Speciation dynamics of New Guinean birds using large scale museomics

Ingo Achim Müller

Speciation dynamics of New Guinean birds using large scale museomics

Ingo Achim Müller

Abstract

The overarching theme of this thesis was to investigate speciation histories and biogeographical patterns of New Guinean birds, by using whole genomes primarily extracted from natural history collections.

In **Chapter I**, I detected hidden diversity within the Lesser Melampitta, a species, which is widely distributed across different mountain ranges throughout New Guinea. In contrast, I found no strong differentiation between populations in the Greater Melampitta, a species characterised by a much more scattered distribution. I hypothesized that this unexpected pattern in the latter species may have been the result of a relatively recent collapse into its current fragmented distribution. Moreover, this chapter was instrumental in establishing and optimising bioinformatic workflows that I applied in the subsequent chapters. In **Chapter II** we investigated hybridisation patterns within a lineage of whistlers (*Pachycephala*). The study revealed a complex network of interactions between species that appears to be linked to geography. Species inhabiting islands and landmasses that were connected in the past during periods of lower sea levels were less differentiated and showed higher signals of introgression than species inhabiting more remote oceanic islands. **Chapter III** focussed on speciation dynamics within a genus of honeyeaters (*Melidectes*). The results support that most species within the genus have formed as a consequence of geographic isolation. However, two taxa with partially overlapping distributions in the central mountains of New Guinea, which are known to hybridise, exhibit a very complex genetic structure that does not follow the current species classification, as the hybridising taxa appear to be genetically indistinguishable. Yet, I recovered some genetic signals of past differentiation and consequently present this complex as a rare empirical example of ephemeral speciation. In addition, I investigated how future climate change may potentially result in new speciation events. For **Chapter IV**, I investigated the extent of reproductive isolation and hybridisation patterns in the Birds-of-paradise genus *Paradisaea* and discuss the placement of various species along the speciation continuum. Unlike species with more isolated distributions, such as those found on islands surrounding New Guinea, three species on New Guinea's mainland exhibit low genetic differentiation and a pattern of more or less unrestricted gene flow across much of the lowlands. An exception to this are the Lesser and Raggiana Birds-of-paradise that show marked genetic differentiation and the absence of ongoing gene flow. The population structure and differentiation of *Paradisaea* species on the New Guinean mainland thus closely resembles that of a ring species.

In the broader context, this thesis has demonstrated the value that historical DNA offers when studying speciation dynamics, especially for species and populations that are difficult to sample. Furthermore, it also highlights the impact of system-specific mechanisms which calls for caution when generalising conclusions within the field of speciation.

Keywords: *speciation, museomics, birds, phylogeography, biogeography, population genetics, hybridisation.*

Department of Zoology

Stockholm University, 106 91 Stockholm

SPECIATION DYNAMICS OF NEW GUINEAN BIRDS USING LARGE SCALE MUSEOMICS

Ingo Achim Müller

Speciation dynamics of New Guinean birds using large scale museomics

Ingo Achim Müller

"But the Raven still beguiling all my fancy into smiling,
Straight I wheeled a cushioned seat in front of bird, and bust and door;
Then, upon the velvet sinking, I betook myself to linking
Fancy unto fancy, thinking what this ominous bird of yore—
What this grim, ungainly, ghastly, gaunt, and ominous bird of yore
Meant in croaking 'Nevermore.'"
— **EDGAR ALLAN POE** (1845)

"No"
— **DAVID LYNCH** (2007)
upon being asked to elaborate on his species concept (probably)

The thesis is based on the following articles, which are referred to in the text by their Roman numerals:

I   **Müller, I.A.**, Thörn, F., Rajan, S., Ericson, P.G., Dumbacher, J.P., Maiah, G., Blom, M.P.K., Jønsson, K.A. & Irestedt, M. (2024). Species-specific dynamics may cause deviations from general biogeographical predictions–evidence from a population genomics study of a New Guinean endemic passerine bird family (*Melampittidae*). *Plos one*, 19(5), e0293715.

II   Irestedt, M., **Müller, I.A.**, Thörn, F., Joseph, L., Nylander, J., Guinet, B., van der Valk, T. & Jønsson, K. (2024). Reticulate and hybrid speciation is promoted by environmental instability in an Indo-Pacific species complex of whistlers (Aves: *Pachycephala*). *Submitted manuscript.*

III   **Müller, I.A.**, Thörn, F., Rajan, S., Olsen, R.A., Ericson, P.G., Peona, V., Smith, B.T., Maiah, G., Koane, B., Iova, B., Blom, M.P.K, Irestedt, M. & Jønsson, K.A. (2025). Ephemeral speciation in a New Guinean honeyeater complex (Aves: *Melidectes*). *Molecular Ecology*. e17760.

IV   **Müller, I.A.**, Waqar, A., Thörn, F., von Rintelen, T., Frahnert, S., Richter, K., Cracraft, J., Irestedt, M. & Blom, M.P.K. (2025). Speciation around a mountain chain: Hybridisation dynamics in a lek-mating Birds-of-paradise species complex (Paradisaea). *Manuscript.*

Candidate contributions to thesis articles*

	I	II	III	IV
Conceived the study	Significant	Minor	Significant	Significant
Designed the study	Significant	Minor	Substantial	Significant
Collected the data	Substantial	Minor	Substantial	Substantial
Analysed the data	Substantial	Significant	Substantial	Substantial
Manuscript preparation	Substantial	Significant	Substantial	Substantial

* Contribution Explanation
    Minor: contributed in some way, but contribution was limited.
    Significant: provided a significant contribution to the work.
    Substantial: took the lead role and performed the majority of the work.

In addition, during my doctoral studies, I have been involved as a co-author in the following articles, which are not included in this thesis (listed in chronological order):

Irestedt, M., Thörn, F., **Müller, I.A.**, Jønsson, K.A., Ericson, P.G. & Blom, M.P.K. (2022). A guide to avian museomics: Insights gained from resequencing hundreds of avian study skins. *Molecular Ecology Resources*, 22(7), pp.2672-2684.

Knief, U., **Müller, I.A.**, Stryjewski, K.F., Metzler, D., Sorenson, M.D. & Wolf, J.B. (2024). Evolution of chromosomal inversions across an avian radiation. *Molecular biology and evolution*, *41*(6), p.msae092.

Thörn, F., Soares, A.E., **Müller, I.A.**, Päckert, M., Frahnert, S., van Grouw, H., Kamminga, P., Peona, V., Suh, A., Blom, M.P.K. & Irestedt, M. (2024). Contemporary intergeneric hybridization and backcrossing among birds-of-paradise. *Evolution Letters*, *8*(5), pp.680-694.

Thörn, F., **Müller, I.A.**, Soares, A.E.R., Nagombi, E., Jønsson, K.A., Blom, M.P.K. & Irestedt, M. (2025). Frequent hybridisation between parapatric lekking bird-of-paradise species. *Molecular Ecology*. e17780.

Schnelle, A., Rollins, R.E., **Müller, I.A.**, Irestedt, M., Cecere, J.G., Sánchez Gutiérrez, J., Masero, J.A., Risch, M., Bouwhuis, S. & Liedvogel, M. (2025). Conservation implications through population structure analyses of historical and contemporary genomes in an endangered tern population. *Manuscript*

This doctoral thesis builds partly upon the author's licentiate thesis which he defended on 24th November 2023. Of the papers included in this thesis, Chapters I and III were presented at an earlier stage as part of the licentiate. The contributions from the licentiate thesis are as follows:

The thesis overview ("kappa"): Segment 1.3 was included in the licentiate; the section has been reviewed and updated. It contains about 50% of the licentiate thesis' content and references.

Chapter I: This chapter was included in the licentiate thesis; the manuscript has since undergone peer review and was published in Plos one after minor revisions.

Chapter III: This chapter was included as chapter II in the licentiate thesis; since then, the manuscript has undergone significant changes including modifications based on comments from reviewers and is now published in Molecular Ecology.

Table of Contents

Glossary

Coverage: The proportion of a genome covered by at least one read

Depth: The number of reads mapped at a specific position

DoC (depth-of-coverage): The <u>average</u> number of reads at any position in the genome

D_{xy}: Absolute pairwise divergence

F_{ST}: Fixation index, measure of differentiation between populations ranging from 0 (no differentiation) to 1 (complete differentiation)

H_0: Observed heterozygosity, percentage of heterozygous loci in an individual/population

hDNA: Historic DNA, DNA obtained from specimens stored at natural history collections that were sampled within the last 200 years

Hybridisation: Interbreeding of genetically distinct populations regardless of their taxonomic classification (Allendorf et al., 2001)

PC(A): Principal Component (Analysis)

(P)SMC: (Pairwise) Sequentially Markovian Coalescent

sCF: Site concordance factor. The fraction of sites within an alignment supporting a branch.

wCF: Window concordance factor. The fraction of window trees that support a branch in a specific phylogeny. Note: This is interchangeable with gene concordance factors in the *IQ-TREE 2* documentation, since our "gene trees" are based on window alignments and not necessarily coding regions

1. Introduction

1.1 Speciation dynamics

Speciation is a fundamental concept in evolutionary biology, providing the basic mechanism by which biodiversity is generated (Butlin et al., 2009). Historically considered a single, more or less instantaneous event, contemporary perceptions of speciation have shifted, leading to the predominant view of speciation as a gradual process along a continuum (Stankowski & Ravinet, 2021). Furthermore, the view of speciation as a strictly bifurcating process has been challenged recently. Speciation histories are now often described as network structures due to phases of divergence and secondary contact leading to hybridisation between lineages (Mallet et al., 2016; Posada & Crandall, 2001; Stull et al., 2023). In this thesis, I define speciation as a gradual process in which reproductive isolation becomes increasingly pronounced as more advanced stages along the speciation continuum are reached.

In order to comprehend the fine scale mechanisms that affect the speciation process, it is imperative to differentiate the processes that can result in speciation events. These so-called modes of speciation include allopatric, peripatric, parapatric and sympatric speciation (**Fig. 1**).
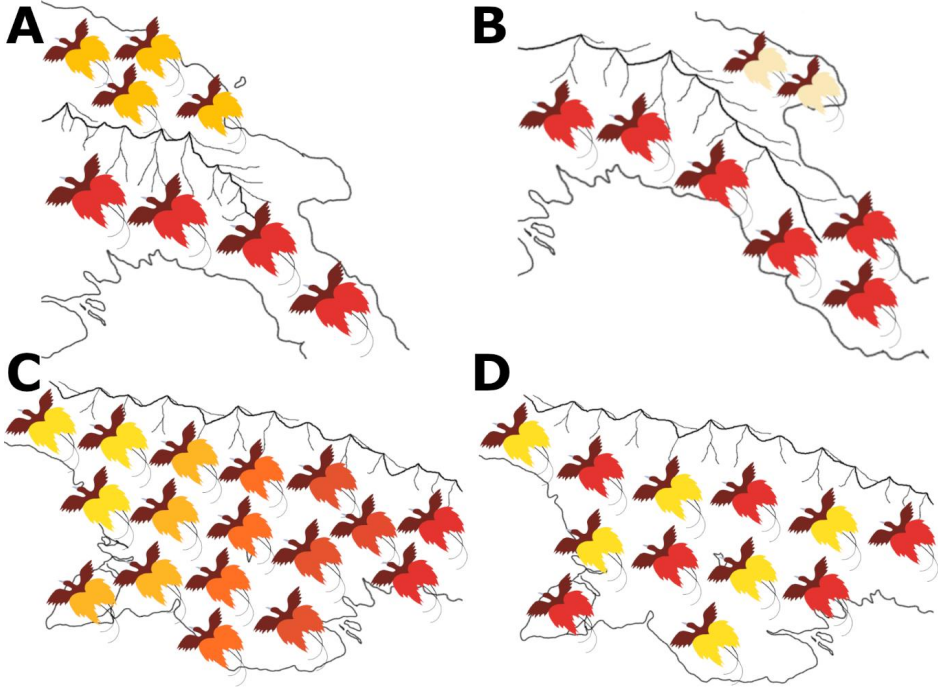
Figure 1. Schematic representation illustrating various modes of speciation. A) Allopatric speciation with complete geographic isolation. B) Peripatric speciation where a distinct species is formed in a smaller population at the periphery of the original population's distribution. C) Parapatric speciation characterised by gradual changes in traits across the distribution. Populations at the extreme ends represent distinct species. D) Sympatric speciation defined as the formation of new species within an entirely overlapping distribution. The map outline was drawn by Pilar Herrera-Egoavil.

