UPPSALA
UNIVERSITET

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 2239

Speciation genomics in *Ficedula* flycatchers

MADELINE CHASE

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2023

ISSN 1651-6214
ISBN 978-91-513-1711-3
URN urn:nbn:se:uu:diva-495937

Dissertation presented at Uppsala University to be publicly examined in Lindahlsalen, Evolutionary Biology Center, Norbyvägen 18A, Uppsala, Friday, 24 March 2023 at 10:00 for the degree of Doctor of Philosophy. The examination will be conducted in English. Faculty examiner: Professor Chris Jiggins (Department of Zoology, Cambridge University).

Abstract
Chase, M. 2023. Speciation genomics in *Ficedula* flycatchers. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology* 2239. 71 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-513-1711-3.

Understanding what evolutionary processes have shaped patterns of genomic differentiation between species is a major aim of speciation genomics. However, disentangling the role of different processes that generate similar patterns remains a substantial challenge. Within this thesis, I aimed to infer the action of different evolutionary processes through population-level genome re-sequencing of closely related species. I explored how processes such as recombination, natural selection, and genetic drift interact to shape the genomic differentiation landscape among multiple species of *Ficedula* flycatcher. Collared flycatcher and pied flycatcher are a pair of closely related species, which hybridize in regions of secondary contact. Reproductive isolation is strong and hybrids appear to be sterile. I compared the differentiation landscape between collared and pied flycatchers with a more distantly related species pair, the red-breasted and taiga flycatchers. This comparison revealed elevated regions of genomic differentiation shared between the two pairs, i.e. shared differentiation peaks, and those unique to a single pair, i.e. lineage-specific differentiation peaks. Since the two species pairs share a negligible portion of genetic variation, shared patterns in the differentiation landscape should be driven and maintained by conserved processes, while lineage-specific patterns should be driven by lineage-specific changes in relevant evolutionary processes. Selective sweep scans suggested that both shared and lineage-specific peaks can result from adaptive evolution and that lineage-specific adaptation is not a sufficient determinant of lineage-specific peaks. Instead, lineage-specific differentiation peaks appeared to be driven by evolutionary changes in the recombination landscape, the dynamics of which had strong impacts on the detection of signatures of linked selection. I also found that adaptation did not play a prominent role on Z-chromosome differentiation. Both the *fast-Z* and *large-Z* effects were apparent within the flycatchers but appeared to be primarily driven by the increased role of genetic drift on the Z-chromosome due to its reduced effective population size compared to the autosomes. I hypothesized that the increased impact of genetic drift could speed up the buildup of genetic incompatibilities of Z-linked and autosomal loci and contribute to reproductive isolation. Finally, using long-read and HiC sequencing data, I generated high-quality reference genomes for the collared flycatcher and pied flycatcher, and provided a first glimpse of the role of structural variation in speciation. I observed an increased prevalence of inversions and translocations on the sex chromosomes and in differentiation peaks. Structural rearrangements may therefore represent an important source of genomic variation contributing to species divergence.

Keywords: recombination, linked selection, selective sweeps, sex chromosomes, structural variation, birds, genome assembly

Madeline Chase, Department of Ecology and Genetics, Evolutionary Biology, Norbyvägen 18D, Uppsala University, SE-75236 Uppsala, Sweden.

ISSN 1651-6214
ISBN 978-91-513-1711-3
URN urn:nbn:se:uu:diva-495937 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-495937)

To our beautiful planet
and all the forms of life we share it with

List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I. **Chase, M. A.**, Ellegren, H., Mugal, C. F. (2021) Positive selection plays a major role in shaping signatures of differentiation across the genomic landscape of two independent *Ficedula* flycatcher species pairs. *Evolution*, *75*(9):2179–2196

II. **Chase, M. A.,** Mugal, C. F. (2022) The role of recombination dynamics in shaping signatures of direct and indirect selection across the *Ficedula* flycatcher genome. *bioRxiv* 2022-08 11.503468 doi:10.1101/2022.08.11.503468 (Manuscript)

III. **Chase, M. A.,** Vilcot, M., Mugal, C. F. Evidence that genetic drift not adaptation drives *fast-Z* and *large-Z* effects in *Ficedula* flycatchers. *bioRxiv* 2022-02 08.527632 doi:10.1101/2023.02.08.527632 (Manuscript)

IV. **Chase, M. A.,** Scofield, D. G., Kraft, F-K., Segami, J. C., Ålund, M., Qvarnström, A. Wheatcroft, D., Mugal, C. F. The combination of HiFi and HiC sequencing technologies enables the investigation of structural variants in speciation of *Ficedula* flycatchers. (Manuscript)

Additional Papers

The following papers were published during the course of my doctoral studies but are not part of this thesis.

Stankowski, S., **Chase, M. A.,** McIntosh, H., Streisfeld, M. A. (2023) Integrating top-down and bottom-up approaches to understand the genetic architecture of speciation across a monkeyflower hybrid zone. *Molecular ecology*

Borges, R., Kotari, I., Bergman, J., **Chase, M. A.,** Mugal, C. F., Kosiol, C. (2022) Traditional phylogenetic models are insensitive to variations in the effective population size. *bioRxiv* 2022.09. 26.509598

Stankowski, S.*, **Chase, M. A.*,** Fuiten, A. M., Rodrigues, M. F., Ralph, P. L., Streisfeld, M. A. (2019) Widespread selection and gene flow shape the genomic landscape during a radiation of monkeyflowers. *PLoS biology* 17 (7) e3000391

* Equal contributions

Contents

Abbreviations

KYA	thousand years ago
MYA	million years ago
BSC	biological species concept
BDMI	Bateson-Dobzhansky-Muller incompatibility
DSB	double strand break
CO	crossover
NCO	non-crossover
gBGC	GC-biased gene conversion
SNV	single nucleotide variant
SV	structural variant
HRI	Hill-Robertson interference
LD	linkage disequilibrium
SFS	site frequency spectrum
DFE	distribution of fitness effects
ILS	incomplete lineage sorting
BGS	background selection
π_N/π_S	nonsynonymous over synonymous polymorphisms
d_N/d_S	nonsynonymous over synonymous substitutions
WGA	whole genome alignment

Introduction

The variety of life present today arose from a common ancestor through a diversification process we call speciation, which is the event of species formation. Despite the universality of this statement, evolutionary biologists still struggle to agree upon a single definition of what a species is, which in part stems from the difficulty of articulating a unifying definition that encompasses all of life. Among sexually reproducing organisms one of the most commonly applied species concepts by speciation researchers is the biological species concept (BSC). The BSC defines species as groups of populations that are reproductively isolated from other populations, while freely exchanging genes among themselves (Mayr, 1963). As well as defining a species, this definition provides an explanation for the process of species formation, positing that speciation is the process by which reproductive isolation evolves.

Reproductive isolation evolves as the result of a myriad of individual 'reproductive barriers.' These barriers are broadly characterized as pre-zygotic and post-zygotic, based on what stage of reproduction they influence, and reduce the frequency of hybridization between budding species (Coyne & Orr 2004). Pre-zygotic barriers include ecogeographic isolation, divergence of mating display or mating timing, differences in reproductive tracts and gamete recognition, among others. Post-zygotic barriers are frequently divided into extrinsic and intrinsic barriers. Extrinsic barriers refer to the fitness of hybrids within their environment, and that they are maladapted compared to the parental species. Intrinsic barriers, on the other hand, restrict gene flow regardless of the external environment. The most frequent source of intrinsic post-zygotic isolation is thought to be the result of Bateson-Dobzhansky-Muller incompatibilities (BDMIs; Seehausen et al., 2014). BDMIs arise through negative epistatic interactions between loci, for example when different alleles have fixed in two species (Fig. 1). When hybrids are formed, the combination of alleles at different loci has never been tested by selection and may be incompatible.

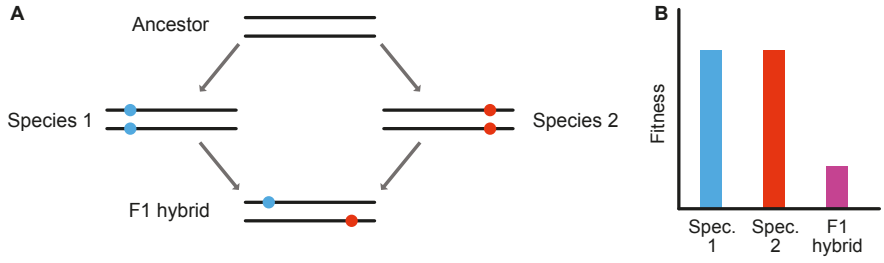

Figure 1: Example of how BDMIs can arise between two species. Shown in panel **A** are representations of two copies of a chromosome for an ancestral species, two descendent species, and their F1 hybrid. Species 1 fixes a derived mutation at one locus (blue circle), and species 2 fixes a derived mutation at a different locus (red circle). When the two species hybridize, the two derived mutations are brought into the same genome in the F1 hybrid. As the combination of these mutations has never been tested before by selection, they may result in genetic incompatibilities. In panel **B** is a hypothetical depiction of the fitness of species 1 (blue), species 2 (red), and the F1 hybrid (purple). Neither species 1 nor species 2 have a difference in relative fitness, but the hybrid experiences reduced fitness because of the genetic incompatibility.

There is substantial interest among speciation researchers in quantifying the cumulative effects of individual reproductive barriers on total reproductive isolation between closely related species. In addition, there are open questions surrounding the order in which pre-zygotic versus post-zygotic barriers arise in the process of speciation, as well as their relative importance to divergence. Intrinsic post-zygotic barriers are thought to be important for the long-term maintenance of species boundaries, due to their perceived permanence compared to ecologically based barriers that may break down if the environment changes (Coyne & Orr 2004). Theoretical work on hybrid zones suggests that pre-zygotic barriers will be insufficient to prevent the merging of two species without any post-zygotic barriers in place (Irwin, 2020). Incompatibilities also often segregate within populations, which suggests that intrinsic post-zygotic barriers could also help to initiate speciation (Coughlan & Matute, 2020). However, because pre-zygotic barriers will act before post-zygotic barriers in an organism's lifecycle, it has been emphasized that the relative importance of pre-zygotic barriers may be much greater than post-zygotic (Sobel & Chen, 2014). Some evidence also suggests that pre-zygotic barriers may evolve before post-zygotic barriers (Seehausen et al., 2014), and could therefore be critical in the earliest stages of speciation. However, caution must be taken when attempting to infer the historical order in which reproductive barriers arose when studying present day species.

In general, studying a historical process such as a past speciation event is a challenge because we are unable to observe the process occurring in real time. One common approach to address this limitation is to compare patterns of

divergence observed from multiple species pairs that are at varying stages of speciation (Wolf & Ellegren, 2017). This can be referred to as reconstructing the 'speciation continuum,' emphasizing the gradual nature of the speciation process. Of course, there are inherent limitations of this approach, as there is no guarantee that two independent pairs have followed the same trajectory during the divergence process, or that any two pairs represent stable species that will not collapse in the future. Nevertheless, this approach has helped expand our knowledge of species divergence and becomes increasingly useful when applied to a diverse array of organisms. Studying closely related species that may continue to exchange genes has also been powerful for understanding the earliest stages of speciation. One outcome from studies of the early stages of speciation was an emphasis of the importance of what is called 'ecological speciation' (Rundle & Nosil, 2005). In this scenario, ecologically based divergent selection is the driving force behind the evolution of reproductive isolation (Schluter, 2001). Often intertwined with ecological speciation is the idea of 'speciation-with-gene-flow,' with evidence suggesting that reproductive isolation could evolve in the face of ongoing gene flow (Nosil, 2008). It appears, however, that many examples of speciation where gene flow is currently observed likely experienced historical periods of allopatry, stressing that geographic isolation may frequently be an important aspect of species divergence.

Finally, although a view of speciation as the gradual evolution of reproductive isolation is common, the definition is not without disagreement. Recent attempts to define the speciation continuum revealed that while many authors use or recognize the term, few agree on what it actually tells us (Stankowski & Ravinet, 2021). Additionally, while some authors have emphasized the importance of quantifying a cumulative value of reproductive isolation between divergent species (Sobel & Chen, 2014; Westram et al., 2022), others have stressed that doing so is unlikely to be either possible or meaningful (Mallet & Mullen, 2022). Speciation can also involve much more than just the evolution of reproductive isolation, and divergence can occur across many axes. Furthermore, a definition fixated on reproductive isolation becomes difficult to apply to asexual organisms. Thus, some authors have advocated for a more multivariate view of speciation that encompasses the multiple axes along which divergence proceeds (Bolnick et al., 2023).

In this thesis, I focus on understanding species divergence at the genomic level by studying several closely related species of the *Ficedula* flycatcher genus. Through this work, I aim to improve our knowledge of how different evolutionary processes interact to shape patterns of genetic diversity and differentiation throughout the speciation process. In the following pages, I outline some of the relevant background information on speciation genomics and describe

how a variety of evolutionary processes and features of the genome are expected to play a role in speciation.

Speciation genomics

Initial studies within speciation genetics sought to identify the loci that were responsible for driving and maintaining species differences, so-called 'speciation genes.' The rise in high-throughput sequencing saw an explosion within this field, and the move from speciation genetics to speciation genomics. The extension of genome sequencing to non-model organisms opened up a new realm of opportunities, and the ability to compare patterns of sequence divergence along the genome of closely related species. Along with the technological advancement came the hope that we could more easily than ever identify the genes that were responsible for speciation.

Early approaches to study genome-wide patterns of differentiation between closely related species or locally adapted populations revealed striking heterogeneity in the extent of divergence when looking spatially along the genome (Ellegren et al., 2012; Martin et al., 2013; Renaut et al., 2013; Turner et al., 2005). These patterns became referred to as the genomic landscape. Because many of the systems within these studies were thought to be currently exchanging genes, the landscapes were interpreted as displaying the interaction between gene flow and divergent selection (Ellegren et al., 2012; Nadeau et al., 2012; Turner et al., 2005). According to this model, the genomic background shows low overall levels of differentiation due to homogenization by gene flow. Peaks of high differentiation, however, were thought to represent loci resistant to gene flow and therefore result from strong divergent selection or genetic incompatibilities. These loci were initially referred to as islands of speciation (Turner et al 2005), and subsequently the more agnostic term, islands of differentiation.

The interpretation of genomic landscapes in light of speciation-with-gene-flow was questioned in a reanalysis of several of the early datasets that put forth this hypothesis (Cruickshank & Hahn, 2014). A major critique of the work was that most studies focused on relative measures of divergence, primarily in the form of F_{ST}, which are sensitive to within species levels of diversity. If differentiation islands represented the few regions of the genome resistant to ongoing gene flow, they would also be expected to show higher *absolute* levels of divergence, estimated by d_{XY} (Fig. 2A; Nachman & Payseur, 2012). Reanalysis, however, revealed that elevated d_{XY} in F_{ST} peaks was rarely observed. On the contrary, peaks of relative divergence often showed lower levels of absolute divergence, as well as within species diversity. Such a pattern is expected if the diversity reducing processes were conserved from a

common ancestor (Fig 2B; Nachman & Payseur, 2012), as d_{XY} is sensitive to variation in ancestral levels of diversity. Moreover, demographic modeling revealed in many cases that ongoing levels of gene flow were negligible. As a result, it appeared the genomic landscape was less informative about the reproductive barrier loci underlying the speciation process than initially thought.

Alternative explanations for the evolution of differentiation islands include lineage-specific selective sweeps after divergence and background selection occurring on a conserved genomic architecture (Ravinet et al., 2017; Wolf & Ellegren, 2017). In the first paper to use the term speciation islands, the authors observed that islands were in fact found close to centromeric regions that typically experience low recombination (Turner et al., 2005). Subsequent studies reported a strong relationship with differentiation and recombination rate, where reduced diversity and increased differentiation coincide with low recombination rate (Burri et al., 2015; Corbett-Detig et al., 2015; Delmore et al., 2018). Natural selection is expected to have a wider impact on levels of linked neutral diversity when recombination rate is low, as neutral variants are unable to escape association with selected sites (Kaplan et al., 1989). The term linked selection has come to refer to the reduction in diversity at neutral sites linked to the focal site under selection, which may result from both purifying and positive selection (balancing selection, that increases genetic diversity, can also have linked effects but is often ignored by this term). This process leads to a lower effective population size (N_e) in regions of low recombination, which can in turn accelerate lineage sorting and cause differentiation islands to arise that have no role in speciation per se (Wolf & Ellegren, 2017).

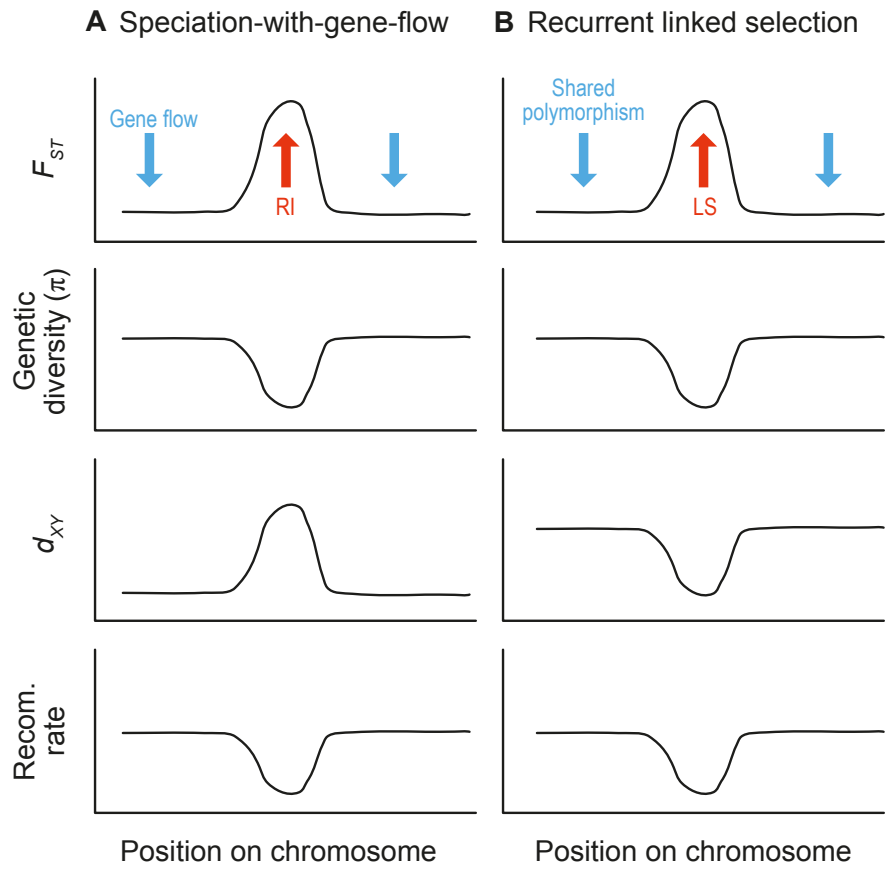

Figure 2: Predicted patterns in the genomic landscape under different scenarios. Shown are predicted patterns of relative genetic differentiation (F_{ST}), genetic diversity (π), absolute genetic divergence (d_{XY}), and recombination rate along a chromosome under two scenarios. In panel **A** are predictions for speciation-with-gene-flow and in panel **B** are predictions for recurrent linked selection (both BGS and positive selection). In panel **A**, the F_{ST} peak contains loci that contribute to reproductive isolation (RI), which drives patterns of elevated differentiation (red arrow). The rest of the chromosome is homogenized through gene flow (blue arrows) and differentiation remains low. The F_{ST} peak observed is also predicted to show lower within species diversity (π), and higher absolute divergence (d_{XY}) between species. We also predict that in the face of gene flow, loci contributing to RI are more likely to establish where recombination rate is low. In panel **B**, the F_{ST} peak is driven by linked selection (LS; red arrow). In this case, reduced N_e leads to faster lineage sorting in the region of the F_{ST} peak, while the rest of the chromosome shows lower levels of differentiation due to increased levels of shared ancestral polymorphism (blue arrows). We also expect to see lower within species diversity (π) and lower absolute divergence (d_{XY}) coinciding with the F_{ST} peak, as well as lower recombination rate.

