flight in a migratory bird using a single bolus injection of FITC-inulin. – American Journal of Physiology-Renal Physiology 305: F823–F829.

Giladi, I., Goldstein, D. L., Pinshow, B. and Gerstberger, R. 1997. Renal function and plasma levels of arginine vasotocin during free flight in pigeons. – Journal of Experimental Biology 200: 3203–3211.

Goetz, S. C. and Anderson, K. V. 2010. The primary cilium: a signalling centre during vertebrate development. – Nat Rev Genet 11: 331–344.

Holstege, H., van der Lee, S. J., Hulsman, M., Wong, T. H., van Rooij, J. G., Weiss, M., Louwersheimer, E., Wolters, F. J., Amin, N., Uitterlinden, A. G., Hofman, A., Ikram, M. A., van Swieten, J. C., Meijers-Heijboer, H., van der Flier, W. M., Reinders, M. J., van Duijn, C. M. and Scheltens, P. 2017. Characterization of pathogenic SORL1 genetic variants for association with Alzheimer's disease: a clinical interpretation strategy. – European Journal of Human Genetics 25: 873–981.

Kimura, M. 1977. Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. – Nature 267: 275–276.

Kirchman, J. J. 2012. Speciation of Flightless Rails on Islands: A DNA-Based Phylogeny of the Typical Rails of the Pacific. – The Auk 129: 56–69.

Kowalczyk, A., Meyer, W. K., Partha, R., Mao, W., Clark, N. L. and Chikina, M. 2019. RERconverge: an R package for associating evolutionary rates with convergent traits. – Bioinformatics 35: 4815–4817.

Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T. and 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.

Liu, Y., Liu, S., Yeh, C.-F., Zhang, N., Chen, G., Que, P., Dong, L. and Li, S. 2018. The first set of universal nuclear protein-coding loci markers for avian phylogenetic and population genetic studies. – Scientific Reports 8: 15723.

Livezey, B. C. 2003. Evolution of Flightlessness in Rails. – American Ornithologists' Union.

Lockhart, P. J., Hulihan, M., Lincoln, S., Hussey, J., Skipper, L., Bisceglio, G., Wilkes, K. and Farrer, M. J. 2004. Identification of the Human Ubiquitin Specific Protease 31 (USP31) Gene: Structure, Sequence and Expression Analysis. – DNA Sequence 15: 9–14.

Machado, J. P., Johnson, W. E., Gilbert, M. T. P., Zhang, G., Jarvis, E. D., O'Brien, S. J. and Antunes, A. 2016. Bone-associated gene evolution and the origin of flight in birds. – BMC Genomics 17: 371.

Metzger, T., Gache, V., Xu, M., Cadot, B., Folker, E. S., Richardson, B. E., Gomes, E. R. and Baylies, M. K. 2012. MAP and kinesin-dependent nuclear positioning is required for skeletal muscle function. – Nature 484: 120–124.

Mi, H., Muruganujan, A., Ebert, D., Huang, X. and Thomas, P. D. 2019. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. – Nucleic Acids Research 47: D419–D426.

Miller, M. A., Pfeiffer, W. and Schwartz, T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. – 2010 Gateway Computing Environments Workshop (GCE). : 1–8.

Møller, A. P. and Erritzøe, J. 1998. Host immune defence and migration in birds. – Evolutionary Ecology 12: 945–953.

Nespolo, R. F., González-Lagos, C., Solano-Iguaran, J. J., Elfwing, M., Garitano-Zavala, A., Mañosa, S., Alonso, J. C. and Altimiras, J. 2018. Aerobic power and flight capacity in birds: a phylogenetic test of the heart-size hypothesis. – Journal of Experimental Biology 221: jeb.162693.

Norris, K. and Evans, M. R. 2000. Ecological immunology: life history trade-offs and immune defense in birds. – Behavioral Ecology 11: 19–26.

Olson, S. L. 1973. Evolution of the rails of the South Atlantic islands (Aves: Rallidae). – Smithsonian Contributions to Zoology.

Pan, S., Lin, Y., Liu, Q., Duan, J., Lin, Z., Wang, Y., Wang, X., Lam, S. M., Zou, Z., Shui, G., Zhang, Y., Zhang, Z. and Zhan, X. 2019. Convergent genomic signatures of flight loss in birds suggest a switch of main fuel. – Nature Communications 10: 1–11.

Partha, R., Kowalczyk, A., Clark, N. L. and Chikina, M. 2019. Robust Method for Detecting Convergent Shifts in Evolutionary Rates. – Molecular Biology and Evolution 36: 1817–1830.

Ranwez, V., Douzery, E. J. P., Cambon, C., Chantret, N. and Delsuc, F. 2018. MACSE v2: Toolkit for the Alignment of Coding Sequences Accounting for Frameshifts and Stop Codons. – Mol Biol Evol 35: 2582–2584.

Ripley, S. D., Lansdowne, J. F. and Olson, S. L. 1977. Rails of the World: A Monograph of the Family Rallidae. – M. F. Feheley Publishers.

Sackton, T. B., Grayson, P., Cloutier, A., Hu, Z., Liu, J. S., Wheeler, N. E., Gardner, P. P., Clarke, J. A., Baker, A. J., Clamp, M. and Edwards, S. V. 2019. Convergent regulatory evolution and the origin of flightlessness in palaeognathous birds. – Science 364: 74–78.

Sato, T., Yamanishi, Y., Kanehisa, M. and Toh, H. 2005. The inference of protein–protein interactions by co-evolutionary analysis is improved by excluding the information about the phylogenetic relationships. – Bioinformatics 21: 3482–3489.

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

Steadman, D. W. 1995. Prehistoric Extinctions of Pacific Island Birds: Biodiversity Meets Zooarchaeology. – Science 267: 1123–1131.

Stern, D. L. 2000. Perspective: Evolutionary Developmental Biology and the Problem of Variation. – Evolution 54: 1079–1091.

Supek, F., Bošnjak, M., Škunca, N. and Šmuc, T. 2011. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. – PLOS ONE 6: e21800.

Taylor, B. 1998. Rails: A Guide to the Rails, Crakes, Gallinules and Coots of the World. – Bloomsbury Publishing.

Trewick, S. A. 1997a. Flightlessness and phylogeny amongst endemic rails (Aves: Rallidae) of the New Zealand region. – Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 352: 429–446.

Trewick, S. A. 1997b. Sympatric flightless rails Gallirallus dieffenbachii and G. modestus on the Chatham Islands, New Zealand; morphometrics and alternative evolutionary scenarios. – Journal of the Royal Society of New Zealand 27: 451–464.

Wallace, G. E. 2002. The Takahe: Fifty Years of Conservation Management and Research. – The Auk 119: 291–293.

Waters, A. M. and Beales, P. L. 2011. Ciliopathies: an expanding disease spectrum. – Pediatr Nephrol 26: 1039–1056.

Wright, N. A., Steadman, D. W. and Witt, C. C. 2016. Predictable evolution toward flightlessness in volant island birds. – PNAS 113: 4765–4770.

Yang, Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. – Mol Biol Evol 24: 1586–1591.

Chapter Six

General Discussion

Transition to flightlessness in rails

Based on the literature review and the results presented in this thesis I propose an overall hypothesis to explain the morphological and molecular evolutionary mechanisms involved in the transition to flightlessness in rails.

The right conditions

The Western Pacific Ocean has provided a natural laboratory of evolution due to the abundance of physically isolated islands and habitat diversity which are known to play an important role in speciation (Diamond 1977, Filardi and Moyle 2005, Trewick and Cowie 2008, Moyle et al. 2009, Gibb and Penny 2010, Garcia-R et al. 2017). Indeed, limited gene flow between habitat patches with abrupt ecological boundaries, such as oceanic islands and natural selection in a new environment promotes evolutionary change (Templeton 1980, Trewick et al. 2017). Rapid and sometimes serial avian colonisation and diversification have been documented across all major archipelagos resulting in numerous insular endemic avian species that evolved over short periods of isolation (Filardi and Moyle 2005, Moyle et al. 2009).

On small isolated islands, the gene flow from the outside and the genetic diversity that reaches the island are often low, producing a "founder effect" (Templeton 1980). The resulting small effective population size usually leads to loss of heterozygosity (see Chapter 3) through genetic drift (Duffy and Vargas 2018). This can result in the accumulation of deleterious mutations that may reduce population viability (Kennedy et al. 2014) and can also result in rapid phenotypic shifts in response to the new environment (Losos et al. 1997).

Isolation tends to speed up fixation of distinct population traits (speciation), and oceanic islands often exhibit conditions suitable for flightlessness. Indeed, islands usually have no terrestrial mammalian predators and fewer competitors for resources so birds can forage without flying (Olson 1973, McNab 1994a, 2002, Wright et al. 2016).

Most rails, despite their noted dispersal capacity, exhibit a terrestrial lifestyle and have short wing lengths relative to their body mass (Taylor 1998, McCall et al. 1998). Relative

short wings, that provide effective predator escape, increase the energetic cost of sustained flight, which, in an island habitat with low resource availability, promotes the walking ecology (McNab 1994a, b, McCall et al. 1998) and therefore predisposes rails to lose the ability to fly.

To summarize, small isolated populations of birds from a lineage predisposed to evolve flightlessness in a habitat conducive to rapid evolution and that promotes the walking ecology generate the optimal conditions for the rapid emergence of multiple flightless species (Fig. 1).

A recent study of *Dryolimnas* rails illustrates this evolutionary phenomenon (Hume and Martill 2019). The Aldabra Atoll, Seychelles, underwent total marine inundation around 127,000 YBP resulting in the loss of all terrestrial fauna. The fossil record revealed a flightless *Dryolimnas* in deposits dated before the inundation. The fact that the flightless subspecies *Dryolimnas cuvieri aldabranus* is found today on the same atoll provides evidence that members of that lineage colonized the atoll at least twice and evolved flightlessness independently on each occasion. This is one of the examples that show that, in the right conditions, rails can rapidly lose the ability to fly (Trewick 1997).

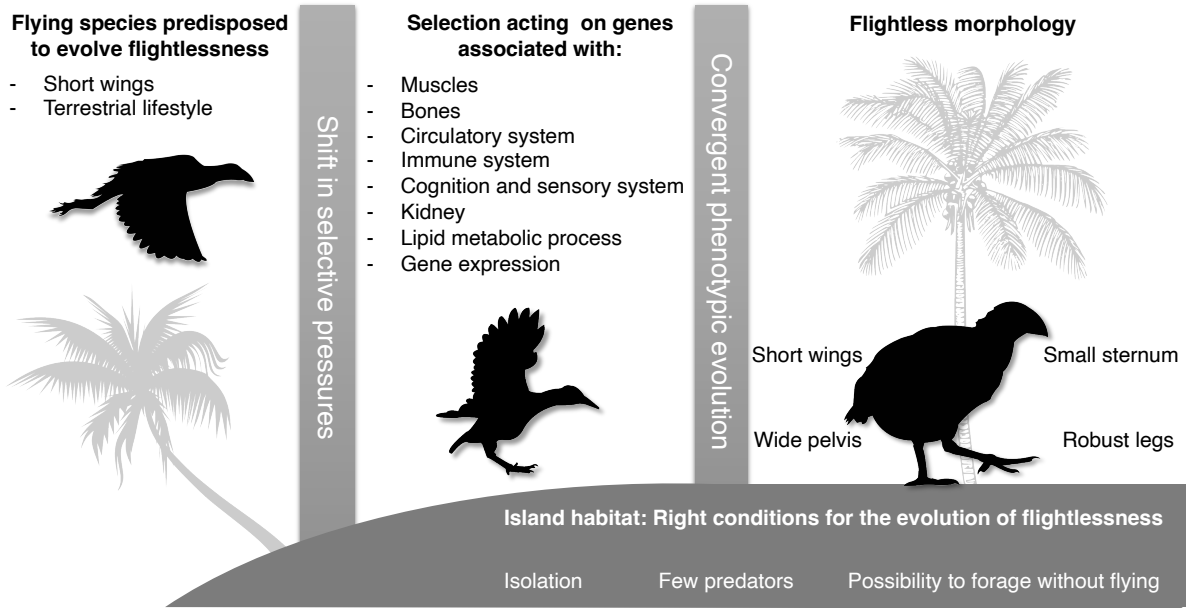


Figure 1: Evolutionary processes involved in the transition from volant to flightless in rails.

Evolutionary mechanisms for the transition to flightlessness

Once released from the physiological constraints of flight, rails undergo evolutionary changes that are apparent at both the morphological and the molecular level.

Limbs, muscles and bones

The pectoral muscles of flightless bird species do not need to sustain power flight and therefore tend to have smaller relative mass (Livezey 2003). It is expected that such phenotypic changes involve relaxation in the selective constraint operating on some genes associated with muscle development and function, resulting in the accumulation of nonsynonymous amino acid replacements (Table 1). Amino acid replacements altering the function of proteins encoded by genes involved in muscle development are indeed observed in flightless species.

As the smaller pectoral muscles need a smaller skeletal area to be anchored, flightless rails exhibit relatively small sterna. The loss of flight also results in shorter wings (Gaspar et al. 2020). Even though volant rails are often reluctant flyers with a predominantly terrestrial lifestyle, flightless rails related to them have wider pelves and more robust legs. I hypothesize that the weight constraint associated with flight maintains the hindlimb and the pelvic girdle to a reduced size in volant species but once the flight constraint is removed, rails evolve bigger legs. In other words, the differences in leg structures between volant and flightless rails are most likely due to the release of weight constraint in flightless rails and not to the fact that volant species spend less time on the ground.

At the molecular level, a set of genes associated with bones and limbs show evidence of relaxed selection in flightless rails (Table 1), and most flightless rails studied exhibit a high number of function-altering amino acid substitutions in proteins associated with cilium formation (including cytoskeleton and microtubule development). Cilia are hairlike microtubule-based structures that project from the surface of the cell and, in vertebrates, are known to have crucial roles in embryonic development (Goetz and Anderson 2010). Proteins involved in cilium development carrying function-altering variants are associated in humans with disorders called ciliopathies that can cause multiple defects including reduced limb development (Waters and Beales 2011, Burga et al. 2017). Even though it is

not a direct correlation, it is likely that the accumulation of deleterious mutations in proteins involved in cilium development plays a role in the morphological shift associated with the loss of flight in rails.

Circulatory system

Selective pressure to maintain a high performance cardiovascular system appears to be reduced in flightless rails. Indeed, flight is the main reason why birds need to have intense aerobic activity and a large heart to deliver oxygen to the pectoral muscles (Berger et al. 1970, Grubb 1983, Nespolo et al. 2018). Without that physiological demand, rails exhibit reduction in metabolic rate (McNab 1994b). Moreover, without flight, rails do not need to regulate their body temperature as much as their volant relatives which, again, releases selective pressure acting on the cardiovascular system (Scott 2011).

Evidence of both relaxed selection and accelerated evolutionary rate was observed in genes associated with the circulatory system in flightless rails. Moreover, the species *A. rogersi* and *P. hochstetteri* exhibited function-altering variants of proteins associated with the circulatory system. These results are consistent with the hypothesis that energy conservation is a basis for the evolution of avian flightlessness (Olson 1973, McNab 1988, 1994a, Diamond 1991, Butler 2016).

Immune system

In flightless rails, genes associated with immune response show evidence of relaxed selection and both accelerated and decelerated rates of molecular evolution. Moreover, all flightless species exhibit deleterious variants in proteins associated with the immune system (Table 1). It therefore appears that volant and flightless rails are optimized for physiological defence against infections in different ways. The immune response associated with flightlessness is more complex than other traits as genes associated with that function exhibit both accelerated and decelerated evolutionary rates in flightless species. I, therefore, hypothesize that the rails' immune system adapts to the walking ecology via a combination of relaxed and purifying selection (decelerated) evolution acting on different genes. The difference between volant and flightless species may be due to the fact that small isolated islands are less exposed to parasites and microbial diseases

as is well documented in human populations (Penman et al. 2017). Indeed, when volant species colonize an island, this involves a small subset of individuals (Templeton 1980) which means they are less likely to carry all of the parasites that the entire species hosts. Moreover, parasitized individuals can be less likely to successfully disperse and colonize distant habitats (Dobson and Mccallum 1997, Duffy and Vargas 2018). Parasitic species that require more than one host are also less likely to be introduced to a completely new environment (Lymbery et al. 2014).

The diversity and loading of parasites that island populations are exposed to varies (Duffy and Vargas 2018) which, in birds, can result in adaptive response in defence mechanisms. If the diversity of disease-causing parasites in the habitat is low, some host species may invest in generalised immune responses rather than in a costly production of specific antibodies (Frank 2000). Based on studies made on migratory birds, immune defences increase with the number of different habitats the birds are exposed to (Møller and Erritzøe 1998, Eikenaar et al. 2020). The loss of flight undeniably reduces the number and diversity of those habitats. As maintaining a strong immune system can be energy consuming (Norris and Evans 2000), it is not surprising to observe evidence of differential selective pressure in genes associated with immune response in flightless species.

Another hypothesis related to the immune response in flightless rails involves the gut microbiome. Indeed, the crosstalk between gut bacteria and innate and adaptive immune systems is crucial for host development (Broom and Kogut 2018, Bodawatta et al. 2022). Gut microbes can influence pathogen populations either directly via competitive exclusion or indirectly through activation of the host immune system (Yitbarek et al. 2018, Huang et al. 2019, Methner et al. 2019, McLaren and Callahan 2020). Volant birds usually exhibit a relatively shorter digestive tract than flightless species (Caviedes-Vidal et al. 2007, Price et al. 2015), which can result in shorter retention time and affect the gut microbiome. As a consequence, the gut environment of flightless species may fluctuate less, which could result in a more stable microbiome and, by extension, improve the immune response (Bodawatta et al. 2022).

Evolutionary changes in immune response may not be a direct consequence of flightlessness but rather an indirect result of the ecological changes associated with the loss of flight (reduced habitat, less environmental diversity etc.). Thus, the selective

pressures acting on immune system genes are less straightforward than those acting directly on form and function associated with flight like limb and muscles development.

Cognition and sensory system

Genes associated with cognition and the sensory system (including memory, eyesight and neuron development) show evidence of both accelerated and relaxed selection in flightless rails and contained deleterious mutations in two out of the four studied flightless lineages (*A. rogersi* and *Z. atra*). As the constraint on body mass is reduced in flightless species, one could expect that they would evolve with bigger heads and brains. However, in rails, cranial size is not associated with flight capacity (Gaspar et al. 2020). This might be because, in birds, having bigger brains costs more energy (Isler and van Schaik 2006). At the neuroanatomical level, no significant differences are observed between birds with and without the ability to fly (Gold and Watanabe 2018). However, it is known that spatial memory capacity is correlated with hippocampus size (Sherry et al. 1989, Krebs et al. 1989, Balda and Kamil 1989). As maintenance of efficient memory is an active and costly process (Dukas 1999), one can assume that the size and the diversity of habitats a bird lives in can result in an adaptive trade-off in terms of memory capacity. Flightless rails often have limited habitat compared to their volant relatives which may explain the differential selective pressure acting on genes associated with memory and visual perception.

Kidney

Genes associated with kidney development are under relaxed selection in flightless rails and proteins involved in that function carry deleterious mutations in the flightless *G. australis* (Table 1). To my knowledge, there has been no comparative analysis of renal function in volant and flightless rails. However, flightless ratites appear to differ from other birds in kidney function (Gerson and Guglielmo 2013) and postrenal modification of ureteral urine (Goldstein 2022). In rails, a comparison of kidneys in the Eurasian coot *Fulica atra* and the lesser moorhen *Gallinula angulate* suggests that the greater medullary thickness in the Eurasian coot is an adaptative advantage for reducing the effect of water scarcity during migration (Deef 2015). The high level of energy expenditure associated with flight results in significant evaporative water loss. This is probably why the evolution

of avian flight is associated with physiological adaptations to minimize their excretion of water (Gerson and Guglielmo 2013). It is therefore reasonable to expect differential selective pressure acting on genes associated with kidney development in a flightless rail.

Gene regulation and post-translational modifications

Numerous genes involved in regulation (DNA, histone and chromatin conformation) and post-translational modifications (protein methylation, demethylation, dealkylation, hydroxylation, lipidation) show evidence of relaxed selection in flightless rails and carry deleterious mutations in all studied flightless rails. This suggests genetic adaptation to flightlessness is not only found in important developmental genes but could also include changes in gene expression and protein modifications. Indeed, mutations in *cis*-regulatory DNA may be the predominant source of morphological evolution (Stern 2000) which means non-coding regions are promising targets to study the molecular basis of phenotypic changes (Hao et al. 2019). The importance of *cis*-regulation in convergent phenotypic evolution in animals has been highlighted in multiple studies (Vavouri et al. 2007, Frankel et al. 2012, Signor et al. 2016, Feigin et al. 2018). Genetic pathways to avian flightlessness have been shown to involve non-coding regions. Indeed, regulatory regions such as ultraconserved non-exonic elements are associated with loss of flight in ratites (kiwis, emus, ostriches, etc.) (Sackton et al. 2019) and in Galapagos cormorants (Burga et al. 2017).

Covalent post-translational modifications of proteins influence the form and function of the mature protein product (Chou 2020), which can result in phenotypic changes. Moreover, regulation of gene expression is influenced by post-translational modification of histones resulting in alteration of the accessibility of the underlying DNA (Peterson and Laniel 2004, Duncan et al. 2014). I, therefore, suggest that genes associated with post-translational modification in rails are likely to be involved in the transition to flightlessness both directly, through modification of protein function, and indirectly through gene regulation.

Lipid metabolic process

Genes involved in lipid metabolic processes are relaxed and carry deleterious substitutions in flightless rails (Table 1). Flight requires high energetic investment as demonstrated by the increased storage of fat by migratory birds before departure and during stopovers (Jenni and Jenni-Eiermann 1998, Weber et al. 1998, Giulivi and Ramsey 2015). Therefore, one can assume that selective pressure to maintain fuel storage in flightless rails is reduced compared to their volant relatives.

Moreover, flight-degenerate birds perform mostly anaerobic metabolism using carbohydrates when volant species predominantly present oxidative metabolism of both carbohydrates and lipids (Ellerby et al. 2003, Pan et al. 2019). Pan et al. (2019) identified two enzymes involved in lipid metabolism that contain convergent amino acid changes in flightless birds.

Table 1: Biological functions enriched in flightless rails and the associated genes under different selective pressures. The columns indicate the genes under relaxed selection in flightless rails, the genes exhibiting an accelerated evolutionary rate in flightless rails compared to the volant rails and cranes, and the genes encoding proteins that carry function-altering variants deleterious in flightless species.

Function	Relaxed	Accelerated evolutionary rate	Proteins carrying function-altering amino acid change in:			
			A. rogersi	*G. australis*	*P. hochstetteri*	*Z. atra*
Muscles development	BVES, GJC1, GTF3C5, HEY2, KLHL40, NPRL3, POPDC3, S1PR1, SMAD1, SVIL, TMOD3	GTF3C5, HOXD10, KLHL40, LRRC40, PRKAR1A, TENM4	BCL9, BCL9L, LMOD2, MYPN, NEBL, NRAP, PDGFRA, RB1, SMAD1, SVIL	CALD1, MYLK, RCSD1		
Circulatory system	BVES, CCNB2, GJC1, HEY2, KCNA5, LAMA4, MAP3K3, NPRL3, PLCD3, POPDC3, PRKDC, S1PR1, SALL4, SLC1A1, SMAD1, ZDHHC21	ANGPTL4, BTK, FOXN1, GJA5, GLO1, HOXB7, JMJD6, KMT2E, LAMA4, LRRC40, MEIS1, NDNF, NTRK3, PRKAR1A, PTK7, SALL4, SFRP2, SMAD5, TENM4, TMOD3, VEZF1, WNT4	ACE2, AGTR2, AVPR1B, DOCK2, FLVCR1, IL18RAP, KMT2E, KNG1, LVRN, NAV2, NPPA, NUP155, RB1, SLC1A1, SLC27A4, SPI1		ACE2, AVPR1B, BVES, CACNA1G	
Immune response	AICDA, BVES, CCNB2, MSH3, PCID2, PRKDC, RSAD2, TMOD3, TPO, VPS33A	BTK, FOXN1, GLO1, HOXB7, JMJD6, KMT2E, MEIS1, REV1, SFRP2, SMAD5, TMOD3, WNT4	APPL2, CLOCK, FAM35A, IL18RAP, PDCD4, RB1, RIPK2, SMPDL3B, SPI1, TRAF3, Tspan32	C1QBP, CASP1, CRH, CYP8B1, DTX3L, FBLN1, FGG, IL18RAP, LIPA, MACF1, MMRN1, SMPDL3B, TICAM1	ACE2, AVPR1B, PIKFYVE, USP6NL	AHSG, C1QC, C5, CBLB, CCR5, CD9, CLOCK, FBXW7, IFNG, IL1R1, NFAM1, PCID2, PCK1, PIK3AP1, PLCL2, RNF26, SEMA7A, SPI1, TRAFD1, TTLL12
Gene regulation and post-translational modification		ALKBH3, ARID1B, ASPH, CHD1L, FAM98B, JMJD6, KDM5A, KMT2E, LEPREL1, MYB, NAP1L1, NDUFAF7, NTMT1, PHF2, PIGF, PRDM16, SCMH1, SETD1B, SPT2, TRMT13, TRMT44, UBN2, USP16, WIPI1, WIPI2, ZCCHC4, ZDHHC12	ARMT1, BCDIN3D, BRPF3, CLOCK, EHMT1, ING5, KMT2E, MECOM, PRDM4, RBM15, SPI1, TADA1, TRMT61B, WBSCR22	SEH1L, HORMAD2, INCENP, KIF2C	SGO2, ATM, CHAF1A, DCLRE1A, DLGAP5, FBH1, HELB, INCENP, MLH3, MMS19, PRDM16, RAD51, SART3, SETD5, SMYD4, TET2, TPRKB, TRMT5, UIMC1, ZNFX1	CABIN1, CHD1, FBH1, KAT6B, MCM2, NAV2, PSME4, SART3
Lipid metabolic process	ABCD2, AOAH, CRABP1, DECR1, ENPP6, PLCD3		AGK, ATP8B4, NAAA, OSBPL9, PDZD8, PITPNM2, PLIN2, PNPLA2, PSAP, RFT1, SCARB2, SLC27A4, SMPD4, SMPDL3B, SQLE	ACOX2, ACSM3, ACSM4, CRH, CUBN, FAM135B, GDPD2, LIPA, MTMR1, PIGN, PIGV, PLA2G6, SMPDL3B		AGK, AGRN, ANKRD52, BDKRB2, CDH2, CETP, DRD4, EMC1, GRIN2A, IP6K3, MAGI2, MOGAT1, MTM1, NAAA, OSBPL10, PCK1, PEX26, PIGG, PIGV, PIGY, PITPNM3, PNPLA2, SLC2A1, SMPD4, ST8SIA4, TTC8, VTG2
Cognition and sensory system	CDH8, CECR2, CHRNA3, CHRNB4, CRH, GJC1, HEY2, JPH3, KALRN, LRIG3, MME, MYO3A, NOX3, ROR1, SLC1A1, SLC24A2, STX2	CACNA1G, CADPS, CHRNB3, DNAJC19, DRD4, GRIN1, GUCA1A, OPN3, PDCL, SLC17A6, SLC24A1	GFAP, SLC1A1, SLC24A1			AGRN, CHL1, FBXW7, IFNG, NAE1, PM20D1, SEMA7A, SORL1, SPG20, TNR
Bones and limbs development		DMRT2, FOXN1, GDF5, HOXB7, HOXD10, MEIS1, NPPC, SALL4, SFRP2, SMAD5				
Cilium development and function				CALD1, CCSAP, CEP162, DNAH3, DZIP1, FMN1, FREM2, GAS2L3, GDPD2, INCENP, KIF25, KIF2C, LAMA5, MACF1, MYLK, NRAP, PCLO, PHACTR4, SASS6, SPAG1, TCTN1, TCTN2, TTLL2, WDPCP	ANK3, CKAP2, INCENP	CENPJ, CLIP1, KATNAL1, SLAIN2
Kidney function		AGTR2, JMJD6, PTK7, SMAD5, UPK3A, WNT4		LAMA5, FMN1, FREM2, WDPCP		

Genes of interest

Introns

I established in Chapter 4 and 5 that CDSs (coding sequences) associated with gene regulation, and particularly those involved in histone and chromatin conformation that affect DNA accessibility, show signs of accelerated evolutionary rate, and carry function-altering amino acid substitutions in flightless rails. It, therefore, seems likely that gene expression and their regulation are key to understanding the evolution of flightlessness. Indeed, non-coding regions are known to play a role in differential gene expression and were associated with convergent phenotypic changes in multiple organisms (Vavouri et al. 2007, Frankel et al. 2012, Signor et al. 2016, Burga et al. 2017, Feigin et al. 2018, Sackton et al. 2019).