Under allopatry (**Fig. 1A**), divergence between populations is driven by complete geographic isolation. This mode of speciation is widely regarded as the most prevalent in nature (Hernández-Hernández et al., 2021). It facilitates independent evolution within isolated populations without gene flow. Peripatric speciation (**Fig. 1B**) is a relatively similar process and often difficult to distinguish, but the isolated population represents a smaller group of individuals at the limits of the original metapopulation's distribution. Parapatric speciation (**Fig. 1C**) occurs within a continuous population that exhibits clinal differentiation along its distribution. A well-known example of parapatric speciation are ring species, where neighbouring populations on either side of a distributional barrier are able to interbreed freely, yet upon contact of terminal populations, reproductive isolation becomes evident. Empirical examples of ring species tend to be highly contentious, particularly in the context of applying stringent defining criteria (Kuchta & Wake, 2016). Finally, sympatric speciation (**Fig. 1D**) occurs when groups of individuals diverge within a shared distribution, in the absence of any apparent geographic barriers. The frequency with which this scenario results in speciation in nature remains a subject of considerable debate, particularly among birds, given the frequent gene flow between nascent species that can counteract the emergence of reproductive barriers (Bird et al., 2012; Phillimore et al., 2008). The concept of species sorting involves the development of divergent traits during periods of allopatry. However, if these differences are not sufficiently pronounced, the lineages cannot coexist upon secondary contact (Anderson & Matute, 2025). This dynamic may culminate in the fusion of these incipient species, or the extinction of either one or both lineages (Templeton, 1981).

When speciation is viewed as a continuous process of gradually increasing reproductive isolation, hybridisation is a very commonly involved process. In this thesis, I adhere to the definition of hybridisation as set by Allendorf et al. (2001), namely as the interbreeding of genetically distinct populations, irrespective of their taxonomic classification. This enables the utilisation of the term 'hybridisation' for interactions across the majority of stages on the speciation continuum, with the exception of earliest stages, where populations have not yet undergone differentiation. The precise manner in which hybridisation impacts speciation can exhibit significant variation, encompassing the potential for processes to either reverse, decelerate or accelerate speciation through a range of mechanisms (Abbott et al., 2013). Extensive hybridisation between lineages has been shown to lead the homogenisation of genomes and the breakdown of reproductive barriers (Coyne, 2007). This suggests the possibility of reversed speciation (Rosenblum et al., 2012). It can thus be hypothesised that less extensive hybridisation could slow down divergence processes. Conversely, accelerated speciation may be facilitated through the production of non-viable or sterile hybrid offspring, thereby reinforcing pre-mating isolation between different (incipient) species (Presgraves et al., 2003).

1.2 Challenges and advantages of using museum specimens for genomic studies

The application of museomics, genomic studies utilising DNA from historical specimens (hDNA) stored in natural history collections, represents a recently emerging field that offers exciting opportunities to study poorly understood organisms that are either extinct or challenging to sample today, e.g. due to low population sizes or distributions in remote areas

of the world (Fong et al., 2023; Lalueza-Fox, 2022). Museum collections are of great relevance to these organisms, as a considerable proportion of them are thought to have become extinct or endangered around the time that most historical collections were established (Wandeler et al., 2007). Research conducted on hDNA is often regarded as a discrete discipline, separate from that of ancient DNA (aDNA). This distinction is largely attributed to the significant difference in DNA quality. While hDNA is obtained from samples that have been actively collected (mostly within the past < 200 years) and preserved in natural history collections, aDNA is typically obtained from archaeological or fossilised specimens that have been collected long after the organism's death (Billerman & Walsh, 2019; Raxworthy & Smith, 2021).Whole genome sequencing of hDNA samples thus tends to be less challenging than for aDNA, as hDNA typically yields higher concentrations, less degradation and contains higher proportions of endogenous DNA.

The study of DNA from museum material has been ongoing for more than 35 years (Pääbo & Wilson, 1988; Thomas et al., 1990). However, the term 'museomics' is more recent (Fong et al., 2023; Schuster & Miller, 2008) and coincides with technological advances in sequencing strategies. These advances enabled the expansion of sequencing data from PCR fragments to complete mitochondrial (Guschanski et al., 2013) or nuclear genomes (van der Valk et al., 2019). This approach enables the assessment of genetic diversity and population dynamics in endangered groups, especially where individuals or populations have been lost due to declining populations or habitat reduction (Meineke et al., 2018; Wandeler et al., 2007). The utilisation of museomics to investigate extinct species has been demonstrated in a variety of eukaryotic taxa, including plants (Sotuyo et al., 2022; Zedane et al., 2016), sharks (Agne et al., 2022), trouts (Delling et al., 2023) and pigeons (Murray et al., 2017). Moreover, the broad temporal scope of museum collections offers a valuable opportunity to study population changes and genome evolution over time; temporal genomics (Cavill et al., 2024; Liu et al., 2025).

In addition, the relatively easy sampling process (compared to in situ collection) facilitates the implementation of dense sampling schemes, which are essential to the study of speciation dynamics at finer scales. Such finer-scale dynamics include, but are not limited to, population substructure, levels of introgression across distributions, and patterns of divergent selection. Investigating these processes thus extends the relevance of museomics beyond the field of conservation genetics into the study of broader evolutionary histories and speciation processes. For instance, the inclusion of gourd species (Cucurbitaceae) from global herbarium collections has enabled researchers to infer the biogeographic history and ancestral origin of this family (Schaefer et al., 2008). Museomic approaches have also facilitated the resolution of phylogenetic relationships and demographic histories of diverse organisms, including guenon primates (Guschanski et al., 2013), moths (Call et al., 2021), and Indo-Pacific passerine birds (Reeve et al., 2023). More recently, whole genome analyses of hDNA have revealed hybridisation events between distinct Birds-of-paradise genera (Blom et al., 2024; Thörn et al., 2024, 2025). Specifically, the focal region of this thesis, namely the island of New Guinea, with its wild terrain and unpredictable weather, poses significant challenges for field sampling. Consequently, conducting exhaustive genomic studies on organisms with extensive distributions is not feasible without utilising historical samples from museum collections.

However, working on hDNA does present a number of challenges. In comparison with DNA from contemporary samples, hDNA displays damage patterns characterised by deamination, stronger fragmentation and lower DNA concentrations (Dabney et al., 2013; Raxworthy &

Smith, 2021). Deamination, a process that predominantly results in C → U → T and G → A substitutions in hDNA, has been shown to lead to erroneous base-calls, which in turn introduce potential biases in downstream analyses. Deamination patterns have been shown to become more prevalent with increasing age of the sample (Raxworthy & Smith, 2021; Sawyer et al., 2012). It is evident that an increase in fragmentation leading to shorter read lengths consequently render genome assemblies more challenging and result in elevated costs per base pair in sequencing. Lower concentrations also increase the risk of cross-contamination, as well as contamination from modern or amplified DNA. Fortunately, a range of techniques have been developed both for laboratory protocols and bioinformatic processing steps with the aim of accommodating these challenges. The incorporation of USER enzyme during the preparation of libraries has proven to be effective in the reduction of deamination patterns in historical and ancient DNA (Briggs et al., 2010). The impact of reduced read length on the quality of assembly can be mitigated through the utilisation of sufficient depth-of-coverage during sequencing and by mapping against high-quality and phylogenetically close reference genomes. Contamination issues can be mitigated by working in dedicated laboratory facilities for historical samples, which have strict routines and regulations to minimise the risk of contamination. Additionally, bioinformatic approaches can be employed to check samples for potential contamination (Irestedt et al., 2022; Shapiro et al., 2019).

1.3 The avifauna of New Guinea

New Guinea is the largest tropical island and the second largest island in the world. Its diverse nature and complex topology with many different types of habitats offers a promising region for studies of speciation dynamics. Major mountain ranges extend from the northwestern Bird's Head Peninsula, through the large Central Mountain Range, down to the mountain range of the southeastern Papuan Peninsula (**Fig. 2**). In addition, several peripheral mountain ranges occur on the island, such as the Huon Peninsula, which is separated from the central mountains by a valley averaging 20 km in width (Davies, 2012).
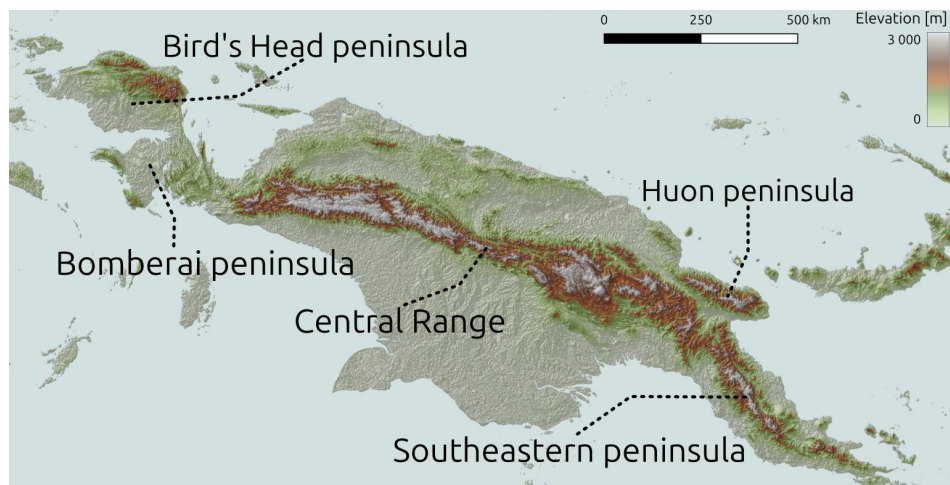

Figure 2. *Map of New Guinea and its surrounding islands. Marked regions reflect areas that are commonly referred to in the thesis.*

The avifauna of New Guinea is extremely diverse, with around 900 species (Lepage et al., 2014). This represents approximately 8% of the world's known bird species, within an island that covers roughly 0.1% of the world's land mass. Furthermore, the region's avian biodiversity is well documented, as most of the bird taxa present in New Guinea have probably already been described (Diamond & Bishop, 2023). It is not only the number of species or the high endemism that make New Guinea's birds special, but also their extraordinary diversity. From the iconic Birds-of-paradise, the toxic *Pitohui*, to the impressive cassowaries, New Guinea's birds represent a spectrum of evolutionary histories, ecological niches, behaviours, and cultural significance (Diamond & Bishop, 2023).

The island's isolation, varied topography, and exceptional avian diversity make it an ideal location for the study of bird evolution. Earlier biologists, such as Ernst Mayr and Jared Diamond, valued the island and have used the avifauna of New Guinea and its surrounding islands as examples when developing and testing theories and hypotheses about speciation and the build-up of biodiversity. Examples include Mayr's founder effect and his subsequent description of peripatric speciation (Mayr, 1954, 1999; Meyer, 2005), the development of the equilibrium model of island biogeography (MacArthur & Wilson, 1967), and Jared Diamond's testing of it (Diamond, 1973).

Despite the challenges of conducting fieldwork on New Guinea, expeditions undertaken during the last centuries have amassed a considerable collection of New Guinean birds in natural history collections. These collections now provide an invaluable resource for biological studies of the island's avifauna.

An overview of the arguments in favour of New Guinea as an ideal study region for evolutionary studies is given below:

Diverse habitats: The mountains of New Guinea are a mosaic of habitats. Ranging from tropical lowlands to arid highlands above 5000m a.s.l., the island provides a gradient of environmental conditions well suited for biodiversity research. The wide gradient of habitats within relatively short geographical spans allows for the study of species adaptation, migration, and evolution in different ecological niches.

Isolation and speciation: Its deep valleys, such as the one separating the Huon Peninsula from the Central Mountain Range, act as natural barriers that promote speciation. Such geographically isolated regions can harbour endemic lineages, providing valuable insights into the processes that drive species divergence in mountainous regions.

Climatic variation: The mountains of New Guinea exhibit a range of climatic conditions, from wet and humid lowlands to the colder climates of the highlands. This climatic variability over short geographical distances can be used to study the effects of climate on species distributions and adaptation.

Historical significance: New Guinea has attracted the attention of many generations of naturalists and biogeographers, including famous researchers such as Alfred Russel Wallace and Ernst Mayr. Expeditions over the years have collected a large number of avian specimens, providing a valuable resource for the study of the island's avian history.

2. Aims of the thesis

The overarching theme of my thesis was to use whole genomes of mainly New Guinean birds to study speciation histories, phylogeographic patterns and the processes that have shaped their current distributions. The projects constituting my thesis benefit from large datasets of whole genome sequences, mainly obtained from hDNA samples, dating from 1879 onwards.

In **Chapter I**, I investigated the population structure and demographic history of the family Melampittidae, which consists of two monotypic genera with different habitat connectivity. I expected to find a strong relationship between connectivity and genetic divergence, with more fragmented populations showing greater genetic divergence. This chapter also served to establish and optimise the bioinformatic workflows that I used in the following chapters.

In **Chapter II**, we examined hybridisation patterns within a lineage of whistlers (*Pachycephala*) inhabiting regions of varying degrees of environmental stability and isolation. We aimed to resolve phylogenetic relationships and investigate whether taxa inhabiting environmentally unstable regions show stronger signals of introgression than taxa inhabiting more stable oceanic islands.