While linked selection unrelated to ongoing gene flow can play a substantial role in shaping the genomic landscape for many species, the importance of background selection versus positive selection remains somewhat unclear. Observations that variation in patterns of diversity and differentiation were remarkably correlated among both closely and distantly related species, particularly within avian systems with a well conserved karyotype, (Burri et al., 2015; Delmore et al., 2018; Dutoit et al., 2017) suggested that background selection was potentially more important. That the very same regions would show elevated differentiation among species pairs with different ecological niches was unexpected due to parallel adaptation. Therefore, repeated differentiation peaks were hypothesized to more likely result from background selection, while lineage specific differentiation islands may be more likely to result from adaptive evolution (Berner & Salzburger, 2015; Burri, 2017). However, the drastic reductions in diversity observed within many differentiation islands may be difficult to explain by background selection alone. Indeed, studies based on simulation and mathematical modeling demonstrated that selective sweeps were necessary to explain all of the variation in diversity along the genome (Rettelbach et al., 2019; Schrider, 2020; Stankowski et al., 2019). Therefore, more work is necessary to understand what processes drive shared patterns of genomic differentiation between independent species. Comparative studies between multiple species pairs at varying stages of divergence can help us to understand how different evolutionary processes and sources of variation contribute to the speciation process.

Recombination rate and speciation

Meiotic recombination is a fundamental evolutionary process that plays a major role in shaping genome-wide variation. Through the reshuffling of genetic variation, recombination creates new combinations of alleles and is thought to provide one of the major advantages of sexual reproduction. By breaking apart linkage between sites, recombination can increase the efficacy of selection on them. On the other hand, this same function can break apart co-adapted sites, reducing fitness. Breaking apart variation also means that recombination can oppose the speciation process, by contributing to the homogenization of variation between diverging populations. Therefore, the evolution of the suppressed recombination has received a great deal of attention within speciation genomics because this can contribute to species divergence in the face of gene flow. Below, I outline some of the mechanisms behind meiotic recombination, how variation occurs across the genome, and what some of the evolutionary consequences of this variation can be.

Molecular mechanisms of recombination

Meiotic recombination is induced by DNA double-strand breaks (DSBs; Gray & Cohen 2016). There are several methods of DSB repair, which can result in a crossover (CO), the reciprocal exchange of genetic material between homologous chromosomes, or a non-crossover (NCO). In most species the majority of DSBs are not repaired through COs, but by NCO events (Zelkowski et al., 2019). Nevertheless, at least one CO per chromosome appears to be nearly ubiquitously necessary for the proper segregation of chromosomes during meiosis (Zickler & Kleckner, 2015). In addition to the reciprocal exchange of genetic material between chromosomes, DSBs can also be repaired by gene conversion. When gene conversion occurs, the genetic material from one of the homologous chromosome pairs is copied into the other chromosome. A specific instance of gene conversion occurs when the locus contains a GC/AT polymorphism, which is known as GC-biased gene conversion (gBGC). In many species a bias is observed towards preferentially replacing the AT copy of the chromosome with the GC copy. This fixation bias has led to the observation of a positive correlation between recombination rate and GC content across (Duret & Galtier, 2009), and can interfere with fitness (Glémin, 2010; Xue et al., 2016). DSB repair can also generate novel mutations, including several types of structural rearrangements, through mechanisms such as non-allelic homologous recombination (NAHR). NAHR is particularly associated with repetitive regions, such as transposable elements (TEs), where recombination occurs between sequences with high similarity but that are not alternate copies of the same allele (Klein & O'Neill, 2018).

Variation in recombination rate

Recombination rate can vary along different scales, including between separate species, between different populations of the same species, between the sexes, and within a single individual spatially along the genome (Stapley et al., 2017). This genome-wide variation can also occur at both the fine-scale and the broadscale. While fine-scale variation in recombination rate is primarily the result of hotspots, broadscale variation depends more on the local genomic architecture, such as the proximity to centromeres. Within many species, there appears to be a greater proportion of COs around the telomeres, with broad regions around the centromeres (pericentromeric regions) experiencing little recombination (Zelkowski et al., 2019). Even in species where centromeres are not centrally positioned on the chromosome, the chromosome centers appear to experience reduced recombination (Haenel et al., 2018). Additionally, CO interference appears to place an upper limit on the number of COs that can occur on a single chromosome (Gray & Cohen, 2016). Because at least one CO per chromosome appears to be necessary for proper segregation during meiosis, we often observe a negative relationship between average

recombination rate and chromosome size, with smaller chromosomes experiencing a higher average rate of recombination (Kawakami et al., 2014). Consequently, these smaller chromosomes also often exhibit higher GC content due to a greater occurrence of gBGC (Weber et al., 2014).

Although it may seem counterintuitive, the fine-scale and broadscale patterns of recombination can evolve at different rates between species (Smukowski & Noor, 2011). In many mammals, the location of recombination hotspots is governed by the zinc-finger protein, PRDM9 (Baudat et al., 2010; Myers et al., 2010; Parvanov et al., 2010). Within species with PRDM9 mediated hotspots, fine-scale recombination patterns can turnover rapidly (Ponting, 2011), with few shared recombination hotspots even between closely related species and sometimes between different populations of the same species (Stevison et al., 2016; Wooldridge & Dumont, 2023). On the other hand, broadscale patterns appear evolutionarily more conserved (Smukowski & Noor, 2011). The degree of evolutionary conservation of recombination rate also varies drastically among different lineages. For instance, recombination landscapes appear to be remarkably more conserved among birds compared to mammals, particularly at the fine-scale (Kawakami et al., 2017; Singhal et al., 2015). The increased conservation of fine-scale recombination among birds is attributed to the lack of PRDM9 in hotspot determination. Hotspots instead appear to localize to functional regions such as transcription start sites, transcription stop sites, and CpG islands (Singhal et al., 2015). At the broadscale, the high degree of synteny among even highly divergent species of birds does appear to contribute to a relatively stable recombination landscape (Ellegren, 2010). However, comparisons of zebra-finch and collared flycatcher recombination rates did show differences in broadscale patterns (Singhal et al., 2015). Changes in recombination rate between closely related species could also be selected for during the speciation process as a potential mechanism for reinforcement of reproductive barriers (Ortiz-Barrientos et al., 2016).

Consequences of recombination rate variation

Variation in recombination rate has a range of evolutionary consequences. Reduced recombination is thought to aid in the evolution of reproductive isolation between closely related species in the face of ongoing gene flow (Ortiz-Barrientos et al., 2016), and genomic signatures of gene flow have been observed to negatively correlate with recombination rate in several systems (Christmas et al., 2021; Martin et al., 2019; Ortiz-Barrientos et al., 2016; Schumer et al., 2018). At the same time, in the absence of gene flow, regions of lower recombination may be expected to evolve elevated differentiation between species at a faster rate (Wolf & Ellegren, 2017), leading to the occurrence of differentiation islands. We also expect that genomic signatures of

selection such as selective sweeps will show a negative relationship with recombination rate. Selective sweep signatures are based on the reduction in neutral diversity at sites linked to the focal site under selection, and are therefore greatly dependent on the recombination rate (Smith & Haigh, 1974). The strength of a selective sweep dramatically reduces as recombination rate increases (Kaplan et al., 1989). Thus, our ability to detect genomic signatures of selection can be greatly influenced by the local recombination rate.

Recombination rate is not only key in our ability to detect signatures of linked selection, but can also directly impact the efficacy of natural selection. Linkage between multiple polymorphic sites that have fitness effects can lead to selective interference, also known as Hill-Robertson interference (HRI; Hill & Robertson, 1966). When beneficial mutations at different sites are found on different haplotypes, this can reduce the efficacy of selection for both sites. In this case, regions of low recombination are expected to have a lower efficacy of selection due to the tighter linkage between selected sites; because recombination releases the association, the efficacy of selection is predicted to increase with increasing recombination rate.

The role of structural variants in speciation

An important source of genetic variation comes in the form of structural variants (SVs), which are defined as variants impacting 50bp or more. In contrast to single nucleotide variants (SNVs), mutations that affect a single basepair, there have been fewer speciation genomic studies focusing on the role of SVs. When studies do investigate the role of SVs, there is a tendency to focus on a specific sub-category, rather than comprehensively investigating the variety of SVs. Although we use the umbrella term ‘structural variant’, this refers to several distinct mutation types. SVs can be broadly divided into two categories: unbalanced variants, which change the genome size, and balanced variants, which leave the genome size unchanged (Mérot et al., 2020). Under the class of unbalanced variants fall insertions and deletions, as well as tandem duplications and whole genome duplications. Balanced variants include translocations and inversions, where a portion of the genome is moved to a new region or the orientation of the region is inverted, respectively. Additionally, chromosomal fissions or fusions are considered balanced SVs.

SVs can impact a much larger portion of the genome compared to SNVs and it has been suggested that they encompass more of the genetic diversity among individuals (Catanach et al., 2019; Pang et al., 2010). It also appears that at least some types of SVs experience a higher mutation rate than SNVs (Conrad & Hurles, 2007; Katju & Bergthorsson, 2013). Other work, however, has found a lower mutation rate for SVs compared to SNVs (López-Cortegano et

al., 2023), thus this is an area that merits further investigation. SVs are thought to have large phenotypic and fitness effects, which are likely often deleterious (Feulner & De-Kayne, 2017). Beyond detrimental effects, however, several studies have also implicated SVs as playing a large role in adaptation and speciation (Dorant et al., 2020; Hämälä et al., 2021; Weissensteiner et al., 2020). In particular, inversions have received much attention, because the lack of recombination within heterozygotes prevents the shuffling of alleles between the alternate orientations (Jay et al., 2018; Lowry & Willis, 2010; Matschiner et al., 2022). This has led in some cases to the evolution of co-adapted sets of alleles that are captured within an inversion, given the term 'super-gene'. Additionally, it appears that inversions can help to maintain intrinsic genetic incompatibilities in the face of gene flow upon secondary contact of two divergent species (Hooper et al., 2019; Noor et al., 2001).

Although inversions appear to play an important role in species divergence, it is unclear to what extent this represents a detection bias. Studies investigating the complete diversity of SVs segregating within a species and divergent between species are lacking in comparison to studies focused on differentiation based on SNVs. The reason for this is, in part, due to the greater difficulty in detecting SVs compared to SNVs. Previously, population level sequencing efforts have mainly been feasible only with short-reads, which are generally much smaller than the size of SVs. SV detection from short reads can be performed based on indirect methods such as local PCA to detect 'haploblocks' with increased linkage disequilibrium (LD), or directly through analysis of read-mapping patterns (Mérot et al., 2020). However with short-read data these methods are typically unable to identify large SVs (Mérot et al., 2020). Compared to SNVs, SVs also suffer more greatly from reference bias. While the problem of reduced detection due to the large size of SVs is somewhat mitigated through sequencing with long-reads, reference bias problems can remain in read mapping-based detection approaches. An approach that minimizes the problems of reference bias is to detect SVs based on whole genome alignments (WGA) of multiple genome assemblies (Mahmoud et al., 2019). This approach is, however, also limited due to the availability of reference quality genome assemblies, which has previously been prohibitive for population level investigations. The increased availability of long-read sequencing at the population scale is beginning to reduce this limitation. More and more, there is a move towards generation of pan-genome graphs, where assemblies of multiple individuals within a species (or between closely related species) are aligned and structural variation can be detected through pairwise comparisons of assemblies (Eizenga et al., 2020; Hickey et al., 2022; Ruggieri et al., 2022). This approach will likely increase the ability to reliably detect SVs among species, enabling a better understanding of their diversity and contribution to speciation.

The role of sex chromosomes in speciation

Sex chromosomes have a number of features that lead to distinct evolutionary consequences compared to the autosomes and they can disproportionately contribute to speciation. Within this section, I will refer only to the Z/W sex determination system, where males are ZZ and females are ZW. The properties discussed are also relevant to an X/Y system with the sexes reversed. In the Z/W sex determination system (as is the case in birds), for each four copies of autosomes there are only three copies of the Z-chromosome in the population due to being present in only one copy in females. Under the assumption of equal reproductive variance between the two sexes, the ratio of Z:A diversity is therefore expected to be 0.75 (Vicoso & Charlesworth, 2009). In the absence of selection, this ratio can vary between 9/16 to 9/8 based on variation in reproductive success between males and females (Irwin, 2018; Vicoso & Charlesworth, 2009). Male-biased mutation rates can also impact the Z:A diversity ratio by increasing diversity on the Z-chromosome. Evidence from a number of species suggests that male-biased mutation rates are common among birds and should be corrected for when estimating the Z:A diversity ratio (Irwin, 2018), because the Z-chromosome is more often exposed to the male biased mutation rate compared to the autosomes. Outside of the pseudoautosomal region (PAR), the Z-chromosome only recombines within males. Thus, linked selection is also expected to have a greater impact on diversity on the Z-chromosome and may reduce the Z:A diversity ratio even more than expected (Hammer et al., 2010).

The hemizygosity of the Z-chromosome leads to two key phenomena related to species divergence. These are the *fast-Z* effect and the *large-Z* effect. Because Z-linked recessive mutations are immediately exposed to selection in the heterogametic sex, the efficacy of selection on adaptive mutations can be increased on the Z-chromosome. The Z-chromosome may therefore experience a faster rate of evolution compared to the autosomes. Although the *fast-Z* effect may be driven by an increased efficacy of selection, there is also evidence to suggest that in birds this is primarily driven by an increased role of genetic drift due to the lower N_e (Hayes et al., 2020; Mank et al., 2010; Wang et al., 2014). In this case, lineage sorting occurs more rapidly and the Z-chromosome will accumulate genetic divergence faster than the autosomes, while also experiencing a reduced efficacy of selection. Whereas the *fast-Z* effect describes the increased evolutionary rate experienced by the Z-chromosome, the *large-Z* effect refers to the increased propensity of the Z-chromosome to evolve genetic incompatibilities. The *large-Z* effect is also driven by the hemizygosity of the Z-chromosome and is related to Haldane's rule, which states that the heterogametic sex tends to evolve sterility and inviability at a faster rate compared to the other sex. As with the *fast-Z* effect, the extent to which the *large-Z* effect is driven by adaptive evolution compared to genetic drift is

equivocal. Moreover, it can be challenging to distinguish between the two phenomena, as they can lead to similar patterns and are not necessarily mutually exclusive.

Methods

In this section, I outline the key methods applied in my thesis. Detailed descriptions of methods can be found within each paper.

Study system: *Ficedula* flycatchers

Throughout this thesis, I investigate the history of genomic divergence among multiple species of *Ficedula* flycatchers, a group of passerine birds within the family Muscicapidae. In **papers I**, **II**, and **III**, I focus on patterns of genomic differentiation between two species pairs, the collared (*F. albicollis*) and pied (*F. hypoleuca*) flycatchers, and the red-breasted (*F. parva*) and taiga (*F. albicilla*) flycatchers. Within **paper IV**, I focus only on collared flycatcher and pied flycatcher.

Collared flycatcher and pied flycatcher are two closely related species (Fig. 3), with a divergence time of less than 1 MYA (Nadachowska-Brzyska et al., 2013). The two species belong to a group of four black-and-white flycatchers, including the atlas flycatcher (*F. speculigara*) and semicollared flycatcher (*F. semitorquata*), which all rapidly diverged from one another (Nater et al., 2015). The collared flycatcher and pied flycatcher have been a model system for speciation research, with an emphasis on studying hybridization and the evolution of reproductive isolation between the two species where their ranges overlap in regions of secondary contact. Both male and female F1 hybrids appear to be sterile, and there is no evidence for ongoing backcrossing (Ålund et al., 2013; Svedin et al., 2008); there is also evidence for pre-mating barriers by assortative mating (Sæther et al., 2007). With their estimated divergence time of less than one MYA, hybrid sterility has evolved rapidly between collared and pied flycatchers, with complete sterility in birds occurring between species with an average divergence of 7 MYA (Price & Bouvier, 2002). Demographic modelling does, however, predict historical gene flow between the two species during the divergence process (Nadachowska-Brzyska et al., 2013).

Red-breasted flycatcher and taiga flycatcher are sister species, which are more deeply diverged from one another than collared and pied flycatcher (Fig.3;

Hung & Zink, 2014) and form a sister group to the black-and-white flycatchers. Despite their deeper divergence, red-breasted flycatcher and taiga flycatcher were taxonomically considered to be the same species until 2004, although suggestions that they be given species status were made much earlier. It is unknown to what extent there is an opportunity for hybridization to occur between the two species, if their ranges overlap during the breeding season. Songs of the two species are different (Svensson et al., 2005), leading to the possibility for premating isolation if there is an opportunity for hybridization. Previous genetic analyses of the two species have been limited to a few mitochondrial and nuclear genes, with no genome-wide analyses (Hung & Zink, 2014; Zink et al., 2008).

Previous genomic investigations of speciation in *Ficedula* flycatchers have focused on the group of four black-and-white flycatchers (Burri et al., 2015; Nadachowska-Brzyska et al., 2013, 2016; Nater et al., 2015), which share a recent common ancestor and large amounts of shared ancestral polymorphisms. These studies revealed that among different comparisons of the four species patterns of genome-wide variation were strikingly similar, with the same genomic regions repeatedly showing elevated differentiation. These studies did have the limitation that the similarities in the genomic landscape observed could result primarily from shared ancestry rather than through the action of conserved evolutionary processes. By incorporating data from a more distantly related species comparison with red-breasted flycatcher and taiga flycatcher, I can address this limitation here.