One approach to detecting a non-coding molecular basis of flightlessness is through examination of the intron characteristics. As they contain *cis*-regulatory elements (Rodriguez-Trelles et al. 2003, Stemmler et al. 2005, Sagai et al. 2005), introns are good targets to study differential gene expression between volant and flightless species. A preliminary analysis to identify introns exhibiting signals of accelerated and decelerated rate of molecular evolution in flightless rails was carried out (see Appendix C6). Gene enrichment analysis revealed an overrepresentation of biological functions similar to those observed in coding regions (Appendix C6 Table 1).

Candidate genes

Results from Chapters 4 and 5 as well as the results from the introns analysis were combined to identify genes of interest to study the evolution of rail flightlessness (Table 2 and Fig. 2). Genes that present significant differences between volant and flightless species for multiple tests are expected to be good candidates; especially when they show evidence of relaxed selection and accelerated evolutionary rate and when they carry function-altering mutations in flightless species. According to these criteria, five genes that presented significant deviations in at least five tests were identified as promising targets to play a role in the transition to flightlessness in rails: ZFAND4, CNKSR3, NUP205, ZGPAT and ZK1290.5 (the first five genes in Table 2). The gene “Zinc Finger AN1-type containing 4” ZFAND4 presented the most distinctive signal as it exhibited significant results for CodeML M0-M2, both relaxed tests,

showed evidence of accelerated evolutionary rate in the CDS region and exhibited deleterious mutations in all flightless rails but *G. australis*. To my knowledge, this gene has not previously been reported to be associated with avian flight capacity, but it is linked to beak shape in Darwin's ground finches (Lawson and Petren 2017). ZFAND4 encodes for a ubiquitin-like protein and is associated with increased susceptibility in acute lung injury in mice (Leikauf et al. 2013). In humans, an important paralogue of this gene is Ubiquitin C (UBC) which is involved in innate immunity (Rajsbaum et al. 2014, Rajsbaum and García-Sastre 2014). The gene CNKSR family member 3 (CNKSR3) is involved in the regulation of transepithelial sodium transport (Ziera et al. 2009) and could be linked to kidney function. The gene Nucleoporin 205 (NUP205) is involved in RNAs and proteins transport through nuclear pores (Grossman et al. 2012). Zinc Finger and G Patch Domain-Containing Protein (ZGPAT) encodes a transcription regulator protein (Xin et al. 2020). Finally, the gene ZK1290.5 encodes an uncharacterized oxidoreductase.

Table 2: Candidate genes showing significant signs of selective pressure in flightless rails. Each column represents a selective pressure acting on the genes based on a series of phylogeny corrected tests. The indicated length is the alignment length used for the different tests and "Chromosome" is the location of the gene in the reference species *Balearica regulorum*.

Gene	Length (bp)	Chromosome	CodeML M0-M2	CodeML relaxed	Positive	HYPHY relaxed	Intensified	Accelerated CDS	Decelerated CDS	Accelerated introns	Decelerated introns	Function-altering variants			
												A. rogersi	*G. australis*	*P. hochstetteri*	*Z. atra*
ZFAND4	1800	7	**1**	**1**	0	**1**	0	**1**	0	0	0	**1**	0	**1**	**1**
CNKSR3	1131	3	**1**	**1**	0	0	**1**	0	0	0	0	0	**1**	**1**	**1**
NUP205	2073	1	**1**	0	**1**	**1**	0	0	0	0	**1**	**1**	0	0	**1**
ZGPAT	525	16	**1**	**1**	0	**1**	0	0	0	0	0	**1**	0	**1**	**1**
ZK1290.5	780	8	**1**	**1**	0	**1**	0	0	0	0	0	0	**1**	**1**	**1**
ACCS	1092	2	**1**	**1**	0	**1**	0	0	0	0	0	**1**	0	**1**	0
AGTR2	1119	11	**1**	0	**1**	0	**1**	**1**	0	0	0	**1**	0	0	0
BIRC6	3213	3	**1**	**1**	0	**1**	0	0	0	0	0	**1**	0	**1**	0
C9ORF64	651	Z	**1**	**1**	0	**1**	0	**1**	0	0	0	**1**	0	0	0
CTNND1	2586	5	**1**	**1**	0	**1**	0	0	0	0	0	**1**	0	0	**1**
EHMT1	744	20	**1**	**1**	0	**1**	0	0	0	0	0	**1**	0	0	**1**
IDUA	1716	Z	**1**	**1**	0	**1**	0	0	0	0	**1**	**1**	0	0	0
IL1R1	771	1	**1**	**1**	0	0	**1**	0	0	0	**1**	0	0	0	**1**
MREG	390	6	**1**	0	0	0	**1**	0	0	0	**1**	**1**	0	0	**1**
NR1H4	1266	25	**1**	0	**1**	0	**1**	0	0	0	0	**1**	0	0	**1**
SALL4	3216	16	**1**	**1**	0	**1**	0	**1**	0	**1**	0	0	0	0	0
SMPD4	867	17	**1**	**1**	0	0	**1**	0	0	0	0	**1**	0	0	**1**
SOGA3	1713	3	**1**	**1**	0	**1**	0	**1**	0	0	0	0	**1**	0	0
SVIL	2844	2	**1**	**1**	0	**1**	0	0	0	0	0	**1**	0	0	**1**
AICDA	597	1	**1**	**1**	0	**1**	0	0	0	0	0	0	0	0	**1**
ALD	906	1	**1**	**1**	0	**1**	0	0	0	0	0	**1**	0	0	0
ALDH18A1	1431	Na	**1**	**1**	0	**1**	0	0	0	0	0	0	0	0	**1**
AMN1	420	1	**1**	**1**	0	**1**	0	0	0	0	0	0	0	0	**1**
ANKRD52	1332	29	**1**	0	**1**	**1**	0	0	0	0	0	0	0	0	**1**
ASAP2	2421	3	**1**	**1**	0	**1**	0	0	0	0	0	**1**	0	0	0
ASPH	1104	2	**1**	**1**	0	**1**	0	**1**	0	0	0	0	0	0	0
BMPR2	2052	6	**1**	**1**	0	**1**	0	**1**	0	0	0	0	0	0	0
BVES	1074	3	**1**	**1**	0	**1**	0	0	0	0	0	0	0	**1**	0
CCDC50	414	9	**1**	**1**	0	**1**	0	0	0	0	0	0	0	0	**1**
CCDC85A	1044	3	**1**	**1**	0	**1**	0	**1**	0	0	0	0	0	0	0
CECR2	285	1	**1**	**1**	0	**1**	0	0	0	0	0	**1**	0	0	0
CEND1	393	5	**1**	**1**	0	**1**	0	0	0	0	0	**1**	0	0	0
CETN2	411	11	**1**	**1**	0	**1**	0	0	0	0	0	**1**	0	0	0
CLTC	4923	17	**1**	**1**	0	**1**	0	0	0	0	0	0	0	0	**1**
COL9A3	963	16	**1**	**1**	0	0	**1**	0	0	0	0	**1**	0	0	0
CORO2A	966	Z	**1**	**1**	0	**1**	0	0	0	0	0	**1**	0	0	0
CRH	513	2	**1**	**1**	0	**1**	0	0	0	0	0	0	**1**	0	0
CYP2J6	1407	9	**1**	**1**	0	**1**	0	0	0	0	0	0	0	**1**	0
CYP8B1	1530	2	**1**	**1**	0	0	**1**	0	0	0	0	0	**1**	0	0
DAAM2	2334	3	**1**	**1**	0	**1**	0	0	0	**1**	0	0	0	0	0
DECR1	939	2	**1**	**1**	0	**1**	0	0	0	**1**	0	0	0	0	0
DOCK10	3198	9	**1**	**1**	0	**1**	0	0	0	0	0	**1**	0	0	0
EFCAB7	1563	8	**1**	**1**	0	0	**1**	**1**	0	0	0	0	0	0	0
ERGIC2	1134	1	**1**	**1**	0	**1**	0	**1**	0	0	0	0	0	0	0
FAM162B	297	3	**1**	**1**	0	**1**	0	0	0	0	0	**1**	0	0	0
FEM1C	1140	Z	**1**	**1**	0	**1**	0	0	0	0	0	0	0	0	**1**
FGG	1257	4	**1**	0	0	**1**	0	0	**1**	0	0	0	**1**	0	0
GALNT9	735	17	**1**	**1**	0	**1**	0	0	0	0	0	0	0	0	**1**
GAS2L3	2115	1	**1**	**1**	0	**1**	0	0	0	0	0	0	**1**	0	0
GDF5	1503	16	**1**	**1**	0	**1**	0	**1**	0	0	0	0	0	0	0
GJD4	1245	2	**1**	**1**	0	**1**	0	0	0	**1**	0	0	0	0	0
GLE1	939	20	**1**	0	0	0	**1**	0	**1**	0	0	**1**	0	0	0
GNB5	990	12	**1**	**1**	0	**1**	0	**1**	0	0	0	0	0	0	0
GPHN	903	5	**1**	0	**1**	**1**	0	0	0	0	0	0	0	**1**	0
GPR55	429	9	**1**	0	0	**1**	0	0	**1**	0	0	**1**	0	0	0
GPR83	1101	11	**1**	**1**	0	**1**	0	**1**	0	0	0	0	0	0	0

Gene	Length (bp)	Chromosome	CodeML M0-M2	CodeML relaxed	Positive	HYPHY relaxed	Intensified	Accelerated CDS	Decelerated CDS	Accelerated introns	Decelerated introns	Function-altering variants			
												A. rogersi	*G. australis*	*P. hochstetteri*	*Z. atra*
GTF3C5	414	20	1	1	0	1	0	1	0	0	0	0	0	0	0
HS3ST6	618	15	1	1	0	1	0	0	0	0	0	0	0	0	1
HS6ST3	591	1	1	1	0	1	0	1	0	0	0	0	0	0	0
HSF2BP	906	1	1	1	0	1	0	0	0	0	0	0	1	0	0
IDS	1470	11	1	1	0	0	1	0	0	0	0	0	0	1	0
IL16	1794	12	1	0	0	0	1	0	1	0	0	0	0	1	0
IRS4	2151	11	1	1	0	1	0	0	0	0	0	1	0	0	0
ITPR3	4638	25	1	1	0	1	0	0	0	1	0	0	0	0	0
JKAMP	435	5	1	1	0	1	0	0	0	0	0	1	0	0	0
KDM5B	1170	25	1	1	0	1	0	0	0	0	0	0	0	0	1
KIF13A	4275	2	1	1	0	1	0	0	0	1	0	0	0	0	0
KLF13	834	12	1	1	0	0	1	1	0	0	0	0	0	0	0
KLF4	483	Z	1	1	0	1	0	0	0	0	0	1	0	0	0
KLHL40	237	2	1	1	0	1	0	1	0	0	0	0	0	0	0
KRT4	1821	29	1	1	0	1	0	0	0	0	0	0	0	0	1
KRT42	546	24	1	0	1	1	0	0	0	0	0	0	1	0	0
LAMA4	780	3	1	1	0	1	0	1	0	0	0	0	0	0	0
LCORL	4452	4	1	1	0	1	0	1	0	0	0	0	0	0	0
LEPREL1	1173	9	1	0	0	1	0	1	0	0	0	0	0	0	1
LRIG3	2037	1	1	1	0	1	0	1	0	0	0	0	0	0	0
LVRN	528	Z	1	1	0	1	0	0	0	0	0	1	0	0	0
LY96	279	2	1	1	0	0	1	0	0	0	0	0	0	0	1
MAGI2	1581	10	1	1	0	1	0	1	0	0	0	0	0	0	0
MAP7	1737	3	1	0	0	1	0	0	0	0	0	1	0	1	0
MBD5	3876	6	1	1	0	0	0	1	0	0	0	0	0	1	0
MYCBP2	1269	1	1	1	0	1	0	1	0	0	0	0	0	0	0
MYO3A	2115	2	1	1	0	1	0	0	0	0	0	1	0	0	0
NDUFAF7	1080	3	1	0	0	1	0	1	0	0	0	0	1	0	0
NOMO1	3276	15	1	1	0	1	0	0	0	0	1	0	0	0	0
PANK1	900	7	1	1	0	1	0	1	0	0	0	0	0	0	0
PAQR8	1059	3	1	1	0	1	0	0	0	0	0	0	0	0	1
PCID2	822	1	1	1	0	1	0	0	0	0	0	0	0	0	1
PCSK6	639	12	1	1	0	1	0	0	0	0	0	1	0	0	0
PGBD5	1230	3	1	1	0	1	0	1	0	0	0	0	0	0	0
PIGY	213	4	1	1	0	1	0	0	0	0	0	0	0	0	1
PLCD3	225	24	1	1	0	1	0	0	0	0	0	1	0	0	0
PPFIBP1	939	1	1	0	0	1	0	1	0	0	0	0	0	1	0
PRR35	1887	15	1	1	0	1	0	1	0	0	0	0	0	0	0
RAPGEF2	4263	4	1	0	0	0	1	0	1	1	0	0	0	0	0
RFX6	2442	3	0	0	0	0	0	1	0	0	0	1	0	1	1
SDC3	756	22	1	1	0	0	1	0	0	0	0	0	0	0	1
SGO2	2778	6	0	0	0	0	0	0	1	0	0	1	0	1	1
SGSM3	1812	1	1	1	0	0	1	0	0	0	0	0	0	1	0
SLC1A1	621	Z	1	1	0	1	0	0	0	0	0	1	0	0	0
SLC2A1	1344	21	1	1	0	1	0	0	0	0	0	0	0	0	1
SMAD1	657	4	1	1	0	1	0	0	0	0	0	1	0	0	0
SORL1	2469	23	1	1	0	1	0	0	0	0	0	0	0	0	1
SRGAP3	1356	10	1	1	0	1	0	0	0	0	1	0	0	0	0
TAF4	2985	16	1	1	0	1	0	0	0	0	0	1	0	0	0
TCFL5	1557	16	1	0	0	1	0	0	0	0	0	1	0	0	1
TENM4	2085	1	1	1	0	1	0	1	0	0	0	0	0	0	0
TMEM179	696	5	1	1	0	1	0	1	0	0	0	0	0	0	0
TMOD3	1050	12	1	1	0	1	0	1	0	0	0	0	0	0	0
TUBA8	972	25	1	1	0	1	0	0	0	0	0	1	0	0	0
UBR4	11145	21	1	1	0	1	0	0	0	0	0	0	0	0	1
UQCRC1	1368	10	1	1	0	1	0	0	0	0	0	0	0	0	1
VPS33A	1635	17	1	1	0	1	0	0	0	0	0	0	0	1	0
WIZ	1548	30	1	1	0	1	0	0	0	0	0	0	0	0	1
WWC2	528	4	1	1	0	1	0	0	0	0	0	1	0	0	0
XPNPEP1	1833	7	1	0	1	0	1	0	0	0	0	0	1	0	0
XYLT1	1512	15	1	1	0	1	0	0	0	0	0	0	0	1	0
ZER1	2301	20	1	1	0	1	0	1	0	0	0	0	0	0	0

The gene Spalt-Like Transcription Factor 4 (SALL4) showed evidence of accelerated evolution rate in both coding and non-coding regions and was also under relaxed selection (according to both CodeML and HYPHY tests). This gene is known to be involved in limb development in mice (Akiyama et al. 2015) and in regulating hematopoietic cell development in humans (Yang et al. 2012 p. 4). These biological functions could plausibly be linked with the ability to fly in birds.

Of the 118 candidate genes (that were significant for four tests or more, Table 2), 37 had been identified as good candidates in the gene enrichment analyses (i.e., genes present in both Table 1 and 2). This means these genes exhibit molecular signals associated with flightlessness according to different phylogeny corrected tests and are associated with biological functions related to flight capacity (Table 3). They, therefore, appear to be ideal candidates for further investigation into their role in the genetic pathway to flightlessness (see Future research section).

Table 3: Genes associated with biological function involved in flight capacity and that show multiple signs of differential selective pressures in flightless rails.

Function	Associated genes
Muscle development	BVES, GTF3C5, KLHL40, SMAD1, SVIL, TMOD3, TENM4
Circulatory system	AGTR2, BVES, LAMA4, LVRN, PLCD3, SALL4, SLC1A1, SMAD1, TENM4, TMOD3
Immune system	AICDA, BVES, CRH, CYP8B1, FGG, PCID2, TMOD3, VPS33A
Gene regulation and post-translational modification	ASPH, EHMT1, LEPREL1, NDUFAF7, SGO2
Lipid metabolic process	ANKRD52, CRH, DECR1, MAGI2, PIGY, PLCD3, SLC2A1, SMPD4
Cognition and sensory system	CECR2, CRH, LRIG3, MYO3A, SLC1A1, SORL1
Bone and limb development	GDF5, SALL4
Cilium development and function	GAS2L3
Kidney function	AGTR2

Interrelations between test results

Tests results for each gene were compared to one another using the Phi-coefficient (see Chapter 4) that measures the association between two variables (Ekström 2011) (Fig. 2). As reported in

Chapter 4, the two tests for relaxed selection showed the highest association with a Phi-coefficient of 0.57. The CodeML M0-M2 test is designed to assess the potential difference of selective pressures between two species groups. It is therefore not surprising to observe a positive association between this test and both relaxed selection tests as well as the intensified selection test. To a lesser extent, CodeML M0-M2 was also associated with positive selection and both accelerated and decelerated CDS. This indicates CodeML M0-M2 is a good test to use with gene alignments to detect uncharacterized selective pressure between two groups of taxa.

The results for accelerated and decelerated intron evolution exhibited no association with any of the other results implying that selective pressure acting on introns is independent of selective pressure acting on coding regions.

Finally, in proteins carrying function-altering amino acid replacement in the flightless rails *A. rogersi* and *Z. atra* were the most correlated variables, and this was expected based on the separate phylogenetic sampling. Indeed, no volant species that are closely related to either of these species were available in the analysis, so *P. melanotus* was used as a reference for both. Consequently, function-altering amino acid replacement shared by *A. rogersi* and *Z. atra* could reflect mutations that occurred in their common ancestor and not be caused by convergent evolution to flightlessness.

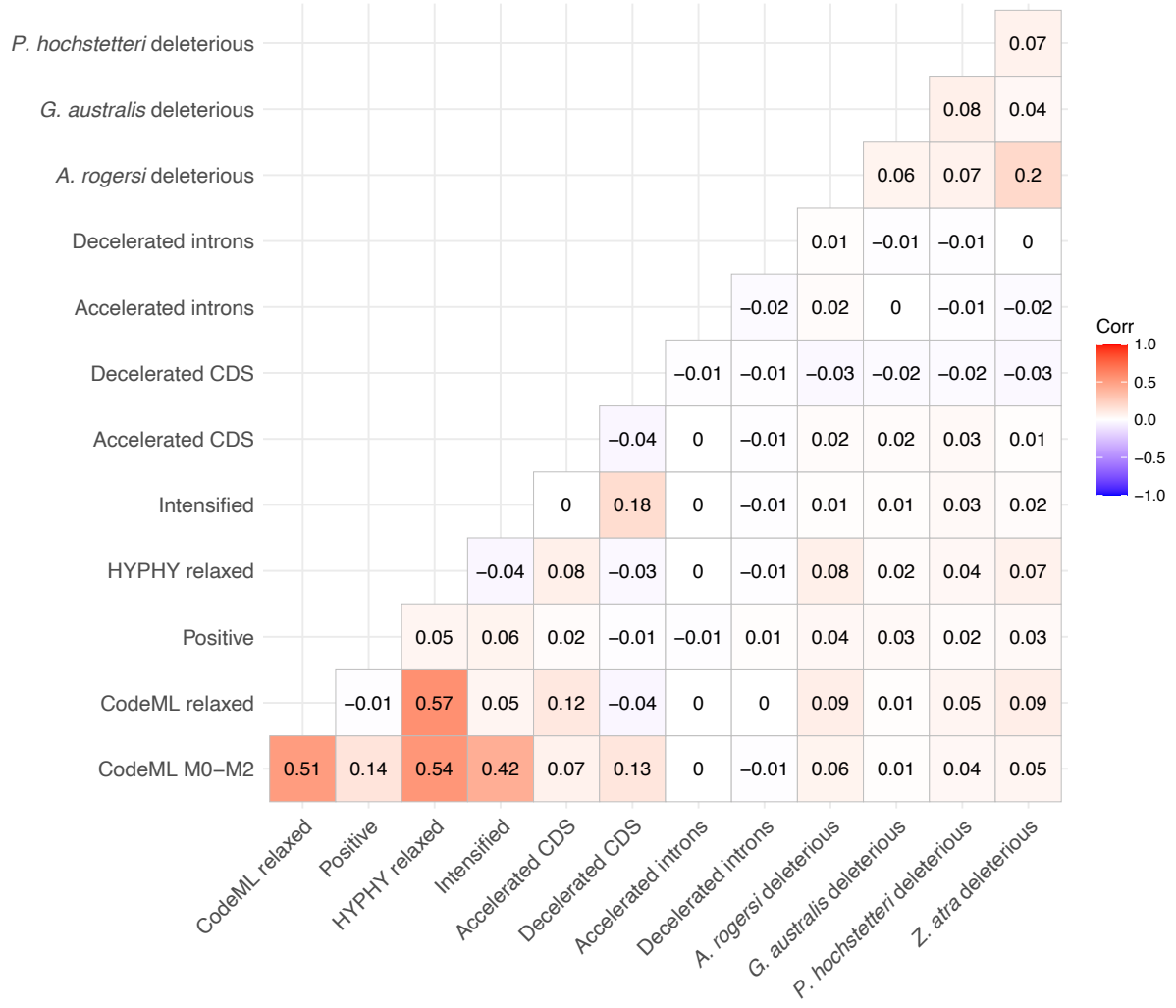


Figure 2: Variable association matrix showing interrelations between all the tests carried out on rails genes. A positive Phi-coefficient (red) indicates a strong association between two test results and a Phi-coefficient close to zero indicates an absence of association.

Levels of convergent evolution

The current study shows that genomic regions evolve in a convergent manner in flightless rails (see Chapters 4 and 5). There are however different ways to characterize convergent evolution (Hao et al. 2019) and I decided in this thesis to investigate these evolutionary mechanisms from a small scale, the amino acid level, to a much broader scale.

Amino acid level

Convergent amino acid substitutions are sometimes associated with similar phenotypic changes in independent lineages (Besnard et al. 2009, Vizueta et al. 2019). In rails, however, a limited number of amino acid substitutions showed evidence of correlation with the flightless character (see Chapter 5). Moreover, this genetic convergence involved only single amino acids, compared to longer convergent sequences in Vizueta et al. (2019), and no more than one per gene was found compared to multiple convergent substitutions in the same gene found in Besnard et al. (2009). These results do not provide strong evidence that the transition to flightlessness is caused by similar amino acid changes across flightless rails.

Gene and protein level

At the gene level, hundreds of genomic regions showed evidence of convergent molecular evolution (see Chapter 4). Genes displayed a similar response to selective pressure (relaxed, positive, or intensified selection) in flightless lineages despite the independent evolution of the loss of flight. And at the protein level, numerous sequences showed an accelerated evolutionary rate in correlation with the flightless trait (see Chapter 5). However, the set of genes exhibiting different selective pressure in flightless rails based on branch-site models (CodeML and HYPHY) showed many differences with the set of genes exhibiting an accelerated evolutionary rate in flightless species (Fig. 2). From a technical point of view, this reflects the fact that the branch-site models detect selection based on the d_N/d_S ratio while the relative evolutionary rate test calculates the accumulation of all substitutions in each lineage to detect convergence based on relative branch length.

Results based on proteins with function-altering amino acid substitutions in flightless rails show only limited evidence of convergent molecular evolution and rather suggest lineage-specific evolutionary solutions (see Chapter 5).

