



UPPSALA
UNIVERSITET

*Digital Comprehensive Summaries of Uppsala Dissertations
from the Faculty of Science and Technology 1326*

Speciation genomics

A perspective from vertebrate systems

NAGARJUN VIJAY



ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2016

ISSN 1651-6214
ISBN 978-91-554-9425-4
urn:nbn:se:uu:diva-265342

Dissertation presented at Uppsala University to be publicly examined in Lindahlsalen, Evolutionary Biology Centre, EBC, Norbyvägen 14-18, Uppsala, Friday, 22 January 2016 at 09:15 for the degree of Doctor of Philosophy. The examination will be conducted in English. Faculty examiner: Associate Professor C. Alex Buerkle (University of Wyoming).

Abstract

Vijay, N. 2016. Speciation genomics. A perspective from vertebrate systems. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology* 1326. 52 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-554-9425-4.

Species are vital entities in biology. Species are generally considered to be discrete entities, consisting of a group of (usually interbreeding) individuals that are similar in phenotype and genetic composition, yet differ in significant ways from other species. The study of speciation has focussed on understanding general evolutionary mechanisms involved in the accumulation of differences both at the genetic and phenotypic level. In this thesis, I investigate incipient speciation, an early stage of divergence towards evolutionary independence in closely related natural populations. I make ample use of recent advances in sequencing technology that allow 1) characterizing phenotypic divergence at the level of the transcriptome and 2) delineate patterns of genetic variation at genome-scale from which processes are inferred by using principles of population genetic theory.

In the first paper, we assembled a draft genome of the hooded crow and investigated population differentiation across a famous European hybrid zone. Comparing sequence differentiation peaks between and within the colour morphs, we could identify regions of the genome that show differentiation only between colour morphs and that could be related to gene expression profiles of the melanogenesis pathway coding for colour differences.

The second paper expands on the first paper in that it includes crow population samples from across the entire Palaearctic distribution spanning two additional zones of contact between colour morphs. The results suggest that regions associated with selection against gene flow between colour morphs were largely idiosyncratic to each contact zone and emerged against a background of conserved 'islands of differentiation' due to shared linked selection.

The third paper focusses on five killer whale ecotypes with distinct feeding and habitat specific adaptations. Differing levels of sequence differentiation between these ecotypes places them along a speciation continuum and provides a unique temporal cross-section of the speciation process. Using genome scans we identified regions of the genome that show ecotype specific differentiation patterns which might contain candidate genes involved in adaptation.

In the fourth and final paper, I assumed a comparative genomic perspective to the problem of heterogeneous genomic differentiation during population divergence. The relatively high correlations in the diversity landscapes as well as differentiation patterns between crow, flycatcher and Darwin's Finch populations is best explained by conservation in broad-scale recombination rate and/or association with telomeres and centromeres conducive to shared, linked selection.

Keywords: evolution, speciation, genomics, vertebrate, adaptation, selection, linked selection, crow, killer whale, hybrid zone, transcriptomics, population genetics, behaviour, colouration

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

© Nagarjun Vijay 2016

ISSN 1651-6214

ISBN 978-91-554-9425-4

urn:nbn:se:uu:diva-265342 (<http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-265342>)

List of Papers

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

- I. Poelstra JW*, **Vijay N***, Bossu C*, Lantz H, Ryll B, Müller I, Baglione V, Unneberg P, Wikelski M, Grabherr MG, Wolf JBW (2014) The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science*, 344, 1410–1414.
- II. **Vijay N***, Bossu C*, Poelstra JW, Weissensteiner M, Suh A, Kryukov A, Wolf JBW (manuscript) Idiosyncratic patterns and processes of genomic differentiation in replicate contact zones in crows.
- III. Foote AD*, **Vijay N***, Ávila-Arcos MC, Baird RW, Durban JW, Fumagalli M, Gibbs RA, Hanson MB, Korneliussen TS, Martin MD, Robertson KM, Vieira FG, Vinař T, Wade P, Worley KC, Morin PA, Gilbert MTP, Wolf JBW (submitted) Genome culture coevolution drives rapid divergence in killer whales.
- IV. **Vijay N** & Wolf JBW (manuscript) Genomic signatures of species diversification – a comparative perspective.