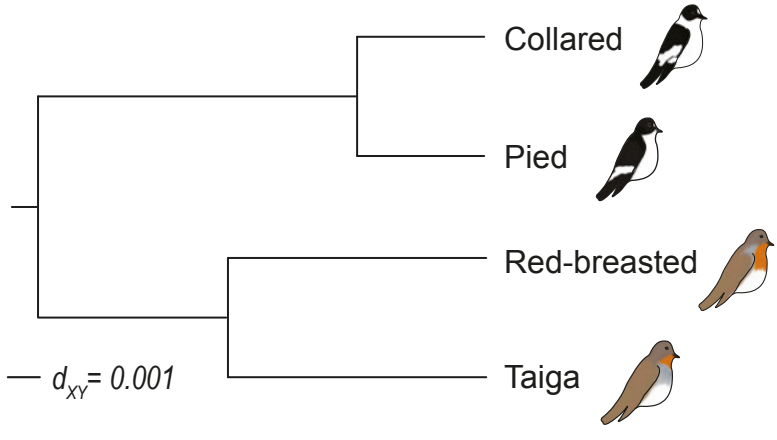

Figure 3: Topology of relationships among four flycatcher species. Shown are depictions of the four flycatcher species (collared flycatcher, pied flycatcher, red-breasted flycatcher and taiga flycatcher) that are the focus of this thesis. Branch lengths are based on estimates of d_{XY} between species pairs estimated in **paper I**.

Life history traits and demographic history of the four species

The four flycatcher species studied here differ in several important life history traits. Both collared flycatcher and pied flycatcher exhibit polygynous mating, while red-breasted flycatcher is found to exhibit monogamous mating (data not available for taiga flycatcher; Storchová & Hořák, 2018). Collared flycatcher on average has a slightly smaller clutch size (6 eggs) compared to pied flycatcher (6.5 eggs), while the red-breasted flycatcher has a smaller average clutch size (5.5 eggs) than both collared flycatcher and pied flycatcher (Storchová & Hořák, 2018). At the same time, the red-breasted flycatcher chicks mature somewhat faster (21 days) than both collared flycatcher (27 days) and pied flycatcher (25 days) chicks (Storchová & Hořák, 2018). Additionally, both the red-breasted flycatcher and taiga flycatcher are on average smaller in mass (10.8 g) than collared flycatcher (12.7 g) and pied flycatcher (13.8 g; Tobias et al., 2022).

Of the four species, taiga flycatcher has the largest range size (~10000000 km^2), followed by pied flycatcher (~7000000 km^2), red-breasted flycatcher (~6000000 km^2), and finally collared flycatcher (~2000000 km^2; Tobias et al., 2022). However, the present-day range size does not reflect what is known about the historical fluctuation in population size and variation in genetic diversity among the species. Pied flycatcher has been observed to have the lowest genetic diversity of the four species and shows historically lower estimates of population size. The remaining three species display similar levels of genetic diversity, but have different fluctuations in population size. Both the collared flycatcher and red-breasted flycatcher are inferred to have recent declines in population size, while the taiga flycatcher is inferred to have a recent population expansion. The island populations of collared flycatcher in the Baltic Sea are known to have undergone very recent population bottlenecks, with colonization of Gotland occurring approximately 150 years ago (Lundberg & Alatalo, 1992).

The avian genome

Bird genomes have many characteristics that make them an interesting system in which to investigate the genomics of species divergence. The avian karyotype shows remarkable evolutionary stability, with most species having approximately n = 40 chromosomes (Ellegren, 2010). Interchromosomal rearrangements appear to have been rare in most avian lineages in comparison to other organisms (Griffin et al., 2007), which may be related to the comparatively low repeat content of bird genomes with an average of 10% (Kapusta & Suh, 2017). Bird genomes are also relatively small and have low variation in genome size compared to other lineages, with a range between 0.96 Mb – 2.2 Mb (Kapusta et al., 2017).

Most of the genome in birds is composed of what are referred to as macrochromosomes. Although these chromosomes typically comprise most of the sequence of the genome, there are generally many more microchromosomes than macrochromosomes. The distinction between chromosome sizes is somewhat arbitrary, and a convention developed within chicken defines microchromosomes as less than 20Mb, intermediate chromosomes as 20Mb to 50Mb and macrochromosomes as ~50Mb to 200Mb (Hillier et al., 2004). Chromosome size appears to have important evolutionary consequences, as microchromosomes generally exhibit higher average recombination rates owing to the obligatory one cross over per chromosome, higher GC content presumed to result from increased gBGC, and a higher mutation rate (Bolívar et al., 2016). At the same time, microchromosomes display a lower d_N/d_S ratio compared to macrochromosomes, suggestive of increased evolutionary constraint (Axelsson et al., 2005). Partly owing to the high GC content of microchromosomes, the smallest chromosomes have historically been challenging to sequence and assemble, therefore are missing from many avian reference genomes including the collared flycatcher.

Another important feature of avian genomes is that they display female heterogamety, having evolved a ZW sex determination system. The W chromosome is highly degenerated compared to the Z-chromosome and has been challenging to assemble in many species; the few W-chromosome assemblies that exist show a much greater repeat content compared to autosomes (Peona et al., 2021; Smeds et al., 2015).

Population genomic dataset

In **papers I**, **II** and **III**, I work with a population genomic dataset collated from previously published data, as well as newly sequenced individuals. The dataset includes 95 collared flycatchers from the Swedish island of Gotland (Nadachowska-Brzyska et al., 2021), 11 pied flycatchers from the Swedish mainland (Burri et al., 2015), 15 red-breasted flycatchers from western Russia, 65 taiga flycatchers from eastern Russia and Mongolia, and one snowy-browed flycatcher (Burri et al., 2015), included as an outgroup. All samples were sequenced with Illumina, and raw reads were aligned to the collared flycatcher reference genome, FicAlb1.5 (Kawakami et al., 2014). I then performed variant calling for all individuals with GATK (McKenna et al., 2010).

To obtain high confidence SNVs, I applied a range of filtering criteria. Firstly, best practice filtering thresholds were applied following recommendations set by GATK. Genotypes with less than 5x coverage or greater than 200x coverage were set to missing, as well as genotypes with a genotype quality (GQ) lower than 30. Sites overlapping with annotated repeats from the collared

flycatcher reference genome were removed (Suh et al., 2018). Additionally, I developed an approach to identify regions representing collapsed duplications within the reference genome based on excessively high heterozygosity. Sites within collapsed regions were then removed.

Because having a reliable estimate of the total number of genotyped sites is critical for obtaining per site measures of genetic diversity and divergence, in **paper I** the invariant sites were estimated from the alignment files by obtaining the genomic regions from each individual with sufficient coverage to identify a variable site. Repeats and collapsed duplications were also masked in the invariant sites to replicate the SNV filtering process. In **papers II** and **III**, I applied a more sophisticated approach and obtained 'all-sites' VCF files from GATK. In this way, invariant sites can be processed in as similar a method as variant sites as possible; however, some statistics used to filter SNVs are still only estimated for variable positions.

Estimating recombination rate

Recombination is a key evolutionary process that plays a major role in shaping genomic signatures of selection and patterns of genetic diversity. Estimates of local variation in recombination rate are therefore critical for understanding patterns in the genomic landscape. In **paper I**, I make use of a previously established recombination map for collared flycatcher, estimated from a four generation pedigree (Kawakami et al., 2014). Having pedigree information for one of the species studied here is an asset, as recombination events within the pedigree can be directly inferred. However, because a pedigree is only available for one species, I am unable to address with this information whether evolutionary changes in recombination rate have occurred that can explain any differences in selection signatures. Because of the intensive tracking of multiple generations required to generate a linkage map, it is intractable to obtain these data from all the species studied here. To this end, in **paper II** I use a separate approach that estimates recombination rate indirectly from population resequencing data based on patterns of linkage disequilibrium (LD) for both collared flycatcher and taiga flycatcher.

Statistical phasing

To improve the accuracy of LD based recombination rate estimates, I performed statistical phasing on population re-sequencing data for collared flycatcher and taiga flycatcher separately. Statistical phasing is improved by increased sample sizes (Nadachowska-Brzyska et al., 2019), therefore I used collared flycatcher and taiga flycatcher as the representative species for the branch separating the two groups, since they had 95 and 65 samples sequenced

respectively. I performed phasing for both species separately using all available samples with shapeit2 (Delaneau et al., 2012), providing genome-wide values of rho from previous estimates for collared flycatcher and values of theta from estimates of diversity for both species. The statistical phasing was performed independently twice, and the haplotypes inferred from both runs were then used to estimate recombination rate to account for stochasticity associated with the phasing step.

LD based recombination rate

The software LDhelmet was used to estimate recombination rate by taking the 25 unrelated samples with the least amount of missing data for both collared flycatcher and taiga flycatcher, because the software is limited to 50 haplotypes. LDhelmet builds upon the software LDhat for estimating recombination rate from phased haplotypes, introducing several improvements (Chan et al., 2012). Both software use a Bayesian approach, computing a pairwise composite likelihood for each pair of SNPs under the coalescent with recombination. LDhelmet improves this likelihood computation by incorporating a quadra-allelic mutation matrix and ancestral allele priors for each site. Additionally, the inclusion of missing data, which is common in sequencing data, causes little computational cost. Positive selection has been shown to produce spurious recombination hotspots (Chan et al., 2012; Reed & Tishkoff, 2006), however this was improved within LDhelmet. Based on simulations, LDhelmet is able to produce more robust estimates of recombination rate in the presence of positive selection, but does suffer from underestimating the recombination rate in the case of recent, strong selective sweeps (Chan et al., 2012).

The identification of recombination hotspots from LD has been shown to suffer from greatly reduced statistical power in response to recent population bottlenecks (Dapper & Payseur, 2018), and LDhelmet is shown to somewhat underestimate recombination rates in the event of a bottleneck (Chan et al., 2012). Nevertheless, simulations have demonstrated that while fine-scale estimates of recombination are sensitive to demography, at the broadscale the estimates are more robust (Raynaud et al., 2022). Thus, I focus primarily on variation in recombination rate between species at the broader scale, averaging recombination rate into 200-kb windows along the genome.

LDhelmet generates the population scaled recombination rate, ρ, which is equal to $4N_e r$. It is possible to convert this measure to cM/Mb with the help of a genetic map. Assuming that the map length for collared flycatcher and taiga flycatcher is equal (a reasonable assumption based on the close relationship between the species), I used the collared flycatcher pedigree-based recombination map to convert the population scaled recombination rate to cM/Mb for both species (Kawakami et al., 2014).

Estimates of diversity and differentiation

Genome-wide variation in patterns of genetic diversity and differentiation can reveal information regarding the history of natural selection. To estimate variation across the genome, I divide the genome into non-overlapping windows of a certain physical length and estimate our summary statistics for the nucleotides contained within each window. I then move along the genome by a given step size and calculate the same statistics in each window. This approach is referred to as a sliding window analysis. When the window size is the same as the step size, the windows are nonoverlapping. In this way, I can observe genome-wide variation in measures of diversity and divergence, in addition to obtaining a single point estimate averaged across the whole genome.

Nucleotide diversity: π

One estimate of genetic diversity is π, which is a measure of pairwise sequence differences within a sample of individuals. To obtain a per site measure, this is normalized by the total length of the sequence. π is an estimate of the population scaled mutation rate, $4N_e\mu$, where N_e is the effective population size and μ is the mutation rate. Directional selection is expected to decrease π. However, population bottlenecks will also lead to a reduction in π. I estimate π using the allele frequencies, p and q, and the formula:

$$\pi = \frac{\sum_{i=1}^{s} 2p_i q_i}{L}$$

where L is the total length of callable sites in the region and s is the number of polymorphic sites. I estimate π in this way in **papers I**, **III**, and **IV**.

Allele frequency differentiation: F_{ST}

The summary statistic F_{ST} provides a relative measure of differentiation between two populations or species. The measure is referred to as relative because it is affected by the level of within species diversity, and increases as diversity within species decreases. There are several methods to estimate F_{ST}, and here I use the approach by (Weir & Cockerham, 1984). Regions of the genome with locally elevated F_{ST} values have regularly been referred to as 'peaks' or 'islands' of differentiation. Identifying statistically significant F_{ST} peaks is a challenge, and I applied two different methods to identify peaks in my thesis. In **paper I**, I performed a permutation test, where I randomized the position of each genomic window and applied a smoothing algorithm to the permuted F_{ST} values. I then converted the smoothed F_{ST} values to Z-scores based on the mean and standard deviation for each chromosome, and finally obtained a p-value for the estimate of F_{ST} for each window based on how many

permutations the permuted z-transformed F_{ST} was as high or higher than the observed value.

For **papers II**, **III**, and **IV** I instead identified F_{ST} peaks as windows with a z-transformed F_{ST} value greater than 2, i.e. windows with F_{ST} greater than two standard deviations from the chromosome mean. Performing these tests by chromosome was important, as the average F_{ST} value is higher for microchromosomes compared to macrochromosomes. Additionally, F_{ST} is on average higher on the Z-chromosome compared to the autosomes.

Absolute sequence divergence: d_{XY}

Whereas F_{ST} is a relative measure of differentiation, d_{XY} is referred to as an absolute measure of divergence. This is because d_{XY} is not affected by variation in levels of within species diversity. d_{XY} is a measure of pairwise sequence divergence between two species, and is analogous to π. In **paper I** and **IV**, I calculate d_{XY} using a similar approach to π, using the allele frequencies, p and q, in species 1 and species 2 with the following formula:

$$d_{XY} = \frac{\sum_{i=1}^{s} p_{1i}q_{2i} + p_{2i}q_{1i}}{L}$$

where L is again the total length of callable sites and s is the number of polymorphic and divergent sites.

While π is a measure of the population scaled mutation rate, $d_{XY} = 4N_e\mu + 2\mu t$ where t is the time since divergence, and $4N_e\mu$ refers to the diversity within the ancestral population. Therefore, although d_{XY} is unaffected by variation in diversity within species, it is affected by variation in ancestral levels of diversity, and this will make a comparatively larger contribution when the divergence time is short. Lower d_{XY} can occur in a genomic region where diversity has been persistently reduced since the common ancestor of the two species. Examining the variation in d_{XY} has been proposed as an approach to determine whether F_{ST} peaks truly represent reductions in effective migration or result from the long-term action of linked selection. If an F_{ST} peak is resistant to gene flow, then we would expect to also observe elevated d_{XY} in the same region. However, if an F_{ST} peak reflects the long-term reduction in diversity within a region due to ongoing linked selection, then we would expect to observe lower d_{XY} within the region.

Identifying selective sweeps

A selective sweep occurs when a new mutation, or a mutation segregating at low frequency in a population, experiences positive selection and sweeps to fixation. In the process, neutral diversity at linked sites is reduced, which leaves the distinctive sweep signature of a drastic reduction in genetic diversity. The size of the region with low diversity depends on the relationship between the strength of selection and the local recombination rate. As the recombination rate increases, the sweep signature is weakened. Because the signature of a selective sweep depends on the distortion of patterns of diversity at neutral sites, in **paper II** I refer to this as a signature of indirect selection.

Within this thesis I apply several methods to detect selective sweeps that utilize the site frequency spectrum (SFS), however other methods also exist that are based on signatures of increased LD or extended haplotype homozygosity (EHH). Selective sweeps are expected to cause a shift in the SFS with an increase in both rare variants and high frequency derived variants.

Ancestral allele identification

Knowledge of the ancestral state for variable sites can increase the power to detect selective sweeps (Huber et al., 2016). I applied a parsimony approach to polarize sites, including snowy-browed flycatcher as one outgroup, and I polarized sites variable within collared and pied flycatcher by using the red-breasted and taiga group as a second outgroup and vice versa. The ancestral allele was identified when two of the three groups (snowy browed flycatcher, collared and pied flycatcher, red-breasted and taiga flycatcher) were fixed for the same allele. With this approach, I am therefore unable to polarize divergent sites between the two species pairs.

CLR statistic

For each species, selective sweeps were detected using SweepFinder2 (DeGiorgio et al., 2016), which computes the composite likelihood ratio (CLR) statistic as a measure of selective sweeps (Nielsen et al., 2005). The test takes the site frequency spectrum and determines whether the SFS at a particular site is best explained by a neutral model, or a model including positive selection. The genome-wide background SFS from the data is supplied to build the neutral model. Basing the neutral model from the data, rather than using a standard neutral model reduces the false positive rate due to demography, although strong recent bottlenecks are still problematic for this approach (Stephan, 2019). As recommended, for each species I removed sites that were fixed for the inferred ancestral allele, as this can increase the false positive rate

and may reflect regions with low diversity due to low mutation rates (Huber et al., 2016).

To determine a critical threshold of the CLR statistic to identify statistically significant sweep signatures, I ran simulations using SLiM3 (Haller & Messer, 2019). Because background selection (BGS) can also reduce genetic diversity, I ran simulations incorporating BGS into a chromosome with varying recombination rate and gene density, based on one chromosome from the collared flycatcher reference genome. I used estimates of the distribution of fitness effects (DFE) and N_e from previous estimates for collared flycatcher (Bolívar et al., 2018; Nater et al., 2015). After generating simulated data, I ran SweepFinder2 to determine the critical CLR value for positive selection. Selective sweeps were identified for the autosomes in **paper I** and **II**, and for the Z-chromosome in **paper III**.

Tajima's D and Fay and Wu's H

Other methods for selective sweep detection based on the SFS include Tajima's D and Fay and Wu's H, which I use in **paper I**. Tajima's D is calculated based on the difference between two estimates of the population mutation rate θ, θ_π (based on pairwise sequence differences) and θ_W (based on the number of segregating sites). Under neutrality, both estimates of θ should be the same, therefore the difference should be zero. Following a selective sweep, we would instead expect a negative value of Tajima's D, because θ_π will be more strongly reduced than θ_W due to an excess of rare variants. However, Tajima's D is also strongly affected by demography, with a negative value also reflecting a population expansion. Thus, it is important to compare to a genome-wide background.

Fay and Wu's H provides another estimate for selective sweeps, which takes a similar approach to Tajima's D, but relies upon the expectation that sweeps should show an excess of high frequency derived alleles. The test therefore requires polarized data.

Signatures of direct selection

Estimating π_N/π_S

The ratio of nonsynonymous to synonymous polymorphisms (π_N/π_S) segregating within a species can inform us about the strength of purifying selection. Generally, we assume that beneficial mutations rapidly reach fixation and are therefore rarely observed as polymorphisms. Thus, we expect that most nonsynonymous mutations segregating within populations are weakly deleterious.