Biological function level

Adaptative phenotypic convergence may be influenced by function-level changes rather than shared specific gene variations (Roda et al. 2013, Berens et al. 2015, Yang et al. 2015, Zhang et al. 2016, Sun et al. 2018). This means convergent phenotype can result from selective

pressures acting on sets of genes and/or sets of amino acids within genes that are different in each species (Hao et al. 2019). The genes on which selection acts will depend largely on the variants that emerge by chance. Or in other words, convergent evolution will follow diverse genetic pathways to end up with similar phenotypic results. Function-level convergence was observed in flightless rails. Indeed, genes that showed evidence of relaxed selection and accelerated evolutionary rate (both in CDSs and introns) in flightless rails were enriched in similar biological functions (Table 1) despite including different sets of genes. The proteins carrying function-altering amino acid changes were different in each flightless rail but showed similarities in GO terms which implies that selection acts on each species through different genetic pathways but with similar functional effects and so ultimately appears as a convergent phenotypic response.

Morphological level

The morphological convergent evolution of flightless birds is well demonstrated (Livezey 2003, Roots 2006, Gaspar et al. 2020). In rails, flightless species share morphological characters that distinguish them from their volant relatives. Flightless rails typically exhibit smaller sterna and wings than volant taxa along with wider pelves and more robust femora (Livezey 2003, Gaspar et al. 2020). Moreover, it has been shown that these differences are independent of phylogeny and instead demonstrate convergent morphological adaptation associated with a walking ecology (see Chapter 2).

Takahē

The flightless takahē *Porphyrio hoschetteri* is the world's largest living rail and is endemic to the South Island of Aotearoa New Zealand. Takahē were once widely distributed across the landscape in many different environments but habitat destruction and hunting reduced the range and by the end of the 19th century the takahe was believed to be extinct (Trewick and Worthy 2001). A small population (maybe as few as two individuals) was however rediscovered in 1948 and since then, that species has been under monitoring and conservation technics including captive breeding, island translocation and wild releases (Wallace 2002, Hegg et al. 2012). The recent extreme bottleneck and the slowly recovering population (445 in 2020 -www.doc.govt.nz) means the takahē is still critically endangered.

In the present study, the takahē was observed to be out of the ordinary in many ways. First, it is the largest rail, which makes it morphologically different from other rails. It is heavier, has a bigger head, wider pelvis and longer legs than any other rallids (Gaspar et al. 2020). Second, its genome assembly showed an extremely high level of homozygosity (See chapter 3). This is likely due to its population collapse and to the extreme bottleneck event (Baker et al. 1995, Burga et al. 2017, White et al. 2018). Finally, the proportion of deleterious mutations observed in takahē was much higher than in other flightless species despite the close phylogenetic relationship with its reference species *P. melanotus* (See Chapter 5). It therefore appears that deleterious mutations are less effectively filtered through selective processes in takahē which is likely to result from the small population and the resulting inbreeding depression (Grueber et al. 2010).

Genomics provides a powerful tool to support intensive management of threatened species, therefore, the takahē genome assembly may be a valuable resource for future conservation research and action for this species. Indeed, genomics increases the precision and accuracy of estimations that involve neutral loci such as migration rates and effective population size (Luikart et al. 2003, Allendorf et al. 2010). Furthermore, studying populations at the genome-scale can address important questions about inbreeding depression (Allendorf et al. 2010). Investigating thousands of loci can increase the accuracy of kinship estimates (Santure et al. 2010) and genes associated with variation in breeding success can be identified (Wright et al. 2020).

Future research

The present research provides detailed methods to study the molecular basis of a convergent trait and has identified multiple genes and biological functions that are associated with the transition to flightlessness in rails. In this section, I describe the potential changes, improvements and further analyses that could be done to increase the quality and robustness of the presented results.

Sampling

Prior to this study, genome assemblies of four rails were available: *Gallirallus okinawae*, *Fulica atra*, *Zapornia atra* and *Atlantisia rogersi. Gallirallus okinawae* was quickly excluded

from further analysis as the assembly quality was much lower than the others (see Chapter 3). Four additional rail genomes (two pairs of volant-flightless species) were sequenced and assembled: *Gallirallus australis*, *Gallirallus philippensis*, *Porphyrio hochstetteri* and *Porphyrio melanotus*, bringing a total of seven good quality genomes.

To study the transition to flightlessness the ideal sampling would include multiple closely related volant and flightless species pairs, so it is likely that molecular differences between these pairs are associated with the difference in their ability to fly and not caused by phylogeny. The two pairs of New Zealand rails *P. hochstetteri – P. melanotus* and *G. australis – G. philippensis* fit this criterion, but *F. atra*, *A. rogersi* and *Z. atra* do not. Thus, it would be interesting to obtain at least three more genome assemblies that are closely related to these species to improve this research. For instance, the Tasmanian nativehen *Tribonyx mortierii* is a flightless rail species that is relatively close to *F. atra* and would therefore be interesting to sample. Alternatively, the Argentinean giant coot (*F. gigantea*) in which adults are considered functioning flightless, or one of the extinct New Zealand flightless coots (*Fulica prisca*, *F. chathamensis*) would be suitable if recent bones fossils of good quality are available. The spotless crake *Zapornia tabuensis* is a volant rail with a broad distribution across the south pacific (including Aotearoa New Zealand) that is part of the same clade as *Zapornia atra* (Garcia-R et al. 2014) and would therefore be a suitable addition to the genome sampling. Finally, the dot-winged crake *Laterallus spiloptera* would form a volant – flightless pair with *A. rogersi* as they are closely related (Stervander et al. 2019). This would generate 5 pairs of volant – flightless rails and would remove a lot of uncertainty caused by the current genome sampling.

Another thing to consider is the sex of sampled species. Indeed, the four genomes we assembled came from males which means the W chromosome was not present. In the future, it would be of interest to add female specimens.

Type of selective pressure

In the current research, I focus on genes exhibiting relaxed selection. This is because I investigate the molecular basis of a lost character (flight) which, to my knowledge, mostly reduces selective pressure rather than applies new selective pressure. For that reason, I expect most genes associated with the loss of flight will be under relaxed selection rather than positive

or intensified selection. Similarly, I focus on genomic regions with an accelerated evolutionary rate in flightless rails rather than regions with decelerated evolutionary rates. Again, I hypothesize that the transition to flightlessness would reduce the selective pressure on genes associated with flight and therefore result in an accelerated mutation rate rather than a decelerated mutation rate that could be a sign of purifying selection.

This methodology is backed up by the GO enrichment analyses that show that overrepresented biological functions associated with flight (muscle, bones, circulatory system development, etc.) are found in greater numbers in relaxed and accelerated genes than in intensified and decelerated genes.

However, to better understand the genetic pathways to flightlessness, more effort should be put into evaluating the role of genes under positive and intensified selection in flightless rails as well as those with decelerated evolutionary rates. The loss of flight directly impacts the selective constraints on genes associated with flight but also results in a different lifestyle that can impact selective constraints on other traits and indirectly play a role in the evolutionary consequences of flightlessness. For instance, the smaller and less diversified habitat of flightless rails is a consequence of the loss of flight and could play an important role in the evolution of their immune system (Møller and Erritzøe 1998, Duffy and Vargas 2018, Eikenaar et al. 2020). The evolutionary trajectory of the selective pressure acting on the immune system in flightless rails appears complex as genes associated with that function are under relaxed selection but also show evidence of both accelerated and decelerated evolutionary rates (see Chapter 4 and 5). It, therefore, appears that different immune genes are under different types of selection. Thus, flightless rails do not simply show a relaxation of immunity but rather exhibit a complex evolutionary shift in the way they deal with parasites and diseases compared to their volant relatives.

To further study the impact of flightlessness on the immune response, the Avian Immunome DB (AVIMM) could be used to target specifically the genomic regions that are known to be associated with that biological function in birds (Mueller et al. 2020)

Non-coding regions

Investigating the molecular signal of a convergent trait at the genomic scale can be done in many ways. In the present study, I targeted the coding regions because these sequences are usually well conserved between species and can easily be retrieved when the genomes are annotated. Moreover, the mechanisms involved in the evolution of exons are well known and can easily be quantified in a phylogenetic context (see Chapters 4 and 5).

However, the coding regions represent a small part of the genomes and a growing number of studies show the importance of non-coding regions in phenotypic adaptations (Stern 2000, Signor et al. 2016, Feigin et al. 2018, Sackton et al. 2019, Hao et al. 2019). Investigating the non-coding genomic regions is therefore required to understand the genetic pathway to avian flightlessness in more detail.

Intron evolutionary rate in volant and flightless rails was evaluated to find evidence of convergence associated with flightlessness (see Introns section in Appendix C6). This work has uncovered some promising preliminary results unveiling a small part of the non-coding basis of flightlessness. To streamline the analysis, all the introns from the same gene were concatenated into one sequence which could mask some convergent signal if only a small section of that sequence exhibits accelerated evolutionary rate (by dilution effect). To tackle that issue, each intron could be considered separately or intron alignments could be split into subsections as presented in Rubin et al. (2019).

At the genomic scale, a way to identify regions associated with loss of flight would be to search for ultraconserved non-coding elements showing accelerated molecular evolution in flightless species. This method has been successfully used to study phenotypic adaptions in multiple organisms (Prabhakar et al. 2008, Booker et al. 2016, Ramírez et al. 2021) and more specifically to study the evolution of avian flightlessness in ratites (Sackton et al. 2019) and cormorants (Burga et al. 2017).

Transcriptomic

Differential gene expression can play an important role in convergent phenotypic adaptation (Berens et al. 2015, Yang et al. 2015). In ratites, Sackton et al. (2019) showed that

morphological convergence associated with loss of flight has a substantial shared regulatory component. Moreover, I showed in Chapter 5 that the genomic regions associated with gene regulation exhibit an accelerated evolutionary rate in flightless rails. For these reasons, I suggest comparative transcriptomic analyses could be carried out on rails to identify genes with different expression levels associated with loss of flight. A first step could be to focus on the *P. melanotus – P. hochstetteri* pair that is genetically very close which means the differences in expression are likely to be involved in phenotypic changes rather than being caused by phylogeny. However, there are significant challenges that would need to be overcome to do transcriptomic analyses in highly endangered species such as the *P. hochstetteri* where invasive sampling is not possible.

Candidate genes

Here I present numerous candidate genes that could be involved in the evolution of flightlessness. To confirm their role in phenotypic changes, further investigations are required.

A good starting point would be to focus on the 37 genes that were identified as being good candidates for a role in flightlessness (Table 3). Gathering more information on the genetic pathways they are involved in or the phenotypic changes they cause when being mutated in model species could be of great help. Retrieving these gene sequences in more species would also allow confirmation of their association with flightlessness and increase the robustness of phylogeny-based tests.

I identified different sets of proteins carrying function-altering substitutions in flightless rails. I however only assess the functional impact of amino acid substitutions and not the impact of indels. Despite the relatively low number of variants caused by insertions and deletions (3.2% of all variations in the Galapagos flightless cormorant), these changes can have a significant influence on protein function (Burga et al. 2017). For that reason, it would be of great value in future research to investigate this aspect of protein variants in flightless rails. The functional impact of protein variants was assessed using PROVEAN by collecting a set of homologous proteins related to the reference sequence, aligning them, and assigning a score to each variation based on the alignment (Choi et al. 2012). To confirm PROVEAN results more practically, different approaches are possible. First, a functional validation can be carried out as in Burga et al. (2017) by genetically modifying a model species (in this case *Caenorhabditis*

elegans) to confirm that a studied variant is sufficient to affect the protein function. Second, protein structure prediction can be used to facilitate mechanistic understanding of their function based on amino acid sequences (Kuhlman and Bradley 2019, Jumper et al. 2021). Thus, the deleterious impact of a protein variant can theoretically be assessed by evaluating the structural changes caused by the mutations. However, predicting the effect of variants is still a major challenge and, to our knowledge, most of the research done in this field involves human diseases (Ittisoponpisan et al. 2019) or model species (Hwang et al. 2021, Bodawatta et al. 2022).

References

Akiyama, R., Kawakami, H., Wong, J., Oishi, I., Nishinakamura, R. and Kawakami, Y. 2015. Sall4-Gli3 system in early limb progenitors is essential for the development of limb skeletal elements. – Proceedings of the National Academy of Sciences of the United States of America.

Allendorf, F. W., Hohenlohe, P. A. and Luikart, G. 2010. Genomics and the future of conservation genetics. – Nat Rev Genet 11: 697–709.

Baker, A. J., Daugherty, C. H., Colbourne, R. and McLennan, J. L. 1995. Flightless brown kiwis of New Zealand possess extremely subdivided population structure and cryptic species like small mammals. – PNAS 92: 8254–8258.

Balda, R. P. and Kamil, A. C. 1989. A comparative study of cache recovery by three corvid species. – Animal Behaviour 38: 486–495.

Berens, A. J., Hunt, J. H. and Toth, A. L. 2015. Comparative Transcriptomics of Convergent Evolution: Different Genes but Conserved Pathways Underlie Caste Phenotypes across Lineages of Eusocial Insects. – Molecular Biology and Evolution 32: 690–703.

Berger, M., Hart, J. S. and Roy, O. Z. 1970. Respiration, oxygen consumption and heart rate in some birds during rest and flight. – Z. Vergl. Physiol. 66: 201–214.

Besnard, G., Muasya, A. M., Russier, F., Roalson, E. H., Salamin, N. and Christin, P.-A. 2009. Phylogenomics of C(4) photosynthesis in sedges (Cyperaceae): multiple appearances and genetic convergence. – Mol Biol Evol 26: 1909–1919.

Bodawatta, K. H., Hird, S. M., Grond, K., Poulsen, M. and Jønsson, K. A. 2022. Avian gut microbiomes taking flight. – Trends in Microbiology 30: 268–280.

Booker, B. M., Friedrich, T., Mason, M. K., VanderMeer, J. E., Zhao, J., Eckalbar, W. L., Logan, M., Illing, N., Pollard, K. S. and Ahituv, N. 2016. Bat Accelerated Regions Identify a Bat Forelimb Specific Enhancer in the HoxD Locus. – PLOS Genetics 12: e1005738.

Broom, L. J. and Kogut, M. H. 2018. The role of the gut microbiome in shaping the immune system of chickens. – Veterinary Immunology and Immunopathology 204: 44–51.

Burga, A., Wang, W., Ben-David, E., Wolf, P. C., Ramey, A. M., Verdugo, C., Lyons, K., Parker, P. G. and Kruglyak, L. 2017. A genetic signature of the evolution of loss of flight in the Galapagos cormorant. – Science 356: eaal3345.

Butler, P. J. 2016. The physiological basis of bird flight. – Philosophical Transactions of the Royal Society B: Biological Sciences 371: 20150384.

Caviedes-Vidal, E., McWhorter, T. J., Lavin, S. R., Chediack, J. G., Tracy, C. R. and Karasov, W. H. 2007. The digestive adaptation of flying vertebrates: High intestinal paracellular absorption compensates for smaller guts. – PNAS 104: 19132–19137.

Choi, Y., Sims, G. E., Murphy, S., Miller, J. R. and Chan, A. P. 2012. Predicting the functional effect of amino acid substitutions and indels. – PLoS One 7: e46688.

Chou, K.-C. 2020. Progresses in Predicting Post-translational Modification. – International Journal of Peptide Research & Therapeutics 26: 873–888.

Deef, L. 2015. Histo-and Ultrastructural Aspects Concerning Renal Corpuscle in Fulica atra and Gallinula angulata (Aves: Gruiformes). – International Journal of Advanced Research 3: 1833–1838.

Diamond, J. M. 1977. Continental and Insular Speciation in Pacific Land Birds. – Systematic Biology 26: 263–268.

Diamond, J. 1991. A New Species of Rail from the Solomon Islands and Convergent Evolution of Insular Flightlessness. – Auk 108: 461–470.

Dobson, A. P. and Mccallum, H. 1997. The role of parasites in bird conservation. – Clayton, D H, Moore, J Host-parasite evolution 155–173.

Duffy, D. C. and Vargas, F. H. 2018. From the Vagile to the Sedentary: Disease Implications and New Host Relationships on Islands. – In: Parker, P. G. (ed), Disease Ecology: Galapagos Birds and their Parasites, Social and Ecological Interactions in the Galapagos Islands. Springer International Publishing, pp. 113–135.

Dukas, R. 1999. Costs of Memory: Ideas and Predictions. – Journal of Theoretical Biology 197: 41–50.

Duncan, E. J., Gluckman, P. D. and Dearden, P. K. 2014. Epigenetics, plasticity, and evolution: How do we link epigenetic change to phenotype? – Journal of Experimental Zoology Part B: Molecular and Developmental Evolution 322: 208–220.

Eikenaar, C., Hessler, S. and Hegemann, A. 2020. Migrating birds rapidly increase constitutive immune function during stopover. – Royal Society Open Science 7: 192031.

Ekström, J. 2011. The Phi-coefficient, the Tetrachoric Correlation Coefficient, and the Pearson-Yule Debate.

Ellerby, D. J., Cleary, M., Marsh, R. L. and Buchanan, C. I. 2003. Measurement of Maximum Oxygen Consumption in Guinea Fowl Numida meleagris Indicates That Birds and Mammals Display a Similar Diversity of Aerobic Scopes during Running. – Physiological and Biochemical Zoology 76: 695–703.

Feigin, C. Y., Newton, A. H., Doronina, L., Schmitz, J., Hipsley, C. A., Mitchell, K. J., Gower, G., Llamas, B., Soubrier, J., Heider, T. N., Menzies, B. R., Cooper, A., O'Neill, R. J.

and Pask, A. J. 2018. Genome of the Tasmanian tiger provides insights into the evolution and demography of an extinct marsupial carnivore. – Nat Ecol Evol 2: 182–192.

Filardi, C. E. and Moyle, R. G. 2005. Single origin of a pan-Pacific bird group and upstream colonization of Australasia. – Nature 438: 216–219.

Frank, S. A. 2000. Specific and Non-specific Defense against Parasitic Attack. – Journal of Theoretical Biology 202: 283–304.

Frankel, N., Wang, S. and Stern, D. L. 2012. Conserved regulatory architecture underlies parallel genetic changes and convergent phenotypic evolution. – PNAS 109: 20975–20979.

Garcia-R, J. C., Gibb, G. C. and Trewick, S. A. 2014. Deep global evolutionary radiation in birds: Diversification and trait evolution in the cosmopolitan bird family Rallidae. – Molecular Phylogenetics and Evolution 81: 96–108.

Garcia-R, J. C., Joseph, L., Adcock, G., Reid, J. and Trewick, S. A. 2017. Interisland gene flow among populations of the buff-banded rail (Aves: Rallidae) and its implications for insular endemism in Oceania. – Journal of Avian Biology 48: 679–690.

Gaspar, J., Gibb, G. C. and Trewick, S. A. 2020. Convergent morphological responses to loss of flight in rails (Aves: Rallidae). – Ecology and Evolution 10: 6186–6207.

Gerson, A. R. and Guglielmo, C. G. 2013. Measurement of glomerular filtration rate during flight in a migratory bird using a single bolus injection of FITC-inulin. – American Journal of Physiology-Renal Physiology 305: F823–F829.

Gibb, G. C. and 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.

Giulivi, C. and Ramsey, J. 2015. On fuel choice and water balance during migratory bird flights. – Int Biol Rev 2015: 58.

Goetz, S. C. and Anderson, K. V. 2010. The primary cilium: a signalling centre during vertebrate development. – Nat Rev Genet 11: 331–344.

Gold, M. E. L. and Watanabe, A. 2018. Flightless birds are not neuroanatomical analogs of non-avian dinosaurs. – BMC Evol Biol 18: 190.

Goldstein, D. L. 2022. Chapter 19 - Renal and extrarenal regulation of body fluid composition. – In: Scanes, C. G. and Dridi, S. (eds), Sturkie's Avian Physiology (Seventh Edition). Academic Press, pp. 411–443.

Grossman, E., Medalia, O. and Zwerger, M. 2012. Functional Architecture of the Nuclear Pore Complex. – Annual Review of Biophysics 41: 557–584.

Grubb, B. R. 1983. Allometric relations of cardiovascular function in birds. – American Journal of Physiology-Heart and Circulatory Physiology 245: H567–H572.

Grueber, C. E., Laws, R. J., Nakagawa, S. and Jamieson, I. G. 2010. Inbreeding Depression Accumulation across Life-History Stages of the Endangered Takahe. – Conservation Biology 24: 1617–1625.

Hao, Y., Qu, Y., Song, G. and Lei, F. 2019. Genomic Insights into the Adaptive Convergent Evolution. – Current Genomics 20: 81–89.

Hegg, D., Greaves, G., Maxwell, J. M., MacKenzie, D. I. and Jamieson, I. G. 2012. Demography of takahe (Porphyrio hochstetteri) in Fiordland: environmental factors and management affect survival and breeding success. – New Zealand Journal of Ecology 36: 75–89.

Huang, T., Peng, X.-Y., Gao, B., Wei, Q.-L., Xiang, R., Yuan, M.-G. and Xu, Z.-H. 2019. The Effect of Clostridium butyricum on Gut Microbiota, Immune Response and Intestinal Barrier Function During the Development of Necrotic Enteritis in Chickens. – Frontiers in Microbiology 10: 2309.

Hume, J. P. and Martill, D. 2019. Repeated evolution of flightlessness in Dryolimnas rails (Aves: Rallidae) after extinction and recolonization on Aldabra. – Zool J Linn Soc 186: 666–672.

Hwang, E., Kim, H., Truong, A. D., Kim, S.-J. and Song, K.-D. 2021. Suppression of the TLR3 mediated pro-inflammatory gene expressions by PPDPF in chicken DF-1 cells. – Journal of Animal Science and Technology 23: 165–175.

Isler, K. and van Schaik, C. P. 2006. Metabolic costs of brain size evolution. – Biology Letters 2: 557–560.

Ittisoponpisan, S., Islam, S. A., Khanna, T., Alhuzimi, E., David, A. and Sternberg, M. J. E. 2019. Can Predicted Protein 3D Structures Provide Reliable Insights into whether Missense Variants Are Disease Associated? – Journal of Molecular Biology 431: 2197–2212.

Jenni, L. and Jenni-Eiermann, S. 1998. Fuel Supply and Metabolic Constraints in Migrating Birds. – Journal of Avian Biology 29: 521–528.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P. and Hassabis, D. 2021. Highly accurate protein structure prediction with AlphaFold. – Nature 596: 583–589.

Kennedy, E. S., Grueber, C. E., Duncan, R. P. and Jamieson, I. G. 2014. Severe Inbreeding Depression and No Evidence of Purging in an Extremely Inbred Wild Species—the Chatham Island Black Robin. – Evolution 68: 987–995.

Krebs, J. R., Sherry, D. F., Healy, S. D., Perry, V. H. and Vaccarino, A. L. 1989. Hippocampal specialization of food-storing birds. – PNAS 86: 1388–1392.

Kuhlman, B. and Bradley, P. 2019. Advances in protein structure prediction and design. – Nat Rev Mol Cell Biol 20: 681–697.

Lawson, L. P. and Petren, K. 2017. The adaptive genomic landscape of beak morphology in Darwin's finches. – Molecular Ecology 26: 4978–4989.

Leikauf, G. D., Concel, V. J., Bein, K., Liu, P., Berndt, A., Martin, T. M., Ganguly, K., Jang, A. S., Brant, K. A., Dopico, R. A., Upadhyay, S., Cario, C., Di, Y. P. P., Vuga, L. J., Kostem, E., Eskin, E., You, M., Kaminski, N., Prows, D. R., Knoell, D. L. and Fabisiak, J. P. 2013. Functional Genomic Assessment of Phosgene-Induced Acute Lung Injury in Mice. – Am J Respir Cell Mol Biol 49: 368–383.

Livezey, B. C. 2003. Evolution of Flightlessness in Rails. – American Ornithologists' Union.

Losos, J. B., Warheitt, K. I. and Schoener, T. W. 1997. Adaptive differentiation following experimental island colonization in Anolis lizards. – Nature 387: 70–73.

Luikart, G., England, P. R., Tallmon, D., Jordan, S. and Taberlet, P. 2003. The power and promise of population genomics: from genotyping to genome typing. – Nat Rev Genet 4: 981–994.

Lymbery, A. J., Morine, M., Kanani, H. G., Beatty, S. J. and Morgan, D. L. 2014. Co-invaders: The effects of alien parasites on native hosts. – International Journal for Parasitology: Parasites and Wildlife 3: 171–177.

McCall, R. a, Nee, S. and Harvey, P. H. 1998. The role of wing length in the evolution of avian flightlessness. – Evolutionary Ecology 12: 569–580.

McLaren, M. R. and Callahan, B. J. 2020. Pathogen resistance may be the principal evolutionary advantage provided by the microbiome. – Philosophical Transactions of the Royal Society B: Biological Sciences 375: 20190592.

McNab, B. K. 1988. Food habits and the basal rate of metabolism in birds. – Oecologia 77: 343–349.

McNab, B. K. 1994a. Energy Conservation and the Evolution of Flightlessness in Birds. – The American Naturalist 144: 628–642.

McNab, B. K. 1994b. Resource Use and the Survival of Land and Freshwater Vertebrates on Oceanic Islands. – The American Naturalist 144: 643–660.

McNab, B. 2002. Minimizing energy expenditure facilitates vertebrate persistence on oceanic islands. – Ecology Letters 5: 693–704.

Methner, U., Friese, A. and Rösler, U. 2019. Competitive exclusion: A tool to combat extended-spectrum β-lactamase-producing Escherichia coli strains in chickens. – Research in Veterinary Science 123: 124–128.

Møller, A. P. and Erritzøe, J. 1998. Host immune defence and migration in birds. – Evolutionary Ecology 12: 945–953.

Moyle, R. G., Filardi, C. E., Smith, C. E. and Diamond, J. 2009. Explosive Pleistocene diversification and hemispheric expansion of a “great speciator.” – PNAS 106: 1863–1868.

Mueller, R. C., Mallig, N., Smith, J., Eöry, L., Kuo, R. I. and Kraus, R. H. S. 2020. Avian Immunome DB: an example of a user-friendly interface for extracting genetic information. – BMC Bioinformatics 21: 502.