Reprints were made with permission from the respective publishers.

I also contributed to the following articles that were submitted or published during my graduate studies.

1. **Vijay N***, Poelstra JW*, Künstner A, Wolf JBW (2013) Challenges and strategies in transcriptome assembly and differential gene expression quantification. A comprehensive in silico assessment of RNA-seq experiments. *Molecular Ecology*, 22, 620–634.
2. Zamani N, Russell P, Lantz H, Hoepfner MP, Meadows JRS, **Vijay N**, Mauceli E, di Palma F, Lindblad-Toh K, Jern P, Grabherr MG (2013) Unsupervised genome-wide recognition of local relationship patterns. *BMC Genomics*, 14, 347.
3. Foote AD, Liu Y, Thomas GWC, Vinař T, Alföldi J, Deng J, Dugan S, Elk CEV, Hunter ME, Joshi V, Khan Z, Kovar C, Lee SL, Lindblad-Toh K, Mancina A, Nielsen R, Qin X, Qu J, Raney BJ, **Vijay N**, Wolf JBW, Hahn MW, Muzny DM, Worley KC, Gilbert MTP, Gibbs RA (2015) Convergent evolution of the genomes of marine mammals. *Nature Genetics*, 47-3, 272-275.
4. Shafer ABA*, Wolf JBW*, Alves PC, Bergström L, Bruford MW, Brännström I, Colling G, Dalen L, Meester LD, Eklom R, Fawcett KD, Fior S, Hajibabaei M, Hill JA, Hoebel AR, Höglund J, Jensen EL, Krause J, Kristensen TN, Krützen M, McKay JK, Norman AJ, Ogdén R, Österling EM, Ouborg NJ, Piccolo J, Popović D, Primmer CR, Reed FA, Roumet M, Salmons J, Schenker T, Schwartz MK, Segelbacher G, Senn H, Thaulow J, Valtonen M, Veale A, Vergeer P, **Vijay N**, Vilà C, Weissensteiner M, Wennerström L, Wheat CW, Zieliński P (2015) Genomics and the challenging translation into conservation practice. *Trends in Ecology & Evolution*, 30, 78–87.
5. Poelstra JW, **Vijay N**, Hoepfner MP and Wolf JBW. (2015) Transcriptomics of colour patterning and colouration shifts in crows. *Molecular Ecology*, 24, 4617–4628.
6. Poelstra JW, Müller I, **Vijay N**, Baglione V, Wolf JBW (submitted) Covariance between colouration, corticosterone response and gene expression patterns suggests pleiotropy in the melanocortin system in carrion and hooded crows.

* These authors contributed equally to the study.

Contents

1	Introduction	7
1.1	Adaptation and Speciation	8
1.2	Effect of spatial distribution.....	10
1.3	Genetics of Speciation	11
1.3.1	Introduction.....	11
1.3.2	Speciation genes	12
1.3.3	Speciation in the Genomics era	14
1.4	Methods.....	17
1.4.1	Genome assembly	17
1.4.2	Transcriptomics	18
1.4.3	Populations genomics	19
1.5	Study system	20
1.5.1	Avian system – Corvids.....	21
1.5.2	Marine system – Orca	22
1.5.3	Comparative avian system	23
2	Research Aims	24
2.1	General Aims	24
2.2	Specific Aims.....	24
3	Summaries of Papers	26
3.1	Concluding Remarks and Future Prospects	33
4	Svensk sammanfattning	35
4.1	Bakgrund.....	35
4.2	Matrial och metoder	35
4.2.1	Studiesystem	35
4.2.2	Metoder.....	36
4.3	Resultat	36
4.4	Diskussion.....	36
5	Acknowledgements	38
6	References	40