In **paper II**, I estimate π_N/π_S for collared flycatcher and taiga flycatcher, for sets of genes with varying levels of recombination rate to assess the relationship between recombination rate and the strength of purifying selection. In **paper III**, I estimate π_N/π_S for the four flycatcher species (collared flycatcher, pied flycatcher, red-breasted flycatcher, and taiga flycatcher) for genes on the autosomes and the Z-chromosome, which provides an estimate of the efficacy of selection on the two chromosome types.

In both papers, I estimate π_N/π_S by identifying zero-fold and four-fold degenerate sites in the collared flycatcher coding sequences. Zero-fold degenerate sites are defined as positions where a mutation to any other nucleotide would lead to a change in the amino acid sequence, and four-fold degenerate sites are defined as positions where a mutation to any other nucleotide does not alter the amino acid sequence. All zero-fold and four-fold degenerate sites were extracted from the collared flycatcher coding sequence (CDS), using the reconstructed ancestral genome sequence, which were then used to estimate π_N and π_S respectively. To account for the impact of gBGC, the strength of which is positively correlated with recombination rate, I subsampled the variable sites for those with GC conservative polymorphisms (i.e. G/C or A/T polymorphisms).

Estimating d_N/d_S

The ratio of nonsynonymous to synonymous divergence (d_N/d_S) provides information both on the rate of positive and negative selection within a lineage. d_N/d_S is a macroevolutionary measure of natural selection, and applying the measure to closely related species may lead to inaccurate inferences of selection. This is due to the contribution of segregating polymorphisms to the estimates of divergence (Mugal et al., 2020), which make a larger contribution when divergence is short. In particular, nonsynonymous substitutions may be inflated, as a larger proportion of nonsynonymous polymorphisms are unlikely to reach fixation. As an approach to mitigate this problem in closely related species, Mugal et al., (2020) suggest to incorporate population polymorphism data, and estimate d_N/d_S for varying sample sizes with masking polymorphic sites from the sequence alignment. One can then examine the degree of sample size dependence of the d_N/d_S estimates.

In **paper II** of this thesis, I estimate d_N/d_S for collared flycatcher and taiga flycatcher, using genes that are one-to-one orthologues with zebra finch, which is used as an outgroup. Due to the recent divergence time of collared flycatcher and taiga flycatcher (< 4 coalescent units), I estimated d_N/d_S for varying sample sizes of both species to investigate sample size dependence in the values when masking polymorphic sites in the sequence.

For **paper III**, I was primarily interested in the difference between d_N/d_S on the autosomes compared to the Z-chromosome, as an indicator of the strength of selection. For this purpose, I estimated d_N/d_S for the branch separating collared flycatcher and zebra finch, using chicken as an outgroup. Due to the longer branch length, investigating the sample size dependence was not necessary.

In both **paper II** and **paper III**, I estimated d_N/d_S with the software bio++ (Dutheil & Boussau, 2008), which allows to estimate nonhomogeneous models with different substitution rates for different substitution types. In this way, I can account for the impact of gBGC by limiting the estimates of d_N/d_S to GC conservative mutations. For both papers, estimates of d_N/d_S were obtained for all substitutions and GC conservative only.

Estimating the DFE and adaptive substitution rate

Incorporating both polymorphism and divergence data can inform us about the proportion of adaptive substitutions occurring within a lineage. We assume that when advantageous mutations arise, they rapidly reach fixation and therefore do not contribute to polymorphism data. Within the McDonald-Kreitman (MK) framework (McDonald & Kreitman, 1991), we can then estimate the proportion of substitution data that is the result of adaptive evolution by using polymorphism data as a neutral reference.

In **papers II** and **III**, I apply an MK approach to infer the adaptive substitution rate, ω_a, using the software DFE-alpha (Eyre-Walker & Keightley, 2009; Keightley & Eyre-Walker, 2007). DFE-alpha infers the distribution of fitness effects (DFE) from the SFS data for a population. The approach requires input from two site classes, a neutral site class and a site class that experiences selection, which are assumed to have identical mutation rates. Differences in the shape of the SFS between the two site classes can therefore inform us about the DFE. After estimating the DFE from the data, the incorporation of divergence data for synonymous and nonsynonymous sites can then be used to separate the substitution data into substitutions that were fixed due to adaptive and non-adaptive evolution. The rate of adaptive substitutions is referred to as ω_a, from which I can also obtain the proportion of adaptive substitutions, α.

There is uncertainty associated with the estimation of the DFE, and the estimation is known to suffer from biases when population size fluctuates. This problem is inherent in the combination of two measures which are dependent on the effective population size but are impacted at different timescales (polymorphism and divergence data). The DFE-alpha method incorporates population size change using a step function, with an instantaneous change

from one population size to another. However, previous work has shown that this still suffers from bias, as polymorphism and divergence data do not reach the new equilibrium at the same rate (Müller et al., 2022). In particular, recent population bottlenecks lead to an underestimation of ω_a and α, while recent expansions lead to an overestimation. Therefore, it is important to interpret the results of the DFE in light of the known demographic history.

Correlation with recombination rate

To investigate the role of recombination rate in shaping signatures of direct selection, in **paper II**, I estimate the above statistics in bins of genes with varying recombination rates for both collared flycatcher and taiga flycatcher. Gene averaged recombination rates were divided into three bins of approximately equal numbers of genes, referred to as low, intermediate, and high recombination rate. Average values of π_N/π_S, d_N/d_S, and DFE estimates were then calculated for each cluster of genes. Although clustering multiple genes reduces the sensitivity of recombination rate variation, this approach is necessary for estimating the DFE, as a single gene does not provide enough data to apply this method. Additionally, in many cases there are few polymorphisms or substitutions available within just one gene that estimates of π_N/π_S and d_N/d_S for single genes are also too noisy.

Detecting genomic signatures of gene flow

Gene flow between species will lead to discordant patterns of evolutionary history in the genome. Previous work has shown that collared flycatcher and pied flycatcher are not sister species, but that pied flycatcher is in fact sister to the atlas flycatcher (Nater et al., 2015), however some gene trees place collared flycatcher and pied flycatcher as sister. This signature is indicative of gene flow between the two species. Discordant gene trees can also result from incomplete lineage sorting (ILS), in which case equal proportions of different discordant trees is expected (Hudson, 1983). An approach to test for asymmetry in gene tree discordance is referred to as the ABBA-BABA test, otherwise known as Patterson's D statistic (Green et al., 2010). This approach takes a four-taxon tree, with two ingroup sister species (P1 and P2), a third species with which gene flow is predicted to have occurred (P3), and an outgroup species (O). The sites with the same allele shared by P2 and P3 are referred to as ABBA sites and the sites with the same allele shared by P1 and P3 are referred to as BABA sites. Under ILS, the number of ABBA and BABA sites is predicted to be the same, however in the case of gene flow there would be an excess of either ABBA or BABA. The D statistic is calculated with the following equation:

$$D = \frac{\sum(1-p_1) * p_2 * p_3 \; - \; \sum p_1 * (1-p_2) * p_3}{\sum(1-p_1) * p_2 * p_3 \; + \; \sum p_1 * (1-p_2) * p_3}$$

where p_x gives the allele frequency of the derived allele in species 1, 2, and 3.

Therefore, when the statistic is significantly different from zero, there is evidence for gene flow. In **paper III**, I estimate this value for multiple population comparisons of collared flycatcher and pied flycatcher, to compare levels of gene flow on the autosomes and the Z-chromosome. By using multiple populations, I am able to estimate when gene flow occurred in relation to population splits.

Patterson's D is robust to detect gene flow at the genome-wide scale, however it suffers from high variance when estimated in smaller regions and is not appropriate to measure variation across the genome (Martin et al., 2015). For this purpose, other similar statistics have been developed that are better suited. In this case, I apply the f_d statistic developed by (Martin et al., 2015).

Genome assembly

Genome assembly presents a significant challenge for biologists, because to sequence the DNA we must fragment the genome into shorter pieces and reassemble the genome computationally. Repetitive regions are often challenging, or simply impossible, to assemble as their genomic location may be ambiguous. Additionally, the use of Illumina sequencing leads to an underrepresentation of GC rich regions in assemblies, as these are poorly sequenced. In birds the impact of this sequencing bias is evident in the lack of many genes in early assemblies, some of which have since been identified as having high GC content.

Assembly with long-read data

Long-read data has led to a substantial improvement of the genome assemblies for many organisms (Kim et al., 2022; Peona et al., 2021). Initially, the error rates for long reads were so high that additional polishing with Illumina data was required, however with PacBio HiFi sequencing we can now generate long reads that have 99.9% accuracy. The availability of HiFi sequencing therefore makes high-quality genome assemblies widely accessible. In **paper IV**, I generate a new assembly for the collared flycatcher and the first assembly for the pied flycatcher, using a combination of HiFi and HiC sequencing. By using a female individual from both species, I am also able to assemble a near complete W-chromosome for collared flycatcher and pied flycatcher.

Scaffolding with chromosomes with HiC technology

The past decade has seen an explosion in our ability to sequence not only the linear representation of a genome, but also to probe the 3D structure and organization of chromatin within the nucleus. The introduction of HiC by Lieberman-Aiden et al., (2009) substantially increased the accessibility of sequencing chromatin interactions by providing a method to identify unbiased interactions genome-wide.

The basis behind using chromatin interactions to scaffold genomes, is that the extent of interaction between loci scales with their linear distance. Thus, using the contact matrix, scaffolding tools can join contigs into scaffolds based on the amount of interactions between loci. However, this is not without limitations, and manual curation of HiC scaffolded assemblies is generally required to correct errors such as misjoins, missed joins, and incorrectly oriented contigs (Howe et al., 2021). Additionally, as HiC sequencing is currently performed with Illumina sequencing, it also suffers from a bias in regions of extreme base composition. Therefore, scaffolding high GC content regions can still be a significant challenge.

In **paper IV** I use HiC sequencing data from the same individual of collared flycatcher and pied flycatcher for which I generated a high-quality contig level assembly with PacBio HiFi. With these combined technologies, I am able to create a chromosome scale assembly for both species, including both the Z- and W-chromosomes.

Structural variant detection

Genome-wide detection of SVs has been challenging to perform with the use of short-read data. SVs will typically span multiple short reads, and often reside in proximity to repetitive elements, which complicates their detection. Reference bias is also a greater challenge for SV detection compared to SNV detection, which can lead to problems when identifying SVs even from mapping long read sequences to a single reference genome. As an alternative, whole genome alignments (WGA) of multiple genomes can be used to detect SVs. The challenge in this case, is that high quality genome assemblies are not often available for multiple individuals of closely related species or populations.

In **paper IV**, I take advantage of the increased accessibility of long-read sequencing, and generate high-quality genome assemblies for both collared flycatcher and pied flycatcher. I then perform WGA of both species with MUMMER (Marçais et al., 2018), and use the software MUM&co (O'Donnell &

Fischer, 2020) to identify SVs between these two individuals. MUM&co is able to identify a diverse set of SVs, including insertions and deletions (indels), tandem duplications, inversions, and translocations. Compared to other software for identifying SVs from WGA data, MUM&co has a lower false positive rate for inversions.

With a genome-wide callset of a wide range of SVs, I am then able to address how SVs are distributed across the genome, whether any SVs are overrepresented on any chromosomes, and how SVs are associated to the differentiation landscape between collared flycatcher and pied flycatcher. To do so, I compared the estimates of diversity within species and divergence between species for genomic windows overlapping with each category of SV and for windows with no SV overlap. In addition, I identified F_{ST} peaks between species, and tested whether any of the SV categories were enriched within F_{ST} peaks compared to the genomic background with no SV overlap.

Research aims

The overall aim of this thesis is to contribute to our understanding of the evolutionary processes that drive speciation. Using genomic data, I explore how processes such as recombination, natural selection, and gene flow interact to shape genome-wide patterns of genetic diversity and differentiation between species. I aim to improve our understanding of how different forms of natural selection shape the genomic differentiation landscape, and how variation in recombination rate impacts our detection of signatures of selection. Through this work, I contribute to our understanding of the genomics of divergence between closely related species.

The specific aims for each paper are described below:

Paper I – To address whether different evolutionary processes are responsible for driving patterns of differentiation that are shared between multiple independent species comparisons versus patterns of differentiation that are lineage-specific.

Paper II – To understand how recombination rate shapes genomic signatures of indirect and direct selection and to investigate whether evolutionary changes in recombination rate between species are reflected in signatures of selection.

Paper III – To investigate the role of the Z-chromosome compared to autosomes in speciation, and the contributions of adaptation versus genetic drift.

Paper IV – To generate high-quality long-read genome assemblies for two closely related species, and to use population scale long-read sequencing to investigate the role of structural variants between species in the divergence process.

Summary of the papers

Paper I

Positive selection plays a major role in shaping signatures of differentiation across the genomic landscape of two independent *Ficedula* flycatcher species pairs

The extent to which genome-wide patterns of between species differentiation inform us about the loci underlying speciation has been a debated topic. A nearly ubiquitous observation within speciation genomics is that patterns of genomic differentiation between closely related species (referred to as the genomic differentiation landscape) are highly heterogeneous with distinct peaks of high differentiation and valleys of low differentiation. Early investigations of the genomic differentiation landscape interpreted these patterns in light of the interacting forces of divergent natural selection and gene flow. Under this interpretation, peaks of elevated differentiation (so-called 'differentiation islands') represent the loci that contribute to reproductive isolation, whereas the genomic background levels of differentiation are homogenized by gene flow. However, this interpretation was called into question by subsequent observations. Firstly, reanalysis demonstrated that for many species pairs that were thought to be currently exchanging genes, there was no evidence for ongoing gene flow. Secondly, measures of absolute sequence divergence did not show the expected pattern for a region resistant to gene flow standing out of a background homogenized by gene flow. Finally, examination of the differentiation landscape across multiple closely related species began to reveal that the same genomic regions would exhibit elevated differentiation repeatedly. These regions often coincided with regions of low recombination rate, and it was hypothesized that the differentiation landscape thus reflected a history of linked selection largely resulting from background selection.

One limitation of previous studies of multiple species differentiation landscapes, including studies of *Ficedula* flycatchers, is that species were so closely related that it is challenging to discriminate between patterns which are shared due to shared ancestry versus conserved evolutionary processes. In this paper, I address this limitation and ask whether different evolutionary processes are responsible for driving patterns of differentiation that are similar among multiple species pairs compared to lineage-specific patterns of

differentiation. To do so, I use population re-sequencing data from four *Ficedula* flycatcher species representing two independent species pairs: collared and pied flycatchers and red-breasted and taiga flycatchers. Because the red-breasted flycatcher and taiga flycatcher share a common ancestor deeper in history than do collared flycatcher and pied flycatcher, I am able to reconstruct three distinct evolutionary timepoints with these species comparisons. The comparison of the two species pairs represents a timescale at which gene flow and shared ancestral polymorphisms are negligible. Thus, any shared signature within the genomic landscape between these comparisons should be the result of conserved evolutionary processes.

Estimating F_{ST} across the genome for the two species, I was able to identify peaks of differentiation in the same genomic regions (shared peaks) in both comparisons, as well as lineage-specific peaks that were unique to one or the other species comparison. I then identified signatures of selective sweeps in each of the four species and compared the overlap between selective sweeps and different categories of F_{ST} peaks. There was an apparent trend to find selective sweep signatures more often within F_{ST} peaks than the genomic background, particularly within shared F_{ST} peaks. Additionally, I used a gene annotation and estimates of recombination rate for collared flycatcher to understand how the underlying genomic architecture shapes patterns of differentiation. This revealed that although shared F_{ST} peaks and collared and pied flycatcher specific F_{ST} peaks were found in regions of remarkably reduced recombination rate, red-breasted and taiga specific F_{ST} peaks were not. This provides indirect evidence that changes in recombination rate could help to generate lineage specific signatures of natural selection, while there did not appear to be a substantial difference in the form of natural selection driving shared versus lineage-specific peaks.

Paper II

The role of recombination dynamics in shaping signatures of direct and indirect selection across the *Ficedula* flycatcher genome

Recombination is known to play an important role in shaping genome-wide patterns of genetic diversity and differentiation. Signatures of natural selection, such as selective sweeps, are dependent on linkage between sites leading to stronger signatures where recombination is low. Additionally, linkage between sites can generate Hill-Robertson interference (HRI), where multiple selected sites interfere with one another leading to a reduction in the efficacy of selection when recombination is low. Although recombination rate has been demonstrated to play an important role in shaping the genomic landscape in many species, we still have a limited understanding of how the dynamics of

recombination rate evolution between species impacts these patterns. This gap in our understanding stems from the fact that recombination rate estimates are often not available for multiple closely related species. Many studies which have investigated the role of recombination rate in patterns of between species differentiation have been limited to estimates of recombination rate from a single species, which in some cases even stems from other species than the ones studied.

In this paper, I use population re-sequencing data from 95 collared flycatcher samples and 65 taiga flycatcher samples to estimate LD based recombination rates for both species. I incorporate population re-sequencing from pied flycatcher and red-breasted flycatcher to estimate indirect signatures of selection in the form of F_{ST} peaks. Additionally for collared flycatcher and taiga flycatcher I identify selective sweep signatures with SweepFinder2. Comparing estimates of recombination rate in both species with signatures of indirect selection revealed that lineage-specific signatures of selection often showed a reduction in recombination rate only in the corresponding species. On the other hand, evolutionarily conserved signatures of selection showed lower recombination rate in both species. This demonstrates that even in a pair of species thought to have highly conserved recombination rate, changes in recombination can underlie lineage-specific selection signatures.

In addition to examining the role of recombination rate dynamics on signatures of indirect selection, which rely on linkage between selected and neutral sites to detect a signal, I also examined evidence for a relationship between recombination rate and the efficacy of direct selection on protein-coding sequences. Evidence for HRI occurring across the genome is highly variable among different organisms. Here, by estimating π_N/π_S, d_N/d_S and ω_a for genes clustered by three different recombination rate bins, I found no impact of recombination rate on signatures of direct selection in either collared flycatcher or taiga flycatcher. The absence of any signature of HRI persisted even when I restricted the analysis to genes which showed similar levels of recombination between both species, and would thus be expected to show a stronger signal for cumulative measures such as d_N/d_S. Therefore, I conclude that the evolutionary processes that shape signatures of indirect versus direct selection differ. In the case of indirect selection, I found that the genomic background and recombination rate is of great importance to observe a signal; for direct selection, however, likely factors such as gene expression patterns play a much greater role in this system.