Chapter III focussed on the speciation dynamics within a genus of honeyeaters (*Melidectes*), of which two species have been described to hybridise in nature. In addition to reconciling previously unresolved phylogenetic relationships, we also aimed to quantify the degree of hybridisation and differentiation between lineages. We expected to find evidence of allopatric speciation in isolated lineages and between taxa separated by altitude, but also increased signals of hybridisation in areas where taxa co-occur.

For **Chapter IV**, I examined patterns of hybridisation in the Bird-of-paradise genus *Paradisaea*. Based on morphological assessments, hybridisation occurs frequently in regions where mainland species meet. We aimed to achieve higher phylogeographic resolution of relationships within and between species, to assess the extent of hybridisation, and to estimate the degree of divergence between different lineages and populations.

3. Material & Methods

In this section I describe the material and methods that are shared between the chapters. Analyses that are more specific to individual projects are described in more detail within each chapter and the corresponding supplementary material.

3.1 DNA extraction and whole genome sequencing

DNA from modern samples (**Chapters I - III**) was extracted using a KingFisher Cell and Tissue DNA Kit, and a KingFisher Duo Prime Purification System (ThermoFisher Scientific) according to the manufacturer's protocol. Preparations of single libraries per individual were performed at the National Genomics Infrastructure (NGI) in Stockholm. For historical samples (**Chapters I - IV**), I extracted DNA from the toepads adhering to our laboratory's protocol for

DNA extraction from historical avian samples (Irestedt et al., 2022). Extraction was performed using a QIAamp DNA Micro Kit (Qiagen) following the manufacturer's recommendations, but with the addition of dithiothreitol (DTT) to improve ligation yields. Our protocol for library preparation as described in Irestedt et al. (2022) is a modification of the protocol of Meyer & Kircher (2010). As part of this modified protocol, I mitigated deamination patterns typical of historical and ancient DNA, particularly at the ends of DNA fragments, by incorporating USER enzyme during library preparation (Briggs et al., 2010). I prepared four libraries for each historical sample to increase library complexity. Each library carried unique indexing primer pairs to allow for pooled sequencing. I processed up to 96 libraries per sequencing batch (i.e. up to 24 individuals per batch). Sequencing was conducted at the National Genomics Infrastructure (NGI) in Stockholm either using an Illumina NovaSeq 6000 platform running 200 cycles (2 × 100 bp) on single lanes of S4 flow cells (**Chapters I - IV**) or using a NovaSeq X Plus platform running 300 cycles (2 × 150 bp) on 25B flow cells (**Chapter IV**).

3.2 Data processing

I processed obtained reads through the reproducible *Nextflow* (Di Tommaso et al., 2017) workflow *nf-polish* (https://github.com/MozesBlom/nf-polish). This workflow performs deduplication, adapter and quality trimming, merging of paired reads, and removal of low-complexity reads where a single nucleotide constitutes >50% of a sequence. Following this step, I mapped both paired and unpaired reads to the corresponding project's reference genome, which represents an outgroup for all included species. I performed the read mapping using *nf-µmap* (https://github.com/IngoMue/nf-umap), a pipeline that I developed as part of this thesis. I used its default mapping algorithm *bwa-mem2* (Vasimuddin et al., 2019) and generated reports on mapping statistics using *Qualimap 2* (Okonechnikov et al., 2016) and characteristic hDNA/aDNA damage patterns using *DamageProfiler* (Neukamm et al., 2021) as part of the workflow. *bwa-aln* is the preferred algorithm for aDNA studies (Dolenz et al., 2024). However, as our hDNA reads have an average length of just over 100 bp, for which *bwa-mem* is reported to perform best (Li, 2013), I decided to use the latter algorithm (Li, 2013). *bwa-mem2* uses the same algorithm as *bwa-mem* and produces identical alignments, but with much faster performance (Vasimuddin et al., 2019).

3.3 Population structure

I assessed population structure through *nf-GL_popstructure* (https://github.com/FilipThorn/nf-GL_popstructure, **Chapters III+IV**) using a genotype likelihood-based approach as implemented in *ANGSD*, which is well suited for data with low depth-of-coverage (DoC) (Korneliussen et al., 2014). In **Chapters I, III+IV**, I performed principal component analyses (PCA) using *PCAngsd* (Meisner & Albrechtsen, 2018) and admixture analyses using *NGSAdmix* (Skotte et al., 2013). I performed these analyses for the entire datasets, but also for subsets of the data to focus on specific taxa. I also estimated admixture proportions in ten replicates for each value of ancestral populations (K). I transformed covariance matrices and visualised PCAs using *Rstudio* (Posit team, 2023; R Core Team, 2021) and the *tidyverse* package (Wickham et al., 2019). I plotted admixture results using *RStudio*, *pong* (Behr et al., 2016)*, StructuRly* (Criscuolo & Angelini, 2020) and *QGIS* (QGIS Development Team, 2025).

A major challenge in admixture analysis is to determine optimal values for K (Garcia-Erill & Albrechtsen, 2020). I applied various approaches, changes in log-likelihood (Evanno et al., 2005), assessing convergence between replicates with *pong*, or, for hard-called variants (**Chapters II+IV**), using cross-validation errors estimated by *ADMIXTURE* (Alexander et al., 2009). However, results from these approaches varied considerably and were largely inconsistent, not only between approaches but also between subsets, often suggesting best values for K that were unlikely given the biogeography of the systems. Another problem with most approaches, apart from cross-validation errors, is the fact that K = 1 cannot be tested as the most likely scenario. I argue that the most reliable method for determining the best K is obtained through cross-validation errors, as it is also the most elaborate approach. Unfortunately, this method cannot be applied on admixture analyses through the genotype likelihood-based *NGSAdmix*. I therefore advise for caution when considering admixture results and strongly recommend comparing several values of K around the proposed optimum with results obtained from additional analyses as well as biogeographical expectations.

3.4 Phylogenomic inference

3.4.1 Mitochondrial phylogenetic reconstruction

Since most of the reference genomes in my chapters lacked mitochondrial assemblies, I obtained mitochondrial genomes from polished reads using *nf_mito-mania* (https://github.com/FilipThorn/nf_mito-mania). In this workflow, I assembled mitochondrial backbones using *MITObim* (Hahn et al., 2013) with the mitochondrial assembly of a closely related species as a starting seed. During the implemented variant calling, sites that fall below a depth-of-coverage (DoC) of 20*x* or with more than three times the average mitochondrial DoC of an individual are removed. Finally, the pipeline generates mitochondrial consensus sequences for each individual.

As part of the variant calling, the pipeline also forces diploid variant calls of the mitochondria which can be levied to identify cases of potential cross-contamination. As mitochondria are haploid, an excessive number of heterozygous positions could indicate contamination between samples. Since I would expect at least a small number of differences between mitochondrial sequences within an organism, the initial choice of a threshold (> 15 sites) was rather arbitrary, but certainly on the conservative side. Furthermore, the presence of nuclear mitochondrial DNA segments (NUMTs) could introduce additional heterozygous positions into these assemblies. I therefore decided to manually inspect the distribution of heterozygous positions across each mitochondrial genome when an individual library showed more than 15 heterozygous positions. Groups of heterozygous positions that appear to be restricted to the same region across individuals most likely represent NUMTs and therefore do not reflect cross-contamination. However, if I found a large number of heterozygous positions (> 50) distributed more or less randomly across the mitochondrial sequence, I decided to exclude that library or individual from further analysis.

Next, I used *MAFFT* (Katoh & Standley, 2013) to align each consensus sequence and an appropriate outgroup genome. A potential problem with *MITObim* is that it occasionally produces mitochondrial assemblies that are much longer than expected, as reads from the beginning of an assembly may also be added to the end due to the circularity of mitochondria.

Therefore, I trimmed overhangs in certain individuals by trimming the alignments to the length of the selected reference sequence using *seqtk* (Li, 2012/2024). I also manually confirmed these trimmed sites by examining the alignments using *AliView* (Larsson, 2014).

I determined optimal substitution models for each trimmed alignment using *Modeltest-NG* (Darriba et al., 2020), which typically resulted in a GTR+I+G4 model. Using the best substitution model, I then generated mitochondrial phylogenies using *RAxML-NG* (Kozlov et al., 2019), including bootstrap support estimates based on 100 iterations and implementing ten random initial parsimony trees.

Finally, I visualised the best maximum likelihood phylogeny using *RStudio*, and the *tidyverse* and *ggtree* packages (Yu et al., 2017).

3.4.2 Nuclear phylogenetic reconstruction

To generate nuclear phylogenies, I first performed individual variant calling using the *nf-var* workflow (https://github.com/MozesBlom/nf-var, **Chapters III+IV**), using only individuals with a minimum depth-of-coverage (DoC) greater than 3*x*. *Freebayes* (Garrison & Marth, 2012) was used as variant caller, as I also conducted joint variant calls in **Chapters II-IV**. The advantage of *freebayes* over alternative variant callers is that it allows the use of prior information, such as population/species identity. It has also been shown to yield a higher number of single nucleotide polymorphisms (SNPs) in low to medium DoC samples (Stegemiller et al., 2023). As part of the workflow, I removed loci with an allelic balance between 0 and 0.2, decomposed multiple nucleotide polymorphisms (MNPs) into SNPs, removed indel variation, excluded heterozygous sites due to their low confidence from the relatively low DoC in my samples, and finally implemented minimum and maximum DoC cut-offs. Based on this filtering, the pipeline also calculates the percentage of missing data for each individual. I excluded individuals with more than 40% missing data from phylogenetic reconstructions. In its final step, the pipeline uses variant calls to produce consensus sequences for each individual and scaffold or chromosome.

I used these consensus sequences as input to *nf-phylo* (https://github.com/MozesBlom/nf-phylo, **Chapters I, III+IV**), which can generate different phylogenies for different window sizes across the genome. Through this workflow, I obtained concatenated phylogenies for the autosomes and the Z chromosome using *IQ-TREE 2* (Minh et al., 2020), as well as summary coalescent phylogenies for all autosomes using *ASTRAL-III* (Zhang et al., 2018). More specifically, after generating alignments, the pipeline divides these alignments into windows of specified size, which are sampled at 100 kb intervals. I filtered each window according to the same criteria: I discarded windows if less than half of the individuals within an alignment were represented in less than 50% of all sites, and if less than 80% of the individuals had more than 40% missing data. If a window did not pass these filters, an adjacent window was checked until a window was found that met the criteria. Based on these filtered window alignments, the pipeline inferred maximum likelihood (ML) phylogenies using *IQ-TREE 2* implementing *ModelFinder* (Kalyaanamoorthy et al., 2017) to find optimal substitution models for each window. Next, summary-coalescent phylogenies were generated through *ASTRAL-III* using autosomal window trees, and a sex chromosome phylogeny was generated through *IQ-TREE 2* using window trees from chromosome Z. The pipeline also uses the autosomal window trees

to generate a concatenated autosomal alignment, which is used to infer concatenated phylogenies through *IQ-TREE 2*. As a note, summary-coalescent models have been shown to reflect the species tree more accurately than concatenated phylogenies (Jiang et al., 2020), but a comparison of the two approaches remains useful as branch lengths within concatenated trees are a more direct measure of sequence divergence, although not normalised by effective population size or generation times (Degnan & Rosenberg, 2009). Finally, as part of the workflow, I calculated site and window concordance factors (sCF, wCF/gCF) for all trees using *IQ-TREE 2*. Concordance factors provide a much better assessment of node support than bootstraps, which tend to always show high support for whole-genome data (Minh et al., 2020).

I visualised nuclear phylogenies using *RStudio* with the *tidyverse*, *ggtree* and *phytools* packages (Revell, 2012).

3.5 Genetic divergence and differentiation

In **Chapters I+III**, I obtained estimates of observed heterozygosity (H_0) for each individual using *nf-Hestu* (https://github.com/IngoMue/nf-Hestu). This workflow also uses a genotype likelihood approach via *ANGSD* and generates individual site frequency spectra (SFS) from which H_0 can be estimated directly. In addition to filtering for quality, the pipeline also removes sites below one third and above twice the average DoC of an individual. A clear pattern that I observed in these estimates was the strong positive correlation between H_0 and DoC, which complicates the interpretation of these results. The interpretation of differences in H_0 is confirmed by evaluating heterozygosities of individuals at similar DoCs (as shown in the supplementary figures to **Chapters I+III**). An optimal approach would have been to downsample genomes with higher DoCs, but due to the wide range of DoCs in my samples and the low minimum DoC of a few single individuals, I would have had to discard a large majority of reads to achieve similarly low DoC levels.