Abbreviations

AFLP	Amplified fragment length polymorphism
bp	base pair(s)
DNA	Deoxyribonucleic acid
GWAS	Genome Wide Association Study
kb	kilobasepairs, i.e. 1000 base pairs
PCA	Principal Component Analysis
RNA	Ribonucleic acid
RNAseq	RNA-sequencing (using next generation sequencing technology)
SNP	Single Nucleotide Polymorphism

1 Introduction

The immense biological diversity on earth has fascinated man since the earliest of days, yet awaited scientific classification until the 18th century. In 1758 the Swedish scientist Carl Linnaeus pioneered a classification of organisms based on similarity in morphological characters into species and higher order taxa (Linnaeus, 1758). Almost a hundred years later, Charles Darwin and Alfred Russel Wallace (C. Darwin & Wallace, 1858) were among the first to appreciate that the hierarchical nature of such a classification was in most cases a result of gradual modification from common descent. In his seminal work on the 'Origin of Species by Means of Natural Selection' Charles Darwin provided many lines of evidence that species indeed were not immutable entities, but subject to change – to evolution (C. R. Darwin, 1859). This set the stage for research on the subject of speciation, a process hitherto simply not recognized. In the century that followed, multiple museum collections of specimens from different parts of the world were carefully catalogued and scrutinized (Mallet, 2007). Comparison of divergent forms from different geographical locations lead to a better understanding of the inter-relatedness and continuity of species (Mallet, 2007). The ensuing debate lead to a multitude of species concepts (De Queiroz, 2007). The biological species concept proposed by Dobzhansky and later popularised by Mayr emerged as a leading idea for species delimitation (Dobzhansky, 1935, 1937; Mayr, 1942) in sexually reproducing organisms. Asexual organisms and those with very high rates of hybridisation did not fit well into the biological species concept (Mallet, 2007). The ecological species concept was suggested as an alternative to explain such anomalies (Leigh Van Valen, 1976).

Irrespective of the species concept, species have been defined based on some form of phenotypic or genetic distinctness that gradually accumulates over time (De Queiroz, 2007). Speciation can be thought of as the process of accumulation of such differences up to the point where most would agree to attribute species status. This continuous and gradual accumulation of “genetic” differences has been termed the speciation continuum (Kerry L Shaw & Mullen, 2014) as it finally leads to distinct entities that are reproductively isolated. Assuming a population genetic view of the speciation problem introduced by Ernst Mayr in the mid 20th century (Mayr, 1942), reproductive isolation becomes the main concern to understand how species barriers evolved. Reproductive isolation refers to barriers to gene flow that can be categorized by life history stage (pre-, postzygotic) or evolutionary process

(extrinsic, intrinsic) (Wolf, Lindell, & Backström, 2010). Prezygotic barriers refer to features that inhibit mating or zygote formation (examples: assortative mating by song (Nuttall's white-crowned sparrows [*Zonotrichia leucophrys nuttalli*] mate assortatively with individuals from the same dialect population (Tomback & Baker, 1984)) or plumage colour). Postzygotic barriers refer to barriers that prevent the zygote from forming a functional and viable individual (examples: Hybrid Infertility (*D. melanogaster*/*D. simulans* F1 hybrids suffer from Inviability (Presgraves, 2010)), Low Hybrid Viability). Extrinsic barriers are those that are caused by an interaction with the environment or with other individuals (examples: reduced hybrid viability due to inability to survive in a particular environment). Intrinsic barriers on the other hand refers to those barriers which are independent of the environment, mainly genetic incompatibilities (example: Bateson-Dobzhansky-Muller incompatibilities, a classic example is the formation of melanomas in the hybrids of *Xiphophorus* species that lack the repressor gene that negatively regulate the Tu locus which controls black cell pigment spot formation (Meierjohann, Schartl, & Volff, 2004; Wittbrodt et al., 1989)).