Paper III

Evidence that genetic drift not adaptation drives fast-Z and large-Z effects in *Ficedula* flycatchers

The sex chromosomes have been hypothesized to play a disproportionate role in driving speciation across many taxa. The reason for this disproportionate role is thought to be the hemizygosity of sex chromosomes in the heterogametic sex, which exposes recessive mutations. Two phenomena have been described as a result of the hemizygosity, referred to as the *fast-X* (or *fast-Z*) effect and the *large-X* (or *large-Z*) effect. I will refer here to the Z-chromosome only, as birds have a Z/W sex determination system. In the case of the *fast-Z* effect, the hemizygous state increases the efficacy of selection of recessive advantageous mutations, which leads to a faster rate of evolution. The *large-Z* effect is also expected to result from hemizygosity of the Z chromosome and posits that the Z chromosome will contribute disproportionately to reproductive isolation due to genetic incompatibilities. Distinguishing between a *fast-Z* effect and a *large-Z* effect is a challenge, because they are both expected to cause increased genetic divergence on the Z-chromosome relative to autosomes and they are not mutually exclusive processes. Additionally, it remains equivocal to what extent either the *fast-Z* or the *large-Z* is the result of adaptive evolution compared to an increased rate of genetic drift on the Z-chromosome.

Using population genomic data from several *Ficedula* flycatcher species, I tested for both the *fast-Z* and *large-Z* effect and investigated the evidence for whether these phenomena are driven by adaptation versus genetic drift. Multiple estimates of the effective population size revealed that all species had lower N_e on the Z-chromosome. Estimates of natural selection revealed that the strength of purifying selection was weaker on the Z-chromosome, with all species showing higher levels of π_N/π_S compared to the autosomes. Estimates of d_N/d_S were slightly elevated on the Z-chromosome relative to the autosomes, however variation in demography between species appeared to impact the estimation of ω_a (the adaptive substitution rate). Additionally, although genetic differentiation between the two species pairs was on average higher on the Z-chromosome, signatures of indirect selection in the form of F_{ST} peaks or selective sweeps were underrepresented on the Z-chromosome compared to the autosomes. Thus, an increased prevalence of linked selection due to the lower recombination rate of the Z-chromosome did not appear to explain the *fast-Z* effect. I next investigated genomic signatures of the *large-Z* effect, where I focused on the collared flycatcher and the pied flycatcher, which show evidence for historical signatures of gene flow. The ABBA-BABA test revealed significant signatures of gene flow between the two species on both the autosomes and the Z-chromosome, with a lower effect size on the Z-

chromosome. The reduction in gene flow appeared to be a chromosome-wide effect, and pre-dates the splits of the different populations within the species studied here.

Overall, I observed clear evidence of a *fast-Z* effect in all species, with elevated levels of divergence between species on the Z-chromosome compared to the autosomes, and evidence of the *large-Z* effect between the collared flycatcher and pied flycatcher. There was no evidence for an increased rate of adaptation on the Z-chromosome, but I did observe evidence to suggest a weaker efficacy of purifying selection. I therefore hypothesize that the *fast-Z* and *large-Z* effects within these species are driven by genetic drift rather than adaptation, which enabled the evolution of genetic incompatibilities between closely related species on the Z-chromosome.

Paper IV

The combination of HiFi and HiC sequencing technology enables investigation of the impact of structural variation on speciation in *Ficedula* flycatchers

Structural variants (SVs) can encompass a large portion of the genome. SVs have been shown to have important phenotypic effects, with a number of disease phenotypes in humans connected to SVs and the large size of SVs is generally assumed to translate into a large effect on fitness. Several examples also exist of SVs playing an important role in local adaptation and speciation, with inversions in particular implicated due to their effect of reducing recombination in heterozygotes. Inversions have been shown to capture multiple coadapted loci in several organisms, creating what is referred to as a super gene. The suppression of recombination between different inversion orientations means that the associations between variation at different loci are not broken apart in the face of gene flow. Despite the potentially important role of SVs in speciation, they are often overlooked in comparison to SNVs, and comprehensively studying the diversity of SVs between closely related species remains a major challenge. SV detection continues to be a technical challenge, even though advances in long-read sequencing and its increasing availability for non-model organisms is changing this. Methods for SV detection from short-read data have relied either on indirect inferences or through analysis of read mapping patterns, however these are generally limited to detection of shorter variants due to size of the reads. Mapping long-reads can improve our ability to detect longer SVs compared to using short-read mapping. However, this approach will still suffer from reference bias, which is a greater problem for SV detection compared to SNVs. As an alternative to mapping reads to a single reference genome, it is possible to instead identify SVs from

whole genome alignments of multiple genomes. Whole genome alignment approaches can provide a less biased estimate of the variety of SVs. The challenge then, is to generate high quality genomes from multiple individuals or closely related species.

In this paper, I investigate the role of SVs in divergence between the closely related species collared flycatcher and pied flycatcher. To do so, I use a combination of PacBio HiFi sequencing and HiC chromatin conformation capture technologies, which together allow for the generation of high-quality, chromosome scale genome assemblies. These assemblies represent the first long-read genome assembly for the collared flycatcher, adding a substantial fraction of sequence length for all previously identified chromosomes, and the first genome assembly for the pied flycatcher. By sequencing a female for both species, I am also able to assemble W-chromosome, for which only a partial collared flycatcher assembly was previously available. I then identified SVs between the two species by whole genome alignment, using the script provided by MUM&co to detect SVs. This analysis identified 14,451 SVs between the two assemblies, the majority of which were indels. SV density was higher on microchromosomes compared to macrochromosomes, and the Z- and W-chromosomes showed a disproportionate frequency of inversions and translocations. The inclusion of population level HiFi sequencing for both collared flycatcher and pied flycatcher enabled the combination of SV detection with SNV based patterns of genetic differentiation. This revealed that although average F_{ST} did not vary greatly among different SV categories compared to the genomic background, inversions and translocations were overrepresented in F_{ST} peaks compared to the background. On the contrary, indels were less common in F_{ST} peaks. When I compared the overlap between all SV categories combined there was no significant difference between F_{ST} peaks and the genomic background. With this work, I provide the first look into the role of SVs in divergence between the collared flycatcher and pied flycatcher, two closely related avian species.

Conclusions and future perspectives

In this thesis I investigated the genomic signatures of divergence in multiple *Ficedula* flycatcher species, with the goal to further our understanding of how different evolutionary forces interact to drive speciation. By comparing multiple species pairs at varying stages of divergence, I was able to investigate the processes that generate shared vs lineage-specific patterns in the genomic landscape. More than distinct histories of natural selection, the lineage-specific signatures appeared to be driven by changes in the recombination landscape between lineages. However, although recombination rate did appear to have a strong impact on signatures of indirect selection, such as F_{ST} peaks and selective sweeps, I did not observe a strong relationship between recombination rate and the efficacy of selection. On the other hand, I did observe that the efficacy of selection was weaker on the Z-chromosome, likely due to its lower effective population size compared to the autosomes. The entire Z-chromosome also showed reduced signatures of gene flow compared to autosomes, providing additional support for previous suggestions that the Z-chromosome is a hotspot for genetic incompatibilities between the collared flycatcher and the pied flycatcher. Finally, I also observed a disproportionately large set of inversions and translocations on both the sex chromosomes, aided by new long-read assemblies for both the collared flycatcher and pied flycatcher. These observations provide a potential mechanism behind the buildup of incompatibilities on these chromosomes.

With each question addressed in this thesis, several additional questions to pursue arise. For example, I observed that evolution of the recombination landscape impacted the detection of signatures of selection. However, I can only speculate about what mutations have caused these changes. Structural rearrangements such as inversions and translocations could alter the rate of recombination and explain the patterns I observed. Additionally, shifts in the location of centromeres have been observed between even closely related avian species and may impact the recombination landscape. Investigations of the role of broadscale rearrangements between the collared flycatcher and taiga flycatcher species could help to address the mechanisms behind evolutionary changes in recombination rate. These investigations would be aided in the future by the publication of a reference genome of the taiga flycatcher, which is currently lacking. Another interesting avenue for future research

would be to conduct a thorough investigation of recombination rate variation at both the fine-scale and broadscale among all the four flycatcher species studied in this thesis. This analysis could be accompanied by an investigation of the molecular mechanisms that shape recombination rate variation.

The demographic history and potential history of gene flow between the red-breasted flycatcher and taiga flycatcher is poorly understood compared to what is known about the collared flycatcher and pied flycatcher. There is no evidence of current gene flow between the red-breasted flycatcher and taiga flycatcher, but we do not know whether gene flow occurred at some point during their divergence. Future demographic analyses can address the history of divergence between these two species, including historical population size changes and evidence for a history of gene flow. An interesting question is whether any of the same regions that appear to be important for generating genetic incompatibilities between collared flycatcher and pied flycatcher also contribute to isolation between the red-breasted flycatcher and taiga flycatcher.

Despite substantially improving the genome assembly for collared flycatcher, and providing a novel assembly for pied flycatcher, the smallest microchromosomes were still not identified. Utilizing the IsoSeq data that is available from the same individuals may generally improve the scaffolding of the assembly and allow us to assemble the missing chromosomes that were not completed using the HiC data alone. Additionally, these data can be used to generate a high-quality gene annotation of both species, and potentially identify missing genes from the original collared flycatcher annotation.

Structural variant identification performed here was done by comparing the two reference quality genomes of collared flycatcher and pied flycatcher. However, with the availability of population level long-read sequencing for both species, it may be possible to use a pangenomic approach to SV detection. With this approach, I could assess the diversity of SVs both within and between species, rather than between only two individuals. I can then address whether the same regions that show more frequent rearrangements between species also show higher diversity in rearrangements within species. Coupled with RNAseq data from each individual, we can also address whether any SVs identified between species are associated with differential gene expression between species. The data available from this project opens a wealth of opportunities for further investigation into the divergence of these two species, and the ability to investigate an understudied form of variation through the analysis of SVs.

Svensk sammanfattning

Mångfalden av liv på vår planet uppstod från en gemensam förfader genom en process som vi kallar artbildning. Denna grundläggande process leder till bildandet av distinkta utvecklingslinjer, där en gång sammankopplade populationer kan utvecklas oberoende av varandra. Hur detta går till är fortfarande föremål för pågående forskning. Inom sexuellt reproducerande organismer tänker vi ofta på denna process som en gradvis minskning av genflödet mellan grupper av individer, eller utvecklingen av reproduktiv isolering. Vi förväntar oss att genetiska skillnader som ligger bakom denna minskning av genflödet ackumuleras och bidrar till så kallade reproduktionsbarriärer. Reproduktionsbarriärer delas ofta in i *prezygotiska* och *postzygotiska*, beroende på om de begränsar genflödet före eller efter befruktning av ägget. Ett exempel på en prezygotisk barriär är när två fågelarter utvecklar olika sånger, och således inte längre attraheras av en partner av den andra arten. Ett exempel på en postzygotisk barriär är när två individer av närbesläktade arter parar sig och skapar en hybridavkomma, men hybriden är steril. Många andra exempel finns på både pre- och postzygotiska barriärer, och i båda fallen kan de utvecklas till att helt eller endast delvis begränsa genflödet mellan arter. Denna funktion belyser en viktig aspekt av artbildningsprocessen: att den representerar ett kontinuum mellan en helt sammankopplad population och helt isolerade arter. Eftersom detta är en process som kan ta miljontals år, kan vi inte studera den i realtid. Istället måste vi använda ett historiskt tillvägagångssätt för att försöka sluta oss till hur artbildning har skett mellan olika organismer, vilket vi ofta gör genom att studera närbesläktade arter. Att fokusera på flera olika par av arter med olika divergenstider kan också användas för att försöka rekonstruera processen. Ett tillvägagångssätt är att sekvensera och jämföra genomen från flera individer från olika arter, eftersom variationsmönster över genomet återspeglar den historiska inverkan av olika evolutionära processer.

Ett huvudsyfte inom artbildningsforskning är att förstå hur dessa olika evolutionära processer har format variation inom och mellan arter, med målet att i slutändan öka vår förståelse för hur artbildningen i sig fortskrider. Till exempel kan vi undersöka mönster av genetisk differentiering mellan närbesläktade arter, och observera hur detta varierar längs genomet. Vissa regioner kan uppvisa anmärkningsvärt ökade nivåer av differentiering, vilket kan vara ett tecken på att där finns en reproduktiv barriär eller att regionen bidrar till

anpassning inom en (eller potentiellt båda) arterna. Andra processer kan dock skapa liknande mönster. Till exempel är variation i rekombinationshastighet, det vill säga hur ofta genetiskt material utbyts mellan de två kopiorna av varje kromosom, känd för att vara positivt korrelerad med nivån av genetisk mångfald. När rekombinationshastigheten är låg är den genetiska mångfalden också ofta lägre, vilket i sin tur kan leda till en ökning av den genetiska differentieringen mellan två arter i denna region av arvsmassan. Denna förhöjda differentiering kan då vara helt orelaterad till artbildningsprocessen. Således är det fortfarande en utmaning att särskilja mönster från olika evolutionära processer över genomet.

I den här avhandlingen studerar jag artbildningsgenomik hos flera par av flugsnappararter från släktet *Ficedula* i ordningen tättingar. Halsbandsflugsnappare och svartvit flugsnappare är närbesläktade arter som hybridiserar i regioner där deras utbredningsområden överlappar. Hybriderna verkar dock vara helt sterila. Det har också utvecklats ekologiska skillnader mellan arterna, såsom skillnader i sång och fjäderdräkt. Mindre flugsnappare och tajgaflugsnappare är ett annat par av arter som är mer avlägset besläktade med varandra och mellan vilka det inte förekommit något känt fall av hybridisering. Med hjälp av genomiska data från nära 200 individer undersökte jag differentieringsmönster i arvsmassan mellan de olika artparen. Inom fåglar är antalet och strukturen av kromosomer i arvsmassan mycket bevarade över miljontals år. Tack vare detta, och med de jämförelsevis nära släktskapen som undersöks här, kan vi direkt jämföra samma regioner av arvsmassan mellan olika arter. På så sätt observerade jag att nivåerna av genetisk differentiering mellan halsbandsflugsnappare och svartvit flugsnappare varierade längs genomet i ett liknande mönster som mellan mindre flugsnappare och tajgaflugsnappare. På många kromosomer hade samma region lokalt förhöjd differentiering mellan arter i båda paren. Men jag observerade också kromosomer där differentieringsmönstren var unika. En huvudfråga är om det finns bevis för att det är andra evolutionära processer som driver dessa unika mönster jämfört med de gemensamma mönstren. Jag fann att både de gemensamma och de unika mönstren visade tecken på att bidra till anpassning inom minst en art. Å andra sidan verkade de unika mönstren drivas av evolutionen av rekombinationshastighet mellan halsband/svartvit flugsnapparlinjen och mindre/tajgaflugsnapparlinjen, snarare än att drivas av olika bakgrunder av naturligt urval. Således verkar det vara avgörande när man tolkar olika mönster för selektion mellan arter att också beakta evolutionära förändringar i rekombinationshastigheten.

Jag undersökte även Z-kromosomen, som är en av de två könskromosomerna, och dess roll i skillnaderna mellan dessa två artpar. Fåglar har ett könsbestämningssystem som vi kallar Z/W, vilket innebär att hanar ZZ och honor är ZW, jämfört med X/Y-systemet där honor är XX och hanar är XY. I fåglar kallar vi honor för det *heterogametiska* könet, vilket hänvisar till det faktum att de

har två olika könskromosomer. Könskromosomer kan spela en oproportionerligt stor roll vid artbildning, eftersom de ackumulerar genetiska skillnader i snabbare takt och kan samla på sig genetiska inkompatibiliteter mellan arter som leder till postzygotisk isolering. Även om högre nivåer av differentiering ofta observeras på Z-kromosomen, säger det inte nödvändigtvis vilken process som ligger bakom. Till exempel kan positiv selektion vara mer effektiv i det heterogametiska könet eftersom alla mutationer är omedelbart synliga för selektion, medan i hanar kan recessiva mutationer maskeras av den andra kopian av kromosomen (eftersom fåglarna är diploida och har två kopior av varje kromosom). Å andra sidan förväntas den genetisk driften, som är den slumpmässiga fluktuationen i frekvensen av olika varianter (alleler), vara starkare på Z-kromosomen. Jag observerade att genetisk differentiering mellan arter hos flugsnapparna var högre på Z-kromosomen, och den genetiska mångfalden minskade. Jag fann också att spår av historiskt genflöde i arvsmassan mellan halsbandsflugsnappare och svartvit flugsnappare var svagare på Z-kromosomen jämfört med autosomer. Generellt tyder dessa bevis på att Z-kromosomen spelar en viktig roll i artbildning och att dessa mönster är resultatet av en ökad roll av genetisk drift på Z-kromosomen snarare än en högre effektivitet av positivt urval.

Slutligen undersökte jag en form av genetisk variation som hittills varit outforskad inom detta system och som kallas för *strukturell variation* (SV). De allra flesta genomstudier av artbildning har använt sig av *enbaspolymorfier*, som är mutationer i ett enda baspar. SV är dock mutationer som per definition påverkar minst 50 bp och kan ha extremt stora fenotypiska effekter och konditionseffekter. I vissa fall har de också kopplats direkt till artbildning. Även om vi refererar till SV som en kategori, omfattar dessa i själva verket flera olika typer av mutationer som kan ha olika roller i artbildning. Dessa variationer är underutforskade på grund av svårigheten att identifiera dem med tidigare sekvenseringstekniker, där mängden DNA som gick att sekvensera på en gång ofta var mycket kortare än själva mutationen. Med avancerad teknologi kan vi sekvensera längre segment, och möjligheten att undersöka mångfalden av SV:er inom och mellan närbesläktade arter ökar. Här karakteriserade jag SV:er mellan en individ var av halsbandsflugsnappare och svartvit flugsnappare och kombinerade dessa data med undersökningar av genomomfattande differentiering på populationsnivå. Jag fann att specifika kategorier av SV:er ökade i könskromosomerna och hittades i regioner av arvsmassan med förhöjd differentiering – högre än vad som är förväntat av slumpen. Dessa spännande resultat tyder på att vissa former av SV kan bidra till genetiska inkompatibiliteter mellan dessa närbesläktade arter.

Kindly translated by Linnéa Smeds

Summary

The diversity of life on our planet arose from a common ancestor through a process we refer to as speciation. This fundamental process leads to the formation of distinct evolutionary lineages, where once interconnected populations can evolve independently of one another. How this occurs is still the subject of ongoing research. Within sexually reproducing organisms, we often think of this process as involving the gradual reduction in gene flow between groups of individuals, or the evolution of reproductive isolation. We expect that genetic differences accumulate that underlie this reduction in gene flow, and contribute to what are known as reproductive barriers. Reproductive barriers are often divided into 'pre-zygotic' and 'post-zygotic', which refers to whether they restrict gene flow before or after fertilization of the egg occurs. An example of a pre-zygotic barrier would be two bird species evolving different songs, and are thus no longer being attracted to mates of the other species; an example of a post-zygotic barrier would be two individual of closely related species mating and creating a hybrid offspring, but the hybrid being sterile. Many other examples exist of both pre- and post-zygotic barriers, and in both cases they can evolve to fully or only partially restrict gene flow between species. This feature highlights an important aspect of the speciation process, which is that it represents a continuum between a completely interconnected population and completely isolated species. Because this is a process that can take millions of years to complete, we cannot study this in real time. Rather, we have to take a historical approach to try to infer how speciation has occurred across different organisms, which we often do by studying closely related species. Focusing on multiple species pairs, which have varying divergence times, can also allow us to try to reconstruct the process. One approach is to sequence and compare the genomes of multiple individuals from different species, because patterns of variation across the genome reflect the historical action of different evolutionary processes.