I have assessed genetic differentiation between populations, specifically through the fixation index (F_{ST}) in **Chapters III+IV**, but with different approaches. The estimates are therefore only comparable within each chapter, but not between projects. In **Chapter III**, I estimated F_{ST} for different species and population pairs through *ANGSD*. Filtering steps were similar to those implemented for H_0 estimates and are described in detail in the supplement to **Chapter III**. F_{ST} was calculated genome-wide using the weighted Hudson's estimator from (Bhatia et al., 2013) as well as through overlapping 100 kb windows with a step size of 20 kb across the genome. In **Chapter IV**, I relied on hard-called variants to estimate F_{ST} using *vcftools* (Danecek et al., 2011). This tool uses Weir and Cockerham's weighted F_{ST} and was calculated both as a genome-wide estimate and as a window-based estimate with the same window and step size as in **Chapter III** (Weir & Cockerham, 1984). I plotted the results of both approaches using *RStudio* and the *tidyverse* package.

4. Results & Discussion

4.1 Distribution patterns do not reflect patterns of genetic diversity in Melampittidae

Chapter I focussed on the family Melampittidae, which consists of two species, *Melampitta lugubris* (Lesser Melampitta) and *Megalampitta gigantea* (Greater Melampitta), separated into two monotypic genera. To varying degrees, both these species have fragmented, partly discontinuous distributions throughout New Guinea's mountain ranges and a number of peripheral ranges. *M. lugubris* is relatively widespread and is common at higher elevations throughout most of the Central Mountain Range. The more elusive *M. gigantaea* has a much more fragmented distribution at lower elevations and is known from only six localities that are widely scattered across the entire island (**Fig. 3A**). *M. gigantaea* is also described as having limited dispersal capabilities and exhibits a strong habitat preference for areas of karst limestone and sinkholes, where it has been described to construct its nests (Diamond, 1983).
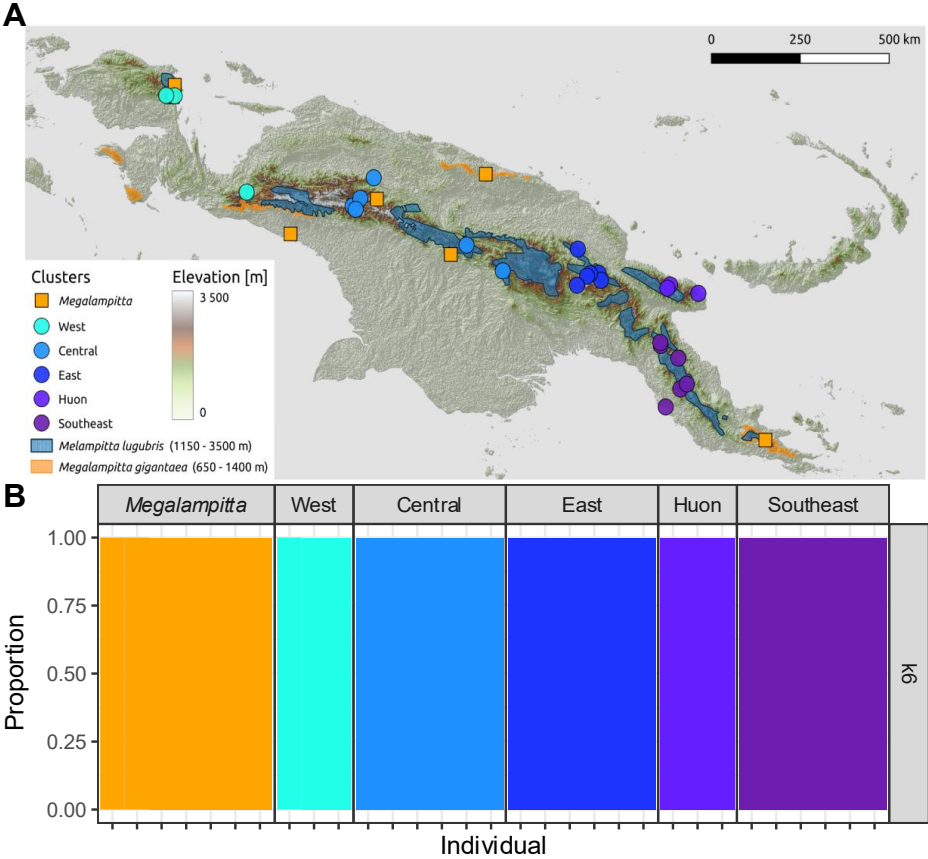

Figure 3. (Caption on the next page)

Figure 3. (previous page) A) Map showing sampling sites of each individual and the distribution of each species (M. gigantea in orange, M. lugubris in blue). B) Admixture proportions for K = 6, showing geographic subdivisions in M. lugubris, but no substructure within M. gigantaea. Figures taken from Chapter I.

To study these two bird species, I re-sequenced the genomes of the seven known *M. gigantaea* specimens stored in natural history collections, as well as 24 *M. lugubris* specimens spanning their entire distribution. By comparing the population structure of these two species, I investigated to what extent habitat specialisation, as exhibited by *M. gigantaea*, affects population connectivity. Phylogenetic reconstruction was broadly consistent between mitochondrial and nuclear data (for both concatenated and summary-coalescent trees) and showed three major clades with relatively deep divisions within *M. lugubris*, corresponding to previously described subspecies in distinct geographic regions. In contrast, branch lengths within *M. gigantaea* were much shorter between individuals, reflecting lower genetic diversity within this species. Population genomic analyses (**Fig. 3B**), as well as acoustic variation, were consistent with the phylogenetic signal supporting a single *M. gigantaea* population, whereas *M. lugubris* showed more pronounced population structure between distinct geographic regions. Estimates of absolute sequence divergence (D_{xy}) were also consistent with this observation, as estimates between the most geographically distant *M. gigantaea* individuals were between 4-10 times lower than for any comparison within *M. lugubris*. The divergence time estimates of the major *M. lugubris* clades ranged from 3-5 mya and are consistent with estimates of major mountain uplift in New Guinea (Davies, 2012; Hall, 2002; Pigram & Davies, 1987). Pairwise Sequentially Markovian Coalescent (PSMC) estimates of changes in effective population sizes through time differed between the identified clades but were shared between individuals from the same cluster. In particular, *M. gigantaea* appears to have undergone a steady population decline from ~200 Kya to ~40 Kya. I propose that a relatively recent collapse of *M. gigantaea* into its fragmented habitats is the most likely explanation of the observed unexpectedly low diversity and lack of population structure. This suggests that the different populations may not have been isolated for long enough to develop significant genetic differences. At the same time, the deep divergence and lack of admixture within *M. lugubris* are at a level that I consider sufficient for treating the major clades as three distinct species. Finally, I discuss these results within the framework of taxon cycles, in which lineages move to higher elevations with increasing age to face less competition from young species colonising lower altitudes (Ricklefs & Bermingham, 2002; Wilson, 1961). Instead of moving to higher elevations, *M. gigantaea* may have moved to karst habitats, an alternative type of suboptimal habitat that may serve as a refuge for older lineages (Erwin, 1981). In summary, this chapter has provided insights into the evolutionary histories that have shaped the current distribution of the family Melampittidae. The results show that population structures of organisms need to be studied carefully, as they may not follow general predictions. The study also serves to illustrate the importance of natural history collections for studies of evolutionary histories of poorly understood and rare species.

4.2 Reticulate relationships of *Pachycephala* driven by habitat connectivity

In **Chapter II** we examined a young monophyletic lineage of whistlers (*Pachycephala*). We generated a genomic dataset of 44 individuals representing 7 described species and 21 subspecies. Taxa within this clade have colonised different landmasses and islands which differ in their environmental stability. While some taxa within this clade are endemic to remote oceanic islands, others are distributed on islands or landmasses on the Sahul continent that have been repeatedly connected to each other during Pleistocene periods of lower sea levels (**Fig. 4A**) (Jønsson et al., 2014). The aim of this project was to investigate the phylogenetic relationships and spatio-temporal hybridisation patterns of this lineage.

Overall, the results revealed a complex network of species interactions that appear to be linked to environmental instability. First, we observed strong mitonuclear incongruence among our phylogenetic results (**Fig. 4B+C**). While our autosomal phylogeny was mostly consistent with current taxonomic classifications, nodes generally showed low support (**Fig. 4C**). The mitochondrial data, on the other hand, had well supported nodes, but the topology differed significantly from the autosomal phylogeny and did not cluster individuals belonging to recognised species in monophyletic groups (**Fig. 4B**). In general, we recovered low congruence between more than 3000 phylogenies constructed from distinct autosomal windows. Despite exhibiting distinct plumages (**Fig. 4**), our introgression analyses revealed extensive signals of hybridisation between different taxa within this recent radiation. We found a strong relationship between geographic proximity and levels of hybridisation, but also that species inhabiting the environmentally unstable Sahul region show particularly high signals of introgression and lower differentiation. However, we found one Sahul species that does not seem to hybridise as extensively as other Sahul taxa. We speculate that this may be due to its adaptation to the distinctly unique ecological niche it occupies. In addition, we explore the hybrid origin and potential hybrid speciation in a lineage inhabiting an island at the intersection of two oceanic radiations. We also discuss the limitations of asymmetric tests for introgression, such as D-statistics, when studying complex systems with multidirectional gene flow that is largely symmetric. Overall, our results support a growing body of literature suggesting that reticulate speciation is more common than previously thought. This has implications for our understanding of how species arise and persist through time.

Figure 4. A) *Sampling map for all individuals using different colours for each species and different shapes for different subspecies. Green areas around current landmasses show the extent of land bridges during glacial maxima at 140 metres below present sea level. Phylogenetic trees for mitochondrial (**B**) and autosomal (**C**) genomes showing high discordance. Support values in **B**) reflect bootstrap values. Support values in **C**) show ultrafast bootstraps/window concordance factors/site concordance factors. Figures taken from **Chapter II**.*

18

4.3 *Melidectes*, a rare example of ephemeral speciation

In **Chapter III**, I investigated speciation dynamics and hybridisation in the New Guinean honeyeater genus *Melidectes*. Most of the six described species of *Melidectes* are distributed allopatrically across various mountain ranges or at different altitudes. However, two species co-occur in the Central Mountain Range, and hybridisation between them has been described (Mayr & Gilliard, 1952, 1954). Despite the occurrence of hybridisation, plumage differentiation and altitudinal adaptations have been hypothesised to be sufficient in maintaining the species boundary between these two taxa (Mayr & Gilliard, 1952). I analysed re-sequenced genomes of 124 individuals of *Melidectes*, including a de novo assembly of *M. torquatus*, in order to study the modes of speciation that have shaped the group's current diversity, hybridisation and potential for future speciation due to climate change.

In agreement with my expectations, allopatric isolation appears to be the main driver of speciation within *Melidectes*. I also found lower genetic diversity and higher differentiation in species with smaller and more isolated distributions. However, contrary to previous hypotheses, my results suggest that *M. belfordi* and *M. rufocrissalis*, which were thought to be two separate species that hybridise, in fact form a single genetic unit despite phenotypic differences. Specifically, admixture proportions follow a geographic cline, with populations at the extreme ends of the distributions being the most differentiated and exhibiting the lowest levels of admixture (**Fig. 5B**).