Study systems from across the tree of life are being investigated at different stages of speciation to understand the relative importance and temporal sequence of mechanisms by which such barriers are formed (Seehausen et al., 2014). With advances in the field of genetics, it has become possible to understand the genetic basis of traits that are thought to contribute to reproductive isolation. This has given rise to further questions such as (1) Are certain gene classes involved in speciation more often than others? (2) Is speciation caused by multiple genes of small effect or single genes of large effect? (3) Do the genetics involved in speciation differ between sympatric and allopatric geographic modes, or more generally by the degree of gene flow during speciation?

1.1 Adaptation and Speciation

The process of speciation is accompanied by divergence in phenotypic characters such as morphology and behaviour (Gompert et al., 2013). Differences in DNA sequence also accumulate at the level of individual base pairs as well as of structural genetic changes all the way to the karyotype (Andolfatto, Depaulis, & Navarro, 2001). The role of these phenotypic and genetic differences in speciation and their interaction with each other is still being understood (Mohamed A F Noor & Feder, 2006). However, since most heritable phenotypic differences are thought to be caused by differences in the DNA sequence, genetic and phenotypic changes ought to be closely related (Eichler et al., 2010). The study of taxa that are in the earliest stages of speciation provides an ideal scenario for linking changes in phenotypes demonstrably involved in reproductive isolation between populations to changes in

DNA sequence (Via, 2009). Still, it has been difficult to establish in how far speciation and adaptation are independent or interact in subtle ways (Barton & Hewitt, 1989). Recent studies that estimated rates of speciation have suggested that speciation occurs at a constant rate and that adaptive changes are an independent process (Hedges, Marin, Suleski, Paymer, & Kumar, 2015). Yet, under the ecological speciation-with-gene flow scenario that has recently gained much attention adaptation and speciation are intimately linked (Schluter & Conte, 2009).

Ecological speciation with gene flow models posit that regions under divergent selection will be resistant to gene flow between the ecotypes under divergent selection (Feder, Egan, & Nosil, 2012). The genes under selection should diverge faster than the rest of the genome which is free to allow gene flow. Over time these genes under divergent selection will either recruit adjoining genes to promote the speciation process or may even pull the entire genome with them (Nosil, Funk, & Ortiz-Barrientos, 2009). The formation of localised clusters of tightly linked, diverged loci caused due to reduced gene flow at physically linked sites is called Divergence Hitchhiking (DH) (Via, 2012). The genome-wide effects of divergent selection is called Genome Hitchhiking (GH) (Feder, Gejji, Yeaman, & Nosil, 2012). Theoretical models and simulations have been used to model this process of divergence of genomes during speciation (Flaxman, Feder, & Nosil, 2013).

Another potential driver of speciation is sexual selection, which is defined as a component of natural selection arising owing to variation in mating or fertilization success (Andersson, 1994; Ritchie, 2007). Comparative phylogenetic methods have been used to show that sexual selection probably escalates the rate of speciation by promoting pre-mating reproductive isolation (Seddon et al., 2013). However, the evidence for the role of sexual selection in driving speciation is not conclusive (Huang & Rabosky, 2014; Ritchie, 2007). In the well studied radiation of cichlid fishes in the African great lakes, sexual selection along with ecology have been implicated (Wagner, Harmon, & Seehausen, 2012). Similarly, the relative importance of sexual selection and direct ecological selection in the rapid evolution of *Drosophila* species on Hawaiian islands is controversial (N. H. Barton, 1984).

In some study systems, single genes that contribute to adaptation have also been shown to promote reproductive isolation. For example, the deletion of a tissue specific enhancer of the PITX1 gene has been shown to result in pelvic loss in threespine stickleback fish (Chan et al., 2010). This change leads to two different alleles of the PITX1 gene that promote reproductive isolation between the two forms. Although many such single genes of large effect have been identified, the idea of polygenic adaptation through subtle changes in allele frequencies at multiple loci has been gaining importance (Berg & Coop, 2014; Yeaman, 2015). Future studies in many different taxa that identify actual causative variants involved in speciation and adaptation

will need to be examined before we get a picture of the effect size distribution of genetic variants (Rockman, 2012).