One major aim in speciation research is to understand how these different evolutionary processes have shaped variation within and between species, with the goal to ultimately increase our understanding of how speciation itself proceeds. For example, we can estimate patterns of genetic differentiation between closely related species, and observe how this varies along the genome. Some regions may show remarkably increased levels of differentiation, which

could be a sign that this region harbors a reproductive barrier or contributes to adaptation within one (or potentially both) species. Other processes however could leave a similar signature. For example, variation in recombination rate, or how often genetic material is exchanged between the two copies of each chromosome, is known to be positively correlated with levels of genetic diversity. When recombination rate is low, genetic diversity is also often lower, which can in turn lead to an in increase in the genetic differentiation between two species at this genomic region. This elevated differentiation may then be entirely unrelated to the speciation process. Thus, disentangling signatures of different evolutionary processes across the genome remains a challenge.

Within this thesis, I study speciation genomics of multiple species pairs of *Ficedula* flycatcher, a passerine bird. The collared flycatcher and pied flycatcher are closely related species, which hybridize in overlapping regions of their ranges. The hybrids, however, appear to be completely sterile. Other ecological differences have evolved too, such as divergence in song and plumage. The red-breasted flycatcher and taiga flycatcher are another species pair, which are more distantly related to one another, and are not known to hybridize. Using genomic data from close to 200 individuals, I investigated genomic patterns of differentiation between the different species pairs. Within birds in particular, the number and structure of chromosomes in the genome is highly conserved across millions of years. Therefore, with the comparatively close relationships investigate here, we can directly compare the same genomic regions within different species. With this, I observed that levels of genetic differentiation between collared and pied flycatcher varied along the genome in a similar pattern to red-breasted and taiga flycatcher. On many chromosomes, the same region would show locally elevated differentiation between species in both pairs. However, I also observed chromosomes where the patterns of differentiation were unique. A major question here is whether there is evidence that different evolutionary processes are important in driving these unique patterns compared to the shared patterns. I found that both the shared and the unique patterns showed signs of contributing to adaptation within at least one species. On the other hand, rather than being driven by very distinct histories of natural selection, the unique patterns appeared to be driven by evolution in the recombination rate between the collared and pied flycatcher lineage and the red-breasted and taiga flycatcher lineage. Thus, when interpreting differing signatures of selection between species it seems to be critical to also consider evolutionary changes in the recombination rate.

Additionally, I examined the role of the Z-chromosome, which is one of the two sex chromosomes, in divergence between these two species pairs. Birds have a sex determination system we refer to as Z/W, which means that, compared to an X/Y system where females are XX and males are XY, within birds males are ZZ and females are ZW. In this case, we call females the

'heterogametic' sex, to refer to the fact that they have two different sex chromosomes. Sex chromosomes can play a disproportionately large role in speciation, as they accumulate genetic differentiation at a faster rate and can accrue genetic incompatibilities between species leading to intrinsic post-zygotic isolation. Although higher levels of differentiation are frequently observed on the Z-chromosome, this does not tell us necessarily the process responsible. For example, positive selection can be more efficient in the heterogametic sex because all mutations are immediately visible to selection, whereas, because the birds are diploid and have two copies of each chromosome, recessive mutations might be masked by the other copy of the chromosome in males. On the other hand, genetic drift, which is the random fluctuation in the frequency of different variants (alleles), is expected to be stronger on the Z-chromosome. I observed in the flycatchers that genetic differentiation between species was higher on the Z-chromosome, and genetic diversity was reduced. I also found that genomic signatures of historical gene flow between collared flycatcher and pied flycatcher were weaker on the Z-chromosome compared to autosomes. Generally, this evidence suggests that the Z-chromosome does play an important role in speciation and that these patterns are the result of an increased role of genetic drift on the Z-chromosome rather than a higher efficacy of positive selection.

Finally, I investigated a form of genetic variation that has been previously unexplored within this system, that is referred to as 'structural variation'. The vast majority of speciation genomic studies have made use of single nucleotide variants (SNVs), which are mutations at a single basepair. SVs, however, are mutations that impact by definition at least 50bp and can have extremely large phenotypic and fitness effects. In some cases, they have also been linked directly to speciation. Although we refer to SVs as one category, this in fact encompasses several different kinds of mutations which may have differing roles in speciation. This variation has been underexplored due to the difficulty of identifying them with earlier sequencing techniques, where the amount of DNA we could sequence at once would be much shorter than the mutation itself. With advancing technologies, we can sequence longer segments, and our ability to probe the diversity of SVs within and between closely related species is increasing. Here, I characterized SVs between an individual each of collared flycatcher and pied flycatcher and combined this data with population level investigations of genome-wide differentiation. I found specific categories of SVs were increased on the sex chromosomes, and were found in genomic regions of elevated differentiation higher than expected by chance. These exciting results suggest that some forms of SVs could contribute to genetic incompatibilities between these closely related species.

Acknowledgements

I would first like to say a big thank you to my supervisor, **Carina**. We took an unconventional path to you becoming my main supervisor, but I am very glad that we got there. Thank you so much for all your support and guidance throughout my PhD, for all the time that you put into the projects, and for always being so ready to discuss any issues or questions that I had. I really respect how meticulous you are with your work, and I much appreciate how that has reflected in the quality of our work together. I have really learned so much during this time and become comfortable with topics that previously seemed completely mystifying to me, which is thanks to your supervision.

I would also like to thank my co-supervisor, **Hans**. Thank you for your enthusiastic response when I contacted you out of the blue from the U.S., asking if there might be a position to come and work in your lab. I am really happy to have had the opportunity to come to Uppsala, and Sweden, and I think that was really a moment that changed the course of my life.

I want to also thank all the group members I have had the opportunity to work with during the PhD, past and present. **Rebekka**, it's been great having journal clubs with you and having your help to understand more of the maths and theory. I'm always impressed with how well you explain your work to be understandable by a bunch of biologists. **Jesper**, you always bring entertainment, and a hint of unpredictability, to every meeting and have given lots of insightful feedback on my projects over the past few years. I always enjoy when you come into the office and we spend the whole afternoon chatting. **Linnéa**, I think that you are such an impressive scientist, and you are a great role model for how to be successful at work while still maintaining balance in your life. I always appreciated when you would remind us of when we have holidays, and how helpful you are when I have a question about where some ancient data might be in the uppmax project. **Maurine**, we didn't get to interact in person much outside of zoom but I've very much enjoyed collaborating with you! **Homa**, you are so good at explaining complex topics in simple and understandable ways, and I really enjoyed getting to work with you when I was starting out. **Jente**, I think you took on a valuable task by organizing weekly afterworks, and I always enjoyed the effort that you put into your

FAWNA emails. I still think of you whenever I see a goose and wonder what species it is… **Agnese** you are so wise in the ways of statistical phasing!

Thank you to my office mate and fellow PhD student **Erik**. We started on this journey at the same time, and how impressive that we both managed to get the same amount of extension to finish within one week of each other. I'm so glad we shared an office over this time and got to chat each morning when I would show up to work several hours after you. I really appreciated it when we were both going through the thesis process, and could commiserate together, and it was especially great when you started your farm and I could buy eggs from you! And of course, a massive thank you for all your help and guidance when I had the most daunting task in the PhD, having to work in the lab for two weeks.

Karin, although I started a bit after you, we always felt like part of the same cohort. You are such a kind person, and you are so impressive! I'm sure you are going to continue to accomplish more amazing things in the future.

Over the past five years there have been so many fikas and lunches, and it took me a while to adjust to taking a proper lunch break! But I'm glad that I did and got to share this time with everyone and would like to thank all the other PhD students over the years. **Lars**, it's always nice to chat with you over lunch, and you're great at the Spex planning (and especially at playing the happiest leek). **David**, I enjoyed getting to TA with you in molecular evolution and how friendly you always were when we would chat in the kitchen. **Gabriel** thanks for all the discussions on structural variants and genome assembly! **Aleix** and **Milena**, it's nice to have fresh (but also returning) faces among the PhD students! Good luck to you both in your PhD. **McKenna** thank you for all the words of encouragement leading up to the thesis submission, I guess it was clear I was just hanging in there, haha. **Shadi** thanks for always being so kind and friendly. **Venkat** it was always nice to hang out and have a beer with you, both times that you were here in the program! Ghazhal, you were a great office mate when I started and you're a very caring person.

Thank you to everyone else in the evolutionary biology program. **Anna**, I always enjoyed your enthusiasm, especially for karaoke, and sense of humor. **Arild**, thank you for welcoming me into your journal clubs in the last year. **Elina** thank you for always being so welcoming and kind. **Niclas,** you've always been so friendly and helpful and I really appreciate it. **Vaishali** it was really nice to have you join the program and chat with you in the halls! **Doug** thank you for all your help with uppmax questions and for your work on the annotation! **Axel** I always enjoyed your dry sense of humor. **Veera,** thanks for the encouraging discussions about writing the thesis! It helped to feel like I would really manage this. **Martyna** it's always nice to catch up with you at

lunch, you're very kind! **Dasha** you bring a great energy to the program when you're around. **Thad** you always have some crazy stories to share over lunch or fika! To **Frida**, **Annette**, and **Jenny**, thank you all so much for the help and support with everything administrative, nothing would be managed without you.

Thank you to my collaborators outside of the EvoBio program. **David Wheatcroft**, working with you is such a joy because you have such a warm and welcoming attitude. Thank you for having us all to Tovetorp where I finally got to see the flycatchers in person! **Fanny**, even though I was very stressed about working on the RNA extractions it was really such a fun time getting to work with you! I really enjoyed your energy and the atmosphere you create. **Murielle** and **Anna** it was really nice to work with you both, and it definitely helped me to learn more about the ecology of the flycatchers. **Caro** it was a lot of fun staying up chatting and comparing flycatcher results with you at Tovetorp. I'm glad we got to work together, and that we became friends.

There have been so many wonderful friends I have met throughout my time in Uppsala, and as more and more of us spread out away from Sweden at least we know this means we have people in so many corners of the world. **Aaron**, I'm so grateful for your friendship over the years here, right from the start pretty much! We've had so many late-night talks, very occasionally with slightly raised voices, and usually a beer in our hands. I know you're not always the biggest fan of birds, but you've managed to be a supportive friend to me in spite of me studying them, and I appreciate that. **Sara,** my nail polish buddy! I'm glad we could always share what color we were wearing, and in general thank you for always being such a sweet person. **Adrian** I'm really glad that you guys moved here, and not just because you crush it on the music round at trivia. You're a great friend, and I'm so glad that we're also neighbors! **Ramsha,** of course also glad that you moved here too haha. Thank for all the great baked goods and movie nights, and always being up to host people! And so glad you were there for the legendary Valborg last year! Thank you both for looking after the cats so much and making the sacrificial offering to our sometimes spiteful little beast. **Dan**, we never quite know if or when you will show up to an event but you always arrive just when you mean to…and its always great! **Mercè**, it feels a bit intimidating to write your acknowledgement, knowing that you started writing yours in pretty much your first week. I am really grateful for the time we've spent together, all the Valborgs, Midsommars, Halloweens, Thanksgivings, and Christmases! I think we've spent almost every holiday together by now! We've huddled under the heat lamps outside at Palermo when it was -10 degrees, and sipped Three Towns out in the courtyard of Uplands in the summertime (with **Aaron** too!). Thank you for all the long chats, and no matter where we end up, I hope for more in the future. **Felix**, I'm so glad to have met you this past year, you are

a great person and a great addition to the trivia team! I'll never question your music knowledge again after the Miley Cyrus/Dua Lipa disaster. **Theo**, I think that you really have an unbeaten enthusiasm for fondue, I've never seen it make anyone quite as happy as it did you. **Karin Steffen** you have such a joyful enthusiasm, it's been so fun to get to know you! **Alex** I really enjoyed the times when we shared journal clubs and group meetings, I learned so much from that. You're such a kind and welcoming person. It was always a lot of fun chatting at FAWNA, and I will miss your homebrewed wines! **Lore** I'm super glad to have moved in with you when I first came to Sweden, and not just because I had literally no backup plan. It can be very lonely moving to a new country where you don't know anyone (as many of us know) and having you as a roommate in the beginning helped make things easier. I really hope that one day soon they develop a cure for cat allergies, because you deserve it.

Ele, how dare you and **Julian** move away before my defense, I will never forgive you. But really, you are such a lovely person and friend, and I'm glad we got to know each other better. Especially happy to have received my very own illustrious knit headband ;) **Julian**, you are an inspiration to all to be more chill and get less stressed about things (after all, it doesn't help anything so why do people even get stressed?). And to both of you, **Ele** and **Julian**, thank you so much for all the amazing weekend trips, cheesy nights, board games, looking after the cats, and everything, but especially thank you for a group chat completely full of cat photos. Hoping for so much more of the same in the future. And to **Pippi** you are a floofy little monster and I love you so much.

Valentina one of my favorite memories with you is when you fiercely defended us during the fork scavenger hunt at the program days after another team picked up *our* fork because they misunderstood *our* clue, yelling 'it's an ichthyosaur not a plesiosaur!'. I have so enjoyed getting to know you, you're a lovely and kind person (but you will not stand for the misidentification of fossils), and you are an amazing scientist! And you always bring the best board games. **Marco**, you are a joyful person to be around, even if you've heard it many times, it's hard not to enjoy your Swedish cat joke (I think it actually gets better each time you hear it). And to both of you, **Vale** and **Marco**, thank you so much for all the times that you helped to watch our cats, even when it meant having to watch Tiki like a hawk to make sure she actually took her medicine! To **Carl Gustav**, you were the queen of Götgatan and I'm grateful for all the times you graced us with your floofy presence.

One constant over the past few years that has been a guiding light through difficult times has been the regular pub quiz nights at William's (and before at Pitcher's). I have a huge thank you to say to all the members of our trivia team, Quizz my ass. **Aaron, Adrian, Dan, Felix, Mercè, Philipp, Ramsha,**

Sara, Theo, thank you all for the many glorious wins and also many bitter defeats! And of course, a big thank you to the quiz master himself, **Chris**, for always being an amazing host, so that we would have a great time win or lose (but mostly when we win).

To people from back in Oregon:

Matt, I am really glad to have started working in your lab years ago during my undergrad (nearly ten?!), starting out with labeling plant pots. When I began university, I wasn't really sure what I wanted to study, but by the time I joined your lab I was at least reasonably sure I was interested in biology. It was through the research experience in your lab though that I really became convinced that I wanted to study evolution, and that eventually led me here to write this book. So, thank you for that opportunity and for supporting me along the way.

Annika, absolutely wild that we go all the way back to Mr. Schaff's math class in the sixth grade. Thank you for sticking through with me these years and being an amazing friend. I owe a lot to our last names being next to each other in the alphabet so that we would get sat next to each other in class. Maybe someday we should start to co-write another epic fantasy story. **Ben**, you've really been a lifeline to me in the past, I'm lucky to have you as a friend (I think I owe that to the alphabet as well, for you and Annika). For both of you, thank you for so much, all the years of Playmobil pirate ship battles and competitive chopped watching. I miss living in the same city! And thank you to **Nina** for all her Zoom appearances and for jumping onto my face at 2am <3.

Bridget and **Tyler**, thank you guys as well for our long friendship and miss you both! It was so fun to see you both in Chicago when you moved there, but even better when you moved back to Eugene. I loved our trivia nights at 16 tonnes!

Kaylee, first of all thank you for painting a beautiful cover for my thesis! I'm so grateful that we happened to move into the same house 12 years ago, you were literally one of the reasons I didn't want to switch universities. You've been such an amazing friend, even when I'm terrible at answering your FaceTimes. I miss you so much! **Eric,** we've known each other almost as long, and I'm so happy to have you as a friend. I loved how you would vibe with the Beef, and so miss going out with you or watching hours of Game of Thrones followed by some Smash Bros! Complete with intense trash talking from **Kaylee** of course. Love you both!

Jack and **Donna** for a long time your house was like a second home to me, thank you for giving me so much support through the years. And I'm glad that I could see you here in Sweden when you visited!

Debbie and **Michael**, thank you for all your support, all the brunches at Morning Glory, and for always welcoming me in your home!

A big thank you to my family. **Mum**, thank you for always supporting me in every way that you can. Your place was always open for a cup of tea or some cheese on toast. **Lucian** and **Rachael**, I'm so glad you guys got to come visit Uppsala while I was here! And in keeping with the theme of acknowledging everybody's cats, to **Archie** and **Hunter**, who I enjoyed getting to see on skype causing chaos. **Grandma, Simon, Lee, Sarah, Ralph,** thank you all for everything!

And lastly to my own little family here. **Tiki** and **Nemo,** thank you for being the best cats ever (sorry to everyone else). I'm not sure if I would really use the word 'bästis' to describe you guys, but you have your moments where you tolerate each other. I can't believe we ever lived together without cats around. **Philipp**, thank you so much for all your incredible support, how would I have ever managed without it all. I feel so happy to have you in my life, and to have you on my team. Thank you for all the dinners you made when I was too stressed to, all the figures you helped me make when I couldn't figure out how to make ggplot do what I wanted, and for helping keep me (mostly) sane along this journey. I'm glad that we will face whatever the future holds at each other's side.