These gradual admixture proportions and higher levels of differentiation between terminal populations assigned to "*M. belfordi*" than between populations assigned to "*M. belfordi*" and "*M. rufocrissalis*" (**Fig. 5C**) do not support that this complex currently consists of two distinct species. Nevertheless, I recovered signals of past differentiation within this group from mitochondrial data and F_{ST} scans (**Fig. 5A+C**). The mitochondrial phylogeny revealed relationships more consistent with the current taxonomy, showing two deeply divergent clades composed largely, but not exclusively, of individuals from specimens assigned to "*M. belfordi*" and "*M. rufocrissalis*", respectively (**Fig. 5A**). Together with the observation that *M. belfordi* individuals outside the overlapping distribution show less nuclear differentiation from individuals identified as *M. belfordi* within the overlapping range than from *M. rufocrissalis* with the same distribution (**Fig. 5C**), I present this system as a rare empirical example of ephemeral speciation. In essence, our findings are consistent with a scenario where secondary contact between two incipient species has led to widespread hybridisation and a subsequent fusion of lineages (Rosenblum et al., 2012). Finally, by correlating genetic variation with environmental factors and future climate predictions, I discuss the potential for future speciation events within the *M. belfordi/rufocrissalis* complex.
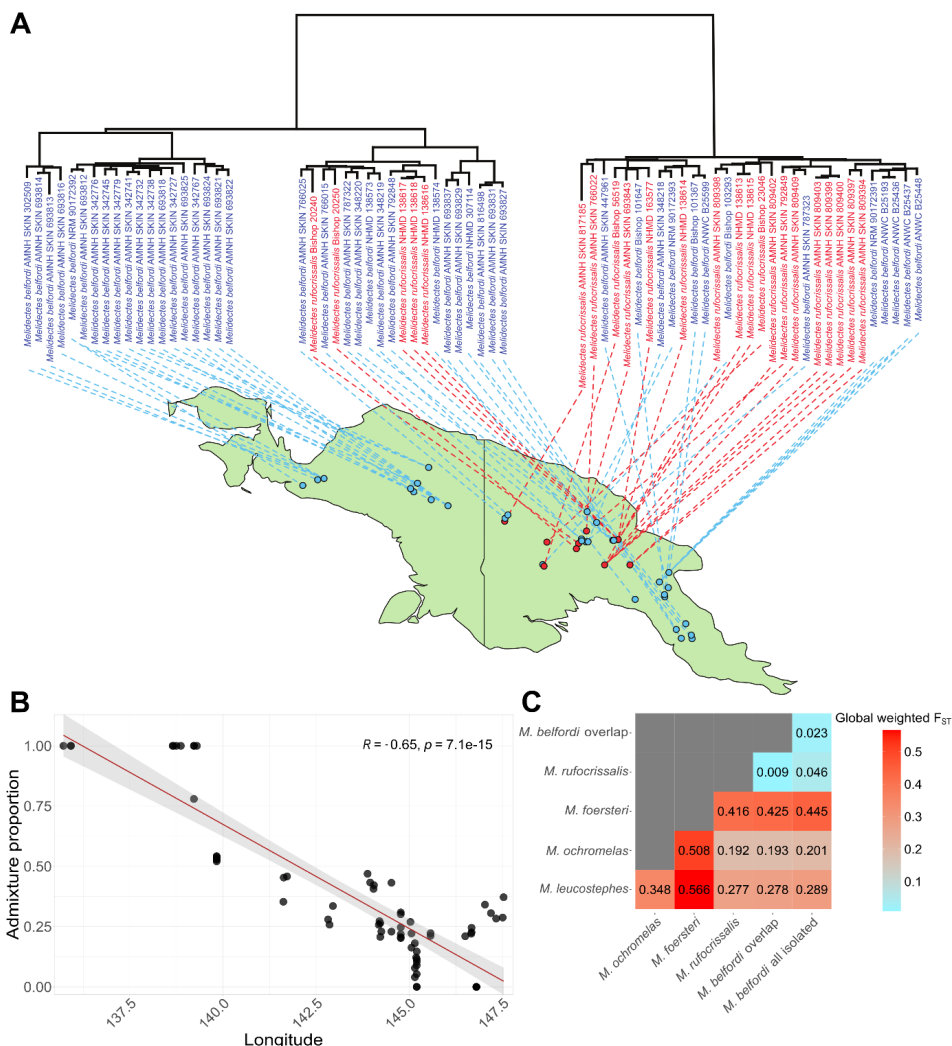
Figure 5. A) *A subset of mitochondrial phylogenetic relationships focussing on M. belfordi (blue) and M. rufocrissalis (red). Individuals are projected onto a map to show their sampling site.* **B)** *Admixture proportions along longitude for M. belfordi and M. rufocrissalis individuals at K=2.* **C)** *Global weighted F_{ST} estimates for M. belfordi and M. rufocrissalis and basal sister species. M. belfordi is divided into populations where only M. belfordi occurs ('all isolated') and where M. belfordi occurs together with M. rufocrissalis ('overlap'). Figure taken from* **Chapter III**.

4.4 A ring-like speciation pattern in *Paradisaea*

In **Chapter IV**, I surveyed the evolutionary diversity in distinct lineages and patterns of hybridisation in the lek-mating Birds-of-paradise genus *Paradisaea* and discuss the placement of species/populations along the speciation continuum. Strong sexual selection in Birds-of-paradise has led to the evolution of exceptionally rich diversity in plumages and mating rituals. As such traits should act as pre-zygotic barriers to gene flow, it is fascinating that hybridisation is surprisingly common within this family (Blom et al., 2024; Thörn et al., 2024, 2025). I re-sequenced genomes of 108 historical specimens to specifically assess phylogeographic relationships, levels of inter- and intraspecific hybridisation, and genetic differentiation within the genus *Paradisaea* and its sister genus *Paradisornis*.
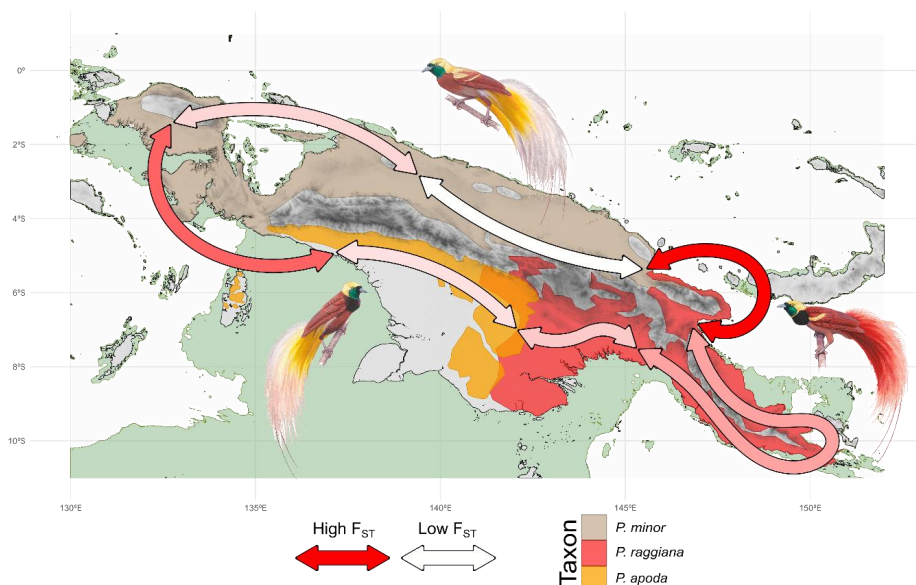

Figure 6. *Simplified schematic of pairwise global weighted F_{ST} estimates between populations of P. minor (beige), P. apoda (orange) and P. raggiana (red). Shaded areas indicate the distribution of each species according to the IUCN Red List. Higher F_{ST} values are indicated by stronger red arrows while low F_{ST} values are indicated by white arrows. Dark green areas surrounding landmasses (grey) represent previously connected regions during periods of lowest estimated sea level in the last 1 million years. The bird illustrations were drawn by Szabolcs Kókay. Figure taken from* ***Chapter IV****.*

Overall, I found that species within this group have reached different stages along the speciation continuum. I found strong differentiation among species with allopatric distributions, but lesser degrees of differentiation and high levels of admixture among species with at least partially overlapping distributions, despite apparent differences in plumage characters. In particular, the widespread mainland taxa *P. minor*, *P. apoda*, and *P. raggiana* show an almost perfectly clinal ancestry along a geographical gradient and clear patterns of isolation by distance. These results challenge current species boundaries within the group. However, despite the low differentiation between most neighbouring populations (**Fig. 6**), I detected stronger differentiation and no signal of hybridisation between *P. minor* and *P. raggiana* across their eastern contact zone. Furthermore, in this comparison, I also revealed

strong peaks of differentiation within a gene related to carotenoid production and perception. I discuss that the observed patterns of differentiation within this complex are reminiscent of the differentiation pattern expected in a ring species. However, in the absence of information on whether gene flow has truly been continuous along the distribution ring and across time, I am currently reluctant to make a definitive statement on whether this system can be considered a true ring species. Nonetheless, the study is an excellent example that highlights how biogeography and sexual selection can jointly drive divergence and may eventually promote speciation.

In addition, I genetically verified that female individuals within this group are more migratory, supporting previous hypotheses about how hybridisation between different species with unique lek-mating behaviours may be female-driven and act as a mechanism to maintain genetic diversity (Greenwood, 1980).

5. Concluding remarks

My thesis demonstrates the value of historical museum collections for genomic studies of speciation dynamics, especially for species and populations that are difficult to sample in the wild. Across my four chapters, I was able to characterise a range of speciation dynamics, ranging from an unexpected lack of any detectable divergence (**Chapter I**), a complicated reticulated history of divergence, triggered by secondary contact due to environmental instability (**Chapter II**), a case of ephemeral speciation where hybridisation has homogenised initially divergent lineages (**Chapter III**), to the description of different levels of reproductive isolation within a single genus and a potential case of ring speciation (**Chapter IV**). Overall, I have obtained results that were highly unexpected based on previous species descriptions and biogeographic predictions. I thus would like to emphasise the importance of system-specific mechanisms and call for caution when generalising conclusions within the field of speciation.

Thanks to the unprecedentedly large whole-genome datasets from historical samples, I was able to detect patterns of fine-scale admixture, species interactions and their phylogeographic histories. These patterns would have been easily missed with less dense sampling schemes and would not have provided such complete pictures of speciation histories within the taxa studied.

Nevertheless, my thesis raises a number of open questions that deserve further investigation. In many cases, the patterns observed contradict morphological assessments and the ecology of the birds studied. Reassessments of morphological variation based on genetically distinct populations may reveal significant differences in traits that may play a more important isolating role than previously considered. In many cases, descriptions of morphological variation within taxa, observations of mating events, vocal recordings or assessments of migratory behaviour are also sorely lacking, which would hopefully add further support and clarity to my genetic results. But even the genetic results could be further enhanced by improved sampling as there are still regions of interest from which we currently lack individuals. One example is the putative contact zone between *Paradisaea minor* and *Paradisaea apoda*. Samples from this region could provide a better understanding of the extent of gene flow between these taxa and could thus allow us to make a more conclusive statement as to whether this complex truly represents a rare case of ring speciation. Similarly, the striking lack of genetic differences between *Megalampitta gigantaea* populations could be further explored by sampling at a denser population level.

This also brings me to the limitations of museum collections; however extensive they may be, sampling localities will still be limited to areas that have been accessible or of interest to previous expeditions, emphasising the necessity of further fieldwork. However, beyond these limitations, museomic data has the potential for further improvement by leveraging genomic techniques.

The main limitation of museomic data remains the short fragment lengths and relatively low depth-of-coverage which impose a substantial constraint on the range of downstream analyses that are possible with this particular type of data. Despite the currently increased sequencing cost, obtaining higher and more consistent depth-of-coverage from historical samples allows for greater confidence in the identification of variant sites, especially heterozygous positions. These are essential for a number of popular analyses, including coalescent-based approaches to demographic modelling (e.g. SMC [Sequentially Markovian Coalescent] methods) and assessments of genetic diversity in conservation research (e.g. global heterozygosity estimates, runs of homozygosity). Additional laboratory procedures, such as size selection, which could yield longer sequencing reads, would facilitate the recovery of regions that are difficult to assemble, such as repetitive regions. These regions often contain transposable elements or structural variants which are becoming increasingly relevant in studying genome evolution. The availability of higher quality reference genomes, in particular pangeomes which capture the genetic variation present within studied groups, holds considerable potential in facilitating the identification and evaluation of the evolutionary consequences of such structural variation. These outlined improvements, when implemented collectively, can also enable genome-wide studies at much finer resolutions. For instance, by phasing haplotypes one could trace how haplotype blocks introgress through hybridisation. This approach also allows us to identify regions that are shielded from introgression and may potentially harbour reproductive barriers maintaining species barriers despite prevalent gene flow.

In summary, my thesis provides a broad view on the rich diversity in speciation dynamics and showcases the vast potential of museomic approaches for studying speciation. It provides a wealth of resources for further museomic research to obtain a more comprehensive understanding of the mechanisms that ultimately shape the biodiversity of our planet.

6. References

Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., Boughman, J., Brelsford, A., Buerkle, C. A., Buggs, R., Butlin, R. K., Dieckmann, U., Eroukhmanoff, F., Grill, A., Cahan, S. H., Hermansen, J. S., Hewitt, G., Hudson, A. G., Jiggins, C., … Zinner, D. (2013). Hybridization and speciation*. *Journal of Evolutionary Biology*, *26*(2), 229–246. https://doi.org/10.1111/j.1420-9101.2012.02599.x

Agne, S., Naylor, G. J. P., Preick, M., Yang, L., Thiel, R., Weigmann, S., Paijmans, J. L. A., Barlow, A., Hofreiter, M., & Straube, N. (2022). Taxonomic Identification of Two Poorly Known Lantern Shark Species Based on Mitochondrial DNA From Wet-Collection Paratypes. *Frontiers in Ecology and Evolution*, *10*. https://doi.org/10.3389/fevo.2022.910009

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664. https://doi.org/10.1101/gr.094052.109

Allendorf, F. W., Leary, R. F., Spruell, P., & Wenburg, J. K. (2001). The problems with hybrids: Setting conservation guidelines. *Trends in Ecology & Evolution*, *16*(11), 613–622. https://doi.org/10.1016/S0169-5347(01)02290-X

Anderson, S. A. S., & Matute, D. R. (2025). Secondary Sympatry as a Sorting Process. *Ecology Letters*, *28*(4), e70108. https://doi.org/10.1111/ele.70108

Behr, A. A., Liu, K. Z., Liu-Fang, G., Nakka, P., & Ramachandran, S. (2016). pong: Fast analysis and visualization of latent clusters in population genetic data. *Bioinformatics*, *32*(18), 2817–2823. https://doi.org/10.1093/bioinformatics/btw327

Bhatia, G., Patterson, N., Sankararaman, S., & Price, A. L. (2013). Estimating and interpreting F_{ST}: The impact of rare variants. *Genome Research*, *23*(9), 1514–1521. https://doi.org/10.1101/gr.154831.113

Billerman, S. M., & Walsh, J. (2019). Historical DNA as a tool to address key questions in avian biology and evolution: A review of methods, challenges, applications, and future directions. *Molecular Ecology Resources*, *19*(5), 1115–1130. https://doi.org/10.1111/1755-0998.13066

Bird, C. E., Fernandez-Silva, I., Skillings, D. J., & Toonen, R. J. (2012). Sympatric Speciation in the Post "Modern Synthesis" Era of Evolutionary Biology. *Evolutionary Biology*, *39*(2), 158–180. https://doi.org/10.1007/s11692-012-9183-6

Blom, M. P. K., Peona, V., Prost, S., Christidis, L., Benz, B. W., Jønsson, K. A., Suh, A., & Irestedt, M. (2024). Hybridization in birds-of-paradise: Widespread ancestral gene flow despite strong sexual selection in a lek-mating system. *iScience*, *27*(7). https://doi.org/10.1016/j.isci.2024.110300

Briggs, A. W., Stenzel, U., Meyer, M., Krause, J., Kircher, M., & Pääbo, S. (2010). Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Research*, *38*(6), e87–e87. https://doi.org/10.1093/nar/gkp1163

Butlin, R., Bridle, J., & Schluter, D. (2009). *Speciation and patterns of biodiversity*. Cambridge University Press.