1.2 Effect of spatial distribution

The spatio-temporal distribution of individuals, populations and species is not random. Large scale patterns are thought to be determined by factors such as the biome to which the species has adapted (Wiens & Graham, 2005) and finer scale patterns are the result of factors such as ecology and distance. Genetic structure of populations is influenced by their relative geographic locations with neighbouring demes exchanging more migrants than more distant demes. The ensuing pattern of decreasing genetic similarity with geographic distance is known as isolation by distance (Wright, 1943). Populations that differ in their ecology/adaptation also show differences in genetic structure. This pattern determined by environment is known as isolation by ecology (I. J. Wang & Bradburd, 2014; Wang & Summers, 2010). Since environmental differences very often covary with distance, it can be hard to disentangle the effects of distance and environment (Aaron B A Shafer & Wolf, 2013). Similarly, inferring processes from species distribution is confounded by geography (Warren, Cardillo, Rosauer, & Bolnick, 2014).

Presence of barriers such as mountains and rivers strongly influence genetic structure. However, we are yet to get a complete understanding of the impact of such barriers (Hewitt, 2001). Interesting strategies that span simulations, experimental and theoretical models have been used to get a better understanding of the impact of obstacles on genetic structure (Moebius, Murray, & Nelson, 2015). The geographic modes of speciation and their relative importance has been the focus of much debate (Mayr, 1942). Mayr (Mayr, 1942) used examples of isolated populations such as cave animals and island endemics to argue that elimination of genes due to drift had important consequences for speciation. Apart from random loss of genes, the “founder effect” in which the gene variants of a small founder population contribute to the formation of a large isolated population can lead to reproductive barriers. Hence, Mayr argued that speciation required geographical isolation (Provine, 2004). The importance of founder events in speciation is a hotly debated topic with Mayr's views about the impact of founder events on genetic variation (Mayr, 1954) being challenged by Lewontin (Lewontin, 1965). For a more fruitful debate i suggest to consider geographic modes as extreme boundaries of a continuous distribution of gene flow probabilities (high gene flow in sympatry, low gene flow in allopatry).

Distinctive distribution of populations can be used to study the process of speciation and adaptation (Barton & Hewitt, 1989). For example, hybridization between taxa that have recently come in contact at suture zones has been studied as natural laboratories (Hewitt, 2001). Islands of land (Grant &

Grant, 2003), water (Kocher, 2004) and sky (Robin, Sinha, & Ramakrishnan, 2010) are all being explored as productive research avenues. Ring species are another interesting system that provide a unique opportunity to infer variation in time by looking at variation in space. Ring species are defined as a chain of intermediate forms around a barrier that intergrade towards two distinct reproductively isolated forms at the end of the range where they meet (Irwin, Bensch, & Price, 2001). Although the greenish warbler (*Phylloscopus trochiloides*) was considered to be a good example of a ring species, recent genetic work has shown the presence of distinct genetic clusters that suggests geographic isolation and presence of hybrids between the most divergent forms (Alcaide, Scordato, Price, & Irwin, 2014). The spatial organisation of species can be used to test for the role of gene flow and selection against gene flow in speciation (S. H. Martin et al., 2013). Genomic studies across super-species complexes (Alcaide et al., 2014; M. Kronforst et al., 2013; Renaut et al., 2013) have started to provide a better understanding of how different parts of the genome are affected by the spatial distribution of species.