References

Ålund, M., Immler, S., Rice, A. M., & Qvarnström, A. (2013). Low fertility of wild hybrid male flycatchers despite recent divergence. *Biology Letters*, *9*(3), 20130169. https://doi.org/10.1098/rsbl.2013.0169

Axelsson, E., Webster, M. T., Smith, N. G. C., Burt, D. W., & Ellegren, H. (2005). Comparison of the chicken and turkey genomes reveals a higher rate of nucleotide divergence on microchromosomes than macrochromosomes. *Genome Research*, *15*(1), 120–125. https://doi.org/10.1101/gr.3021305

Baudat, F., Buard, J., Grey, C., Fledel-Alon, A., Ober, C., Przeworski, M., Coop, G., & de Massy, B. (2010). PRDM9 Is a Major Determinant of Meiotic Recombination Hotspots in Humans and Mice. *Science*, *327*(5967), 836–840. https://doi.org/10.1126/science.1183439

Berner, D., & Salzburger, W. (2015). The genomics of organismal diversification illuminated by adaptive radiations. *Trends in Genetics*, *31*(9), 491–499. https://doi.org/10.1016/j.tig.2015.07.002

Bolívar, P., Mugal, C. F., Nater, A., & Ellegren, H. (2016). Recombination Rate Variation Modulates Gene Sequence Evolution Mainly via GC-Biased Gene Conversion, Not Hill–Robertson Interference, in an Avian System. *Molecular Biology and Evolution*, *33*(1), 216–227. https://doi.org/10.1093/molbev/msv214

Bolívar, P., Mugal, C. F., Rossi, M., Nater, A., Wang, M., Dutoit, L., & Ellegren, H. (2018). Biased Inference of Selection Due to GC-Biased Gene Conversion and the Rate of Protein Evolution in Flycatchers When Accounting for It. *Molecular Biology and Evolution*, *35*(10), 2475–2486. https://doi.org/10.1093/molbev/msy149

Bolnick, D. I., Hund, A. K., Nosil, P., Peng, F., Ravinet, M., Stankowski, S., Subramanian, S., Wolf, J. B. W., & Yukilevich, R. (2023). A multivariate view of the speciation continuum. *Evolution*, *77*(1), 318–328. https://doi.org/10.1093/evolut/qpac004

Burri, R. (2017). Interpreting differentiation landscapes in the light of long-term linked selection. *Evolution Letters*, *1*(3), 118–131. https://doi.org/10.1002/evl3.14

Burri, R., Nater, A., Kawakami, T., Mugal, C. F., Olason, P. I., Smeds, L., Suh, A., Dutoit, L., Bureš, S., Garamszegi, L. Z., Hogner, S., Moreno, J., Qvarnström, A., Ružić, M., Sæther, S.-A., Sætre, G.-P., Török, J., & Ellegren, H. (2015). Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of Ficedula flycatchers. *Genome Research*, *25*(11), 1656–1665. https://doi.org/10.1101/gr.196485.115

Catanach, A., Crowhurst, R., Deng, C., David, C., Bernatchez, L., & Wellenreuther, M. (2019). The genomic pool of standing structural variation outnumbers single nucleotide polymorphism by threefold in the marine teleost Chrysophrys

auratus. *Molecular Ecology*, *28*(6), 1210–1223. https://doi.org/10.1111/mec.15051
Chan, A. H., Jenkins, P. A., & Song, Y. S. (2012). Genome-Wide Fine-Scale Recombination Rate Variation in Drosophila melanogaster. *PLOS Genetics*, *8*(12), e1003090. https://doi.org/10.1371/journal.pgen.1003090
Christmas, M. J., Jones, J. C., Olsson, A., Wallerman, O., Bunikis, I., Kierczak, M., Peona, V., Whitley, K. M., Larva, T., Suh, A., Miller-Struttmann, N. E., Geib, J. C., & Webster, M. T. (2021). Genetic Barriers to Historical Gene Flow between Cryptic Species of Alpine Bumblebees Revealed by Comparative Population Genomics. *Molecular Biology and Evolution*, *38*(8), 3126–3143. https://doi.org/10.1093/molbev/msab086
Conrad, D. F., & Hurles, M. E. (2007). The population genetics of structural variation. *Nature Genetics*, *39*(7), Article 7. https://doi.org/10.1038/ng2042
Corbett-Detig, R. B., Hartl, D. L., & Sackton, T. B. (2015). Natural Selection Constrains Neutral Diversity across A Wide Range of Species. *PLOS Biology*, *13*(4), e1002112. https://doi.org/10.1371/journal.pbio.1002112
Coughlan, J. M., & Matute, D. R. (2020). The importance of intrinsic postzygotic barriers throughout the speciation process. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1806), 20190533. https://doi.org/10.1098/rstb.2019.0533
Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, *23*(13), 3133–3157. https://doi.org/10.1111/mec.12796
Dapper, A. L., & Payseur, B. A. (2018). Effects of Demographic History on the Detection of Recombination Hotspots from Linkage Disequilibrium. *Molecular Biology and Evolution*, *35*(2), 335–353. https://doi.org/10.1093/molbev/msx272
DeGiorgio, M., Huber, C. D., Hubisz, M. J., Hellmann, I., & Nielsen, R. (2016). S weep F inder 2: Increased sensitivity, robustness and flexibility. *Bioinformatics*, *32*(12), 1895–1897. https://doi.org/10.1093/bioinformatics/btw051
Delaneau, O., Marchini, J., & Zagury, J.-F. (2012). A linear complexity phasing method for thousands of genomes. *Nature Methods*, *9*(2), Article 2. https://doi.org/10.1038/nmeth.1785
Delmore, K. E., Ramos, J. S. L., Doren, B. M. V., Lundberg, M., Bensch, S., Irwin, D. E., & Liedvogel, M. (2018). Comparative analysis examining patterns of genomic differentiation across multiple episodes of population divergence in birds. *Evolution Letters*, *2*(2), 76–87. https://doi.org/10.1002/evl3.46
Dorant, Y., Cayuela, H., Wellband, K., Laporte, M., Rougemont, Q., Mérot, C., Normandeau, E., Rochette, R., & Bernatchez, L. (2020). Copy number variants outperform SNPs to reveal genotype–temperature association in a marine species. *Molecular Ecology*, *29*(24), 4765–4782. https://doi.org/10.1111/mec.15565
Duret, L., & Galtier, N. (2009). Biased Gene Conversion and the Evolution of Mammalian Genomic Landscapes. *Annual Review of Genomics and Human Genetics*, *10*(1), 285–311. https://doi.org/10.1146/annurev-genom-082908-150001
Dutheil, J., & Boussau, B. (2008). Non-homogeneous models of sequence evolution in the Bio++ suite of libraries and programs. *BMC Evolutionary Biology*, *8*(1), 255. https://doi.org/10.1186/1471-2148-8-255
Dutoit, L., Vijay, N., Mugal, C. F., Bossu, C. M., Burri, R., Wolf, J., & Ellegren, H. (2017). Covariation in levels of nucleotide diversity in homologous regions of the avian genome long after completion of lineage sorting. *Proceedings of*

the Royal Society B: Biological Sciences, *284*(1849), 20162756. https://doi.org/10.1098/rspb.2016.2756

Eizenga, J. M., Novak, A. M., Sibbesen, J. A., Heumos, S., Ghaffaari, A., Hickey, G., Chang, X., Seaman, J. D., Rounthwaite, R., Ebler, J., Rautiainen, M., Garg, S., Paten, B., Marschall, T., Sirén, J., & Garrison, E. (2020). Pangenome Graphs. *Annual Review of Genomics and Human Genetics*, *21*(1), 139–162. https://doi.org/10.1146/annurev-genom-120219-080406

Ellegren, H. (2010). Evolutionary stasis: The stable chromosomes of birds. *Trends in Ecology & Evolution*, *25*(5), 283–291. https://doi.org/10.1016/j.tree.2009.12.004

Ellegren, H., Smeds, L., Burri, R., Olason, P. I., Backström, N., Kawakami, T., Künstner, A., Mäkinen, H., Nadachowska-Brzyska, K., Qvarnström, A., Uebbing, S., & Wolf, J. B. W. (2012). The genomic landscape of species divergence in Ficedula flycatchers. *Nature*, *491*(7426), Article 7426. https://doi.org/10.1038/nature11584

Eyre-Walker, A., & Keightley, P. D. (2009). Estimating the Rate of Adaptive Molecular Evolution in the Presence of Slightly Deleterious Mutations and Population Size Change. *Molecular Biology and Evolution*, *26*(9), 2097–2108. https://doi.org/10.1093/molbev/msp119

Feulner, P. G. D., & De-Kayne, R. (2017). Genome evolution, structural rearrangements and speciation. *Journal of Evolutionary Biology*, *30*(8), 1488–1490. https://doi.org/10.1111/jeb.13101

Glémin, S. (2010). Surprising Fitness Consequences of GC-Biased Gene Conversion: I. Mutation Load and Inbreeding Depression. *Genetics*, *185*(3), 939–959. https://doi.org/10.1534/genetics.110.116368

Gray, S., & Cohen, P. E. (2016). Control of Meiotic Crossovers: From Double-Strand Break Formation to Designation. *Annual Review of Genetics*, *50*(1), 175–210. https://doi.org/10.1146/annurev-genet-120215-035111

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M. H.-Y., Hansen, N. F., Durand, E. Y., Malaspinas, A.-S., Jensen, J. D., Marques-Bonet, T., Alkan, C., Prüfer, K., Meyer, M., Burbano, H. A., … Pääbo, S. (2010). A Draft Sequence of the Neandertal Genome. *Science*, *328*(5979), 710–722. https://doi.org/10.1126/science.1188021

Griffin, D. K., Robertson, L. B. W., Tempest, H. G., & Skinner, B. M. (2007). The evolution of the avian genome as revealed by comparative molecular cytogenetics. *Cytogenetic and Genome Research*, *117*(1–4), 64–77. https://doi.org/10.1159/000103166

Haenel, Q., Laurentino, T. G., Roesti, M., & Berner, D. (2018). Meta-analysis of chromosome-scale crossover rate variation in eukaryotes and its significance to evolutionary genomics. *Molecular Ecology*, *27*(11), 2477–2497. https://doi.org/10.1111/mec.14699

Haller, B. C., & Messer, P. W. (2019). SLiM 3: Forward Genetic Simulations Beyond the Wright–Fisher Model. *Molecular Biology and Evolution*, *36*(3), 632–637. https://doi.org/10.1093/molbev/msy228

Hämälä, T., Wafula, E. K., Guiltinan, M. J., Ralph, P. E., dePamphilis, C. W., & Tiffin, P. (2021). Genomic structural variants constrain and facilitate adaptation in natural populations of Theobroma cacao, the chocolate tree. *Proceedings of the National Academy of Sciences*, *118*(35), e2102914118. https://doi.org/10.1073/pnas.2102914118

Hammer, M. F., Woerner, A. E., Mendez, F. L., Watkins, J. C., Cox, M. P., & Wall, J. D. (2010). The ratio of human X chromosome to autosome diversity is

positively correlated with genetic distance from genes. *Nature Genetics*, *42*(10), Article 10. https://doi.org/10.1038/ng.651
Hayes, K., Barton, H. J., & Zeng, K. (2020). A Study of Faster-Z Evolution in the Great Tit (Parus major). *Genome Biology and Evolution*, *12*(3), 210–222. https://doi.org/10.1093/gbe/evaa044
Hickey, G., Monlong, J., Novak, A., Eizenga, J. M., Consortium, H. P. R., Li, H., & Paten, B. (2022). *Pangenome Graph Construction from Genome Alignment with Minigraph-Cactus* (p. 2022.10.06.511217). bioRxiv. https://doi.org/10.1101/2022.10.06.511217
Hill, W. G., & Robertson, A. (1966). The effect of linkage on limits to artificial selection. *Genetics Research*, *8*(3), 269–294. https://doi.org/10.1017/S0016672300010156
Hillier, L. W., Miller, W., Birney, E., Warren, W., Hardison, R. C., Ponting, C. P., Bork, P., Burt, D. W., Groenen, M. A. M., Delany, M. E., Dodgson, J. B., Chinwalla, A. T., Cliften, P. F., Clifton, S. W., Delehaunty, K. D., Fronick, C., Fulton, R. S., Graves, T. A., Kremitzki, C., … Project management: (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*, *432*(7018), Article 7018. https://doi.org/10.1038/nature03154
Hooper, D. M., Griffith, S. C., & Price, T. D. (2019). Sex chromosome inversions enforce reproductive isolation across an avian hybrid zone. *Molecular Ecology*, *28*(6), 1246–1262. https://doi.org/10.1111/mec.14874
Howe, K., Chow, W., Collins, J., Pelan, S., Pointon, D.-L., Sims, Y., Torrance, J., Tracey, A., & Wood, J. (2021). Significantly improving the quality of genome assemblies through curation. *GigaScience*, *10*(1), giaa153. https://doi.org/10.1093/gigascience/giaa153
Huber, C. D., DeGiorgio, M., Hellmann, I., & Nielsen, R. (2016). Detecting recent selective sweeps while controlling for mutation rate and background selection. *Molecular Ecology*, *25*(1), 142–156. https://doi.org/10.1111/mec.13351
Hudson, R. R. (1983). Testing the Constant-Rate Neutral Allele Model with Protein Sequence Data. *Evolution*, *37*(1), 203–217. https://doi.org/10.2307/2408186
Hung, C.-M., & Zink, R. M. (2014). Distinguishing the effects of selection from demographic history in the genetic variation of two sister passerines based on mitochondrial–nuclear comparison. *Heredity*, *113*(1), Article 1. https://doi.org/10.1038/hdy.2014.9
Irwin, D. E. (2018). Sex chromosomes and speciation in birds and other ZW systems. *Molecular Ecology*, *27*(19), 3831–3851. https://doi.org/10.1111/mec.14537
Irwin, D. E. (2020). Assortative Mating in Hybrid Zones Is Remarkably Ineffective in Promoting Speciation. *The American Naturalist*, *195*(6), E150–E167. https://doi.org/10.1086/708529
Jay, P., Whibley, A., Frézal, L., Rodríguez de Cara, M. Á., Nowell, R. W., Mallet, J., Dasmahapatra, K. K., & Joron, M. (2018). Supergene Evolution Triggered by the Introgression of a Chromosomal Inversion. *Current Biology*, *28*(11), 1839-1845.e3. https://doi.org/10.1016/j.cub.2018.04.072
Kaplan, N. L., Hudson, R. R., & Langley, C. H. (1989). The "hitchhiking effect" revisited. *Genetics*, *123*(4), 887–899. https://doi.org/10.1093/genetics/123.4.887
Kapusta, A., & Suh, A. (2017). Evolution of bird genomes—A transposon's-eye view. *Annals of the New York Academy of Sciences*, *1389*(1), 164–185. https://doi.org/10.1111/nyas.13295

Kapusta, A., Suh, A., & Feschotte, C. (2017). Dynamics of genome size evolution in birds and mammals. *Proceedings of the National Academy of Sciences*, *114*(8), E1460–E1469. https://doi.org/10.1073/pnas.1616702114

Katju, V., & Bergthorsson, U. (2013). Copy-number changes in evolution: Rates, fitness effects and adaptive significance. *Frontiers in Genetics*, *4*. https://www.frontiersin.org/articles/10.3389/fgene.2013.00273

Kawakami, T., Mugal, C. F., Suh, A., Nater, A., Burri, R., Smeds, L., & Ellegren, H. (2017). Whole-genome patterns of linkage disequilibrium across flycatcher populations clarify the causes and consequences of fine-scale recombination rate variation in birds. *Molecular Ecology*, *26*(16), 4158–4172. https://doi.org/10.1111/mec.14197

Kawakami, T., Smeds, L., Backström, N., Husby, A., Qvarnström, A., Mugal, C. F., Olason, P., & Ellegren, H. (2014). A high-density linkage map enables a second-generation collared flycatcher genome assembly and reveals the patterns of avian recombination rate variation and chromosomal evolution. *Molecular Ecology*, *23*(16), 4035–4058. https://doi.org/10.1111/mec.12810

Keightley, P. D., & Eyre-Walker, A. (2007). Joint Inference of the Distribution of Fitness Effects of Deleterious Mutations and Population Demography Based on Nucleotide Polymorphism Frequencies. *Genetics*, *177*(4), 2251–2261. https://doi.org/10.1534/genetics.107.080663

Kim, J., Lee, C., Ko, B. J., Yoo, D. A., Won, S., Phillippy, A. M., Fedrigo, O., Zhang, G., Howe, K., Wood, J., Durbin, R., Formenti, G., Brown, S., Cantin, L., Mello, C. V., Cho, S., Rhie, A., Kim, H., & Jarvis, E. D. (2022). False gene and chromosome losses in genome assemblies caused by GC content variation and repeats. *Genome Biology*, *23*(1), 204. https://doi.org/10.1186/s13059-022-02765-0

Klein, S. J., & O'Neill, R. J. (2018). Transposable elements: Genome innovation, chromosome diversity, and centromere conflict. *Chromosome Research*, *26*(1), 5–23. https://doi.org/10.1007/s10577-017-9569-5

Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O., Sandstrom, R., Bernstein, B., Bender, M. A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L. A., Lander, E. S., & Dekker, J. (2009). Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science*, *326*(5950), 289–293. https://doi.org/10.1126/science.1181369

López-Cortegano, E., Craig, R. J., Chebib, J., Balogun, E. J., & Keightley, P. D. (2023). Rates and spectra of de novo structural mutations in Chlamydomonas reinhardtii. *Genome Research*, *33*(1), 45–60. https://doi.org/10.1101/gr.276957.122

Lowry, D. B., & Willis, J. H. (2010). A Widespread Chromosomal Inversion Polymorphism Contributes to a Major Life-History Transition, Local Adaptation, and Reproductive Isolation. *PLOS Biology*, *8*(9), e1000500. https://doi.org/10.1371/journal.pbio.1000500

Lundberg, A., & Alatalo, R. V. (1992). *The Pied Flycatcher*. A&C Black.