Call, E., Mayer, C., Twort, V., Dietz, L., Wahlberg, N., & Espeland, M. (2021). Museomics: Phylogenomics of the Moth Family Epicopeiidae (Lepidoptera) Using Target Enrichment. *Insect Systematics and Diversity*, *5*(2), 6. https://doi.org/10.1093/isd/ixaa021

Cavill, E. L., Morales, H. E., Sun, X., Westbury, M. V., van Oosterhout, C., Accouche, W., Zora, A., Schulze, M. J., Shah, N., Adam, P.-A., Brooke, M. de L., Sweet, P.,

Gopalakrishnan, S., & Gilbert, M. T. P. (2024). When birds of a feather flock together: Severe genomic erosion and the implications for genetic rescue in an endangered island passerine. *Evolutionary Applications*, *17*(7), e13739. https://doi.org/10.1111/eva.13739

Coyne, J. A. (2007). Sympatric speciation. *Current Biology*, *17*(18), R787–R788.

Criscuolo, N. G., & Angelini, C. (2020). StructuRly: A novel shiny app to produce comprehensive, detailed and interactive plots for population genetic analysis. *PLOS ONE*, *15*(2), e0229330. https://doi.org/10.1371/journal.pone.0229330

Dabney, J., Meyer, M., & Pääbo, S. (2013). Ancient DNA Damage. *Cold Spring Harbor Perspectives in Biology*, *5*(7), a012567. https://doi.org/10.1101/cshperspect.a012567

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Darriba, D., Posada, D., Kozlov, A. M., Stamatakis, A., Morel, B., & Flouri, T. (2020). ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. *Molecular Biology and Evolution*, *37*(1), 291–294. https://doi.org/10.1093/molbev/msz189

Davies, H. L. (2012). The geology of New Guinea-the cordilleran margin of the Australian continent. *Episodes Journal of International Geoscience*, *35*(1), 87–102.

Degnan, J. H., & Rosenberg, N. A. (2009). Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution*, *24*(6), 332–340. https://doi.org/10.1016/j.tree.2009.01.009

Delling, B., Thörn, F., Norén, M., & Irestedt, M. (2023). Museomics reveals the phylogenetic position of the extinct Moroccan trout. *Journal of Fish Biology*, *102*(3), 619–627. https://doi.org/10.1111/jfb.15299

Di Tommaso, P., Chatzou, M., Floden, E. W., Barja, P. P., Palumbo, E., & Notredame, C. (2017). Nextflow enables reproducible computational workflows. *Nature Biotechnology*, *35*(4), Article 4. https://doi.org/10.1038/nbt.3820

Diamond, J., & Bishop, K. D. (2023). What's so special about New Guinea birds? *Bulletin of the British Ornithologists' Club*, *143*(2), 212–236. https://doi.org/10.25226/bboc.v143i2.2023.a6

Diamond, J. M. (1973). Distributional Ecology of New Guinea Birds: Recent ecological and biogeographical theories can be tested on the bird communities of New Guinea. *Science*, *179*(4075), 759–769. https://doi.org/10.1126/science.179.4075.759

Diamond, J. M. (1983). Melampitta gigantea: Possible Relation between Feather Structure and Underground Roosting Habits. *The Condor*, *85*(1), 89–91. https://doi.org/10.2307/1367895

Dolenz, S., van der Valk, T., Jin, C., Oppenheimer, J., Sharif, M. B., Orlando, L., Shapiro, B., Dalén, L., & Heintzman, P. D. (2024). Unravelling reference bias in ancient DNA datasets. *Bioinformatics*, *40*(7), btae436. https://doi.org/10.1093/bioinformatics/btae436

Erwin, T. L. (1981). Taxon Pulses, Vicariance, and Dispersal; an Evolutionary Synthesis Illustrated by Carabid Beetles. In *Vicariance biogeography: A critique* (pp. 159–196). Columbia University Press.

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: A simulation study. *Molecular Ecology*, *14*(8), 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Fong, J. J., Blom, M. P. K., Aowphol, A., McGuire, J. A., Sutcharit, C., & Soltis, P. S. (2023). Editorial: Recent advances in museomics: revolutionizing biodiversity research.

Frontiers in Ecology and Evolution, *11*.
https://www.frontiersin.org/articles/10.3389/fevo.2023.1188172

Garcia-Erill, G., & Albrechtsen, A. (2020). Evaluation of model fit of inferred admixture proportions. *Molecular Ecology Resources*, *20*(4), 936–949. https://doi.org/10.1111/1755-0998.13171

Garrison, E., & Marth, G. (2012). *Haplotype-based variant detection from short-read sequencing* (No. arXiv:1207.3907). arXiv. https://doi.org/10.48550/arXiv.1207.3907

Greenwood, P. J. (1980). Mating systems, philopatry and dispersal in birds and mammals. *Animal Behaviour*, *28*(4), 1140–1162. https://doi.org/10.1016/S0003-3472(80)80103-5

Guschanski, K., Krause, J., Sawyer, S., Valente, L. M., Bailey, S., Finstermeier, K., Sabin, R., Gilissen, E., Sonet, G., Nagy, Z. T., Lenglet, G., Mayer, F., & Savolainen, V. (2013). Next-Generation Museomics Disentangles One of the Largest Primate Radiations. *Systematic Biology*, *62*(4), 539–554. https://doi.org/10.1093/sysbio/syt018

Hahn, C., Bachmann, L., & Chevreux, B. (2013). Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—A baiting and iterative mapping approach. *Nucleic Acids Research*, *41*(13), e129. https://doi.org/10.1093/nar/gkt371

Hall, R. (2002). Cenozoic geological and plate tectonic evolution of SE Asia and the SW Pacific: Computer-based reconstructions, model and animations. *Journal of Asian Earth Sciences*, *20*(4), 353–431. https://doi.org/10.1016/S1367-9120(01)00069-4

Hernández-Hernández, T., Miller, E. C., Román-Palacios, C., & Wiens, J. J. (2021). Speciation across the Tree of Life. *Biological Reviews*, *96*(4), 1205–1242. https://doi.org/10.1111/brv.12698

Irestedt, M., Thörn, F., Müller, I. A., Jønsson, K. A., Ericson, P. G. P., & Blom, M. P. K. (2022). A guide to avian museomics: Insights gained from resequencing hundreds of avian study skins. *Molecular Ecology Resources*, *22*(7), 2672–2684. https://doi.org/10.1111/1755-0998.13660

Jiang, X., Edwards, S. V., & Liu, L. (2020). The Multispecies Coalescent Model Outperforms Concatenation Across Diverse Phylogenomic Data Sets. *Systematic Biology*, *69*(4), 795–812. https://doi.org/10.1093/sysbio/syaa008

Jønsson, K. A., Irestedt, M., Christidis, L., Clegg, S. M., Holt, B. G., & Fjeldså, J. (2014). Evidence of taxon cycles in an Indo-Pacific passerine bird radiation (Aves: Pachycephala). *Proceedings of the Royal Society B: Biological Sciences*, *281*(1777), 20131727. https://doi.org/10.1098/rspb.2013.1727

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587–589. https://doi.org/10.1038/nmeth.4285

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780.

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics*, *15*(1), 356. https://doi.org/10.1186/s12859-014-0356-4

Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., & Stamatakis, A. (2019). RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*, *35*(21), 4453–4455. https://doi.org/10.1093/bioinformatics/btz305

Kuchta, S. R., & Wake, D. B. (2016). Wherefore and Whither the Ring Species? *Copeia*, *104*(1), 189–201. https://doi.org/10.1643/OT-14-176

Lalueza-Fox, C. (2022). Museomics. *Current Biology*, *32*(21), R1214–R1215. https://doi.org/10.1016/j.cub.2022.09.019

Larsson, A. (2014). AliView: A fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*, *30*(22), 3276–3278. https://doi.org/10.1093/bioinformatics/btu531

Lepage, D., Vaidya, G., & Guralnick, R. (2014). Avibase–a database system for managing and organizing taxonomic concepts. *ZooKeys*, *420*, 117.

Li, H. (2013). *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM* (No. arXiv:1303.3997). arXiv. https://doi.org/10.48550/arXiv.1303.3997

Li, H. (2024). *Lh3/seqtk* [C]. https://github.com/lh3/seqtk (Original work published 2012)

Liu, X., Milesi, E., Fontsere, C., Owens, H. L., Heinsohn, R., Gilbert, M. T. P., Crates, R., Nogués-Bravo, D., & Morales, H. E. (2025). Time-lagged genomic erosion and future environmental risks in a bird on the brink of extinction. *Proceedings of the Royal Society B: Biological Sciences*, *292*(2043), 20242480. https://doi.org/10.1098/rspb.2024.2480

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

Mallet, J., Besansky, N., & Hahn, M. W. (2016). How reticulated are species? *BioEssays*, *38*(2), 140–149. https://doi.org/10.1002/bies.201500149

Mayr, E. (1954). *Change of genetic environment and evolution*. George Allen & Unwin Limited. http://www2.hawaii.edu/~khayes/Journal_Club/fall2006/Mayr_1954_in_Evol_as_a_Proc_157-180.pdf

Mayr, E. (1999). *Systematics and the origin of species, from the viewpoint of a zoologist*. Harvard University Press.

Mayr, E., & Gilliard, E. T. (1952). Altitudinal Hybridization in New Guinea Honeyeaters. *The Condor*, *54*(6), 325–337. https://doi.org/10.2307/1364906

Mayr, E., & Gilliard, E. T. (1954). *Birds of central New Guinea: Results of the American Museum of Natural History expeditions to New Guinea in 1950 and 1952. Bulletin of the AMNH; v. 103, article 4*.

Meineke, E. K., Davies, T. J., Daru, B. H., & Davis, C. C. (2018). Biological collections for understanding biodiversity in the Anthropocene. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*(1763), 20170386. https://doi.org/10.1098/rstb.2017.0386

Meisner, J., & Albrechtsen, A. (2018). Inferring Population Structure and Admixture Proportions in Low-Depth NGS Data. *Genetics*, *210*(2), 719–731. https://doi.org/10.1534/genetics.118.301336

Meyer, A. (2005). On the Importance of Being Ernst Mayr. *PLOS Biology*, *3*(5), e152. https://doi.org/10.1371/journal.pbio.0030152

Meyer, M., & Kircher, M. (2010). Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. *Cold Spring Harbor Protocols*, *2010*(6), pdb.prot5448-pdb.prot5448. https://doi.org/10.1101/pdb.prot5448

Minh, B. Q., Hahn, M. W., & Lanfear, R. (2020). New Methods to Calculate Concordance Factors for Phylogenomic Datasets. *Molecular Biology and Evolution*, *37*(9), 2727–2733. https://doi.org/10.1093/molbev/msaa106

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution*, *37*(5), 1530–1534. https://doi.org/10.1093/molbev/msaa015

Murray, G. G. R., Soares, A. E. R., Novak, B. J., Schaefer, N. K., Cahill, J. A., Baker, A. J., Demboski, J. R., Doll, A., Da Fonseca, R. R., Fulton, T. L., Gilbert, M. T. P., Heintzman, P. D., Letts, B., McIntosh, G., O'Connell, B. L., Peck, M., Pipes, M.-L., Rice, E. S., Santos, K. M., … Shapiro, B. (2017). Natural selection shaped the rise and fall of passenger pigeon genomic diversity. *Science*, *358*(6365), 951–954. https://doi.org/10.1126/science.aao0960

Neukamm, J., Peltzer, A., & Nieselt, K. (2021). DamageProfiler: Fast damage pattern calculation for ancient DNA. *Bioinformatics*, *37*(20), 3652–3653. https://doi.org/10.1093/bioinformatics/btab190

Okonechnikov, K., Conesa, A., & García-Alcalde, F. (2016). Qualimap 2: Advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics*, *32*(2), 292–294. https://doi.org/10.1093/bioinformatics/btv566

Pääbo, S., & Wilson, A. C. (1988). Polymerase chain reaction reveals cloning artefacts. *Nature*, *334*(6181), 387–388. https://doi.org/10.1038/334387b0

Phillimore, A. B., Orme, C. D. L., Thomas, G. H., Blackburn, T. M., Bennett, P. M., Gaston, K. J., & Owens, I. P. F. (2008). Sympatric Speciation in Birds Is Rare: Insights from Range Data and Simulations. *The American Naturalist*, *171*(5), 646–657. https://doi.org/10.1086/587074

Pigram, C., & Davies, H. (1987). Terranes and the accretion history of the Papua New Guinea orogeny. *BMR Journal of Australian Geology and Geophysics*, *10*, 193–212.