1.3 Genetics of Speciation

1.3.1 Introduction

Over time genomes evolve, gene families expand and diversify or just become defunct (De Bie, Cristianini, Demuth, & Hahn, 2006). Comparative genomics between species that are separated by a few million years has shown that a majority of the genes are largely conserved in the coding regions (Lindblad-Toh et al., 2011). Differences in the rates of synonymous and non-synonymous changes have been used to understand the evolutionary trajectory of a gene (Miyata, Yasunaga, & Nishida, 1980; Nei & Gojobori, 1986). Attempts to infer selection based on changes at the level of base-pairs within coding regions have seen wide-spread adoption (Yang, 2007) with varied success (Schneider et al., 2009). Despite coding sequence conservation, changes in expression pattern (Brawand et al., 2011) as well as placement in protein interaction networks (Jinho Kim, Kim, Han, Bowie, & Kim, 2012) can alter the function of genes. Studies that measure the phenotype after knocking out genes in closely related species have found that orthologous genes could have developed very different functions (Verster, Ramani, McKay, & Fraser, 2014), a process defined as developmental systems drift. Hence, establishing a link between genotype and phenotype is a difficult process.

Even efforts to map traits and disease-causing variants using methods such as GWAS (Genome-wide association studies) have many limitations (Korte & Farlow, 2013; Maher, 2008). Overall, a multitude of comparative

genomic studies suggest that trait evolution is a complex process and probably involves changes in multiple genes (Berg & Coop, 2014; Rockman, 2012). Yet, it is not random and certain parts of the genome are more likely to harbour functional regions and will therefore be more conserved than others (Graur et al., 2013). Allelic variation in genes with strong pleiotropic effects for example is less likely to contribute to adaptation, while genes with single roles (or freed from constraint by duplication (Kondrashov, 2012)) are expected to be more often involved in adaptation and speciation. This underscores the question “how far evolution may be predictable?” (Stern & Orgogozo, 2009), i.e. whether independent lineages that respond to similar evolutionary conditions undergo parallel genetic changes (Foote et al., 2015; Parker et al., 2013).

1.3.2 Speciation genes

Reproductive isolation between species like any other trait is thought to have a genetic basis. This has led to a search for speciation genes. While numerous definitions of speciation genes (Nosil & Schluter, 2011) have been suggested, most agree that these are functional genomic elements that convey some degree of ecological, sexual, pre- or postmating, pre- or postzygotic isolation (Presgraves, 2010; Wolf, Lindell, et al., 2010). Despite the identification of some good examples of speciation genes (Presgraves, 2010; Wolf, Lindell, et al., 2010), general rules regarding these genes are not well-established. Are certain classes of genes involved in speciation more often than others? Do these genes act independently or along with other genes? Are these genes causing pre-zygotic or post-zygotic reproductive isolation? Are genetic changes in these genes a result of pre-existing allelic variation or de-novo mutations? For the exception of few well characterized genes conveying postzygotic intrinsic isolation in *Drosophila* (e.g. the *Ovd* gene that leads to sterility of F1 hybrid males, *Zhr* gene that leads to in-viability of F1 hybrid females (Presgraves, 2010)), the nature of speciation genes remain largely elusive. A large catalogue of such genes in a variety of taxa across different timescales in the speciation process needs to be put together before we can begin to answer these questions.

The Dobzhansky–Muller model of genetic incompatibilities has built upon the ideas of Bateson (Bateson W, 1909), Dobzhansky (Dobzhansky, 1935) and Muller (Muller, 1942) to become the guiding principle for genetic analysis in speciation (Presgraves, 2010). Classical speciation genetics has focussed on mapping such loci involved in hybrid incompatibilities mainly in *Drosophila* species (Presgraves, 2003) but was later extended to yeast, mice and *Arabidopsis*. It has been possible to obtain three clear general rules based on this extensive work in model species. (1) Reduced hybrid fitness evolves gradually through incompatible epistatic interactions which are advantageous or do not alter the fitness of individuals in one species, but have

harmful effects in hybrid individuals that have a genetic background from another species (Coyne & Orr, 1989, 1997). (2) A large fraction of hybrid incompatibility loci are partially recessive and are located on the sex chromosome (large X effect) (M Turelli & Orr, 1995; Michael Turelli & Orr, 2000). This observation also accounts for the Haldane's rule (hybrids of the heterogametic sex (e.g. XX) suffer more severe problems than hybrid individuals of the homogametic sex (e.g. XY) (Haldane, 1922)) (3) The “faster male” theory postulates that the rapid emergence of hybrid male sterility is caused by the faster divergence of male-specific fertility genes (Swanson, Clark, Waldrip-Dail, Wolfner, & Aquadro, 2001). Moreover, movement of genes to and from the X chromosome have been shown to contribute towards the final two rules of speciation (Moyle, Muir, Han, & Hahn, 2010). Taxon specific rates of accumulation of such incompatibilities can contribute to diversification rate differences. However, such differences in rates of evolution of intrinsic reproductive isolation have not been able to explain differences in the rates of diversification (Rabosky & Matute, 2013).