Nespolo, R. F., González-Lagos, C., Solano-Iguaran, J. J., Elfwing, M., Garitano-Zavala, A., Mañosa, S., Alonso, J. C. and Altimiras, J. 2018. Aerobic power and flight capacity in birds: a phylogenetic test of the heart-size hypothesis. – Journal of Experimental Biology 221: jeb.162693.

Norris, K. and Evans, M. R. 2000. Ecological immunology: life history trade-offs and immune defense in birds. – Behavioral Ecology 11: 19–26.

Olson, S. L. 1973. Evolution of the rails of the South Atlantic islands (Aves: Rallidae). – Smithsonian Contributions to Zoology.

Pan, S., Lin, Y., Liu, Q., Duan, J., Lin, Z., Wang, Y., Wang, X., Lam, S. M., Zou, Z., Shui, G., Zhang, Y., Zhang, Z. and Zhan, X. 2019. Convergent genomic signatures of flight loss in birds suggest a switch of main fuel. – Nature Communications 10: 1–11.

Penman, B. S., Gupta, S. and Shanks, G. D. 2017. Rapid mortality transition of Pacific Islands in the 19th century. – Epidemiology & Infection 145: 1–11.

Peterson, C. L. and Laniel, M.-A. 2004. Histones and histone modifications. – Curr Biol 14: R546-51.

Prabhakar, S., Visel, A., Akiyama, J. A., Shoukry, M., Lewis, K. D., Holt, A., Plajzer-Frick, I., Morrison, H., Fitzpatrick, D. R., Afzal, V., Pennacchio, L. A., Rubin, E. M. and Noonan, J. P. 2008. Human-specific gain of function in a developmental enhancer. – Science 321: 1346–1350.

Price, E. R., Brun, A., Caviedes-Vidal, E. and Karasov, W. H. 2015. Digestive Adaptations of Aerial Lifestyles. – Physiology 30: 69–78.

Rajsbaum, R. and García-Sastre, A. 2014. Unanchored ubiquitin in virus uncoating. – Science 346: 427–428.

Rajsbaum, R., Versteeg, G. A., Schmid, S., Maestre, A. M., Belicha-Villanueva, A., Martínez-Romero, C., Patel, J. R., Morrison, J., Pisanelli, G., Miorin, L., Laurent-Rolle, M., Moulton, H. M., Stein, D. A., Fernandez-Sesma, A., tenOever, B. R. and García-Sastre, A. 2014. Unanchored K48-linked poly-ubiquitin synthesized by the E3-ubiquitin ligase TRIM6 stimulates the interferon-IKKε kinase mediated antiviral response. – Immunity 40: 880–895.

Ramírez, M. J., Magalhaes, I. L. F., Derkarabetian, S., Ledford, J., Griswold, C. E., Wood, H. M. and Hedin, M. 2021. Sequence Capture Phylogenomics of True Spiders Reveals Convergent Evolution of Respiratory Systems. – Systematic Biology 70: 14–20.

Roda, F., Liu, H., Wilkinson, M. J., Walter, G. M., James, M. E., Bernal, D. M., Melo, M. C., Lowe, A., Rieseberg, L. H., Prentis, P. and Ortiz-Barrientos, D. 2013. Convergence and

Divergence During the Adaptation to Similar Environments by an Australian Groundsel. – Evolution 67: 2515–2529.

Rodriguez-Trelles, F., Tarrio, R. and Ayala, F. J. 2003. Evolution of cis-regulatory regions versus codifying regions. – Int. J. Dev. Biol. 47: 665–673.

Roots, C. 2006. Flightless Birds. – Greenwood Publishing Group.

Rubin, B. E. R., Jones, B. M., Hunt, B. G. and Kocher, S. D. 2019. Rate variation in the evolution of non-coding DNA associated with social evolution in bees. – Philosophical Transactions of the Royal Society B: Biological Sciences 374: 20180247.

Sackton, T. B., Grayson, P., Cloutier, A., Hu, Z., Liu, J. S., Wheeler, N. E., Gardner, P. P., Clarke, J. A., Baker, A. J., Clamp, M. and Edwards, S. V. 2019. Convergent regulatory evolution and the origin of flightlessness in palaeognathous birds. – Science 364: 74–78.

Sagai, T., Hosoya, M., Mizushina, Y., Tamura, M. and Shiroishi, T. 2005. Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. – Development 132: 797–803.

Santure, A. W., Stapley, J., Ball, A. D., Birkhead, T. R., Burke, T. and Slate, J. 2010. On the use of large marker panels to estimate inbreeding and relatedness: empirical and simulation studies of a pedigreed zebra finch population typed at 771 SNPs. – Molecular Ecology 19: 1439–1451.

Scott, G. R. 2011. Elevated performance: the unique physiology of birds that fly at high altitudes. – Journal of Experimental Biology 214: 2455–2462.

Sherry, D. F., Vaccarino, A. L., Buckenham, K. and Herz, R. S. 1989. The hippocampal complex of food-storing birds. – Brain Behav Evol 34: 308–317.

Signor, S. A., Liu, Y., Rebeiz, M. and Kopp, A. 2016. Genetic Convergence in the Evolution of Male-Specific Color Patterns in Drosophila. – Current Biology 26: 2423–2433.

Stemmler, M. P., Hecht, A. and Kemler, R. 2005. E-cadherin intron 2 contains cis-regulatory elements essential for gene expression. – Development 132: 965–976.

Stern, D. L. 2000. Perspective: Evolutionary Developmental Biology and the Problem of Variation. – Evolution 54: 1079–1091.

Stervander, M., Ryan, P. G., Melo, M. and Hansson, B. 2019. The origin of the world's smallest flightless bird, the Inaccessible Island Rail Atlantisia rogersi (Aves: Rallidae). – Molecular Phylogenetics and Evolution 130: 92–98.

Sun, Y.-B., Fu, T.-T., Jin, J.-Q., Murphy, R. W., Hillis, D. M., Zhang, Y.-P. and Che, J. 2018. Species groups distributed across elevational gradients reveal convergent and continuous genetic adaptation to high elevations. – PNAS 115: E10634–E10641.

Taylor, B. 1998. Rails: A Guide to the Rails, Crakes, Gallinules and Coots of the World. – Bloomsbury Publishing.

Templeton, A. R. 1980. The theory of speciation via the founder principle. – Genetics 94: 1011–1038.

Trewick, S. A. 1997. Sympatric flightless rails Gallirallus dieffenbachii and G. modestus on the Chatham Islands, New Zealand; morphometrics and alternative evolutionary scenarios. – Journal of the Royal Society of New Zealand 27: 451–464.

Trewick, S. A. and Worthy, T. H. 2001. Origins and prehistoric ecology of takahe based on morphometric, molecular and fossil data. – The takahe – fifty years of conservation management and research, Jamieson, I.G. (eds). University of Otago Press, pp. 31–48.

Trewick, S. A. and Cowie, R. H. 2008. Introduction. Evolution on Pacific islands: Darwin's legacy. – Philosophical Transactions of the Royal Society B: Biological Sciences 363: 3289–3291.

Trewick, S. A., Pilkington, S., Shepherd, L. D., Gibb, G. C. and Morgan-Richards, M. 2017. Closing the gap: Avian lineage splits at a young, narrow seaway imply a protracted history of mixed population response. – Molecular Ecology 26: 5752–5772.

Vavouri, T., Walter, K., Gilks, W. R., Lehner, B. and Elgar, G. 2007. Parallel evolution of conserved non-coding elements that target a common set of developmental regulatory genes from worms to humans. – Genome Biol 8: R15.

Vizueta, J., Macías-Hernández, N., Arnedo, M. A., Rozas, J. and Sánchez-Gracia, A. 2019. Chance and predictability in evolution: The genomic basis of convergent dietary specializations in an adaptive radiation. – Molecular Ecology 28: 4028–4045.

Wallace, G. E. 2002. The Takahe: Fifty Years of Conservation Management and Research. – The Auk 119: 291–293.

Waters, A. M. and Beales, P. L. 2011. Ciliopathies: an expanding disease spectrum. – Pediatr Nephrol 26: 1039–1056.

Weber, T. P., Ens, B. J. and Houston, A. I. 1998. Optimal avian migration: A dynamic model of fuel stores and site use. – Evolutionary Ecology 12: 377.

White, D. J., Ramón-Laca, A., Amey, J. and Robertson, H. A. 2018. Novel genetic variation in an isolated population of the nationally critical Haast tokoeka (Apteryx australis 'Haast') reveals extreme short-range structure within this cryptic and flightless bird. – Conserv Genet 19: 1401–1410.

Wright, N. A., Steadman, D. W. and Witt, C. C. 2016. Predictable evolution toward flightlessness in volant island birds. – PNAS 113: 4765–4770.

Wright, B. R., Farquharson, K. A., McLennan, E. A., Belov, K., Hogg, C. J. and Grueber, C. E. 2020. A demonstration of conservation genomics for threatened species management. – Molecular Ecology Resources 20: 1526–1541.

Xin, J., Chai, Z., Zhang, C., Zhang, Q., Zhu, Y., Cao, H., Yangji, C., Chen, X., Jiang, H., Zhong, J. and Ji, Q. 2020. Methylome and transcriptome profiles in three yak tissues

revealed that DNA methylation and the transcription factor ZGPAT co-regulate milk production. – BMC Genomics 21: 731.

Yang, J., Liao, W. and Ma, Y. 2012. Role of SALL4 in hematopoiesis. – Curr Opin Hematol 19: 287–291.

Yang, Y., Wang, L., Han, J., Tang, X., Ma, M., Wang, K., Zhang, X., Ren, Q., Chen, Q. and Qiu, Q. 2015. Comparative transcriptomic analysis revealed adaptation mechanism of Phrynocephalus erythrurus, the highest altitude Lizard living in the Qinghai-Tibet Plateau. – BMC Evolutionary Biology 15: 101.

Yitbarek, A., Alkie, T., Taha-Abdelaziz, K., Astill, J., Rodriguez-Lecompte, J. C., Parkinson, J., Nagy, É. and Sharif, S. 2018. Gut microbiota modulates type I interferon and antibody-mediated immune responses in chickens infected with influenza virus subtype H9N2. – Beneficial Microbes 9: 417–427.

Zhang, Z., Xu, D., Wang, L., Hao, J., Wang, J., Zhou, X., Wang, W., Qiu, Q., Huang, X., Zhou, J., Long, R., Zhao, F. and Shi, P. 2016. Convergent Evolution of Rumen Microbiomes in High-Altitude Mammals. – Curr Biol 26: 1873–1879.

Ziera, T., Irlbacher, H., Fromm, A., Latouche, C., Krug, S. M., Fromm, M., Jaisser, F. and Borden, S. A. 2009. Cnksr3 is a direct mineralocorticoid receptor target gene and plays a key role in the regulation of the epithelial sodium channel. – The FASEB Journal 23: 3936–3946.

Appendices

Chapter Two

Appendix C2 Table 1: Full morphological dataset, 90 rail species, missing values for: Flying=0; Body lenth=14, Wing length = 0; Body mass = 32; Cranial length = 41; Cranial depth = 41; Cranial width = 36; Sternum length = 34; Sternum depth = 34; Pelvis width = 38; Femur length = 38; Femur width = 38. All units in millimetres except Body mass (grams), values in black were retrieved from Livezey (2003) and values in blue were retrieved from Handbook of the Birds of the World Alive Online (del Hoyo et al. 2015)

Scientific name	In Livezey	Flying	Body length	Wing length	Body mass	Cranial length	Cranial depth	Cranial width	Sternum length	Sternum depth	Pelvis width	Femur length	Femur width
Zapornia akool	*Amaurornis akool*	1	270	126.9	133.5	Na	Na	15.2	Na	Na	Na	Na	Na
Megacrex inepta	*Amaurornis ineptus*	0	370	179.7	967	Na	Na	27.8	Na	Na	Na	Na	Na
Amaurornis isabellina	*Amaurornis isabellinus*	1	375	165.3	Na	Na	Na	20.3	Na	Na	Na	Na	Na
Amaurornis moluccana	*Amaurornis moluccanus*	1	265	136.1	191.5	30.8	17.9	19.6	37.6	11.4	13.8	47.8	7.8
Amaurornis olivacea	*Amaurornis olivaceus*	1	310	162.9	276.5	33	20	21	38.5	8.9	16.4	55.1	9.4
Amaurornis phoenicurus	*Amaurornis phoenicurus*	1	305	162.1	210	31	18.5	19.6	41.1	11.3	14.1	47.9	8.3
Amaurornis moluccana	*Amaurornis ruficrissus*	1	na	140.3	163.5	Na	Na	Na	Na	Na	Na	Na	Na
Aramides albiventris	*Aramides albiventris*	1	na	189.3	466	Na	Na	Na	Na	Na	Na	Na	Na
Aramides cajaneus	*Aramides cajanea*	1	365	180.5	417.5	39.4	23.5	23.3	58	13.1	17.6	62.3	11.4
Aramides ypecaha	*Aramides ypecaha*	1	430	223.7	692.5	42.7	26.6	26	76.4	17.7	22.8	77.4	17.1
Aramidopsis plateni	*Aramidopsis plateni*	0	300	147.4	Na	Na	Na	Na	Na	Na	Na	Na	Na
Atlantisia rogersi	*Atlantisia rogersi*	0	142	54.8	38.75	Na	Na	Na	7.7	2.4	9.6	24.3	4.1
Gallirallus modestus	*Cabalus modestus*	0	195	83.2	Na	25.1	15.2	14.3	11.3	2.7	9.6	30.1	5.1
Mentocrex kioliodes	*Canirallus kioloides*	1	280	132.8	250.5	31.5	19.1	19.7	43.5	10.7	17.9	44.3	8.2
Canirallus oculeus	*Canirallus oculeus*	1	300	174.1	278	33.5	21.4	21.1	48.3	12.3	18.7	53	9.7
Coturnicops noveboracensis	*Coturnicops noveboracensis*	1	175	85.6	55.5	20.5	12.5	13	23.8	9.4	9.5	28.7	4.4
Mustelirallus albicollis	*Crex albicollis*	1	225	108.6	104	28.2	16.7	17.1	33.3	10.2	11.5	42	6.7
Crex crex	*Crex crex*	1	285	137.5	152	28.1	16.4	16.4	43.3	14.1	12.2	45.6	7
Cyanolimnas cerverai	*Cyanolimnas cerverai*	0	290	106	Na	Na	Na	Na	25.4	4.2	Na	Na	Na
Dryolimnas cuvieri abbotti	*Dryolimnas abbotti*	0	na	135.4	Na	32.6	18.8	18.3	41.4	11.1	Na	Na	Na
Dryolimnas cuvieri aldabranus	*Dryolimnas aldabranus*	0	na	116.7	182.5	30.6	18.1	17.6	30.6	7.9	13.9	43.6	7.8
Dryolimnas cuvieri	*Dryolimnas cuvieri*	0	315	150.9	241	33.2	19.9	19.2	44.8	12.4	15.6	51.3	8.9
Eulabeornis castaneoventris	*Eulabeornis castaneoventris*	1	500	210.5	687	Na	Na	Na	50.5	12.3	23.4	72.1	13.9
Fulica alai	*Fulica alai*	1	390	179.6	495	32.3	20.4	20.2	52.7	15.3	15.3	52.7	9.9
Fulica americana	*Fulica americana*	1	385	187.2	610.5	32.8	20.4	19.8	53.8	16	14.8	52.4	10
Fulica armillata	*Fulica armillata*	1	470	195.5	912.5	37.3	23.2	22.4	64.6	16.7	17.8	62.7	13.1
Fulica atra	*Fulica atra*	1	375	203.1	831.5	35.1	21.7	21	59.1	16.7	15.5	54.5	10.9
Fulica cornuta	*Fulica cornuta*	1	495	284.3	1978.5	42.6	24.4	25	84.8	22.5	26	78.9	16.2
Fulica gigantea	*Fulica gigantea*	1	535	266.1	2355	42.4	25.6	26.2	82.4	20.1	25.3	83.2	18.1
Gallicrex cinerea	*Gallicrex cinerea*	1	415	190.2	426	32.3	19.6	19.6	54	14.6	14.7	55.7	9.4
Gallinula angulata	*Gallinula angulata*	1	225	136.4	132	Na	Na	Na	Na	Na	Na	Na	Na
Gallinula galeata	*Gallinula cachinnans*	1	355	171.2	366.5	31.6	19.5	18.6	48.6	14.5	16.4	52.7	9.4
Gallinula chloropus	*Gallinula chloropus*	1	340	167.1	302.5	31.3	19.1	18.1	45.6	14.2	15.7	49.2	9
Gallinula nesiotis comeri	*Gallinula comeri*	0	250	144.1	513	33.7	22.2	20.5	38.6	8.5	20.7	56.9	11
Gallinula galeata	*Gallinula galeata*	1	na	213.7	448.5	Na	Na	Na	Na	Na	Na	Na	Na
Gallinula nesiotis	*Gallinula nesiotis*	0	250	143.5	513	Na	Na	Na	34.6	8.3	Na	Na	Na
Gallinula chloropus	*Gallinula pyrrhorrhoa*	1	na	159.3	Na	Na	Na	Na	Na	Na	Na	Na	Na
Gallinula galeata	*Gallinula sandvicensis*	1	na	175.2	Na	31	19.2	18.5	42.1	12.8	16.6	51.1	9.5
Gallinula tenebrosa	*Gallinula tenebrosa*	1	375	187.8	531.5	34.4	21.6	20	59.7	16.4	19	61.7	11.8
Gallirallus australis	*Gallirallus australis*	0	550	176.6	890	43.1	24.8	24.8	32	6.7	21.4	73.8	14.3
Gallirallus dieffenbachii	*Gallirallus dieffenbachii*	0	320	121	Na	Na	Na	Na	29.1	8.1	17.1	54.6	10.2
Gallirallus owstoni	*Gallirallus owstoni*	0	280	119.8	Na	33.1	18.8	18.4	31.1	8.3	15.5	52.3	8.8

Scientific name	In Livezey	Flying	Body length	Wing length	Body mass	Cranial length	Cranial depth	Cranial width	Sternum length	Sternum depth	Pelvis width	Femur length	Femur width
Lewinia pectoralis	*Gallirallus pectoralis*	1	225	97.3	83.5	25.7	15.6	15.1	29.5	9.7	10.8	36.7	5.7
Gallirallus philippensis	*Gallirallus philippensis*	1	290	133.9	182	30.9	18.1	17.8	41.4	12.6	13.8	47	7.9
Na	*Gallirallus sharpei*	1	na	70	Na	Na	Na	Na	Na	Na	Na	Na	Na
Gallirallus striatus	*Gallirallus striatus*	1	275	115.4	112.5	28.3	16.4	16.2	38.9	12.1	11.9	43.4	6.9
Gallirallus wakensis	*Gallirallus wakensis*	0	235	90.1	Na	25.6	15.6	16.9	21.3	5.4	13.2	36	6.6
Gymnocrex plumbeiventris	*Gymnocrex plumbeiventris*	1	315	189.3	292	Na	25.6	22.7	Na	Na	Na	Na	Na
Gymnocrex rosenbergii	*Gymnocrex rosenbergii*	1	300	194.4	Na	Na	Na	Na	Na	Na	Na	Na	Na
Gallirallus torquatus	*Habropteryx celebensis*	1	na	148.8	Na	Na	Na	Na	Na	Na	Na	Na	Na
Gallirallus insignis	*Habropteryx insignis*	0	330	144.1	Na	Na	Na	22.5	Na	Na	Na	Na	Na
Gallirallus okinawae	*Habropteryx okinawae*	0	320	142.5	433	38.5	22	20.4	34.1	8.5	18.7	61.2	11.7
Gallirallus torquatus	*Habropteryx sulcirostris*	1	na	149.1	Na	Na	Na	Na	Na	Na	Na	Na	Na
Gallirallus torquatus	*Habropteryx torquatus*	1	340	147.2	245.5	32.7	19.2	19.6	39	11.1	15.9	52	9.3
Habroptila wallacii	*Habroptila wallacii*	0	350	167.5	Na	45.8	25.9	27.9	47.8	6.2	23.2	79.5	14.3
Laterallus leucopyrrhus	*Laterallus leucopyrrhus*	1	150	83.3	44.5	22.4	14	14.9	21.3	6.6	9.3	30.5	4.8
Laterallus spilonota	*Laterallus spilonotus*	1	155	67.3	40	Na	Na	Na	Na	Na	Na	Na	Na
Nesoclopeus poecilopterus	*Nesoclopeus poecilopterus*	0	330	161	Na	Na	Na	Na	44	8.5	Na	Na	Na
Pardirallus sanguinolentus	*Ortygonax sanguinolentus*	1	340	133.6	137	Na	Na	Na	40.1	12.3	14.3	48	7.5
Pardirallus nigricans	*Ortygonax nigricans*	1	280	130.3	179	Na	Na	Na	Na	Na	Na	Na	Na
Gallinula pacifica	*Pareudiastes pacificus*	0	230	123.3	Na	Na	Na	Na	24	7	Na	Na	Na
Gallinula silvestris	*Pareudiastes silvestris*	0	265	149	450	Na	Na	Na	Na	Na	Na	Na	Na
Porphyrio albus	*Porphyrio albus*	1	na	229	Na	Na	Na	Na	Na	Na	Na	Na	Na
Porphyrio hochstetteri	*Porphyrio hochstetteri*	0	630	235.5	2718	50.5	31.5	32.6	50.2	7.4	37.1	100.8	21.2
Porphyrio indicus	*Porphyrio indicus*	1	na	225.1	Na	Na	Na	Na	Na	Na	Na	Na	Na
Porphyrio madagascariensis	*Porphyrio madagascariensis*	1	na	245.7	598.5	38.8	25.2	24.1	54.5	15.5	22.1	75.2	12.5
Porphyrio porphyrio	*Porphyrio porphyrio*	1	440	262.4	984.5	40.2	26.8	24.5	61.2	16.7	24.7	76.2	13.3
Porphyrio indicus viridis	*Porphyrio viridis*	1	na	242.9	Na	Na	Na	Na	Na	Na	Na	Na	Na
Porphyrio alleni	*Porphyrula alleni*	1	240	151.9	139.5	28.4	18	18.2	33.3	11.1	14.1	46.7	7.4
Porphyrio flavirostris	*Porphyrula flavirostris*	1	245	134	93	25.7	15.5	15.8	28.7	Na	13.4	38	6.1
Porphyrio martinica	*Porphyrula martinica*	1	315	179.3	226.5	31.4	18.9	19.6	38.9	12.5	15.4	73.5	8.4
Zapornia atra	*Porzana atra*	0	180	83.1	77	23.7	14.6	15.4	15.4	5.75	11.4	34	5.5
Zapornia bicolor	*Porzana bicolor*	1	210	113.5	Na	Na	Na	Na	Na	Na	Na	Na	Na
Porzana carolina	*Porzana carolina*	1	220	106.2	81	23	13.8	14.1	26.6	10.2	10.6	33.5	5.2
Zapornia flavirostra	*Porzana flavirostra*	1	210	102.3	83.5	24.9	15.5	16.2	27.1	8.1	10.7	36.8	5.9
Zapornia monasa	*Porzana monasa*	0	180	78	Na	Na	Na	Na	Na	Na	Na	Na	Na
Zapornia olivieri	*Porzana olivieri*	1	190	103.8	Na	Na	Na	Na	Na	Na	Na	Na	Na
Zapornia palmeri	*Porzana palmeri*	0	150	58.5	Na	20.25	12.8	13	11	3.9	9.8	24.9	4.2
Zapornia pusilla	*Porzana pusilla*	1	180	88.3	37.5	20.65	12.6	12.6	22.2	8.6	8.8	28	4.1
Zapornia sandwichensis	*Porzana sandwichensis*	0	145	67.6	Na	Na	Na	Na	Na	Na	Na	Na	Na
Zapornia tabuensis	*Pozana tabuensis*	1	165	79	42	22.4	13.7	14.1	18.9	6.7	9.3	29.4	4.5
Rallus aquaticus	*Rallus aquaticus*	1	255	119.1	164.5	27.6	16.5	15.8	35	11.4	11.3	41	6.6
Rallus elegans	*Rallus elegans-group*	1	430	163.3	352.5	34.7	19.3	18.6	52.9	16	14.5	58	9.7
Rallus limicola	*Rallus limicola*	1	225	102	83	25	15.1	14.5	30.2	10.7	10.3	36.4	5.7
Rallus longirostris	*Rallus longirostris*	1	330	140.2	295	33.15	18.2	17.8	46.9	15	13.6	54	8.8
Rougetius rougetii	*Rougetius rougetii*	1	300	131.9	195	32.5	Na	18.3	Na	Na	Na	Na	Na
Tribonyx mortierii	*Tribonyx mortierii*	0	465	178.9	1292.5	42.75	24.8	23.9	52.8	10	25.6	82.5	16.3
Tribonyx ventralis	*Tribonyx ventralis*	1	340	208.2	401	34.3	20.5	20.5	52.5	15.6	19.6	55.1	11
Gallirallus lafresnayanus	*Tricholimnas lafresnayanus*	0	465	181.4	Na	Na	Na	Na	Na	7	23.2	68.2	14.1
Gallirallus sylvestris	*Tricholimnas sylvestris*	0	360	138	Na	36.6	21.7	21.3	32.3	8.3	18.5	59.4	11.4

Appendix C1 Table 2 : Subset of the dataset present in Appendix 1 including the 52 rails species with phylogenetic information (Garcia-R et al. 2014) and the 11 morphological traits with the fewest missing values. All units in millimetres except Body mass (grams)