Posada, D., & Crandall, K. A. (2001). Intraspecific gene genealogies: Trees grafting into networks. *Trends in Ecology & Evolution*, *16*(1), 37–45. https://doi.org/10.1016/S0169-5347(00)02026-7

Posit team. (2023). *RStudio: Integrated Development Environment for R*. Posit Software, PBC. http://www.posit.co/

Presgraves, D. C., Balagopalan, L., Abmayr, S. M., & Orr, H. A. (2003). Adaptive evolution drives divergence of a hybrid inviability gene between two species of Drosophila. *Nature*, *423*(6941), Article 6941. https://doi.org/10.1038/nature01679

QGIS Development Team. (2025). *QGIS Geographic Information System*. QGIS Association. https://www.qgis.org

R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Raxworthy, C. J., & Smith, B. T. (2021). Mining museums for historical DNA: Advances and challenges in museomics. *Trends in Ecology & Evolution*, *36*(11), 1049–1060. https://doi.org/10.1016/j.tree.2021.07.009

Reeve, A. H., Kennedy, J. D., Pujolar, J. M., Petersen, B., Blom, M. P. K., Alström, P., Haryoko, T., Ericson, P. G. P., Irestedt, M., Nylander, J. A. A., & Jønsson, K. A. (2023). The formation of the Indo-Pacific montane avifauna. *Nature Communications*, *14*(1), Article 1. https://doi.org/10.1038/s41467-023-43964-y

Revell, L. J. (2012). phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, *3*(2), 217–223. https://doi.org/10.1111/j.2041-210X.2011.00169.x

Ricklefs, R. E., & Bermingham, E. (2002). The concept of the taxon cycle in biogeography. *Global Ecology and Biogeography*, *11*(5), 353–361. https://doi.org/10.1046/j.1466-822x.2002.00300.x

Rosenblum, E. B., Sarver, B. A. J., Brown, J. W., Des Roches, S., Hardwick, K. M., Hether, T. D., Eastman, J. M., Pennell, M. W., & Harmon, L. J. (2012). Goldilocks Meets Santa Rosalia: An Ephemeral Speciation Model Explains Patterns of Diversification Across Time Scales. *Evolutionary Biology*, *39*(2), 255–261. https://doi.org/10.1007/s11692-012-9171-x

Sawyer, S., Krause, J., Guschanski, K., Savolainen, V., & Pääbo, S. (2012). Temporal Patterns of Nucleotide Misincorporations and DNA Fragmentation in Ancient DNA. *PLOS ONE*, *7*(3), e34131. https://doi.org/10.1371/journal.pone.0034131

Schaefer, H., Heibl, C., & Renner, S. S. (2008). Gourds afloat: A dated phylogeny reveals an Asian origin of the gourd family (Cucurbitaceae) and numerous oversea dispersal events. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1658), 843–851. https://doi.org/10.1098/rspb.2008.1447

Schuster, S. C., & Miller, W. (2008). *Mammoth Genome Project PSU*. http://mammoth.psu.edu/museomics.html

Shapiro, B., Barlow, A., Heintzman, P. D., Hofreiter, M., Paijmans, J. L. A., & Soares, A. E. R. (Eds.). (2019). *Ancient DNA: Methods and Protocols* (Vol. 1963). Springer. https://doi.org/10.1007/978-1-4939-9176-1

Skotte, L., Korneliussen, T. S., & Albrechtsen, A. (2013). Estimating Individual Admixture Proportions from Next Generation Sequencing Data. *Genetics*, *195*(3), 693–702. https://doi.org/10.1534/genetics.113.154138

Sotuyo, S., Pedraza-Ortega, E., Martínez-Salas, E., Linares, J., & Cabrera, L. (2022). Insights into phylogenetic divergence of Dalbergia (Leguminosae: Dalbergiae) from Mexico and Central America. *Frontiers in Ecology and Evolution*, *10*. https://doi.org/10.3389/fevo.2022.910250

Stankowski, S., & Ravinet, M. (2021). Defining the speciation continuum. *Evolution*, *75*(6), 1256–1273. https://doi.org/10.1111/evo.14215

Stegemiller, M. R., Redden, R. R., Notter, D. R., Taylor, T., Taylor, J. B., Cockett, N. E., Heaton, M. P., Kalbfleisch, T. S., & Murdoch, B. M. (2023). Using whole genome sequence to compare variant callers and breed differences of US sheep. *Frontiers in Genetics*, *13*. https://doi.org/10.3389/fgene.2022.1060882

Stull, G. W., Pham, K. K., Soltis, P. S., & Soltis, D. E. (2023). Deep reticulation: The long legacy of hybridization in vascular plant evolution. *The Plant Journal*, *114*(4), 743–766. https://doi.org/10.1111/tpj.16142

Templeton, A. R. (1981). Mechanisms of Speciation—A Population Genetic Approach. *Annual Review of Ecology and Systematics*, *12*, 23–48.

Thomas, W. K., Pääbo, S., Villablanca, F. X., & Wilson, A. C. (1990). Spatial and temporal continuity of kangaroo rat populations shown by sequencing mitochondrial DNA from museum specimens. *Journal of Molecular Evolution*, *31*(2), 101–112. https://doi.org/10.1007/BF02109479

Thörn, F., Müller, I. A., Soares, A. E. R., Nagombi, E., Jønsson, K. A., Blom, M. P. K., & Irestedt, M. (2025). Frequent Hybridisation Between Parapatric Lekking Bird-of-Paradise Species. *Molecular Ecology*, e17780. https://doi.org/10.1111/mec.17780

Thörn, F., Soares, A. E. R., Müller, I. A., Päckert, M., Frahnert, S., van Grouw, H., Kamminga, P., Peona, V., Suh, A., Blom, M. P. K., & Irestedt, M. (2024). Contemporary intergeneric hybridization and backcrossing among birds-of-paradise. *Evolution Letters*, *8*(5), 680–694. https://doi.org/10.1093/evlett/qrae023

van der Valk, T., Díez-del-Molino, D., Marques-Bonet, T., Guschanski, K., & Dalén, L. (2019). Historical Genomes Reveal the Genomic Consequences of Recent Population Decline in Eastern Gorillas. *Current Biology*, *29*(1), 165-170.e6. https://doi.org/10.1016/j.cub.2018.11.055

Vasimuddin, Md., Misra, S., Li, H., & Aluru, S. (2019). Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, 314–324. https://doi.org/10.1109/IPDPS.2019.00041

Wandeler, P., Hoeck, P. E. A., & Keller, L. F. (2007). Back to the future: Museum specimens in population genetics. *Trends in Ecology & Evolution*, *22*(12), 634–642. https://doi.org/10.1016/j.tree.2007.08.017

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, 1358–1370.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wilson, E. O. (1961). The Nature of the Taxon Cycle in the Melanesian Ant Fauna. *The American Naturalist*, *95*(882), 169–193.

Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T.-Y. (2017). ggtree: An r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, *8*(1), 28–36. https://doi.org/10.1111/2041-210X.12628

Zedane, L., Hong-Wa, C., Murienne, J., Jeziorski, C., Baldwin, B. G., & Besnard, G. (2016). Museomics illuminate the history of an extinct, paleoendemic plant lineage (Hesperelaea, Oleaceae) known from an 1875 collection from Guadalupe Island, Mexico. *Biological Journal of the Linnean Society*, *117*(1), 44–57. https://doi.org/10.1111/bij.12509

Zhang, C., Rabiee, M., Sayyari, E., & Mirarab, S. (2018). ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*, *19*(6), 153. https://doi.org/10.1186/s12859-018-2129-y

7. Svensk sammanfattning

Studier av artbildning hos fåglar i Nya Guinea med hjälp av storskalig museigenomik

Att förstå hur arter bildas är en central fråga inom evolutionsbiologin, eftersom artbildning är en grundläggande process för att generera biologisk mångfald. Synen på hur artbildning sker har förändrats genom historien, men numera betraktas det oftast som en gradvis process där populationer successivt utvecklar reproduktiv isolering från varandra. De bakomliggande mekanismerna kan dock variera. Det vanligaste scenariot innebär att någon form av barriär uppstår som delar en population och förhindrar genflöde, exempel på detta är allopatrisk och peripatrisk artbildning. Ett annat exempel är parapatrisk artbildning, där gradvisa skillnader inom en geografiskt vidsträckt population leder till att subpopulationer långt ifrån varandra utvecklar så stora skillnader att de blir reproduktivt isolerade. Så kallade ringarter är ett exempel på detta fenomen. Slutligen kan selektion inom en population, exempelvis till följd av anpassning till olika typer av föda, leda till att den ursprungliga populationen delas upp i två eller flera populationer med olika egenskaper, detta kallas sympatrisk artbildning. Denna form av artbildning anses vara ovanlig, särskilt bland fåglar.

I denna avhandling använder jag museigenomik för att studera artbildning, hybridisering och biogeografiska mönster hos fåglar på Nya Guinea. Museigenomik innebär att man utvinner och analyserar DNA från hela genom hos organismer som bevarats i naturhistoriska samlingar. En fördel med museigenomik är att det möjliggör studier av evolutionära processer hos arter som är utdöda eller svåra att få färska DNA-prov från. En nackdel är att DNA i äldre museiprover ofta förekommer i små mängder, är fragmenterat samt uppvisar skademönster som kan påverka analysresultaten. Tack vare framsteg inom sekvenseringsteknologi och bioinformatik har dock många av dessa begränsningar kunnat överkommas. I denna avhandling har användningen av historiska prover från museisamlingar varit central, då det möjliggjort en tillräckligt tät provtagning för att undersöka hybridisering och artbildningsprocesser.

I **kapitel I** undersökte jag populationsstrukturen hos mindre (*Melampitta lugubris*) och större svartpitta (*Megalampitta gigantea*). Hos den mindre svartpittan, som är tämligen vitt spridd över olika bergskedjor på Nya Guinea, upptäckte jag en djup populationsstruktur som tyder på att populationer på olika bergsmassiv har varit åtskilda i flera miljoner år. Hos den större svartpittan fann jag däremot nästan ingen genetisk differentiering mellan populationer, trots att denna art har en mycket mer fragmenterad utbredning. Min hypotes för att förklara detta oväntade mönster är att den större svartpittan historiskt haft en betydligt större utbredning, men att den genomgått en relativt nylig populationskollaps som resulterat i den nuvarande fragmenteringen. I detta kapitel utvecklade jag även bioinformatiska arbetsflöden som används i de efterföljande kapitlen.

I **kapitel II** undersökte vi hybridiseringsmönster hos en evolutionär linje av visslare (*Pachycephala*). Detta artkomplex förekommer både på mer stabila, isolerade oceaniska öar och i den mer dynamiska Sahulregionen, som omfattar bland annat Australien, Nya Guinea och flera mellanliggande öar. Under pleistocena perioder med låga havsnivåer har dessa områden tidvis utgjort en sammanhängande landmassa. Studien visade på ett mycket komplext nätverk av genflöden, där framför allt arter med utbredning i den instabila Sahulregionen uppvisar

tydliga tecken på omfattande hybridisering. Resultaten stöder hypotesen att miljöinstabilitet kan bidra till en ökad frekvens av hybridisering.

I **kapitel III** studerar jag artbildningsdynamik inom ett släkte av honungsätare (*Melidectes*). Resultaten stödjer att de flesta arter inom detta släkte har uppkommit genom geografisk isolering. Två arter med delvis överlappande utbredning i det centrala bergsmassivet på Nya Guinea, som tidigare beskrivits som hybridiserande, uppvisar dock en mycket komplex genetisk struktur som stämmer dåligt överens med den nuvarande artindelningen. Våra resultat visar i stället att dessa arter i själva verket utgör en gemensam genetisk population. I vissa delar av genomet finns dock spår som tyder på att denna population historiskt bestått av två distinkta evolutionära linjer, vilka senare har kommit i sekundär kontakt och smält samman till en gemensam art. Jag presenterar därför detta fall som ett ovanligt empiriskt exempel på efemär artbildning. Genom att korrelera miljödata med genetiska data undersöker jag även möjlig framtida artbildning inom detta komplex under olika scenarier av klimatförändringar.