Mapping reproductive isolation by classical genetics between diverged species under laboratory conditions has produced insights into the genetic mechanisms underlying reproductive isolation. Yet, finding the genetic basis of prezygotic and extrinsic postzygotic isolation which intrinsically link the complex interplay between adaptation and speciation has long been a largely unexplored area. This is understandable, as it requires an understanding of selection pressures in populations under natural conditions, and at the same time access to genetic resources. Theory predicts that genes identified as speciation genes in many cases may have alternative alleles that are adapted to contrasting environments (Chan et al., 2010; Colosimo et al., 2005). It has been argued that such adaptations have contributed to extrinsic reproductive isolation suggesting an interaction between adaptation and speciation (Funk, Nosil, & Etges, 2006). Functional characterisation of the alternative alleles and quantification of the selective advantage offered by each of the alleles in the two different environments has been used to establish the adaptive nature of the genetic change. A special case of ecologically divergent selection leading to speciation is the idea of magic traits (Servedio, Van Doorn, Kopp, Frame, & Nosil, 2011). These are traits under divergent selection that directly cause reproductive isolation without the need for linking a gene coding for an ecologically selected trait to a gene promoting assortative mating.

More generally the role of pleiotropy, i.e. the fact that a single gene affects multiple phenotypes, in speciation is debated (Gavrilets, 2004). Shaw et al., (K. L. Shaw, Ellison, Oh, & Wiley, 2011) describe four examples of scenarios in which a single gene has pleiotropic effects on both a signal and its recognition. Sexual discrimination and isolation between species of *Drosophila* is due to contact pheromones expressed by both sexes. The *desat 1* gene expression determines the levels of the cuticular hydrocarbons (CHCs) which act as pheromones. The same gene is also expressed in the chemosen-

sory hair that is involved in sex pheromone perception. The role of this gene in sexual perception has been validated by the inability of males (that had this gene knocked out) to discriminate males from females.

A model by van Doorn et al. (van Doorn, Edelaar, & Weissing, 2009) suggests that natural selection conveying local adaptation even through a multigenic architecture can act in concert with sexual selection and promote reproductive isolation. This model provides an alternative to accidental occurrence of magic traits that link ecological ability with assortative mating and predicts the emergence of such a link as a consequence of local adaptation.

Finding speciation genes and identifying the processes that have been responsible for promoting speciation in each case has been an arduous task. Moreover, in many cases reproductive isolation might have been caused by certain genes, following which other genes might have acquired changes that can cause reproductive isolation now, but were actually not involved in the process of causing speciation. Disentangling this complex interaction between genes and understanding the role of each individual gene will be extremely challenging. The genomic era promises to make it easier to study speciation genetics, especially in non-model species.

1.3.3 Speciation in the Genomics era

The genomics era is akin to the early 20th century in that large specimen collections are being generated, albeit of a different nature (Mallet, 2007; Seehausen et al., 2014). Careful analysis of this catalogue has the potential to bring about radical changes in the way we think about species and speciation. Early genetic studies using AFLP markers and a handful of markers scattered across the genome have in most cases strengthened the knowledge that has been gained by morphological comparisons (Hillis, 1987). Studies that have sampled multiple populations have also been used to model demographic changes and gene flow between populations. Candidate gene based approaches that leverage the information from studies in model species have been used to identify phenotype causing variations. However, such an approach has limited power and could potentially miss previously uncharacterised genes (Zhu & Zhao, 2007).