Species	Taxa	Flying	Body length	Wing lenth	Body mass	Cranial length	Cranial depth	Cranial width	Sternum length	Sternum depth	Pelvis width	Femur length	Femur width
Aramides cajaneus	*Aramide*	1	365	180.55	417.5	39.35	23.5	23.25	58	13.05	17.6	62.25	11.45
Aramides ypecaha	*Aramide*	1	430	223.75	692.5	42.65	26.6	25.95	76.35	17.65	22.75	77.35	17.1
Mustelirallus albicollis	*Aramide*	1	225	108.65	104	28.15	16.65	17.1	33.25	10.2	11.5	42	6.7
Pardirallus sanguinolentus	*Aramide*	1	340	133.6	137	na	na	na	40.1	12.3	14.3	48	7.5
Pardirallus nigricans	*Aramide*	1	280	130.3	179	na	na	na	na	na	na	na	na
Fulica americana	*Fulica*	1	385	187.2	610.5	32.8	20.35	19.75	53.8	16	14.75	52.35	10
Fulica armillata	*Fulica*	1	470	195.5	912.5	37.3	23.2	22.4	64.6	16.7	17.8	62.7	13.1
Fulica atra	*Fulica*	1	375	203.1	831.5	35.05	21.7	21	59.1	16.65	15.5	54.45	10.9
Fulica cornuta	*Fulica*	1	495	284.35	1978	42.6	24.4	25	84.8	22.5	26	78.9	16.2
Fulica gigantea	*Fulica*	1	535	266.1	2355	42.35	25.55	26.2	82.35	20.05	25.3	83.15	18.05
Gallinula angulata	*Fulica*	1	225	136.4	132	na	na	na	na	na	na	na	na
Gallinula galeata	*Fulica*	1	355	171.2	366.5	31.6	19.45	18.6	48.6	14.5	16.4	52.7	9.4
Gallinula chloropus	*Fulica*	1	340	167.15	302.5	31.25	19.05	18.1	45.55	14.15	15.65	49.2	8.95
Gallinula tenebrosa	*Fulica*	1	375	187.85	531.5	34.35	21.55	19.95	59.7	16.35	18.95	61.7	11.75
Porzana carolina	*Fulica*	1	220	106.2	81	23	13.8	14.1	26.6	10.15	10.6	33.5	5.25
Tribonyx mortierii	*Fulica*	0	465	178.85	1292	42.75	24.75	23.9	52.8	10	25.6	82.45	16.35
Tribonyx ventralis	*Fulica*	1	340	208.15	401	34.3	20.45	20.45	52.45	15.6	19.55	55.05	10.95
Amaurornis moluccana	*Gallicrex*	1	265	136.1	191.5	30.75	17.85	19.6	37.6	11.35	13.75	47.8	7.85
Amaurornis phoenicurus	*Gallicrex*	1	305	162.15	210	31	18.45	19.6	41.1	11.3	14.05	47.85	8.35
Gallicrex cinerea	*Gallicrex*	1	415	190.2	426	32.3	19.55	19.6	54	14.55	14.7	55.65	9.4
Megacrex inepta	*Gallicrex*	0	370	179.7	967	na	na	27.8	na	na	na	na	na
Coturnicops noveboracensis	*Laterallus*	1	175	85.65	55.5	20.45	12.5	12.95	23.8	9.35	9.5	28.7	4.4
Porphyrio hochstetteri	*Porphyrio*	0	630	235.45	2718	50.5	31.45	32.6	50.2	7.35	37.1	100.7	21.25
Porphyrio porphyrio	*Porphyrio*	1	440	262.35	984.5	40.2	26.75	24.45	61.2	16.65	24.65	76.2	13.35
Porphyrio alleni	*Porphyrio*	1	240	151.85	139.5	28.4	17.95	18.15	33.25	11.05	14.1	46.65	7.4
Porphyrio martinica	*Porphyrio*	1	315	179.25	226.5	31.35	18.85	19.55	38.9	12.5	15.4	73.45	8.4
Zapornia akool	*Porzana*	1	270	126.95	133.5	na	na	15.2	na	na	na	na	na
Zapornia flavirostra	*Porzana*	1	210	102.25	83.5	24.9	15.5	16.15	27.1	8.1	10.65	36.8	5.9
Zapornia pusilla	*Porzana*	1	180	88.3	37.5	20.65	12.6	12.6	22.15	8.6	8.75	27.95	4.15
Zapornia tabuensis	*Porzana*	1	165	79	42	22.4	13.65	14.05	18.9	6.7	9.3	29.4	4.55
Aramidopsis plateni	*Rallus*	0	300	151.2	150	na	na	na	na	na	na	na	na
Crex crex	*Rallus*	1	285	137.5	152	28.05	16.4	16.4	43.3	14.1	12.15	45.6	7
Dryolimnas cuvieri	*Rallus*	1	315	150.95	241	33.2	19.9	19.2	44.75	12.4	15.6	51.3	8.9
Eulabeornis castaneoventris	*Rallus*	1	500	210.5	687	na	na	na	50.5	12.3	23.4	72.1	13.9

Species	Taxa	Flying	Body length	Wing lenth	Body mass	Cranial length	Cranial depth	Cranial width	Sternum length	Sternum depth	Pelvis width	Femur length	Femur width
Gallirallus modestus	*Rallus*	0	195	83.2	60	25.05	15.2	14.3	11.25	2.65	9.6	30.05	5.05
Gallirallus australis	*Rallus*	0	550	176.65	890	43.1	24.75	24.8	31.95	6.7	21.35	73.75	14.25
Gallirallus dieffenbachii	*Rallus*	0	320	121	340	na	na	na	29.05	8.1	17.1	54.6	10.2
Gallirallus owstoni	*Rallus*	0	280	119.85	226	33.1	18.8	18.35	31.05	8.25	15.5	52.25	8.85
Gallirallus philippensis	*Rallus*	1	290	133.9	182	30.85	18.1	17.75	41.4	12.55	13.8	46.95	7.9
Gallirallus wakensis	*Rallus*	0	235	90.05	105	25.55	15.55	16.9	21.3	5.35	13.2	36	6.65
Gallirallus insignis	*Rallus*	0	330	144.1	na	na	na	22.5	na	na	na	na	na
Gallirallus okinawae	*Rallus*	0	320	142.5	433	38.5	22	20.4	34.1	8.5	18.7	61.2	11.7
Gallirallus torquatus	*Rallus*	1	340	147.2	245.5	32.65	19.2	19.55	39	11.05	15.85	51.95	9.3
Gallirallus lafresnayanus	*Rallus*	0	465	181.35	na	na	na	na	na	7	23.2	68.2	14.1
Gallirallus sylvestris	*Rallus*	0	360	137.95	470	36.6	21.7	21.25	32.25	8.3	18.45	59.35	11.4
Gallirallus striatus	*Rallus*	1	275	115.4	112.5	28.3	16.4	16.15	38.9	12.1	11.85	43.4	6.95
Habroptila wallacii	*Rallus*	0	350	167.45	na	45.75	25.9	27.9	47.8	6.2	23.15	79.5	14.3
Lewinia pectoralis	*Rallus*	1	225	97.3	83.5	25.7	15.6	15.1	29.5	9.7	10.8	36.7	5.7
Rallus elegans	*Rallus*	1	430	163.25	352.5	34.7	19.3	18.6	52.85	16	14.45	58	9.7
Rallus limicola	*Rallus*	1	225	101.95	83	25	15.1	14.5	30.15	10.7	10.3	36.35	5.7
Rallus longirostris	*Rallus*	1	330	140.15	295	33.15	18.2	17.8	46.9	14.95	13.6	53.95	8.85
Rallus aquaticus	*Rallus*	1	255	119.1	115.7	27.6	16.45	15.8	35	11.4	11.25	40.95	6.65

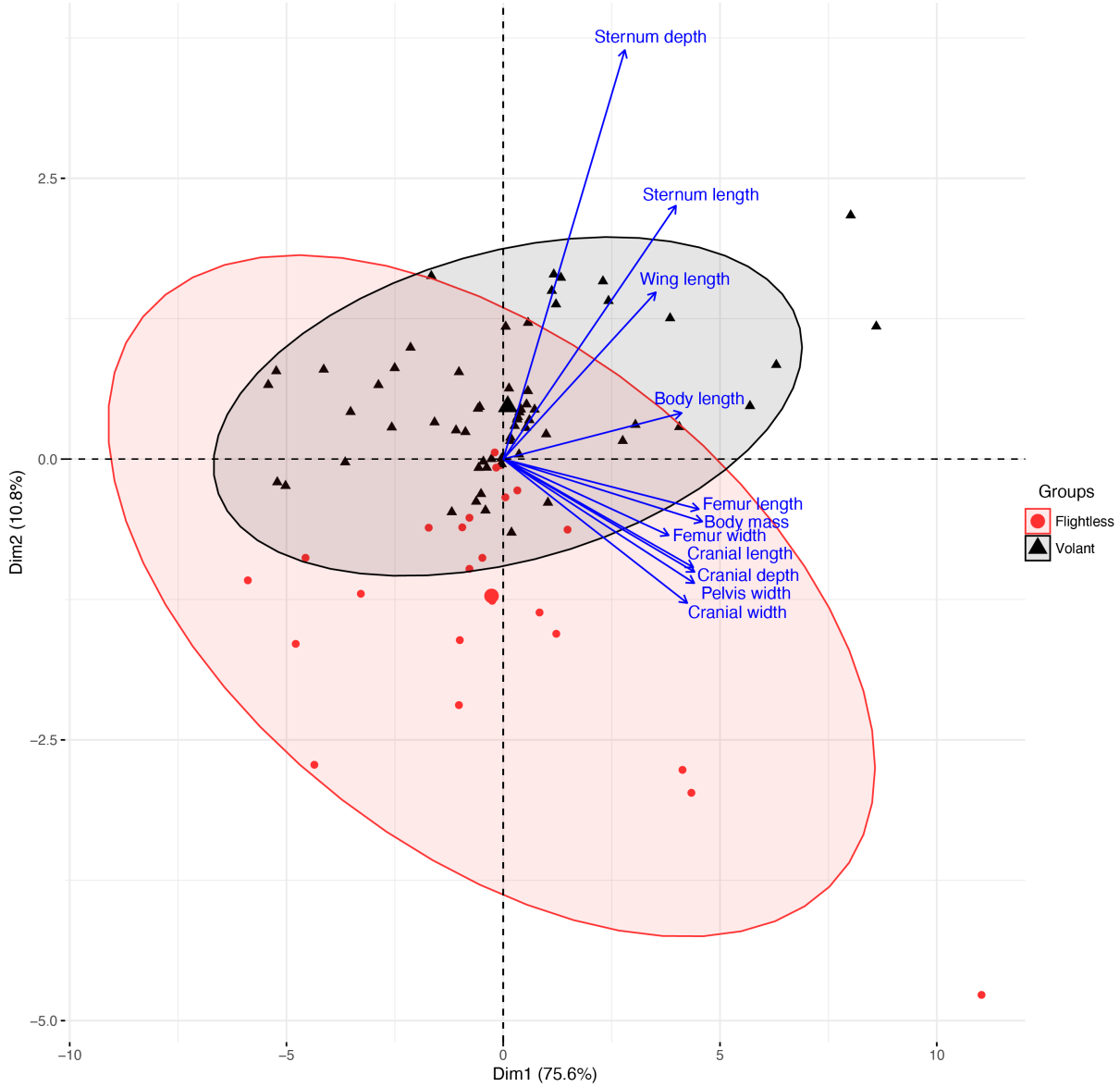
Appendix C2 Table 2: NCBI Access numbers used to investigate the Maximum likelihood phylogeny

Species name	16S	COI	FGB-7	RAG-1	cyt-b
Amaurolimnas concolor		JQ173980.1			
Amaurornis akool		FJ661094.1			JQ342141.1
Amaurornis flavirostra	KC613979.1	KC614036.1	KC613861.1	KC613913.1	KC614062.1
Amaurornis moluccana	KC613981.1	KC614038.1		KC613915.1	KC614064.1
Amaurornis phoenicurus	KC613982.1	JQ342118.1	KC613863.1	KC613916.1	KC614065.1
Anurolimnas fasciatus	KC614006.1	KC614046.1	KC613884.1	KC613942.1	KC614090.1
Anurolimnas viridis	KC614010.1	JQ174052.1	KC613888.1	KC613947.1	KC614094.1
Aramides axillaris	KC613978.1	JN801494.1	KC613860.1	KC613912.1	KC614061.1
Aramides cajanea	KC613983.1	JN801496.1	KC613864.1	KC613917.1	KC614066.1
Aramides mangle	KC613980.1	KC614037.1	KC613862.1	KC613914.1	KC614063.1
Aramides ypecaha	KC613984.1	FJ027148.1	DQ881942.1	AY756084.1	KC614067.1
Aramidopsis plateni					JQ347988.1
Aramus guarauna	DQ485854.1	FJ027151.1	AY695250.1	DQ881798.1	DQ485899.1
Canirallus beankaensis					HQ403671.1
Canirallus kioloides					HQ403670.1
Coturnicops exquisitus		NC_012143.1			
Coturnicops	KC613985.1	DQ433553.1	AY695239.1	KC613918.1	KC614068.1
Crex crex	KC613986.1	GU571355.1	KC613865.1	KC613919.1	KC614069.1
Diaphorapteryx hawkinsi					KC614124.1
Dryolimnas cuvieri	KC613987.1	KC614039.1	KC613866.1	KC613920.1	KC614070.1
Eulabeornis	KC613988.1	KC614058.1	KC613867.1	KC613921.1	KC614071.1
Fulica alai	KC613989.1	JF498857.1	KC613868.1	KC613922.1	KC614072.1
Fulica americana		DQ434598.1	AY695244.1	KC613923.1	DQ485910.1
Fulica ardesiaca	KC613990.1	FJ027587.1	KC613869.1	KC613924.1	KC614073.1
Fulica armillata	KC613995.1	FJ027588.1	KC613874.1	KC613929.1	KC614078.1
Fulica atra	KC613991.1	GU571406.1	KC613870.1	KC613925.1	KC614074.1
Fulica cornuta		FJ027592.1			KC614075.1
Fulica cristata	KC613992.1	KC614040.1	KC613871.1	KC613926.1	
Fulica gigantea		FJ027593.1			
Fulica leucoptera	KC613993.1	KC614060.1	KC613872.1	KC613927.1	KC614076.1
Fulica rufifrons	KC613994.1	FJ027594.1	KC613873.1	KC613928.1	KC614077.1
Gallicrex cinerea	KC613997.1	JQ342129.1	KC613877.1	KC613932.1	KC614080.1
Gallinula angulata	KC613996.1	KC614041.1	KC613875.1	KC613930.1	KC614079.1
Gallinula chloropus		FJ027609.1	AY695245.1	KC613931.1	DQ485911.1
Gallinula galeata		JF498859.1			
Gallinula melanops	KC613998.1	FJ027612.1	KC613878.1	KC613933.1	KC614081.1
Gallinula mortierii	KC613999.1	KC614042.1		KC613934.1	KC614082.1
Gallinula tenebrosa	KC614002.1	JQ174909.1	KC613880.1	KC613938.1	KC614086.1
Gallinula ventralis	KC614003.1		KC613881.1	KC613939.1	KC614087.1
Gallirallus australis	KC614035.1		KC613911.1	KC613977.1	KC614123.1
Gallirallus calayanensis					KC614128.1
Gallirallus dieffenbachii					KC614127.1
Gallirallus insignis					JQ347978.1
Gallirallus lafresnayanus					KC614130.1
Gallirallus modestus					KC614125.1
Gallirallus okinawae		NC_012140.1			NC012140
Gallirallus owstoni	KC614000.1	KC614043.1		KC613935.1	
Gallirallus philippensis			AY695241.1	KC613936.1	DQ485907.1
Gallirallus rovianae					JQ348011.1
Gallirallus striatus	KC614001.1	JQ342122.1	KC613879.1	KC613937.1	KC614085.1

Species name	16S	COI	FGB-7	RAG-1	cyt-b
Gallirallus sylvestris	KC614034.1	KC614057.1	KC613910.1	KC613976.1	KC614122.1
Gallirallus torquatus					JQ347980.1
Gallirallus wakensis					JQ348014.1
Habroptila wallacii					JQ347984.1
Heliornis fulica	DQ485857.1	JQ175018.1	AY695246.1		DQ485902.1
Himantornis haematopus					KC614126.1
Laterallus albigularis		JQ175222.1	AY082411.1	DQ881813.1	
Laterallus exilis	KC614004.1	JQ175223.1	KC613883.1	KC613941.1	KC614089.1
Laterallus jamaicensis	KC614009.1	DQ432997.1	KC613885.1	KC613943.1	KC614091.1
Laterallus melanophaius	DQ485859.1		AY695238.1	KC613944.1	DQ485906.1
Lewinia mirifica	KC614005.1	KC614045.1	KC613882.1	KC613940.1	KC614088.1
Lewinia muelleri	KC614007.1	KC614047.1	KC613886.1	KC613945.1	KC614092.1
Lewinia pectoralis	KC614008.1	KC614048.1	KC613887.1	KC613946.1	KC614093.1
Megacrex inepta					JQ347987.1
Neocrex erythrops	KC614011.1	KC614050.1	KC613889.1	KC613948.1	KC614095.1
Nesoclopeus woodfordi	KC614012.1		KC613891.1	KC613949.1	KC614096.1
Pardirallus maculatus		JQ175674.1		KC613965.1	KC614114.1
Pardirallus nigricans	KC614020.1	KC614054.1	KC613898.1	KC613957.1	KC614104.1
Pardirallus sanguinolentus	KC614025.1	JQ175676.1	KC613904.1	KC613963.1	KC614113.1
Porphyrio alleni	KC614015.1	KC614052.1	KC613893.1	KC613952.1	KC614100.1
Porphyrio hochstetteri	NC_010092.1	NC_010092.1	KC613909.1	KC613974.1	NC010092
Porphyrio martinica	KC614019.1	AY666523.1	KC613897.1	KC613956.1	KC614103.1
Porphyrio porphyrio	DQ485858.1	JQ175970.1	AY695240.1	KC613975.1	DQ485905.1
Porzana albicollis	KC614018.1	JQ175972.2	KC613896.1	KC613955.1	KC614102.1
Porzana carolina	DQ485862.1	DQ433143.1	KC613899.1	KC613958.1	DQ485909.1
Porzana flaviventer		JQ175973.1			
Porzana fluminea	KC614016.1	KC614053.1	KC613894.1	KC613953.1	KC614107.1
Porzana fusca	KC614017.1	JQ342114.1	KC613895.1	KC613954.1	KC614101.1
Porzana parva	KC614022.1		KC613901.1	KC613960.1	KC614106.1
Porzana paykullii	KC614013.1	JQ342128.1	KC613892.1	KC613950.1	KC614097.1
Porzana porzana	KC614023.1	GQ482558.1	KC613902.1	KC613961.1	
Porzana pusilla	KC614021.1	JQ342132.1	KC613900.1	KC613959.1	KC614105.1
Porzana tabuensis	KC614026.1			KC613964.1	
Porzana spiloptera		JN801952.1			
Psophia crepitans	DQ485855.1	JQ176018.1	AY695248.1		DQ485900.1
Rallina eurizonoides	NC_012142.1	NC_012142.1			NC012142
Rallina fasciata	KC614030.1			KC613969.1	KC614118.1
Rallina tricolor	KC614032.1	KC614056.1	KC613907.1	KC613972.1	KC614120.1
Rallus aquaticus	KC614027.1	GU097233.1	EF552781.1	KC613966.1	KC614115.1
Rallus caerulescens	KC614028.1	KC614055.1	KC613905.1	KC613967.1	KC614116.1
Rallus elegans	KC614029.1	AY666315.1	KC613906.1	KC613968.1	KC614117.1
Rallus limicola	KC614031.1	GU097263.1	AY695242.1	KC613970.1	KC614119.1
Rallus longirostris	DQ485861.1	DQ433164.1	AY695243.1	KC613971.1	DQ485908.1
Sarothrura rufa	KC614033.1		KC613908.1	KC613973.1	KC614121.1
Grus americana	KP966312.1	DQ433674.1	AY695254		

Appendix C2 Figure 1: Maximum Likelihood phylogeny based on a 5-gene (COI, cyt-b, 16S, FGB, RAG-1) concatenated analysis, bootstrap supports are indicated in each branch. Colours are indicating the taxon according to Garcia (2014), Dark grey: *Fulica*; Blue: *Porphyrio*; Green: *Rallina*; Red: *Porzana*; Purple: *Laterallus*; Orange: *Gallicrex*; Brown: *Aramides*; Pink: *Rallus*

Appendix C2 Figure 2: Principal component analysis (PCA) plot showing the two first dimension of the multivariate variation among 90 species of rails in terms of morphological traits. Vectors indicate the direction and strength of each trait contribution to the overall distribution. Red dots represent flightless species and black triangles are volant species.



Chapter Four

Appendix C4 Table 1: Relaxed genes. Overrepresentation test for Gene Ontology (GO) biological process using PANTHER (Protein Analysis Through Evolutionary Relationships, http://pantherdb.org) (Mi et al. 2019) with *Gallus gallus* annotations. The "Relaxed" column is the number of relaxed genes associated with the GO term, the p-value as determined by Fisher's exact test and False Discovery Rate (FDR) as calculated by the Benjamini-Hochberg procedure.

GO	GO terms	*Gallus gallus*	Relaxed genes	expected	Fold enrichment	P	FDR
GO:0007043	cell-cell junction assembly	90	4	1.03	3.88	0.022	1.000
GO:0007338	single fertilization	65	4	0.74	5.37	0.008	1.000
GO:0007127	meiosis I	86	4	0.99	4.06	0.019	1.000
GO:0009566	fertilization	82	4	0.94	4.26	0.017	1.000
GO:0061982	meiosis I cell cycle process	87	4	1	4.01	0.020	1.000
GO:0030534	adult behavior	95	4	1.09	3.67	0.026	1.000
GO:0007611	learning or memory	138	5	1.58	3.16	0.024	1.000
GO:0007613	memory	58	4	0.66	6.02	0.005	1.000
GO:0007612	learning	83	4	0.95	4.2	0.017	1.000
GO:0032868	response to insulin	113	4	1.3	3.09	0.044	1.000
GO:0043087	regulation of GTPase activity	265	8	3.04	2.63	0.013	1.000
GO:0044282	small molecule catabolic process	238	7	2.73	2.57	0.022	1.000
GO:0016042	lipid catabolic process	205	6	2.35	2.55	0.033	1.000
GO:0061061	muscle structure development	322	9	3.69	2.44	0.013	1.000
GO:0098657	import into cell	132	5	1.51	3.31	0.020	1.000
GO:0007268	chemical synaptic transmission	251	7	2.88	2.43	0.028	1.000
GO:0099537	trans-synaptic signaling	266	7	3.05	2.3	0.036	1.000
GO:0098916	anterograde trans-synaptic signaling	251	7	2.88	2.43	0.028	1.000
GO:1901990	regulation of mitotic cell cycle phase…	205	6	2.35	2.55	0.033	1.000
GO:0042391	regulation of membrane potential	278	9	3.19	2.82	0.006	1.000
GO:0030510	regulation of BMP signaling pathway	84	4	0.96	4.15	0.018	1.000
GO:0023057	negative regulation of signaling	890	21	10.2	2.06	0.002	0.924
GO:0010648	negative regulation of cell…	887	21	10.17	2.07	0.002	1.000
GO:0044057	regulation of system process	296	9	3.39	2.65	0.008	1.000
GO:0008016	regulation of heart contraction	109	5	1.25	4	0.010	1.000
GO:1903522	regulation of blood circulation	132	6	1.51	3.97	0.005	1.000
GO:0002027	regulation of heart rate	56	4	0.64	6.23	0.005	1.000
GO:0048638	regulation of developmental growth	213	6	2.44	2.46	0.039	1.000
GO:0022604	regulation of cell morphogenesis	199	8	2.28	3.51	0.003	1.000
GO:0060284	regulation of cell development	336	9	3.85	2.34	0.017	1.000
GO:0008360	regulation of cell shape	98	4	1.12	3.56	0.029	1.000
GO:0033044	regulation of chromosome organization	205	6	2.35	2.55	0.033	1.000
GO:0031333	negative regulation of protein…	101	4	1.16	3.46	0.032	1.000
GO:1902904	negative regulation of supramolecular..	117	5	1.34	3.73	0.013	1.000
GO:0010639	negative regulation of organelle…	226	6	2.59	2.32	0.049	1.000
GO:0051494	negative regulation of cytoskeleton…	118	5	1.35	3.7	0.013	1.000
GO:0034765	regulation of ion transmembrane transport	284	8	3.25	2.46	0.019	1.000
GO:1903531	negative regulation of secretion by cell	74	4	0.85	4.72	0.012	1.000
GO:0051048	negative regulation of secretion	83	4	0.95	4.2	0.017	1.000
GO:1901137	carbohydrate derivative biosynthetic…	454	11	5.2	2.11	0.022	1.000
GO:0009100	glycoprotein metabolic process	281	7	3.22	2.17	0.046	1.000

GO	GO terms	*Gallus gallus*	Relaxed genes	expected	Fold enrichment	P	FDR
GO:0006310	DNA recombination	178	7	2.04	3.43	0.005	1.000
GO:0016444	somatic cell DNA recombination	33	4	0.38	10.58	0.001	0.599
GO:0000723	telomere maintenance	65	4	0.74	5.37	0.008	1.000
GO:0006302	double-strand break repair	163	5	1.87	2.68	0.043	1.000
GO:0002562	somatic diversification of immune…	32	4	0.37	10.91	0.001	0.746
GO:0002377	immunoglobulin production	62	4	0.71	5.63	0.007	1.000
GO:0002520	immune system development	470	11	5.39	2.04	0.025	1.000
GO:0002200	somatic diversification of immune…	38	4	0.44	9.18	0.001	0.832
GO:0016447	somatic recombination of…	21	4	0.24	16.62	0.000	0.459
GO:0016445	somatic diversification of…	26	4	0.3	13.42	0.000	0.476
GO:0007015	actin filament organization	232	8	2.66	3.01	0.006	1.000
GO:0007601	visual perception	115	5	1.32	3.79	0.012	1.000
GO:0050890	cognition	163	5	1.87	2.68	0.043	1.000
GO:0050953	sensory perception of light stimulus	121	5	1.39	3.61	0.014	1.000
GO:0051640	organelle localization	361	9	4.14	2.18	0.040	1.000
GO:0051650	establishment of vesicle localization	110	4	1.26	3.17	0.041	1.000
GO:0051648	vesicle localization	113	4	1.3	3.09	0.044	1.000
GO:0051656	establishment of organelle localization	259	7	2.97	2.36	0.032	1.000
GO:0006898	receptor-mediated endocytosis	109	4	1.25	3.2	0.040	1.000
GO:0098739	import across plasma membrane	110	5	1.26	3.97	0.010	1.000
GO:0071805	potassium ion transmembrane…	141	8	1.62	4.95	0.000	0.458
GO:0006813	potassium ion transport	146	8	1.67	4.78	0.000	0.468
GO:0032200	telomere organization	65	4	0.74	5.37	0.008	1.000
GO:0051146	striated muscle cell differentiation	151	6	1.73	3.47	0.009	1.000
GO:0060538	skeletal muscle organ development	103	6	1.18	5.08	0.002	0.940
GO:0042692	muscle cell differentiation	193	6	2.21	2.71	0.026	1.000
GO:0007517	muscle organ development	162	8	1.86	4.31	0.001	0.708
GO:0048839	inner ear development	149	5	1.71	2.93	0.031	1.000
GO:0043583	ear development	164	5	1.88	2.66	0.044	1.000
GO:0014706	striated muscle tissue development	202	11	2.32	4.75	0.000	0.444
GO:0060537	muscle tissue development	216	11	2.48	4.44	0.000	0.396
GO:0007519	skeletal muscle tissue development	98	6	1.12	5.34	0.001	0.855
GO:0048738	cardiac muscle tissue development	107	6	1.23	4.89	0.002	1.000
GO:0050806	positive regulation of synaptic transmission	66	4	0.76	5.29	0.008	1.000
GO:0048167	regulation of synaptic plasticity	103	5	1.18	4.24	0.008	1.000
GO:0042445	hormone metabolic process	106	4	1.21	3.29	0.037	1.000
GO:0035239	tube morphogenesis	437	10	5.01	2	0.037	1.000
GO:0048514	blood vessel morphogenesis	269	7	3.08	2.27	0.038	1.000
GO:0001568	blood vessel development	323	9	3.7	2.43	0.014	1.000
GO:0002440	production of molecular mediator of…	62	4	0.71	5.63	0.007	1.000
GO:0001944	vasculature development	341	9	3.91	2.3	0.019	1.000
GO:0007507	heart development	381	11	4.37	2.52	0.005	1.000
GO:0010817	regulation of hormone levels	305	8	3.5	2.29	0.027	1.000
GO:0035295	tube development	554	13	6.35	2.05	0.014	1.000
GO:0072359	circulatory system development	599	15	6.87	2.18	0.005	1.000

Appendix C4 Table 2: Intensified genes. Overrepresentation test for Gene Ontology (GO) biological process using PANTHER (Protein Analysis Through Evolutionary Relationships, http://pantherdb.org) (Mi et al. 2019) with *Gallus gallus* annotations. The "Intensified" column is the number of intensified genes associated with the GO term, the p-value as determined by Fisher's exact test and False Discovery Rate (FDR) as calculated by the Benjamini-Hochberg procedure.