I **kapitel IV** undersökte jag omfattningen av reproduktiv isolering och hybridiseringsmönster inom paradisfågelsläktet *Paradisaea*, samt diskuterar de olika arternas placering längs ett artbildningskontinuum. Till skillnad från arter som förekommer på öar utanför Nya Guinea, vilka är genetiskt tydligt differentierade från varandra, uppvisar de tre arter som beskrivits från fastlandsdelen av Nya Guinea låg genetisk differentiering och ett mönster av mer eller mindre fria genflöden över stora delar av låglandet. Ett undantag utgörs av området där utbredningen av mindre paradisfågel och raggiparadisfågel möts, där dessa "arter" uppvisar en högre grad av genetisk differentiering och inga tecken på pågående genflöde. Populationsstrukturen och differentieringsmönstren hos *Paradisaea*-arterna på fastlandet liknar därmed i hög grad det man förväntar sig hos en så kallad ringart.

Sammanfattningsvis visar denna avhandling på värdet av historiska museisamlingar för genomiska och evolutionära studier av arter och populationer som är svåra att samla in i naturen. Många av resultaten i avhandlingen var oväntade i förhållande till rådande taxonomiska uppdelningar och biogeografiska förväntningar. Jag vill därför avslutningsvis poängtera att effekterna av systemspecifika mekanismer kräver försiktighet när man drar generella slutsatser om artbildningsprocesser.

8. Deutsche Zusammenfassung

Einblicke in Artbildungsdynamiken neuguineischer Vögel durch groß angelegte Museomik-Datensätze

Das Verständnis, wie Arten geformt werden, ist eine zentrale Frage innerhalb der Evolutionsbiologie, da Artbildung ein grundlegender Prozess ist, um biologische Vielfalt zu generieren. Unsere Perspektive über, wie Artbildung zustande kommt, hat sich im Laufe der Geschichte verändert, aber heutzutage wird Artbildung generell als ein gradueller Prozess betrachtet, in welchem Populationen sukzessiv reproduktive Isolierung voneinander entwickeln. Die zugrundeliegenden Mechanismen dahinter können jedoch variieren. Die am häufigsten vorkommenden Szenarien beinhalten, dass eine gewisse Art von Barriere entsteht, welche Populationen isoliert und Genfluss verhindert. Beispiele für diese Szenarien sind die sogenannte „allopatrische" oder „peripatrische" Artbildung. Ein weiteres Beispiel ist „parapatrische" Artbildung, in welcher schrittweise Unterschiede innerhalb einer geographisch weitverbreiteten Population dazu führen, dass Teilpopulationen, die weit voneinander entfernt sind derart große Unterschiede aufweisen, dass sie reproduktiv isoliert werden. Sogenannte „Ringarten" sind ein berühmtes Beispiel für dieses Phänomen. Außerdem kann Selektion innerhalb einer einzigen Population, beispielsweise als Folge von Anpassungen an verschiedene Nahrungstypen dazu führen, dass sich diese ursprüngliche Population innerhalb eines Habitats in mehrere Populationen mit unterschiedlichen Merkmalen aufteilt, ein Szenario, das als „sympatrische" Artbildung bekannt ist. Diese Form von Artbildung wird allerdings als sehr selten angesehen, insbesondere unter Vögeln.

In dieser Dissertation wende ich Museomik an, um Artbildung, Hybridisierung und biogeographische Muster innerhalb neuguineischer Vögel zu studieren. Museomik bedeutet die Gewinnung und Analyse von gesamt-genomischer DNS aus Organismen, welche in naturhistorischen Sammlungen aufbewahrt sind. Ein Vorteil von Museomik ist die Ermöglichung von Studien evolutionärer Prozesse innerhalb Arten, welche ausgestorben sind oder für welche die Sammlung moderner DNS-Proben schwierig ist. Ein Nachteil ist, dass die DNS älterer Museumsexemplare oft in geringere Konzentrationen ergibt, fragmentiert ist und gewisse Schadensmuster aufweist, welche die Ergebnisse gängiger Analysen beeinflussen kann. Jedoch können wir heutzutage viele dieser Hindernisse dank der Fortschritte in Sequenzierungstechnologien und bioinformatischen Methoden überwinden. In dieser Dissertation war die Anwendung historischer Exemplare ein zentraler Punkt, da uns diese Proben ausreichend dichte Probenahmen ermöglicht haben, um Artbildungsprozesse und Hybridisierung zu untersuchen.

In **Kapitel I** untersuchte ich Populationsstrukturen innerhalb Glanzflöter (*Melampitta* lugubris) und Rußflöter (*Megalampitta gigantaea*). Innerhalb der Glanzflöter, welche weit verbreitet über verschiedene Bergketten Neuguineas verbreitet sind, entdeckte ich deutliche Populationsstruktur, was darauf deutet, dass sich Populationen unterschiedlicher Bergmassive über mehrere Millionen Jahre hinaus isoliert haben. Innerhalb der Rußflöter fand ich hingegen nahezu keinerlei genetische Differenzierung zwischen den Populationen, obwohl diese Art eine weitaus stärker fragmentierte Verbreitung aufweist. Meine Hypothese zur Erklärung dieses unerwarteten Musters ist, dass Rußflöter historisch eine wesentlich größere Ausbreitung hatten, sie allerdings relativ kürzlich einen Populationskollaps erlebt haben, welcher zur derzeitigen

Fragmentierung der Lebensräume geführt hat. In diesem Kapitel entwickelte ich überdies bioinformatische Arbeitsabläufe, welche ich in den folgenden Kapiteln angewendet habe.

In **Kapitel II** untersuchten wir Hybridisierungsmuster innerhalb einer evolutionären Linie von Dickköpfen (*Pacycephalidae*). Dieser Artkomplex kommt sowohl auf stabileren, isolierten ozeanischen Inseln als auch innerhalb der dynamischeren „Sahulregion" vor, welche während der letzten Eiszeit eine verbundene Landmasse zwischen Australien, Neuguinea und mehreren umliegenden Inseln umfasst hatte. Die Studie ergab ein sehr komplexes Netzwerk von Genflüssen, in welchem insbesondere Arten mit Ausbreitungen innerhalb der instabilen Sahulregion deutliche Zeichen ausgiebiger Hybridisierung aufwiesen. Die Ergebnisse unterstützen die Hypothese, dass instabile Umweltbedingungen zu einer erhöhten Frequenz von Hybridisierung führen können.

In **Kapitel III** studierte ich Artbildungsdynamiken innerhalb einer Gattung der Honigfresser (*Melidectes*). Die Ergebnisse belegen, dass sich die meisten Arten innerhalb dieser Gattung durch geographische Isolierung geformt haben. Jedoch zeigten zwei Arten mit teilweise überlappenden Ausbreitungen im zentralen neuguineischen Bergmassiv, welche zuvor als hybridisierend beschrieben waren, recht komplexe genetische Struktur, welche nicht mit der derzeitigen Klassifizierung dieser Arten übereinstimmt. Unsere Resultate zeigen stattdessen, dass diese Arten tatsächlich eine einzige genetische Population repräsentieren. In gewissen Segmenten des Genoms findet man allerdings noch Spuren, die darauf deuten, dass diese Population historisch aus zwei verschiedenen Abstammungslinien bestand, welche später in Sekundärkontakt kamen und wieder in eine einzige Art verschmolzen sind. Ich präsentiere daher diesen Fall als ein seltenes empirisches Beispiel für ephemere oder kurzlebige Artbildung. Überdies untersuchte ich anhand von Korrelationen zwischen Umwelt- und genetischen Daten die Möglichkeit zukünftiger Artbildung innerhalb dieses Komplexes unter verschiedenen Klimawandelszenarien.

In **Kapitel IV** untersuchte ich den Umfang reproduktiver Isolierung und Hybridisierungsmuster innerhalb der Paradiesvogelgattung *Paradisaea* und erörterte die Platzierung der verschiedenen Arten innerhalb des Artbildungskontinuums. Im Gegensatz zu den isolierteren Arten, welche zum Beispiel auf abgelegeneren Inseln außerhalb Neuguineas vorkommen und sich genetisch deutlich voneinander unterscheiden, wiesen drei beschriebene Arten auf Neuguineas Festland geringe genetische Differenzierung auf und zeigten ein Muster mehr oder weniger freien Genflusses über weite Teile des Tieflandes. Eine Ausnahme bildete das Areal, in welchem die Verbreitungsgebiete des Kleinen Paradiesvogels und des Raggi-Paradiesvogels aufeinandertreffen, in welchem diese „Arten" einen höheren Grad an genetischer Differenzierung und keine Zeichen fortlaufenden Genflusses aufweisen. Die Populationsstrukturen innerhalb Arten der Eigentlichen Paradiesvögel (*Paradisaea*) auf dem Festland ähneln daher stark derer, die man innerhalb sogenannter „Ringarten" erwarten würde.

Zusammenfassend demonstriert meine Dissertation den Wert naturhistorischer Museumssammlungen für genomische und evolutionäre Studien von Arten und Populationen, die in freier Wildbahn nur schwer zu sammeln sind. Viele der Ergebnisse meiner Dissertation waren unerwartet angesichts der derzeitigen taxonomischen Klassifizierungen und biogeographischer Erwartungen. Abschließend, möchte ich daher betonen, dass die Auswirkungen systemspezifischer Mechanismen zur Vorsicht mahnen, wenn generelle Schlussfolgerungen zu Artbildungsprozessen gezogen werden.

9. Acknowledgements

A PhD is a tremendous voyage which would have been impossible for me without the support of a large number of people. I hope I am able to remember everyone who has helped me and kept me sane during these four years. My deepest apologies if I have missed any names.

Firstly, I must express my eternal gratitude to my main advisor, **Martin**, for allowing me to embark on this fantastic journey to explore these beautiful New Guinean birds. Thank you for your unending patience and support, for keeping my motivation high throughout the projects and for preparing me well for the rollercoaster of academia. Joining you for birding and ringing on Örskär has been an absolute delight every time I have been able to join you, and it has strengthened my enthusiasm about birds immensely. I am also extremely grateful to **Mozes** for always being an encouraging advisor, for recognising my own accomplishments and making me feel good about them, and for your invaluable support and perspective in many aspects, especially for bioinformatics-related issues. Surprisingly and unfortunately, I have visited you in Berlin more often during the pandemic than when it was over, but nonetheless, I highly appreciated the change in work environment and our personal interactions. **Knud**, although you weren't my advisor on paper, your enthusiasm, valuable feedback and concern about my wellbeing throughout the PhD have been immensely helpful. I would like to thank my co-advisor **Karin** and to my follow-up committee members **Chris W.** and **Ulf** for your positive evaluations and advice throughout my progress meetings.

I am especially grateful to have shared my office with **Filip,** who has been with me for most of this journey. You have been a tremendous support to me, both for work-related issues and through your friendship. It is only thanks to you that I have been able to navigate through the many frustrations we have encountered throughout our PhDs.

I would also like to thank all collaborators who have helped to improve my projects with their own approaches and input. In particular, I'd like to thank you, **Per E.**, for your valuable contributions and the discussions we've had on all my projects. To **Sam R.**, for being not only a great collaborator, but also a fantastic friend and for accompanying me through this PhD from start to finish. Thank you, **Abdul** and **Alma,** for giving me new perspectives and making me question the methods I've been using all this time.

I want to thank all my colleagues at NRM for making this isolated building feel like a welcoming and inspiring place to work. I would like to thank you **Johan**, for your immensely helpful input and our discussions about bioinformatic methods. Thank you, **Bodil**, for saving a helpless bioinformatician in the lab on several occasions. I want to thank the whole NRM PhD crew, **Chenyu**, **Isa**, **Marika**, **Meri**, **Sam L.** and **Victor** for sharing this journey with me and not making it a lonely experience. Thank you so much, **Tim**, for the best hugs and many words of encouragement, which I definitely needed during the final stretch. In general, I would like to thank the current and former members of BIO at NRM for providing a warm (not literally) and encouraging environment. Unfortunately, it would be too much to list you all.

I am also grateful to all my friends and colleagues at Zootis for their positive feedback during conferences and bloodbaths and their great moral support. Among the many names I would like to thank in particular are **Alex R.**, **Bertil**, **Chris C.**, **Ciaran**, **David W.**, **Dick**, **Fanny**, **Gua**, **Javier**, **Jennifer**, **John**, **Kalle**, **Kathi**, **Loke,** **Lore**, **Philip**, **Rachel**, **Rike**, **VJ** and **Zahra**.