Mahmoud, M., Gobet, N., Cruz-Dávalos, D. I., Mounier, N., Dessimoz, C., & Sedlazeck, F. J. (2019). Structural variant calling: The long and the short of it. *Genome Biology*, *20*(1), 246. https://doi.org/10.1186/s13059-019-1828-7

Mallet, J., & Mullen, S. P. (2022). Reproductive isolation is a heuristic, not a measure: A commentary on Westram et al., 2022. *Journal of Evolutionary Biology*, *35*(9), 1175–1182. https://doi.org/10.1111/jeb.14052

Mank, J. E., Nam, K., & Ellegren, H. (2010). Faster-Z Evolution Is Predominantly Due to Genetic Drift. *Molecular Biology and Evolution*, *27*(3), 661–670. https://doi.org/10.1093/molbev/msp282

Marçais, G., Delcher, A. L., Phillippy, A. M., Coston, R., Salzberg, S. L., & Zimin, A. (2018). MUMmer4: A fast and versatile genome alignment system. *PLOS Computational Biology*, *14*(1), e1005944. https://doi.org/10.1371/journal.pcbi.1005944

Martin, S. H., Dasmahapatra, K. K., Nadeau, N. J., Salazar, C., Walters, J. R., Simpson, F., Blaxter, M., Manica, A., Mallet, J., & Jiggins, C. D. (2013). Genome-wide evidence for speciation with gene flow in Heliconius butterflies. *Genome Research*, *23*(11), 1817–1828. https://doi.org/10.1101/gr.159426.113

Martin, S. H., Davey, J. W., & Jiggins, C. D. (2015). Evaluating the Use of ABBA–BABA Statistics to Locate Introgressed Loci. *Molecular Biology and Evolution*, *32*(1), 244–257. https://doi.org/10.1093/molbev/msu269

Martin, S. H., Davey, J. W., Salazar, C., & Jiggins, C. D. (2019). Recombination rate variation shapes barriers to introgression across butterfly genomes. *PLOS Biology*, *17*(2), e2006288. https://doi.org/10.1371/journal.pbio.2006288

Matschiner, M., Barth, J. M. I., Tørresen, O. K., Star, B., Baalsrud, H. T., Brieuc, M. S. O., Pampoulie, C., Bradbury, I., Jakobsen, K. S., & Jentoft, S. (2022). Supergene origin and maintenance in Atlantic cod. *Nature Ecology & Evolution*, *6*(4), Article 4. https://doi.org/10.1038/s41559-022-01661-x

Mayr, E. (1963). Animal Species and Evolution. In *Animal Species and Evolution*. Harvard University Press. https://doi.org/10.4159/harvard.9780674865327

McDonald, J. H., & Kreitman, M. (1991). Adaptive protein evolution at the Adh locus in Drosophila. *Nature*, *351*(6328), Article 6328. https://doi.org/10.1038/351652a0

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

Mérot, C., Oomen, R. A., Tigano, A., & Wellenreuther, M. (2020). A Roadmap for Understanding the Evolutionary Significance of Structural Genomic Variation. *Trends in Ecology & Evolution*, *35*(7), 561–572. https://doi.org/10.1016/j.tree.2020.03.002

Mugal, C. F., Kutschera, V. E., Botero-Castro, F., Wolf, J. B. W., & Kaj, I. (2020). Polymorphism Data Assist Estimation of the Nonsynonymous over Synonymous Fixation Rate Ratio ω for Closely Related Species. *Molecular Biology and Evolution*, *37*(1), 260–279. https://doi.org/10.1093/molbev/msz203

Müller, R., Kaj, I., & Mugal, C. F. (2022). A Nearly Neutral Model of Molecular Signatures of Natural Selection after Change in Population Size. *Genome Biology and Evolution*, *14*(5), evac058. https://doi.org/10.1093/gbe/evac058

Myers, S., Bowden, R., Tumian, A., Bontrop, R. E., Freeman, C., MacFie, T. S., McVean, G., & Donnelly, P. (2010). Drive Against Hotspot Motifs in Primates Implicates the PRDM9 Gene in Meiotic Recombination. *Science*, *327*(5967), 876–879. https://doi.org/10.1126/science.1182363

Nachman, M. W., & Payseur, B. A. (2012). Recombination rate variation and speciation: Theoretical predictions and empirical results from rabbits and mice. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1587), 409–421. https://doi.org/10.1098/rstb.2011.0249

Nadachowska-Brzyska, K., Burri, R., & Ellegren, H. (2019). Footprints of adaptive evolution revealed by whole Z chromosomes haplotypes in flycatchers. *Molecular Ecology*, *28*(9), 2290–2304. https://doi.org/10.1111/mec.15021

Nadachowska-Brzyska, K., Burri, R., Olason, P. I., Kawakami, T., Smeds, L., & Ellegren, H. (2013). Demographic Divergence History of Pied Flycatcher and Collared Flycatcher Inferred from Whole-Genome Re-sequencing Data. *PLOS Genetics*, *9*(11), e1003942. https://doi.org/10.1371/journal.pgen.1003942

Nadachowska-Brzyska, K., Burri, R., Smeds, L., & Ellegren, H. (2016). PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white Ficedula flycatchers. *Molecular Ecology*, *25*(5), 1058–1072. https://doi.org/10.1111/mec.13540

Nadachowska-Brzyska, K., Dutoit, L., Smeds, L., Kardos, M., Gustafsson, L., & Ellegren, H. (2021). Genomic inference of contemporary effective population size in a large island population of collared flycatchers (Ficedula albicollis). *Molecular Ecology*, *30*(16), 3965–3973. https://doi.org/10.1111/mec.16025

Nadeau, N. J., Whibley, A., Jones, R. T., Davey, J. W., Dasmahapatra, K. K., Baxter, S. W., Quail, M. A., Joron, M., ffrench-Constant, R. H., Blaxter, M. L., Mallet, J., & Jiggins, C. D. (2012). Genomic islands of divergence in hybridizing Heliconius butterflies identified by large-scale targeted sequencing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1587), 343–353. https://doi.org/10.1098/rstb.2011.0198

Nater, A., Burri, R., Kawakami, T., Smeds, L., & Ellegren, H. (2015). Resolving Evolutionary Relationships in Closely Related Species with Whole-Genome Sequencing Data. *Systematic Biology*, *64*(6), 1000–1017. https://doi.org/10.1093/sysbio/syv045

Nielsen, R., Williamson, S., Kim, Y., Hubisz, M. J., Clark, A. G., & Bustamante, C. (2005). Genomic scans for selective sweeps using SNP data. *Genome Research*, *15*(11), 1566–1575. https://doi.org/10.1101/gr.4252305

Noor, M. A. F., Grams, K. L., Bertucci, L. A., & Reiland, J. (2001). Chromosomal inversions and the reproductive isolation of species. *Proceedings of the National Academy of Sciences*, *98*(21), 12084–12088. https://doi.org/10.1073/pnas.221274498

Nosil, P. (2008). Speciation with gene flow could be common. *Molecular Ecology*, *17*(9), 2103–2106. https://doi.org/10.1111/j.1365-294X.2008.03715.x

O'Donnell, S., & Fischer, G. (2020). MUM&Co: Accurate detection of all SV types through whole-genome alignment. *Bioinformatics*, *36*(10), 3242–3243. https://doi.org/10.1093/bioinformatics/btaa115

Ortiz-Barrientos, D., Engelstädter, J., & Rieseberg, L. H. (2016). Recombination Rate Evolution and the Origin of Species. *Trends in Ecology & Evolution*, *31*(3), 226–236. https://doi.org/10.1016/j.tree.2015.12.016

Pang, A. W., MacDonald, J. R., Pinto, D., Wei, J., Rafiq, M. A., Conrad, D. F., Park, H., Hurles, M. E., Lee, C., Venter, J. C., Kirkness, E. F., Levy, S., Feuk, L., & Scherer, S. W. (2010). Towards a comprehensive structural variation map of an individual human genome. *Genome Biology*, *11*(5), R52. https://doi.org/10.1186/gb-2010-11-5-r52

Parvanov, E. D., Petkov, P. M., & Paigen, K. (2010). Prdm9 Controls Activation of Mammalian Recombination Hotspots. *Science*, *327*(5967), 835–835. https://doi.org/10.1126/science.1181495

Peona, V., Blom, M. P. K., Xu, L., Burri, R., Sullivan, S., Bunikis, I., Liachko, I., Haryoko, T., Jønsson, K. A., Zhou, Q., Irestedt, M., & Suh, A. (2021). Identifying the causes and consequences of assembly gaps using a multiplatform

genome assembly of a bird-of-paradise. *Molecular Ecology Resources*, *21*(1), 263–286. https://doi.org/10.1111/1755-0998.13252

Peona, V., Palacios-Gimenez, O. M., Blommaert, J., Liu, J., Haryoko, T., Jønsson, K. A., Irestedt, M., Zhou, Q., Jern, P., & Suh, A. (2021). The avian W chromosome is a refugium for endogenous retroviruses with likely effects on female-biased mutational load and genetic incompatibilities. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1833), 20200186. https://doi.org/10.1098/rstb.2020.0186

Ponting, C. P. (2011). What are the genomic drivers of the rapid evolution of PRDM9? *Trends in Genetics*, *27*(5), 165–171. https://doi.org/10.1016/j.tig.2011.02.001

Price, T. D., & Bouvier, M. M. (2002). The Evolution of F1 Postzygotic Incompatibilities in Birds. *Evolution*, *56*(10), 2083–2089. https://doi.org/10.1111/j.0014-3820.2002.tb00133.x

Ravinet, M., Faria, R., Butlin, R. K., Galindo, J., Bierne, N., Rafajlović, M., Noor, M. a. F., Mehlig, B., & Westram, A. M. (2017). Interpreting the genomic landscape of speciation: A road map for finding barriers to gene flow. *Journal of Evolutionary Biology*, *30*(8), 1450–1477. https://doi.org/10.1111/jeb.13047

Raynaud, M., Gagnaire, P.-A., & Galtier, N. (2022). *Performance and limitations of linkage-disequilibrium-based methods for inferring the genomic landscape of recombination and detecting hotspots: A simulation study* (p. 2022.03.30.486352). bioRxiv. https://doi.org/10.1101/2022.03.30.486352

Reed, F. A., & Tishkoff, S. A. (2006). Positive Selection Can Create False Hotspots of Recombination. *Genetics*, *172*(3), 2011–2014. https://doi.org/10.1534/genetics.105.052183

Renaut, S., Grassa, C. J., Yeaman, S., Moyers, B. T., Lai, Z., Kane, N. C., Bowers, J. E., Burke, J. M., & Rieseberg, L. H. (2013). Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nature Communications*, *4*(1), Article 1. https://doi.org/10.1038/ncomms2833

Rettelbach, A., Nater, A., & Ellegren, H. (2019). How Linked Selection Shapes the Diversity Landscape in Ficedula Flycatchers. *Genetics*, *212*(1), 277–285. https://doi.org/10.1534/genetics.119.301991

Ruggieri, A. A., Livraghi, L., Lewis, J. J., Evans, E., Cicconardi, F., Hebberecht, L., Ortiz-Ruiz, Y., Montgomery, S. H., Ghezzi, A., Rodriguez-Martinez, J. A., Jiggins, C. D., McMillan, W. O., Counterman, B. A., Papa, R., & Belleghem, S. M. V. (2022). A butterfly pan-genome reveals that a large amount of structural variation underlies the evolution of chromatin accessibility. *Genome Research*, *32*(10), 1862–1875. https://doi.org/10.1101/gr.276839.122

Rundle, H. D., & Nosil, P. (2005). Ecological speciation. *Ecology Letters*, *8*(3), 336–352. https://doi.org/10.1111/j.1461-0248.2004.00715.x

Sæther, S. A., Sætre, G.-P., Borge, T., Wiley, C., Svedin, N., Andersson, G., Veen, T., Haavie, J., Servedio, M. R., Bureš, S., Král, M., Hjernquist, M. B., Gustafsson, L., Träff, J., & Qvarnström, A. (2007). Sex Chromosome-Linked Species Recognition and Evolution of Reproductive Isolation in Flycatchers. *Science*, *318*(5847), 95–97. https://doi.org/10.1126/science.1141506

Schluter, D. (2001). Ecology and the origin of species. *Trends in Ecology & Evolution*, *16*(7), 372–380. https://doi.org/10.1016/S0169-5347(01)02198-X

Schrider, D. R. (2020). Background Selection Does Not Mimic the Patterns of Genetic Diversity Produced by Selective Sweeps. *Genetics*. https://doi.org/10.1534/genetics.120.303469

Schumer, M., Xu, C., Powell, D. L., Durvasula, A., Skov, L., Holland, C., Blazier, J. C., Sankararaman, S., Andolfatto, P., Rosenthal, G. G., & Przeworski, M. (2018). Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science*, *360*(6389), 656–660. https://doi.org/10.1126/science.aar3684

Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., Peichel, C. L., Saetre, G.-P., Bank, C., Brännström, Å., Brelsford, A., Clarkson, C. S., Eroukhmanoff, F., Feder, J. L., Fischer, M. C., Foote, A. D., Franchini, P., Jiggins, C. D., Jones, F. C., … Widmer, A. (2014). Genomics and the origin of species. *Nature Reviews Genetics*, *15*(3), Article 3. https://doi.org/10.1038/nrg3644

Singhal, S., Leffler, E. M., Sannareddy, K., Turner, I., Venn, O., Hooper, D. M., Strand, A. I., Li, Q., Raney, B., Balakrishnan, C. N., Griffith, S. C., McVean, G., & Przeworski, M. (2015). Stable recombination hotspots in birds. *Science*, *350*(6263), 928–932. https://doi.org/10.1126/science.aad0843

Smeds, L., Warmuth, V., Bolivar, P., Uebbing, S., Burri, R., Suh, A., Nater, A., Bureš, S., Garamszegi, L. Z., Hogner, S., Moreno, J., Qvarnström, A., Ružić, M., Sæther, S.-A., Sætre, G.-P., Török, J., & Ellegren, H. (2015). Evolutionary analysis of the female-specific avian W chromosome. *Nature Communications*, *6*(1), Article 1. https://doi.org/10.1038/ncomms8330

Smith, J. M., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetics Research*, *23*(1), 23–35. https://doi.org/10.1017/S0016672300014634

Smukowski, C. S., & Noor, M. a. F. (2011). Recombination rate variation in closely related species. *Heredity*, *107*(6), Article 6. https://doi.org/10.1038/hdy.2011.44

Sobel, J. M., & Chen, G. F. (2014). Unification of Methods for Estimating the Strength of Reproductive Isolation. *Evolution*, *68*(5), 1511–1522. https://doi.org/10.1111/evo.12362

Stankowski, S., Chase, M. A., Fuiten, A. M., Rodrigues, M. F., Ralph, P. L., & Streisfeld, M. A. (2019). Widespread selection and gene flow shape the genomic landscape during a radiation of monkeyflowers. *PLOS Biology*, *17*(7), e3000391. https://doi.org/10.1371/journal.pbio.3000391

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

Stapley, J., Feulner, P. G. D., Johnston, S. E., Santure, A. W., & Smadja, C. M. (2017). Variation in recombination frequency and distribution across eukaryotes: Patterns and processes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1736), 20160455. https://doi.org/10.1098/rstb.2016.0455

Stephan, W. (2019). Selective Sweeps. *Genetics*, *211*(1), 5–13. https://doi.org/10.1534/genetics.118.301319

Stevison, L. S., Woerner, A. E., Kidd, J. M., Kelley, J. L., Veeramah, K. R., McManus, K. F., Great Ape Genome Project, Bustamante, C. D., Hammer, M. F., & Wall, J. D. (2016). The Time Scale of Recombination Rate Evolution in Great Apes. *Molecular Biology and Evolution*, *33*(4), 928–945. https://doi.org/10.1093/molbev/msv331

Storchová, L., & Hořák, D. (2018). Life-history characteristics of European birds. *Global Ecology and Biogeography*, *27*(4), 400–406. https://doi.org/10.1111/geb.12709

Suh, A., Smeds, L., & Ellegren, H. (2018). Abundant recent activity of retrovirus-like retrotransposons within and among flycatcher species implies a rich

source of structural variation in songbird genomes. *Molecular Ecology*, *27*(1), 99–111. https://doi.org/10.1111/mec.14439

Svedin, N., Wiley, C., Veen, T., Gustafsson, L., & Qvarnström, A. (2008). Natural and sexual selection against hybrid flycatchers. *Proceedings of the Royal Society B: Biological Sciences*, *275*(1635), 735–744. https://doi.org/10.1098/rspb.2007.0967

Svensson, L., Collinson, J. M., Knox, A., Parkin, D. T., & Sangster, G. (2005). Species limits in the Red-breasted Flycatcher. *British Birds*, *98*, 538–541.

Tobias, J. A., Sheard, C., Pigot, A. L., Devenish, A. J. M., Yang, J., Sayol, F., Neate-Clegg, M. H. C., Alioravainen, N., Weeks, T. L., Barber, R. A., Walkden, P. A., MacGregor, H. E. A., Jones, S. E. I., Vincent, C., Phillips, A. G., Marples, N. M., Montaño-Centellas, F. A., Leandro-Silva, V., Claramunt, S., … Schleuning, M. (2022). AVONET: Morphological, ecological and geographical data for all birds. *Ecology Letters*, *25*(3), 581–597. https://doi.org/10.1111/ele.13898

Turner, T. L., Hahn, M. W., & Nuzhdin, S. V. (2005). Genomic Islands of Speciation in Anopheles gambiae. *PLOS Biology*, *3*(9), e285. https://doi.org/10.1371/journal.pbio.0030285

Vicoso, B., & Charlesworth, B. (2009). Effective Population Size and the Faster-X Effect: An Extended Model. *Evolution*, *63*(9), 2413–2426. https://doi.org/10.1111/j.1558-5646.2009.00719.x

Wang, Z., Zhang, J., Yang, W., An, N., Zhang, P., Zhang, G., & Zhou, Q. (2014). Temporal genomic evolution of bird sex chromosomes. *BMC Evolutionary Biology*, *14*(1), 250. https://doi.org/10.1186/s12862-014-0250-8

Weber, C. C., Boussau, B., Romiguier, J., Jarvis, E. D., & Ellegren, H. (2014). Evidence for GC-biased gene conversion as a driver of between-lineage differences in avian base composition. *Genome Biology*, *15*(12), 549. https://doi.org/10.1186/s13059-014-0549-1

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, *38*(6), 1358–1370. JSTOR. https://doi.org/10.2307/2408641

Weissensteiner, M. H., Bunikis, I., Catalán, A., Francoijs, K.-J., Knief, U., Heim, W., Peona, V., Pophaly, S. D., Sedlazeck, F. J., Suh, A., Warmuth, V. M., & Wolf, J. B. W. (2020). Discovery and population genomics of structural variation in a songbird genus. *Nature Communications*, *11*(1), Article 1. https://doi.org/10.1038/s41467-020-17195-4

Westram, A. M., Stankowski, S., Surendranadh, P., & Barton, N. (2022). What is reproductive isolation? *Journal of Evolutionary Biology*, *35*(9), 1143–1164. https://doi.org/10.1111/jeb.14005

Wolf, J. B. W., & Ellegren, H. (2017). Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics*, *18*(2), 87–100. https://doi.org/10.1038/nrg.2016.133

Wooldridge, L. K., & Dumont, B. L. (2023). Rapid Evolution of the Fine-scale Recombination Landscape in Wild House Mouse (Mus musculus) Populations. *Molecular Biology and Evolution*, *40*(1), msac267. https://doi.org/10.1093/molbev/msac267

Xue, C., Chen, H., & Yu, F. (2016). Base-Biased Evolution of Disease-Associated Mutations in the Human Genome. *Human Mutation*, *37*(11), 1209–1214. https://doi.org/10.1002/humu.23065

Zelkowski, M., Olson, M. A., Wang, M., & Pawlowski, W. (2019). Diversity and Determinants of Meiotic Recombination Landscapes. *Trends in Genetics*, *35*(5), 359–370. https://doi.org/10.1016/j.tig.2019.02.002

Zickler, D., & Kleckner, N. (2015). Recombination, Pairing, and Synapsis of Homologs during Meiosis. *Cold Spring Harbor Perspectives in Biology*, *7*(6), a016626. https://doi.org/10.1101/cshperspect.a016626

Zink, R. M., Pavlova, A., Drovetski, S., & Rohwer, S. (2008). Mitochondrial phylogeographies of five widespread Eurasian bird species. *Journal of Ornithology*, *149*(3), 399–413. https://doi.org/10.1007/s10336-008-0276-z

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 2239

Editor: The Dean of the Faculty of Science and Technology

A doctoral dissertation from the Faculty of Science and Technology, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology. (Prior to January, 2005, the series was published under the title "Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology".)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-495937

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2023