GO	GO biological process complete	*Gallus gallus*	Intensified	expected	Fold enrichment	P-value	FDR
GO:0006275	regulation of DNA replication	75	4	0.93	4.3	0.016	1.000
GO:0007568	aging	77	4	0.96	4.19	0.018	1.000
GO:0008283	cell population proliferation	322	11	4	2.75	0.003	0.694
GO:0009628	response to abiotic stimulus	590	15	7.32	2.05	0.012	1.000
GO:0006898	receptor-mediated endocytosis	108	4	1.34	2.98	0.049	1.000
GO:0051260	protein homooligomerization	146	5	1.81	2.76	0.039	1.000
GO:0006260	DNA replication	149	6	1.85	3.24	0.012	1.000
GO:0007018	microtubule-based movement	263	7	3.26	2.14	0.049	1.000
GO:0010970	transport along microtubule	107	4	1.33	3.01	0.048	1.000
GO:0006928	movement of cell or subcellular component	1040	26	12.91	2.01	0.001	0.449
GO:0034101	erythrocyte homeostasis	69	4	0.86	4.67	0.012	1.000
GO:0060249	anatomical structure homeostasis	221	7	2.74	2.55	0.023	1.000
GO:0019725	cellular homeostasis	530	17	6.58	2.58	0.000	0.329
GO:0050801	ion homeostasis	504	16	6.26	2.56	0.001	0.422
GO:0048878	chemical homeostasis	728	19	9.04	2.1	0.003	0.625
GO:0072507	divalent inorganic cation homeostasis	306	10	3.8	2.63	0.006	0.830
GO:0055082	cellular chemical homeostasis	470	13	5.83	2.23	0.009	0.989
GO:0002262	myeloid cell homeostasis	85	4	1.05	3.79	0.024	1.000
GO:0006874	cellular calcium ion homeostasis	272	9	3.38	2.67	0.008	0.929
GO:0006875	cellular metal ion homeostasis	353	11	4.38	2.51	0.005	0.799
GO:0055065	metal ion homeostasis	405	14	5.03	2.79	0.001	0.473
GO:0007204	positive regulation of cytosolic calcium…	167	7	2.07	3.38	0.006	0.833
GO:0055080	cation homeostasis	467	16	5.8	2.76	0.000	0.273
GO:0006873	cellular ion homeostasis	419	13	5.2	2.5	0.003	0.681
GO:0051480	regulation of cytosolic calcium ion…	193	8	2.4	3.34	0.003	0.660
GO:0030003	cellular cation homeostasis	409	13	5.08	2.56	0.002	0.567
GO:0098771	inorganic ion homeostasis	474	15	5.88	2.55	0.001	0.444
GO:0055074	calcium ion homeostasis	281	10	3.49	2.87	0.003	0.635
GO:0072503	cellular divalent inorganic cation…	293	9	3.64	2.47	0.012	1.000
GO:0043087	regulation of GTPase activity	265	8	3.29	2.43	0.020	1.000
GO:0043085	positive regulation of catalytic activity	765	19	9.49	2	0.004	0.742
GO:0051345	positive regulation of hydrolase activity	365	10	4.53	2.21	0.027	1.000
GO:0033044	regulation of chromosome organization	203	6	2.52	2.38	0.044	1.000
GO:2001252	positive regulation of chromosome…	119	6	1.48	4.06	0.004	0.750
GO:0010638	positive regulation of organelle…	410	11	5.09	2.16	0.020	1.000
GO:0010811	positive regulation of cell-substrate adhesion	88	4	1.09	3.66	0.027	1.000
GO:0002685	regulation of leukocyte migration	113	5	1.4	3.57	0.015	1.000
GO:0002687	positive regulation of leukocyte…	67	4	0.83	4.81	0.011	1.000
GO:0040017	positive regulation of locomotion	353	9	4.38	2.05	0.046	1.000
GO:0032101	regulation of response to external stimulus	498	14	6.18	2.27	0.006	0.854
GO:0048584	positive regulation of response to…	1360	35	16.88	2.07	0.000	0.082
GO:1901361	organic cyclic compound catabolic process	248	8	3.08	2.6	0.014	1.000
GO:0019439	aromatic compound catabolic process	237	8	2.94	2.72	0.011	1.000
GO:0007200	phospholipase C-activating G protein-…	72	6	0.89	6.71	0.000	0.314

GO	GO biological process complete	*Gallus gallus*	Intensified	expected	Fold enrichment	P-value	FDR
GO:0009266	response to temperature stimulus	99	4	1.23	3.26	0.038	1.000
GO:0009408	response to heat	57	4	0.71	5.65	0.007	0.850
GO:0001666	response to hypoxia	120	5	1.49	3.36	0.019	1.000
GO:0070482	response to oxygen levels	134	5	1.66	3.01	0.028	1.000
GO:0071214	cellular response to abiotic stimulus	192	6	2.38	2.52	0.036	1.000
GO:0034605	cellular response to heat	34	4	0.42	9.48	0.001	0.421
GO:0071478	cellular response to radiation	128	5	1.59	3.15	0.024	1.000
GO:0036293	response to decreased oxygen levels	122	5	1.51	3.3	0.020	1.000
GO:0006954	inflammatory response	185	7	2.3	3.05	0.010	1.000
GO:0104004	cellular response to environmental stimulus	192	6	2.38	2.52	0.036	1.000
GO:0010506	regulation of autophagy	169	7	2.1	3.34	0.006	0.841
GO:0010507	negative regulation of autophagy	48	4	0.6	6.71	0.004	0.695
GO:0007030	Golgi organization	102	4	1.27	3.16	0.042	1.000
GO:0034097	response to cytokine	473	14	5.87	2.38	0.003	0.691
GO:0006935	chemotaxis	352	9	4.37	2.06	0.046	1.000
GO:1990868	response to chemokine	59	4	0.73	5.46	0.007	0.926
GO:1990830	cellular response to leukemia…	75	5	0.93	5.37	0.003	0.636
GO:0070098	chemokine-mediated signaling pathway	51	4	0.63	6.32	0.005	0.767
GO:1990823	response to leukemia inhibitory factor	76	5	0.94	5.3	0.003	0.641
GO:0042330	taxis	355	9	4.41	2.04	0.047	1.000
GO:0071345	cellular response to cytokine stimulus	426	14	5.29	2.65	0.001	0.433
GO:0060326	cell chemotaxis	128	5	1.59	3.15	0.024	1.000
GO:1990869	cellular response to chemokine	59	4	0.73	5.46	0.007	0.935
GO:0090068	positive regulation of cell cycle process	155	5	1.92	2.6	0.047	1.000
GO:1901989	positive regulation of cell cycle phase…	77	4	0.96	4.19	0.018	1.000
GO:0033365	protein localization to organelle	504	13	6.26	2.08	0.013	1.000
GO:0030705	cytoskeleton-dependent intracellular…	142	5	1.76	2.84	0.035	1.000
GO:0061512	protein localization to cilium	41	5	0.51	9.83	0.000	0.220
GO:0061458	reproductive system development	166	6	2.06	2.91	0.020	1.000
GO:0043065	positive regulation of apoptotic process	296	8	3.67	2.18	0.034	1.000
GO:2001236	regulation of extrinsic apoptotic…	101	5	1.25	3.99	0.010	1.000
GO:0010942	positive regulation of cell death	325	9	4.03	2.23	0.022	1.000
GO:2001233	regulation of apoptotic signaling…	235	9	2.92	3.09	0.003	0.642
GO:2001235	positive regulation of apoptotic…	90	5	1.12	4.48	0.006	0.859
GO:0019932	second-messenger-mediated signaling	123	7	1.53	4.59	0.001	0.425
GO:0019722	calcium-mediated signaling	88	6	1.09	5.49	0.001	0.450
GO:0007548	sex differentiation	170	6	2.11	2.84	0.022	1.000
GO:0046660	female sex differentiation	83	4	1.03	3.88	0.022	1.000
GO:0048608	reproductive structure development	164	6	2.04	2.95	0.019	1.000
GO:0046661	male sex differentiation	104	5	1.29	3.87	0.011	1.000
GO:0008584	male gonad development	85	5	1.05	4.74	0.005	0.796
GO:0045137	development of primary sexual…	135	5	1.68	2.98	0.029	1.000
GO:0008406	gonad development	131	5	1.63	3.08	0.026	1.000
GO:0046546	development of primary male sexual…	86	5	1.07	4.68	0.005	0.788
GO:0032200	telomere organization	65	4	0.81	4.96	0.010	1.000
GO:0007034	vacuolar transport	125	5	1.55	3.22	0.022	1.000
GO:0035725	sodium ion transmembrane transport	106	4	1.32	3.04	0.047	1.000
GO:0006814	sodium ion transport	130	5	1.61	3.1	0.026	1.000
GO:0007186	G protein-coupled receptor signaling pathway	700	20	8.69	2.3	0.001	0.452

GO	GO biological process complete	*Gallus gallus*	Intensified	expected	Fold enrichment	P-value	FDR
GO:0046578	regulation of Ras protein signal transduction	154	7	1.91	3.66	0.004	0.685
GO:0023056	positive regulation of signaling	1082	30	13.43	2.23	0.000	0.078
GO:0010647	positive regulation of cell…	1074	30	13.33	2.25	0.000	0.081
GO:0043123	positive regulation of I-kappaB…	89	4	1.1	3.62	0.028	1.000
GO:0051056	regulation of small GTPase mediated…	188	8	2.33	3.43	0.003	0.693
GO:1902533	positive regulation of intracellular…	612	18	7.6	2.37	0.001	0.428
GO:0009967	positive regulation of signal transduction	972	30	12.06	2.49	0.000	0.023
GO:0051057	positive regulation of small GTPase…	57	4	0.71	5.65	0.007	0.858
GO:0050727	regulation of inflammatory response	161	6	2	3	0.017	1.000
GO:0032103	positive regulation of response to…	201	6	2.49	2.41	0.043	1.000
GO:0042592	homeostatic process	1074	30	13.33	2.25	0.000	0.092
GO:0006310	DNA recombination	177	7	2.2	3.19	0.008	0.916
GO:0000723	telomere maintenance	65	4	0.81	4.96	0.010	1.000
GO:0006302	double-strand break repair	162	6	2.01	2.98	0.018	1.000
GO:0090316	positive regulation of intracellular protein…	84	4	1.04	3.84	0.023	1.000
GO:0032388	positive regulation of intracellular…	109	5	1.35	3.7	0.013	1.000

Appendix C4 Table 3: M0-M2 genes. Overrepresentation test for Gene Ontology (GO) biological process using PANTHER (Protein Analysis Through Evolutionary Relationships, http://pantherdb.org) (Mi et al. 2019) with *Gallus gallus* annotations. The "M0-M2" column is the number of genes significant for M0-M2 associated with the GO term, the p-value as determined by Fisher's exact test and False Discovery Rate (FDR) as calculated by the Benjamini-Hochberg procedure.

GO	GO biological process	*Gallus gallus*	M0-M2	expected	Fold enrichment	P-value	FDR
GO:0007568	aging	77	7	2.85	2.46	0.030	1.000
GO:0010810	regulation of cell-substrate adhesion	147	11	5.43	2.02	0.029	1.000
GO:0070988	demethylation	45	10	1.66	6.01	0.000	0.026
GO:0090132	epithelium migration	63	6	2.33	2.58	0.036	1.000
GO:0090136	epithelial cell-cell adhesion	16	4	0.59	6.77	0.005	0.686
GO:0034446	substrate adhesion-dependent cell…	50	5	1.85	2.71	0.046	1.000
GO:0016339	calcium-dependent cell-cell adhesion via…	21	4	0.78	5.15	0.011	0.930
GO:0071402	cellular response to lipoprotein particle stimulus	17	4	0.63	6.37	0.006	0.784
GO:0061512	protein localization to cilium	41	5	1.52	3.3	0.023	1.000
GO:0051258	protein polymerization	67	6	2.48	2.42	0.046	1.000
GO:0043462	regulation of ATPase activity	38	5	1.4	3.56	0.018	0.976
GO:0032781	positive regulation of ATPase activity	27	5	1	5.01	0.005	0.709
GO:0060191	regulation of lipase activity	41	6	1.52	3.96	0.006	0.779
GO:0010518	positive regulation of phospholipase activity	26	5	0.96	5.2	0.004	0.649
GO:0010517	regulation of phospholipase activity	33	6	1.22	4.92	0.002	0.431
GO:0060193	positive regulation of lipase activity	32	5	1.18	4.23	0.010	0.876
GO:0098900	regulation of action potential	24	4	0.89	4.51	0.017	1.000
GO:0098901	regulation of cardiac muscle cell action…	14	4	0.52	7.73	0.003	0.528
GO:1903900	regulation of viral life cycle	68	6	2.51	2.39	0.048	1.000
GO:0045071	negative regulation of viral genome…	23	4	0.85	4.71	0.015	0.966
GO:0045069	regulation of viral genome replication	44	5	1.63	3.08	0.030	1.000
GO:1903115	regulation of actin filament-based movement	22	5	0.81	6.15	0.002	0.435
GO:2000107	negative regulation of leukocyte apoptotic process	22	4	0.81	4.92	0.013	0.903
GO:2001236	regulation of extrinsic apoptotic signaling…	101	9	3.73	2.41	0.017	1.000
GO:2001238	positive regulation of extrinsic apoptotic…	35	4	1.29	3.09	0.049	1.000
GO:1900076	regulation of cellular response to insulin stimulus	35	4	1.29	3.09	0.049	1.000
GO:0001938	positive regulation of endothelial cell proliferation	59	6	2.18	2.75	0.028	1.000
GO:0044057	regulation of system process	295	23	10.9	2.11	0.001	0.331
GO:0006937	regulation of muscle contraction	92	11	3.4	3.24	0.001	0.329
GO:0090257	regulation of muscle system process	116	12	4.29	2.8	0.002	0.426
GO:0008016	regulation of heart contraction	108	14	3.99	3.51	0.000	0.065
GO:1903522	regulation of blood circulation	131	19	4.84	3.92	0.000	0.005
GO:0086065	cell communication involved in cardiac…	20	6	0.74	8.12	0.000	0.114
GO:0045823	positive regulation of heart contraction	22	4	0.81	4.92	0.013	0.889
GO:1903524	positive regulation of blood circulation	22	4	0.81	4.92	0.013	0.894
GO:0086069	bundle of His cell to Purkinje myocyte…	8	4	0.3	13.53	0.001	0.205
GO:0061337	cardiac conduction	38	6	1.4	4.27	0.004	0.643
GO:0055117	regulation of cardiac muscle contraction	40	8	1.48	5.41	0.000	0.114
GO:0086004	regulation of cardiac muscle cell contraction	18	5	0.67	7.52	0.001	0.325
GO:0002027	regulation of heart rate	56	8	2.07	3.87	0.002	0.419
GO:0006942	regulation of striated muscle contraction	53	8	1.96	4.08	0.001	0.339
GO:0033044	regulation of chromosome organization	203	16	7.5	2.13	0.008	0.806
GO:2001252	positive regulation of chromosome…	119	11	4.4	2.5	0.007	0.827
GO:0010506	regulation of autophagy	169	15	6.25	2.4	0.004	0.566
GO:0010507	negative regulation of autophagy	48	6	1.77	3.38	0.012	0.901
GO:0008214	protein dealkylation	30	7	1.11	6.31	0.000	0.115
GO:0010332	response to gamma radiation	27	4	1	4.01	0.023	1.000
GO:0009408	response to heat	57	6	2.11	2.85	0.024	1.000
GO:0034605	cellular response to heat	34	5	1.26	3.98	0.012	0.895
GO:0043627	response to estrogen	35	4	1.29	3.09	0.049	1.000
GO:0071391	cellular response to estrogen stimulus	15	4	0.55	7.22	0.004	0.595
GO:0002440	production of molecular mediator of immune…	62	8	2.29	3.49	0.003	0.523
GO:0043171	peptide catabolic process	22	4	0.81	4.92	0.013	0.898

GO	GO biological process	*Gallus gallus*	M0-M2	expected	Fold enrichment	P-value	FDR
GO:0007528	neuromuscular junction development	34	4	1.26	3.18	0.045	1.000
GO:0045216	cell-cell junction organization	132	10	4.88	2.05	0.037	1.000
GO:0007043	cell-cell junction assembly	90	10	3.33	3.01	0.003	0.499
GO:0006829	zinc ion transport	25	5	0.92	5.41	0.004	0.573
GO:0071805	potassium ion transmembrane transport	141	12	5.21	2.3	0.013	0.888
GO:0006813	potassium ion transport	146	12	5.4	2.22	0.014	0.962
GO:0008154	actin polymerization or depolymerization	49	6	1.81	3.31	0.013	0.912
GO:0051298	centrosome duplication	33	4	1.22	3.28	0.042	1.000
GO:0030050	vesicle transport along actin filament	21	4	0.78	5.15	0.011	0.924
GO:0099515	actin filament-based transport	23	4	0.85	4.71	0.015	0.962
GO:0010970	transport along microtubule	107	9	3.95	2.28	0.023	1.000
GO:0042073	intraciliary transport	36	5	1.33	3.76	0.015	0.963
GO:2000573	positive regulation of DNA biosynthetic process	42	5	1.55	3.22	0.025	1.000
GO:0090130	tissue migration	68	6	2.51	2.39	0.048	1.000
GO:0035637	multicellular organismal signaling	74	8	2.73	2.93	0.009	0.822
GO:0007612	learning	83	7	3.07	2.28	0.042	1.000
GO:0001990	regulation of systemic arterial blood…	24	4	0.89	4.51	0.017	1.000
GO:0097746	blood vessel diameter maintenance	77	7	2.85	2.46	0.030	1.000
GO:0003044	regulation of systemic arterial blood…	31	4	1.15	3.49	0.035	1.000
GO:0019229	regulation of vasoconstriction	32	6	1.18	5.07	0.002	0.420
GO:0035296	regulation of tube diameter	77	7	2.85	2.46	0.030	1.000
GO:0031929	TOR signaling	27	4	1	4.01	0.023	1.000
GO:0008630	intrinsic apoptotic signaling pathway in…	46	5	1.7	2.94	0.035	1.000
GO:0071875	adrenergic receptor signaling pathway	27	4	1	4.01	0.023	1.000
GO:0007200	phospholipase C-activating G protein…	72	9	2.66	3.38	0.002	0.436
GO:0051482	positive regulation of cytosolic calcium…	33	5	1.22	4.1	0.011	0.913
GO:0090068	positive regulation of cell cycle process	155	12	5.73	2.1	0.019	0.995
GO:1901992	positive regulation of mitotic cell cycle…	58	6	2.14	2.8	0.026	1.000
GO:1901989	positive regulation of cell cycle phase…	77	8	2.85	2.81	0.011	0.912
GO:0003179	heart valve morphogenesis	34	4	1.26	3.18	0.045	1.000
GO:0003231	cardiac ventricle development	84	7	3.1	2.26	0.044	1.000
GO:0003281	ventricular septum development	52	6	1.92	3.12	0.017	1.000
GO:0006892	post-Golgi vesicle-mediated transport	71	8	2.62	3.05	0.007	0.823
GO:0006898	receptor-mediated endocytosis	108	9	3.99	2.26	0.038	1.000
GO:0048193	Golgi vesicle transport	213	16	7.87	2.03	0.010	0.900
GO:0043949	regulation of cAMP-mediated signaling	18	4	0.67	6.01	0.007	0.827
GO:0043122	regulation of I-kappaB kinase/NF-kappaB…	138	12	5.1	2.35	0.012	0.950
GO:0046580	negative regulation of Ras protein signal…	43	5	1.59	3.15	0.028	1.000
GO:0043123	positive regulation of I-kappaB kinase/NF…	89	10	3.29	3.04	0.003	0.494
GO:0051056	regulation of small GTPase mediated signal…	188	16	6.95	2.3	0.003	0.523
GO:0051058	negative regulation of small GTPase…	49	5	1.81	2.76	0.043	1.000
GO:0046578	regulation of Ras protein signal transduction	154	13	5.69	2.28	0.009	0.854
GO:0051057	positive regulation of small GTPase…	57	6	2.11	2.85	0.024	1.000
GO:0035023	regulation of Rho protein signal transduction	78	8	2.88	2.78	0.012	0.946
GO:0035024	negative regulation of Rho protein signal…	20	4	0.74	5.41	0.010	0.872
GO:0002709	regulation of T cell mediated immunity	32	4	1.18	3.38	0.038	1.000
GO:0002763	positive regulation of myeloid leukocyte…	33	4	1.22	3.28	0.042	1.000
GO:0032200	telomere organization	65	11	2.4	4.58	0.000	0.054
GO:0070076	histone lysine demethylation	24	7	0.89	7.89	0.000	0.057
GO:0033169	histone H3-K9 demethylation	12	4	0.44	9.02	0.002	0.435
GO:0016577	histone demethylation	28	7	1.03	6.77	0.000	0.097
GO:0006482	protein demethylation	30	7	1.11	6.31	0.000	0.119
GO:0016556	mRNA modification	28	5	1.03	4.83	0.006	0.796
GO:0000387	spliceosomal snRNP assembly	29	4	1.07	3.73	0.029	1.000
GO:0000154	rRNA modification	29	4	1.07	3.73	0.029	1.000
GO:0060759	regulation of response to cytokine stimulus	69	7	2.55	2.75	0.019	0.978

GO	GO biological process	*Gallus gallus*	M0-M2	expected	Fold enrichment	P-value	FDR
GO:0060249	anatomical structure homeostasis	221	19	8.17	2.33	0.001	0.327
GO:0002262	myeloid cell homeostasis	85	7	3.14	2.23	0.046	1.000
GO:0002833	positive regulation of response to biotic stimulus	77	7	2.85	2.46	0.030	1.000
GO:0045089	positive regulation of innate immune…	56	6	2.07	2.9	0.023	1.000
GO:0002218	activation of innate immune response	18	4	0.67	6.01	0.007	0.834
GO:0001959	regulation of cytokine-mediated signaling pathway	67	7	2.48	2.83	0.016	1.000
GO:0001961	positive regulation of cytokine-mediated…	25	7	0.92	7.58	0.000	0.065
GO:0060760	positive regulation of response to cytokine…	27	7	1	7.02	0.000	0.086
GO:0018105	peptidyl-serine phosphorylation	147	11	5.43	2.02	0.029	1.000
GO:0035150	regulation of tube size	78	7	2.88	2.43	0.032	1.000
GO:0046822	regulation of nucleocytoplasmic transport	75	7	2.77	2.53	0.027	1.000
GO:0046888	negative regulation of hormone secretion	32	4	1.18	3.38	0.038	1.000
GO:0046824	positive regulation of nucleocytoplasmic…	45	7	1.66	4.21	0.002	0.434
GO:0033157	regulation of intracellular protein transport	126	10	4.66	2.15	0.031	1.000
GO:0042307	positive regulation of protein import into…	30	4	1.11	3.61	0.032	1.000
GO:0032388	positive regulation of intracellular transport	109	11	4.03	2.73	0.004	0.563
GO:0090316	positive regulation of intracellular protein…	84	9	3.1	2.9	0.006	0.791
GO:1904591	positive regulation of protein import	32	4	1.18	3.38	0.038	1.000
GO:1990830	cellular response to leukemia inhibitory factor	75	7	2.77	2.53	0.027	1.000
GO:1990823	response to leukemia inhibitory factor	76	7	2.81	2.49	0.029	1.000
GO:0030705	cytoskeleton-dependent intracellular transport	142	14	5.25	2.67	0.001	0.337
GO:0006310	DNA recombination	177	18	6.54	2.75	0.000	0.104
GO:0006312	mitotic recombination	16	6	0.59	10.15	0.000	0.057
GO:0006303	double-strand break repair via…	32	6	1.18	5.07	0.002	0.414
GO:0000723	telomere maintenance	65	11	2.4	4.58	0.000	0.057
GO:0016444	somatic cell DNA recombination	33	8	1.22	6.56	0.000	0.058
GO:0006302	double-strand break repair	162	14	5.99	2.34	0.006	0.777
GO:0002562	somatic diversification of immune…	32	8	1.18	6.77	0.000	0.060
GO:0000724	double-strand break repair via homologous…	91	8	3.36	2.38	0.025	1.000
GO:0000722	telomere maintenance via recombination	12	4	0.44	9.02	0.002	0.428
GO:0000725	recombinational repair	92	8	3.4	2.35	0.026	1.000
GO:0033152	immunoglobulin VDJ recombination	8	5	0.3	16.91	0.000	0.056
GO:0033151	V(D)J recombination	19	5	0.7	7.12	0.001	0.360
GO:0016447	somatic recombination of immunoglobulin…	21	8	0.78	10.31	0.000	0.011
GO:0016445	somatic diversification of immunoglobulins	26	8	0.96	8.33	0.000	0.027
GO:0002377	immunoglobulin production	62	8	2.29	3.49	0.003	0.517
GO:0002200	somatic diversification of immune receptors	38	8	1.4	5.7	0.000	0.100
GO:0051607	defense response to virus	82	7	3.03	2.31	0.040	1.000
GO:0140546	defense response to symbiont	82	7	3.03	2.31	0.040	1.000

Chapter Five

Appendix C5 Table 1. Accelerated evolutionary rate. Overrepresentation test for Gene Ontology (GO) biological process using PANTHER (Protein Analysis Through Evolutionary Relationships, http://pantherdb.org) (Mi et al. 2019) with *Gallus gallus* annotations. The "Accelerated genes" column is the number of genes showing an accelerated evolutionary rate shift in flightless rails associated with the GO term, the p-value as determined by Fisher's exact test and False Discovery Rate (FDR) as calculated by the Benjamini-Hochberg procedure.

GO ID	Function	*Gallus gallus*	Accelerated genes	expected	fold enrichment	P-value	FDR
GO:0070988	demethylation	45	5	0.56	8.91	0.0004	0.39
GO:0032259	methylation	239	9	2.98	3.02	0.0037	0.85
GO:0071103	DNA conformation change	191	6	2.38	2.52	0.0354	1.00
GO:0016577	histone demethylation	28	4	0.35	11.46	0.0006	0.47
GO:0006338	chromatin remodeling	115	6	1.43	4.18	0.0039	0.83
GO:0006325	chromatin organization	489	15	6.1	2.46	0.0015	0.61
GO:0016570	histone modification	292	8	3.64	2.2	0.0328	1.00
GO:0016569	covalent chromatin modification	294	8	3.67	2.18	0.0339	1.00
GO:0006482	protein demethylation	30	4	0.37	10.69	0.0008	0.48
GO:0008214	protein dealkylation	30	4	0.37	10.69	0.0008	0.46
GO:0018126	protein hydroxylation	29	4	0.36	11.06	0.0007	0.47
GO:0006497	protein lipidation	74	4	0.92	4.34	0.0157	1.00
GO:0042158	lipoprotein biosynthetic process	76	4	0.95	4.22	0.0171	1.00
GO:0008213	protein alkylation	104	5	1.3	3.86	0.0112	1.00
GO:0043414	macromolecule methylation	198	9	2.47	3.65	0.0011	0.52
GO:0006479	protein methylation	104	5	1.3	3.86	0.0112	1.00
GO:1902275	regulation of chromatin organization	139	5	1.73	2.89	0.0329	1.00
GO:0061448	connective tissue development	153	6	1.91	3.15	0.0140	1.00
GO:0014706	striated muscle tissue development	202	6	2.52	2.38	0.0442	1.00
GO:0060429	epithelium development	619	16	7.72	2.07	0.0085	1.00
GO:0030855	epithelial cell differentiation	317	13	3.95	3.29	0.0002	0.33
GO:0060173	limb development	145	5	1.81	2.77	0.0382	1.00
GO:0030326	embryonic limb morphogenesis	96	4	1.2	3.34	0.0351	1.00
GO:0035113	embryonic appendage morphogenesis	97	4	1.21	3.31	0.0362	1.00
GO:0060348	bone development	148	5	1.85	2.71	0.0410	1.00
GO:0051216	cartilage development	126	5	1.57	3.18	0.0231	1.00
GO:0048646	anatomical structure formation involved…	602	15	7.51	2	0.0134	1.00
GO:0009888	tissue development	1058	27	13.19	2.05	0.0005	0.43
GO:0010927	cellular component assembly involved…	77	4	0.96	4.17	0.0178	1.00
GO:0035239	tube morphogenesis	448	12	5.59	2.15	0.0150	1.00
GO:0001838	embryonic epithelial tube formation	93	4	1.16	3.45	0.0319	1.00
GO:0016331	morphogenesis of embryonic epithelium	107	4	1.33	3	0.0484	1.00
GO:0001525	angiogenesis	208	8	2.59	3.09	0.0054	0.94
GO:0048514	blood vessel morphogenesis	278	9	3.47	2.6	0.0094	1.00
GO:0001568	blood vessel development	332	11	4.14	2.66	0.0036	0.84
GO:0072175	epithelial tube formation	102	4	1.27	3.15	0.0421	1.00
GO:0001501	skeletal system development	359	9	4.48	2.01	0.0490	1.00
GO:0072359	circulatory system development	608	16	7.58	2.11	0.0050	0.94
GO:0003231	cardiac ventricle development	84	4	1.05	3.82	0.0234	1.00
GO:0001944	vasculature development	348	11	4.34	2.54	0.0051	0.94
GO:2000027	regulation of animal organ morphogenesis	82	5	1.02	4.89	0.0044	0.89
GO:0035265	organ growth	45	4	0.56	7.13	0.0031	0.77
GO:0034101	erythrocyte homeostasis	69	4	0.86	4.65	0.0126	1.00
GO:0030097	Hemopoiesis	408	11	5.09	2.16	0.0203	1.00
GO:0002520	immune system development	470	12	5.86	2.05	0.0190	1.00
GO:0030099	myeloid cell differentiation	157	7	1.96	3.58	0.0043	0.87
GO:0002262	myeloid cell homeostasis	85	4	1.06	3.77	0.0242	1.00
GO:0030218	erythrocyte differentiation	61	4	0.76	5.26	0.0084	1.00

GO ID	Function	*Gallus gallus*	Accelerated genes	expected	fold enrichment	P-value	FDR
GO:0048736	appendage development	146	5	1.82	2.75	0.0391	1.00
GO:0001822	kidney development	191	6	2.38	2.52	0.0354	1.00
GO:0072001	renal system development	199	6	2.48	2.42	0.0417	1.00
GO:0001655	urogenital system development	204	6	2.54	2.36	0.0459	1.00
GO:0001764	neuron migration	78	4	0.97	4.11	0.0185	1.00
GO:0006814	sodium ion transport	130	5	1.62	3.09	0.0259	1.00
GO:0070588	calcium ion transmembrane transport	149	5	1.86	2.69	0.0420	1.00
GO:0007601	visual perception	115	5	1.43	3.49	0.0165	1.00
GO:0050953	sensory perception of light stimulus	121	5	1.51	3.31	0.0199	1.00
GO:0007623	circadian rhythm	93	4	1.16	3.45	0.0319	1.00
GO:0032922	circadian regulation of gene expression	54	4	0.67	5.94	0.0057	0.97
GO:0030010	establishment of cell polarity	95	4	1.18	3.38	0.0340	1.00
GO:0061919	process utilizing autophagic mechanism	168	6	2.09	2.86	0.0209	1.00
GO:0034329	cell junction assembly	177	6	2.21	2.72	0.0260	1.00
GO:0051259	protein complex oligomerization	187	6	2.33	2.57	0.0325	1.00
GO:0008283	cell population proliferation	322	9	4.01	2.24	0.0218	1.00
GO:0007267	cell-cell signaling	569	15	7.09	2.11	0.0066	1.00
GO:0045995	regulation of embryonic development	45	4	0.56	7.13	0.0031	0.75
GO:0030111	regulation of Wnt signaling pathway	249	9	3.1	2.9	0.0048	0.94
GO:0031396	regulation of protein ubiquitination	151	5	1.88	2.66	0.0440	1.00
GO:0042157	lipoprotein metabolic process	92	4	1.15	3.49	0.0309	1.00
GO:0042445	hormone metabolic process	105	4	1.31	3.06	0.0458	1.00
GO:0006914	autophagy	168	6	2.09	2.86	0.0209	1.00
GO:0043632	modification-dependent macro…	441	14	5.5	2.55	0.0016	0.62
GO:0010498	proteasomal protein catabolic process	284	9	3.54	2.54	0.0106	1.00
GO:0030163	protein catabolic process	539	15	6.72	2.23	0.0046	0.92
GO:0006511	ubiquitin-dependent protein catabolic…	425	14	5.3	2.64	0.0011	0.52
GO:0043161	proteasome-mediated ubiquitin…	253	7	3.15	2.22	0.0424	1.00
GO:0051603	proteolysis involved in cellular protein…	484	15	6.03	2.49	0.0014	0.58
GO:0044257	cellular protein catabolic process	504	15	6.28	2.39	0.0020	0.64
GO:0019941	modification-dependent protein…	433	14	5.4	2.59	0.0013	0.59
GO:0009451	RNA modification	143	5	1.78	2.8	0.0364	1.00
GO:0045165	cell fate commitment	190	6	2.37	2.53	0.0347	1.00
GO:0045893	positive regulation of transcription	1171	32	14.6	2.19	0.0000	0.24
GO:0009891	positive regulation of biosynthetic…	1393	35	17.37	2.02	0.0001	0.23
GO:0051254	positive regulation of RNA metabolic…	1270	32	15.83	2.02	0.0002	0.30
GO:0010557	positive regulation of macromolecule…	1325	33	16.52	2	0.0002	0.28
GO:0045944	positive regulation of transcription…	864	25	10.77	2.32	0.0001	0.27
GO:1902680	positive regulation of RNA biosynthetic…	1173	32	14.62	2.19	0.0000	0.16
GO:0031328	positive regulation of cellular…	1374	35	17.13	2.04	0.0001	0.24
GO:1903508	positive regulation of nucleic acid…	1171	32	14.6	2.19	0.0000	0.47
GO:0016055	Wnt signaling pathway	153	7	1.91	3.67	0.0037	0.84
GO:0198738	cell-cell signaling by wnt	154	7	1.92	3.65	0.0039	0.84
GO:0007268	chemical synaptic transmission	251	7	3.13	2.24	0.0410	1.00
GO:1905114	cell surface receptor signaling pathway…	209	10	2.61	3.84	0.0004	0.39
GO:0035567	non-canonical Wnt signaling pathway	37	5	0.46	10.84	0.0002	0.31
GO:0098916	anterograde trans-synaptic signaling	251	7	3.13	2.24	0.0410	1.00
GO:0030178	negative regulation of Wnt signaling pathway	136	5	1.7	2.95	0.0305	1.00
GO:0060828	regulation of canonical Wnt signaling…	194	7	2.42	2.89	0.0124	1.00

Appendix C5 Table 2. Decelerated evolutionary rate. Overrepresentation test for Gene Ontology (GO) biological process using PANTHER (Protein Analysis Through Evolutionary Relationships, http://pantherdb.org) (Mi et al. 2019) with I annotations. The "Decelerated genes" column is the number of genes showing an decelerated evolutionary rate shift in flightless rails associated with the GO term, the p-value as determined by Fisher's exact test and False Discovery Rate (FDR) as calculated by the Benjamini-Hochberg procedure.

GO ID	Function	Gallus gallus	Decelerated genes	Expected	fold Enrichment	P-value	FDR
GO:0007568	aging	77	6	1.78	3.37	0.011	1.000
GO:0007586	digestion	57	4	1.32	3.04	0.049	1.000
GO:0009611	response to wounding	193	9	4.46	2.02	0.050	1.000
GO:0035821	modulation of process of other…	51	4	1.18	3.4	0.035	1.000
GO:0051607	defense response to virus	82	5	1.89	2.64	0.047	1.000
GO:0140546	defense response to symbiont	82	5	1.89	2.64	0.047	1.000
GO:0040007	growth	236	12	5.45	2.2	0.014	1.000
GO:0043087	regulation of GTPase activity	265	18	6.12	2.94	0.000	0.281
GO:0043547	positive regulation of GTPase…	192	13	4.43	2.93	0.001	0.579
GO:0050673	epithelial cell proliferation	56	4	1.29	3.09	0.046	1.000
GO:0006099	tricarboxylic acid cycle	30	4	0.69	5.77	0.007	0.895
GO:0043277	apoptotic cell clearance	109	6	2.52	2.38	0.046	1.000
GO:0006898	receptor-mediated endocytosis	109	6	2.52	2.38	0.046	1.000
GO:0032456	endocytic recycling	43	4	0.99	4.03	0.021	1.000
GO:0016482	cytosolic transport	104	6	2.4	2.5	0.039	1.000
GO:0007160	cell-matrix adhesion	80	5	1.85	2.71	0.043	1.000
GO:0051260	protein homooligomerization	146	8	3.37	2.37	0.024	1.000
GO:0051259	protein complex oligomerization	187	9	4.32	2.08	0.045	1.000
GO:0007018	microtubule-based movement	263	13	6.07	2.14	0.012	0.999
GO:0010970	transport along microtubule	107	6	2.47	2.43	0.043	1.000
GO:0030317	flagellated sperm motility	56	4	1.29	3.09	0.046	1.000
GO:0047497	mitochondrion transport along…	17	4	0.39	10.19	0.001	0.593
GO:0099111	microtubule-based transport	131	7	3.02	2.31	0.037	1.000
GO:0072384	organelle transport along…	51	5	1.18	4.25	0.009	0.952
GO:0051654	establishment of mitochondrion…	19	4	0.44	9.12	0.002	0.647
GO:0034643	establishment of mitochondrion…	17	4	0.39	10.19	0.001	0.571
GO:0031343	positive regulation of cell killing	26	4	0.6	6.66	0.004	0.872
GO:0090559	regulation of membrane permeability	34	4	0.79	5.1	0.010	1.000
GO:0032890	regulation of organic acid transport	28	4	0.65	6.19	0.006	0.856
GO:0044070	regulation of anion transport	39	4	0.9	4.44	0.016	1.000
GO:0060627	regulation of vesicle-mediated…	303	17	7	2.43	0.002	0.610
GO:0050764	regulation of phagocytosis	47	4	1.09	3.69	0.028	1.000
GO:0017157	regulation of exocytosis	125	8	2.89	2.77	0.011	1.000
GO:0045807	positive regulation of endocytosis	51	4	1.18	3.4	0.035	1.000
GO:1903305	regulation of regulated secretory…	80	5	1.85	2.71	0.043	1.000
GO:0045921	positive regulation of exocytosis	45	4	1.04	3.85	0.024	1.000
GO:0032677	regulation of interleukin-8 production	32	4	0.74	5.41	0.009	0.946
GO:0001819	positive regulation of cytokine…	220	12	5.08	2.36	0.007	0.907
GO:0031341	regulation of cell killing	31	4	0.72	5.59	0.008	0.907
GO:0050863	regulation of T cell activation	177	9	4.09	2.2	0.040	1.000
GO:1902105	regulation of leukocyte…	155	8	3.58	2.24	0.032	1.000
GO:0042129	regulation of T cell proliferation	83	5	1.92	2.61	0.049	1.000
GO:0045580	regulation of T cell differentiation	80	6	1.85	3.25	0.013	0.942
GO:1903708	positive regulation of hemopoiesis	88	6	2.03	2.95	0.020	1.000
GO:0045619	regulation of lymphocyte…	97	6	2.24	2.68	0.029	1.000
GO:1902107	positive regulation of leukocyte	88	6	2.03	2.95	0.020	1.000
GO:0045582	positive regulation of T cell	47	4	1.09	3.69	0.028	1.000
GO:0045621	positive regulation of lymphocyte	56	5	1.29	3.87	0.012	1.000
GO:0016485	protein processing	148	8	3.42	2.34	0.025	1.000
GO:1901617	organic hydroxy compound…	116	9	2.68	3.36	0.002	0.692
GO:0046165	alcohol biosynthetic process	73	7	1.69	4.15	0.002	0.649
GO:0016053	organic acid biosynthetic process	187	9	4.32	2.08	0.045	1.000
GO:0006066	alcohol metabolic process	188	10	4.34	2.3	0.015	1.000
GO:0046173	polyol biosynthetic process	44	4	1.02	3.94	0.023	1.000
GO:0007229	integrin-mediated signaling pathway	63	6	1.45	4.12	0.005	0.875

GO ID	Function	Gallus gallus	Decelerated genes	Expected	fold Enrichment	P-value	FDR
GO:0050927	positive regulation of positive…	15	4	0.35	11.55	0.001	0.575
GO:0050926	regulation of positive chemotaxis	16	4	0.37	10.83	0.001	0.605
GO:0048545	response to steroid hormone	110	8	2.54	3.15	0.005	0.836
GO:0071383	cellular response to steroid…	76	8	1.75	4.56	0.001	0.524
GO:0009755	hormone-mediated signaling…	91	9	2.1	4.28	0.000	0.591
GO:0071396	cellular response to lipid	229	15	5.29	2.84	0.000	0.565
GO:0071407	cellular response to organic cyclic	258	13	5.96	2.18	0.011	1.000
GO:0043401	steroid hormone mediated…	45	6	1.04	5.77	0.001	0.578
GO:0030518	intracellular steroid hormone…	32	5	0.74	6.77	0.001	0.606
GO:0051646	mitochondrion localization	34	4	0.79	5.1	0.010	0.998
GO:0000132	establishment of mitotic spindle…	29	4	0.67	5.97	0.006	0.887
GO:0051653	spindle localization	47	4	1.09	3.69	0.028	1.000
GO:0051656	establishment of organelle…	259	12	5.98	2.01	0.023	1.000
GO:0051294	establishment of spindle…	34	4	0.79	5.1	0.010	0.984
GO:0040001	establishment of mitotic spindle…	34	4	0.79	5.1	0.010	0.991
GO:0051293	establishment of spindle…	43	4	0.99	4.03	0.021	1.000
GO:0001954	positive regulation of cell-matrix…	35	4	0.81	4.95	0.011	0.988
GO:0008637	apoptotic mitochondrial changes	38	4	0.88	4.56	0.015	1.000
GO:0007006	mitochondrial membrane…	64	5	1.48	3.38	0.020	1.000
GO:0022602	ovulation cycle process	24	4	0.55	7.22	0.003	0.817
GO:0048232	male gamete generation	236	11	5.45	2.02	0.029	1.000
GO:0007548	sex differentiation	170	10	3.93	2.55	0.008	0.920
GO:0042698	ovulation cycle	27	4	0.62	6.42	0.005	0.818
GO:0046660	female sex differentiation	83	8	1.92	4.17	0.001	0.594
GO:0048608	reproductive structure…	164	9	3.79	2.38	0.017	1.000
GO:0008585	female gonad development	69	6	1.59	3.77	0.007	0.910
GO:0045137	development of primary sexual…	135	9	3.12	2.89	0.005	0.847
GO:0001541	ovarian follicle development	30	4	0.69	5.77	0.007	0.904
GO:0008406	gonad development	131	8	3.02	2.64	0.014	0.968
GO:0046545	development of primary female…	72	7	1.66	4.21	0.002	0.712
GO:0000096	sulfur amino acid metabolic process	26	4	0.6	6.66	0.004	0.860
GO:0044283	small molecule biosynthetic…	317	16	7.32	2.19	0.004	0.884
GO:1901605	alpha-amino acid metabolic…	143	8	3.3	2.42	0.021	1.000
GO:0046395	carboxylic acid catabolic process	158	8	3.65	2.19	0.035	1.000
GO:1901607	alpha-amino acid biosynthetic…	52	4	1.2	3.33	0.037	1.000
GO:0090132	epithelium migration	63	5	1.45	3.44	0.019	1.000
GO:0043542	endothelial cell migration	46	4	1.06	3.77	0.026	1.000
GO:0010631	epithelial cell migration	61	5	1.41	3.55	0.017	1.000
GO:0090130	tissue migration	68	5	1.57	3.18	0.025	1.000
GO:0045216	cell-cell junction organization	133	8	3.07	2.61	0.015	1.000
GO:0034332	adherens junction organization	39	4	0.9	4.44	0.016	1.000
GO:0070830	bicellular tight junction assembly	36	4	0.83	4.81	0.012	1.000
GO:0120193	tight junction organization	41	4	0.95	4.23	0.018	1.000
GO:0034329	cell junction assembly	177	9	4.09	2.2	0.040	1.000
GO:0007043	cell-cell junction assembly	90	6	2.08	2.89	0.022	1.000
GO:0043297	apical junction assembly	45	4	1.04	3.85	0.024	1.000
GO:0120192	tight junction assembly	39	4	0.9	4.44	0.016	1.000
GO:0003018	vascular process in circulatory system	104	6	2.4	2.5	0.039	1.000
GO:0000956	nuclear-transcribed mRNA catabolic…	80	5	1.85	2.71	0.043	1.000
GO:0007569	cell aging	40	5	0.92	5.41	0.003	0.811
GO:0031099	regeneration	52	4	1.2	3.33	0.037	1.000
GO:0002088	lens development in camera-type eye	64	6	1.48	4.06	0.005	0.825
GO:0070306	lens fiber cell differentiation	22	4	0.51	7.87	0.003	0.705
GO:0030522	intracellular receptor signaling…	97	8	2.24	3.57	0.003	0.699
GO:1905710	positive regulation of membrane…	14	4	0.32	12.37	0.001	0.554
GO:0097345	mitochondrial outer membrane…	9	4	0.21	19.25	0.000	0.354
GO:0046902	regulation of mitochondrial…	28	4	0.65	6.19	0.006	0.847
GO:1902108	regulation of mitochondrial…	15	4	0.35	11.55	0.001	0.609
GO:1902110	positive regulation of…	11	4	0.25	15.75	0.000	0.435
GO:1902686	mitochondrial outer membrane…	13	4	0.3	13.33	0.000	0.549
GO:0035794	positive regulation of…	13	4	0.3	13.33	0.000	0.507
GO:0007200	phospholipase C-activating G protein…	72	6	1.66	3.61	0.008	0.959
GO:0015850	organic hydroxy compound transport	108	7	2.49	2.81	0.016	1.000
GO:0048589	developmental growth	233	12	5.38	2.23	0.014	0.968
GO:0042246	tissue regeneration	33	4	0.76	5.25	0.009	0.993
GO:1990138	neuron projection extension	45	4	1.04	3.85	0.024	1.000
GO:0060560	developmental growth involved…	83	5	1.92	2.61	0.049	1.000

Appendix C5 Table 3. Proteins with altered function in each flightless species. GO terms associated with biological function enriched in function-altered proteins for each flightless species. GO terms were manually assigned to a main biological function.

Biological function	*A. rogersi*	*G. australis*	*P. hochstetteri*	*Z. atra*
Immune response	**GO:0002699** positive regulation of immune… **GO:0031347** regulation of defence response **GO:0031349** positive regulation of defence…	**GO:0042060** wound healing **GO:0031347** regulation of defence response **GO:0006954** inflammatory response	**GO:0016032** viral process **GO:0019058** viral life cycle	**GO:0002685** regulation of leukocyte… **GO:0002699** positive regulation of immune… **GO:1903706** regulation of hemopoiesis **GO:1902105** regulation of leukocyte… **GO:0002683** negative regulation of… **GO:0045621** positive regulation of… **GO:0045619** regulation of lymphocyte… **GO:0031347** regulation of defence response
DNA structure: Chromosome, chromatine, histone	**GO:0016573** histone acetylation **GO:0043966** histone H3 acetylation **GO:0065004** protein-DNA complex assembly	**GO:0007059** chromosome segregation **GO:0051303** establishment of chromosome.. **GO:0050000** chromosome localization **GO:0098813** nuclear chromosome… **GO:0000070** mitotic sister chromatid… **GO:0000819** sister chromatid segregation	**GO:0007059** chromosome segregation **GO:0098813** nuclear chromosome… **GO:0070192** chromosome organization… **GO:0031497** chromatin assembly **GO:0006333** chromatin assembly or… **GO:0006325** chromatin organization **GO:0016570** histone modification **GO:0016569** covalent chromatin… **GO:0051276** chromosome organization **GO:0071824** protein-DNA complex… **GO:0065004** protein-DNA complex assembly **GO:0071103** DNA conformation change **GO:0006323** DNA packaging	**GO:0031497** chromatin assembly **GO:0071103** DNA conformation change **GO:0032508** DNA duplex unwinding **GO:0006323** DNA packaging
Lipid metabolic process	**GO:0006869** lipid transport **GO:0010876** lipid localization **GO:0006665** sphingolipid metabolic process	**GO:0006629** lipid metabolic process		**GO:1903509** liposaccharide metabolic… **GO:0006639** acylglycerol metabolic process **GO:0006638** neutral lipid metabolic process **GO:0006643** membrane lipid metabolic… **GO:0006664** glycolipid metabolic process **GO:0046486** glycerolipid metabolic process **GO:0045017** glycerolipid biosynthetic… **GO:0046467** membrane lipid biosynthetic… **GO:0009247** glycolipid biosynthetic process **GO:0006641** triglyceride metabolic process **GO:0010876** lipid localization **GO:0072657** protein localization to… **GO:0051668** localization within membrane **GO:0043113** receptor clustering

Biological function	*A. rogersi*	*G. australis*	*P. hochstetteri*	*Z. atra*
Cilium		**GO:0007017** microtubule-based process **GO:0060271** cilium assembly **GO:0030031** cell projection assembly **GO:0044782** cilium organization **GO:0120031** plasma membrane bounded… **GO:0000226** microtubule cytoskeleton… **GO:0007010** cytoskeleton organization **GO:0030036** actin cytoskeleton organization **GO:0007015** actin filament organization	**GO:0061640** cytoskeleton-dependent… **GO:0007051** spindle organization	**GO:0031122** cytoplasmic microtubule…
Circulatory system	**GO:0002274** myeloid leukocyte activation **GO:0030218** erythrocyte differentiation **GO:0034101** erythrocyte homeostasis **GO:0008015** blood circulation **GO:0003013** circulatory system process **GO:0003073** regulation of systemic arterial… **GO:0008217** regulation of blood pressure		**GO:1903522** regulation of blood circulation	
Muscle development and contraction	**GO:0014706** striated muscle tissue… **GO:0051146** striated muscle cell… **GO:0030239** myofibril assembly **GO:0055002** striated muscle cell…	**GO:0006936** muscle contraction **GO:0003012** muscle system process		
Methylation	**GO:0032259** methylation **GO:0043414** macromolecule methylation **GO:0001510** RNA methylation		**GO:0032259** methylation	
Neurone development	**GO:0050806** positive regulation of synaptic… **GO:0060291** long-term synaptic potentiation			**GO:1901214** regulation of neuron death **GO:0010977** negative regulation of neuron…
Respiratory system		**GO:0060541** respiratory system…		
Kidney		**GO:0001822** kidney development		

Biological function	*A. rogersi*	*G. australis*	*P. hochstetteri*	*Z. atra*
Others	**GO:0006275** regulation of DNA replication	**GO:0030155** regulation of cell adhesion	**GO:0010212** response to ionizing radiation	**GO:0007338** single fertilization
	GO:0051235 maintenance of location	**GO:0051656** establishment of organelle…	**GO:0051321** meiotic cell cycle	**GO:0015849** organic acid transport
	GO:0070085 glycosylation	**GO:0051310** metaphase plate congression	**GO:1903046** meiotic cell cycle process	**GO:0051775** response to redox state
	GO:0031032 actomyosin structure…	**GO:1903046** meiotic cell cycle process	**GO:0000910** cytokinesis	**GO:1901615** organic hydroxy compound…
	GO:0002831 regulation of response to…	**GO:0051321** meiotic cell cycle	**GO:1903047** mitotic cell cycle process	**GO:0051346** negative regulation of…
	GO:0006721 terpenoid metabolic process	**GO:0022402** cell cycle process	**GO:0000278** mitotic cell cycle	**GO:0052548** regulation of endopeptidase…
	GO:0006720 isoprenoid metabolic process	**GO:0140013** meiotic nuclear division	**GO:0022402** cell cycle process	**GO:0045861** negative regulation of…
	GO:0009308 amine metabolic process	**GO:0046394** carboxylic acid biosynthetic…	**GO:0007127** meiosis I	**GO:0052547** regulation of peptidase activity
	GO:0042445 hormone metabolic process	**GO:0016053** organic acid biosynthetic…	**GO:0061982** meiosis I cell cycle process	**GO:0010951** negative regulation of…
	GO:0034754 cellular hormone metabolic…	**GO:0097435** supramolecular fiber…	**GO:0140013** meiotic nuclear division	**GO:0010466** negative regulation of…
	GO:0008033 tRNA processing	**GO:0002831** regulation of response to…	**GO:0000281** mitotic cytokinesis	**GO:0010389** regulation of G2/M transition…
	GO:0009451 RNA modification	**GO:0010810** regulation of cell-substrate…	**GO:0051301** cell division	**GO:1902749** regulation of cell cycle G2/M…
	GO:0016072 rRNA metabolic process	**GO:0000280** nuclear division	**GO:0008283** cell population proliferation	**GO:1901988** negative regulation of cell…
	GO:0034470 ncRNA processing	**GO:0140014** mitotic nuclear division	**GO:0007049** cell cycle	**GO:0031570** DNA integrity checkpoint…
	GO:0006400 tRNA modification	**GO:0044085** cellular component biogenesis	**GO:0043087** regulation of GTPase activity	**GO:0044773** mitotic DNA damage…
	GO:0043543 protein acylation	**GO:0070372** regulation of ERK1 and ERK2…	**GO:0071902** positive regulation of protein…	**GO:1901991** negative regulation of mitotic…
	GO:1902905 positive regulation of supramolecular fiber organization	**GO:0060429** epithelium development	**GO:0090630** activation of GTPase activity	**GO:0000075** cell cycle checkpoint signaling
		GO:0002009 morphogenesis of an…	**GO:0043270** positive regulation of ion…	**GO:0044774** mitotic DNA integrity…
	GO:0006022 aminoglycan metabolic process	**GO:0048729** tissue morphogenesis	**GO:0051302** regulation of cell division	**GO:0007093** mitotic cell cycle checkpoint…
	GO:1903510 mucopolysaccharide…	**GO:0060562** epithelial tube morphogenesis	**GO:0051726** regulation of cell cycle	**GO:0000077** DNA damage checkpoint…
	GO:0006029 proteoglycan metabolic process	**GO:0048285** organelle fission	**GO:0051052** regulation of DNA metabolic…	**GO:0033135** regulation of peptidyl-serine…
	GO:0009100 glycoprotein metabolic process	**GO:0035239** tube morphogenesis	**GO:0080135** regulation of cellular response…	**GO:0030162** regulation of proteolysis
	GO:0009101 glycoprotein biosynthetic…	**GO:0032101** regulation of response to…	**GO:2001020** regulation of response to DNA…	**GO:0050770** regulation of axonogenesis
	GO:0030203 glycosaminoglycan metabolic…	**GO:0035295** tube development	**GO:0010564** regulation of cell cycle process	**GO:0010721** negative regulation of cell…
	GO:0033138 positive regulation of…		**GO:1901987** regulation of cell cycle phase…	**GO:0030308** negative regulation of cell…
	GO:0010927 cellular component assembly…		**GO:0006260** DNA replication	**GO:0048640** negative regulation of…
	GO:0060041 retina development in camera…		**GO:0006261** DNA-dependent DNA…	**GO:0022407** regulation of cell-cell adhesion
	GO:0008585 female gonad development		**GO:0006974** cellular response to DNA…	**GO:0022408** negative regulation of cell-cell…
	GO:0046545 development of primary…		**GO:0042770** signal transduction in…	**GO:0080134** regulation of response to stress
	GO:0043413 macromolecule glycosylation		**GO:0033554** cellular response to stress	**GO:0006282** regulation of DNA repair
	GO:0006487 protein N-linked glycosylation		**GO:0034660** ncRNA metabolic process	**GO:2000779** regulation of double-strand…
	GO:0006486 protein glycosylation		**GO:0006310** DNA recombination	**GO:0043648** dicarboxylic acid metabolic…
	GO:0018205 peptidyl-lysine modification		**GO:0036297** interstrand cross-link repair	**GO:0006066** alcohol metabolic process
	GO:0006473 protein acetylation		**GO:0006259** DNA metabolic process	**GO:0019751** polyol metabolic process
	GO:0006475 internal protein amino acid…		**GO:0016072** rRNA metabolic process	**GO:0046165** alcohol biosynthetic process
	GO:0018394 peptidyl-lysine acetylation		**GO:0000723** telomere maintenance	**GO:0046173** polyol biosynthetic process
	GO:0018393 internal peptidyl-lysine…		**GO:0090304** nucleic acid metabolic proces	**GO:0019221** cytokine-mediated signaling…

Biological function	*A. rogersi*	*G. australis*	*P. hochstetteri*	*Z. atra*
Others			**GO:0006396** RNA processing **GO:0006281** DNA repair **GO:0006302** double-strand break repair **GO:0006364** rRNA processing **GO:0034470** ncRNA processing **GO:0000280** nuclear division **GO:0032200** telomere organization **GO:0000018** regulation of DNA… **GO:0051054** positive regulation of DNA… **GO:0006282** regulation of DNA repair **GO:2000779** regulation of double-strand… **GO:0010569** regulation of double-strand… **GO:0048285** organelle fission	**GO:0032922** circadian regulation of gene… **GO:1902414** protein localization to cell… **GO:0001819** positive regulation of cytokine… **GO:0042770** signal transduction in… **GO:0034728** nucleosome organization **GO:0006334** nucleosome assembly **GO:0051258** protein polymerization

Chapter Six

Appendix C6: Introns

Methods

Whole gene sequences were retrieved from seven rail annotated genomes (see Chapter 3 for more information about genome assembly and annotation), four flightless species *Porphyrio hochstetteri, Gallirallus australis, Zapornia atra, Atlantisia rogersi* and three volant species *Porphyrio melanotus, Gallirallus philippensis* and *Fulica atra* using the "extract regions" tool on Geneious R.11 (https://www.geneious.com). Sequences from each species were organised by gene and the exons were discarded. Only the genes that contained homologous sequences for all seven rail species were used, and intron lists for these were aligned with MAFFT v7.490 (Katoh et al. 2002) using the default settings and ambiguous alignment positions were eliminated using Gblocks v0.91b (Castresana 2000) (options: -b1=4 -b2=7 -b4=50).

The intron alignments were processed with BaseML, a part of the PAML package (Yang 2007), to estimate branch lengths of each taxon using the REV model of nucleotide substitution (model = 7) on the species topology inferred in Chapter 4. Evolutionary rate tests were applied to each intron alignment using *RERconverge* (Kowalczyk et al. 2019) in order to identify the non-coding regions that exhibited significant substitution rate differences between volant and flightless species. *RERconverge* was implemented using a Docker image (https://github.com/nclark-lab/RERconverge). Intron trees were extracted from the BaseML outputs and raw branch lengths were transformed (transform = "sqrt", weighted= T, and scale =T) into relative rates using a projection operator method (Sato et al. 2005). The flightless lineages were defined as foreground branches and the flightless trait was set as terminal. Introns that exhibit significantly different rates of molecular evolution between flightless and volant species ($p<0.05$) were assigned to one of two groups; one where the evolution rate was higher in the volant species (rho <0) and one where the evolution rate was higher in the flightless species (rho >0). An overrepresentation test was carried out on the introns with accelerated evolution rates in flightless rails using the methods described in Chapters 4 and 5.

Results and discussion

Out of 8,916 intron alignments, 417 showed significant relative evolution rate difference between flightless and volant species; 208 intron alignments had a higher rate in flightless species (rho>0) and 209 had a higher rate in volant species (rho<0). To find out if the introns with accelerated evolutionary rates in flightless species were enriched in certain biological functions, an overrepresentation test was carried out (Appendix C6 Table 1). The results showed that some of the overrepresented biological functions were similar to those enriched in CDSs with an accelerated evolutionary rate associated with immune response, chromatin assembly, limb morphogenesis and circulatory system. However, the proportion of enriched functions in accelerated introns that could be associated with flight capacity was much lower than in coding regions (see Chapter 5 and Appendix C6 Table 1).

Relative evolution rate results for coding regions (CDSs) and introns were compared for evidence of correlation (Fig. 2). Only 11 genes showed evidence of accelerated evolutionary rates in flightless rails in both coding (CDSs) and non-coding (introns) regions. It, therefore, appears that selective pressure acting on introns and exons within the same genes are mostly unlinked. Based on this result, one can assume that, at the gene level, selection associated with flightlessness can lead to two evolutionary trajectories. Mutations accumulating in the coding regions resulting in a direct effect on protein structure or in the non-coding regions potentially affecting gene regulation could both influence phenotypic evolution.

It is worth mentioning that the relative evolutionary rate analysis made on introns included only 7 species when the one on CDSs included 12 species. This was done because of the higher variation in non-coding regions than in coding regions, and including genetically more distant species in the analysis would reduce the number and the length of introns successfully aligned. Moreover, one should be careful when comparing intron and CDSs results as introns alignments were nucleotide sequences while CDSs were amino acids. For these reasons, the results presented in this section are considered preliminary. To investigate the non-coding basis of convergent evolution to flightlessness in rails, more detailed analysis at a larger scale should be carried out (see Future research section).

Appendix C6 Table 1: Introns with accelerated evolutionary rate in flightless rails. Overrepresentation test for Gene Ontology (GO) biological process using PANTHER (Protein Analysis Through Evolutionary Relationships, http://pantherdb.org) (Mi et al. 2019) with *Gallus gallus* annotations. The "Accelerated genes" column is the number of genes showing an accelerated evolutionary rate shift in flightless rails associated with the GO term, the p-value as determined by Fisher's exact test and False Discovery Rate (FDR) as calculated by the Benjamini-Hochberg procedure.

GO ID	Function	*Gallus gallus*	Accelerated introns	expected	fold Enrichment	P-value	FDR
GO:0002697	regulation of immune effector process	164	5	1.9	2.63	0.045	1.00
GO:0002703	regulation of leukocyte mediated immunity	104	4	1.2	3.32	0.036	1.00
GO:0002637	regulation of immunoglobulin production	38	4	0.44	9.1	0.001	0.42
GO:0002700	regulation of production of molecular…	79	4	0.91	4.38	0.015	1.00
GO:0006325	chromatin organization	489	13	5.66	2.3	0.008	0.86
GO:0031497	chromatin assembly	79	4	0.91	4.38	0.015	1.00
GO:0006333	chromatin assembly or disassembly	88	4	1.02	3.93	0.021	1.00
GO:0006338	chromatin remodeling	165	9	1.91	4.71	0.000	0.24
GO:0006323	DNA packaging	103	4	1.19	3.36	0.035	1.00
GO:0006334	nucleosome assembly	59	4	0.68	5.86	0.006	0.73
GO:0048646	anatomical structure formation involved in morphogenesis	593	14	6.86	2.04	0.011	0.97
GO:0022612	gland morphogenesis	36	4	0.42	9.6	0.001	0.43
GO:0035108	limb morphogenesis	115	4	1.33	3.01	0.048	1.00
GO:0007435	salivary gland morphogenesis	21	4	0.24	16.46	0.000	0.34
GO:0035107	appendage morphogenesis	116	4	1.34	2.98	0.049	1.00
GO:0007431	salivary gland development	23	4	0.27	15.03	0.000	0.29
GO:0099518	vesicle cytoskeletal trafficking	52	4	0.6	6.65	0.004	0.63
GO:0007268	chemical synaptic transmission	251	7	2.9	2.41	0.029	1.00
GO:0099536	synaptic signaling	286	10	3.31	3.02	0.002	0.52
GO:0099537	trans-synaptic signaling	266	8	3.08	2.6	0.014	1.00
GO:0008217	regulation of blood pressure	94	4	1.09	3.68	0.026	1.00
GO:1905954	positive regulation of lipid localization	69	4	0.8	5.01	0.010	0.91
GO:0032368	regulation of lipid transport	82	5	0.95	5.27	0.003	0.59
GO:0046649	lymphocyte activation	247	8	2.86	2.8	0.009	0.89
GO:0030097	hemopoiesis	408	13	4.72	2.75	0.001	0.44
GO:0002520	immune system development	470	13	5.44	2.39	0.004	0.63
GO:0061515	myeloid cell development	45	5	0.52	9.6	0.000	0.25
GO:0030183	B cell differentiation	51	5	0.59	8.47	0.000	0.31
GO:0030099	myeloid cell differentiation	157	7	1.82	3.85	0.003	0.57
GO:0042113	B cell activation	86	6	1	6.03	0.001	0.32
GO:0002521	leukocyte differentiation	244	7	2.82	2.48	0.026	1.00
GO:0030098	lymphocyte differentiation	172	6	1.99	3.01	0.017	1.00
GO:0045321	leukocyte activation	296	8	3.43	2.34	0.024	1.00
GO:1903131	mononuclear cell differentiation	195	6	2.26	2.66	0.028	1.00
GO:0048534	hematopoietic or lymphoid organ development	443	13	5.13	2.54	0.002	0.54
GO:0007018	microtubule-based movement	263	8	3.04	2.63	0.013	1.00
GO:0030048	actin filament-based movement	74	5	0.86	5.84	0.002	0.53
GO:0001667	ameboidal-type cell migration	143	5	1.65	3.02	0.028	1.00
GO:0010970	transport along microtubule	107	5	1.24	4.04	0.009	0.89
GO:0030036	actin cytoskeleton organization	458	11	5.3	2.08	0.024	1.00
GO:0099111	microtubule-based transport	131	5	1.52	3.3	0.020	1.00
GO:0050817	coagulation	75	4	0.87	4.61	0.013	1.00

GO ID	Function	*Gallus gallus*	Accelerated introns	expected	fold Enrichment	P-value	FDR
GO:1905952	regulation of lipid localization	104	5	1.2	4.15	0.008	0.86
GO:0070201	regulation of establishment of protein…	301	8	3.48	2.3	0.026	1.00
GO:0032386	regulation of intracellular transport	196	7	2.27	3.09	0.009	0.88
GO:0060341	regulation of cellular localization	501	13	5.8	2.24	0.009	0.88
GO:0051272	positive regulation of cellular component…	348	11	4.03	2.73	0.003	0.57
GO:0017157	regulation of exocytosis	125	5	1.45	3.46	0.017	1.00
GO:0051270	regulation of cellular component movement	645	15	7.46	2.01	0.013	1.00
GO:0060627	regulation of vesicle-mediated transport	303	8	3.51	2.28	0.027	1.00
GO:0051046	regulation of secretion	354	12	4.1	2.93	0.001	0.43
GO:0051049	regulation of transport	1037	24	12	2	0.001	0.45
GO:0046822	regulation of nucleocytoplasmic transport	75	4	0.87	4.61	0.013	1.00
GO:0051050	positive regulation of transport	481	15	5.57	2.69	0.001	0.30
GO:1903305	regulation of regulated secretory pathway	80	4	0.93	4.32	0.016	1.00
GO:1903532	positive regulation of secretion by cell	149	9	1.72	5.22	0.000	0.29
GO:0051047	positive regulation of secretion	155	9	1.79	5.02	0.000	0.31
GO:0046883	regulation of hormone secretion	144	5	1.67	3	0.029	1.00
GO:0033157	regulation of intracellular protein transport	126	5	1.46	3.43	0.017	1.00
GO:0045921	positive regulation of exocytosis	45	4	0.52	7.68	0.002	0.53
GO:1903530	regulation of secretion by cell	330	10	3.82	2.62	0.006	0.72
GO:0046887	positive regulation of hormone secretion	71	5	0.82	6.09	0.002	0.47
GO:0006974	cellular response to DNA damage stimulus	548	13	6.34	2.05	0.014	1.00
GO:0007596	blood coagulation	75	4	0.87	4.61	0.013	1.00
GO:0030433	ubiquitin-dependent ERAD pathway	65	4	0.75	5.32	0.008	0.86
GO:0042770	signal transduction in response to DNA damage	110	4	1.27	3.14	0.042	1.00
GO:0036503	ERAD pathway	76	4	0.88	4.55	0.013	1.00
GO:0007599	hemostasis	76	4	0.88	4.55	0.013	1.00
GO:0071824	protein-DNA complex subunit organization	135	6	1.56	3.84	0.006	0.72
GO:0071826	ribonucleoprotein complex subunit organization	165	5	1.91	2.62	0.046	1.00
GO:0034622	cellular protein-containing complex assembly	588	14	6.8	2.06	0.011	0.96
GO:0065004	protein-DNA complex assembly	109	4	1.26	3.17	0.041	1.00
GO:0022618	ribonucleoprotein complex assembly	160	5	1.85	2.7	0.041	1.00
GO:0034728	nucleosome organization	83	6	0.96	6.25	0.001	0.32
GO:0001775	cell activation	344	10	3.98	2.51	0.008	0.87
GO:0045667	regulation of osteoblast differentiation	95	4	1.1	3.64	0.027	1.00
GO:0002181	cytoplasmic translation	62	4	0.72	5.57	0.007	0.81
GO:0032940	secretion by cell	260	7	3.01	2.33	0.034	1.00
GO:0006887	exocytosis	147	5	1.7	2.94	0.031	1.00
GO:0030029	actin filament-based process	497	12	5.75	2.09	0.017	1.00
GO:0007267	cell-cell signaling	569	17	6.58	2.58	0.000	0.32
GO:0051338	regulation of transferase activity	617	16	7.14	2.24	0.003	0.60
GO:0051348	negative regulation of transferase activity	175	6	2.03	2.96	0.018	1.00
GO:0043549	regulation of kinase activity	541	13	6.26	2.08	0.013	1.00
GO:0071900	regulation of protein serine/threonine kinase	249	7	2.88	2.43	0.028	1.00
GO:0050878	regulation of body fluid levels	153	5	1.77	2.82	0.035	1.00
GO:0032784	regulation of DNA-templated transcription, elongation	39	5	0.45	11.08	0.000	0.32
GO:0008593	regulation of Notch signaling pathway	64	4	0.74	5.4	0.008	0.86
GO:0044057	regulation of system process	296	8	3.43	2.34	0.024	1.00
GO:0040017	positive regulation of locomotion	352	10	4.07	2.45	0.009	0.88
GO:2000147	positive regulation of cell motility	342	9	3.96	2.27	0.020	1.00
GO:0032879	regulation of localization	1725	40	19.96	2	0.000	0.17
GO:0035272	exocrine system development	29	4	0.34	11.92	0.001	0.34
GO:1903050	regulation of proteolysis involved in cellular protein…	158	5	1.83	2.73	0.040	1.00
GO:2000060	positive regulation of ubiquitin-dependent…	77	4	0.89	4.49	0.014	1.00
GO:1903364	positive regulation of cellular protein catabolic…	108	4	1.25	3.2	0.040	1.00
GO:1903052	positive regulation of proteolysis involved in…	93	4	1.08	3.72	0.025	1.00
GO:0007498	mesoderm development	81	4	0.94	4.27	0.016	1.00
GO:0061053	somite development	70	5	0.81	6.17	0.002	0.47
GO:0060485	mesenchyme development	188	6	2.18	2.76	0.024	1.00
GO:0043009	chordate embryonic development	260	7	3.01	2.33	0.034	1.00
GO:0016331	morphogenesis of embryonic epithelium	107	4	1.24	3.23	0.039	1.00
GO:0009792	embryo development ending in birth or egg…	277	7	3.21	2.18	0.045	1.00
GO:0001756	somitogenesis	55	4	0.64	6.28	0.005	0.68
GO:0035282	segmentation	73	4	0.84	4.73	0.012	1.00
GO:0034243	regulation of transcription elongation from RNA…	23	4	0.27	15.03	0.000	0.27
GO:0007229	integrin-mediated signaling pathway	63	4	0.73	5.49	0.007	0.82

GO ID	Function	*Gallus gallus*	Accelerated introns	expected	fold Enrichment	P-value	FDR
GO:0050684	regulation of mRNA processing	116	4	1.34	2.98	0.049	1.00
GO:0051053	negative regulation of DNA metabolic process	87	4	1.01	3.97	0.021	1.00
GO:0000018	regulation of DNA recombination	82	4	0.95	4.22	0.017	1.00
GO:0043543	protein acylation	135	5	1.56	3.2	0.023	1.00
GO:0018193	peptidyl-amino acid modification	693	19	8.02	2.37	0.001	0.34
GO:0017038	protein import	125	5	1.45	3.46	0.017	1.00
GO:0033365	protein localization to organelle	507	12	5.87	2.05	0.019	1.00
GO:0034504	protein localization to nucleus	141	7	1.63	4.29	0.002	0.46
GO:0045184	establishment of protein localization	870	21	10.07	2.09	0.002	0.46
GO:0006886	intracellular protein transport	541	14	6.26	2.24	0.006	0.77
GO:0072594	establishment of protein localization to organelle	241	8	2.79	2.87	0.008	0.86
GO:0015031	protein transport	804	19	9.3	2.04	0.004	0.61
GO:0030705	cytoskeleton-dependent intracellular transport	142	7	1.64	4.26	0.002	0.47
GO:0051650	establishment of vesicle localization	110	5	1.27	3.93	0.010	0.96
GO:0006913	nucleocytoplasmic transport	188	6	2.18	2.76	0.024	1.00
GO:0051169	nuclear transport	188	6	2.18	2.76	0.024	1.00
GO:0051648	vesicle localization	113	5	1.31	3.82	0.012	0.99
GO:0006607	NLS-bearing protein import into nucleus	14	4	0.16	24.69	0.000	0.20
GO:0006606	protein import into nucleus	84	5	0.97	5.14	0.004	0.61
GO:0051170	import into nucleus	87	5	1.01	4.97	0.004	0.66
GO:0098916	anterograde trans-synaptic signaling	251	7	2.9	2.41	0.029	1.00
GO:0051052	regulation of DNA metabolic process	242	7	2.8	2.5	0.025	1.00
GO:0006473	protein acetylation	99	4	1.15	3.49	0.031	1.00
GO:0010817	regulation of hormone levels	305	8	3.53	2.27	0.028	1.00
GO:0048872	homeostasis of number of cells	164	8	1.9	4.22	0.001	0.35
GO:0034101	erythrocyte homeostasis	69	5	0.8	6.26	0.002	0.47
GO:0006874	cellular calcium ion homeostasis	272	8	3.15	2.54	0.016	1.00
GO:0002262	myeloid cell homeostasis	85	5	0.98	5.08	0.004	0.63
GO:0072507	divalent inorganic cation homeostasis	306	9	3.54	2.54	0.011	0.96
GO:0006875	cellular metal ion homeostasis	352	9	4.07	2.21	0.024	1.00
GO:0030218	erythrocyte differentiation	61	5	0.71	7.08	0.001	0.39
GO:0055074	calcium ion homeostasis	281	8	3.25	2.46	0.018	1.00
GO:0072503	cellular divalent inorganic cation homeostasis	293	9	3.39	2.65	0.008	0.85

References

Castresana, J. 2000. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. – Mol Biol Evol 17: 540–552.

del Hoyo, J., Elliott, A., Sargatal, J., Christie, D. A. and de Juana, E. 2015. Handbook of the Birds of the World Alive. – Lynx Edicions.

Garcia-R, J. C., Gibb, G. C. and Trewick, S. A. 2014. Deep global evolutionary radiation in birds: Diversification and trait evolution in the cosmopolitan bird family Rallidae. – Molecular Phylogenetics and Evolution 81: 96–108.

Katoh, K., Misawa, K., Kuma, K. and Miyata, T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. – Nucleic Acids Research 30: 3059–3066.

Kowalczyk, A., Meyer, W. K., Partha, R., Mao, W., Clark, N. L. and Chikina, M. 2019. RERconverge: an R package for associating evolutionary rates with convergent traits. – Bioinformatics 35: 4815–4817.

Mi, H., Muruganujan, A., Ebert, D., Huang, X. and Thomas, P. D. 2019. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. – Nucleic Acids Research 47: D419–D426.

Sato, T., Yamanishi, Y., Kanehisa, M. and Toh, H. 2005. The inference of protein–protein interactions by co-evolutionary analysis is improved by excluding the information about the phylogenetic relationships. – Bioinformatics 21: 3482–3489.

Yang, Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. – Mol Biol Evol 24: 1586–1591.