Up until now, it has been difficult to tackle the genetic basis of the phenotypic differences. The sequencing revolution has made it possible to investigate the genetic basis of traits in a variety of species. The genic view of speciation postulated by Wu (Wu & Ting, 2004) has been a motivation to search for speciation genes outside traditional genetic approaches in the laboratory. Despite the numerous definitions of speciation genes (Nosil & Schluter, 2011), efforts have been directed towards finding nucleotide changes that lead to some form of reproductive isolation between species.

As speciation research enters the genomics era, large number of candidate speciation genes are being identified (Nosil & Schluter, 2011; Seehausen et al., 2014). A popular strategy has been to scan genomes for genomic signatures of selection to find genes that could be involved in speciation. Genomic scans across the speciation continuum have found surprisingly variable patterns of genomic divergence that suggest that incipient species can rapidly accumulate substantial divergence (Andrew & Rieseberg, 2013). Locus-specific differences in the magnitude of divergent selection are thought to contribute to genomic heterogeneity in divergence. Genome scans have focussed on identifying regions of the genome that show strong signatures of divergent selection. Based on early studies using genome scans of differentiation measures such as F_{st} between closely related species, speciation genes were thought to be found in certain regions of the genome that had high levels of sequence differentiation (L. M. Turner & Harr, 2014; T. L. Turner, Hahn, & Nuzhdin, 2005; White et al., 2011). These narrow regions harbouring speciation genes have been termed “speciation islands” implying a causal link between the observed pattern of high differentiation and reproductive isolation. However, it is being realised that most traits are probably polygenic (Rockman, 2012) and would not leave such strong signatures of selection in well circumscribed genomic regions. Moreover, processes unrelated to reproductive isolation between population pairs connected via gene flow, such as low levels of genetic diversity, for example caused by background selection are also known to locally increase genetic differentiation (B Charlesworth, Morgan, & Charlesworth, 1993; Cruickshank & Hahn, 2014; M A F Noor & Bennett, 2009).

Whole genome sequencing has allowed the investigation of genome-wide patterns of diversity and differentiation in many species pairs (Ellegren, Smeds, Burri, Olason, Backström, et al., 2012; J W Poelstra et al., 2014; Renaut et al., 2013; T. L. Turner et al., 2005). With increasing recognition of the difficulties in interpreting patterns emerging from genome scans the initial euphoria of trying to find speciation genes in islands of elevated differentiation (M A F Noor & Bennett, 2009) is being replaced by cautious optimism (Cruickshank & Hahn, 2014; Seehausen et al., 2014) towards understanding the roles of selection and drift in shaping the genomic landscape of differentiation (Cutter & Payseur, 2013). We are yet to understand how genetic changes that cause speciation relevant phenotypic differences are distributed across the genome and how they relate to genome-wide patterns of genomic differentiation.

While it seems increasingly clear that broad-scale heterogeneity in the differentiation landscape results in large part from variation in the strength of linked selection shared across populations and follows structural genomic properties (e.g. low recombination, centromere, subtelomeric regions) (Burri et al., 2015; Ellegren, Smeds, Burri, Olason, Backström, et al., 2012), differential gene flow mediated by divergent selection between populations

seems to also contribute (Nosil et al., 2009). With the availability of whole genome sequencing data, it is now possible to identify the actual genomic regions being homogenized by gene-flow or those that resist introgression. Recent studies looking at human populations and Neanderthal genomes have found many regions (Sankararaman et al., 2014) that are thought to have introgressed from Neanderthals into human populations. Many of these regions are thought to be functionally relevant in causing specific phenotypes. Similar studies in various other species has lead to speculation that introgression is probably more common than previously thought (Hedrick, 2013). The role of introgression in transferring adaptive phenotypes makes it even more important. Methods to detect introgression have seen increasing sophistication with development of tests such as the ABBA-BABA test (Green et al., 2010), 5-population test (Pease & Hahn, 2015) and S^* (Vernot & Akey, 2014).

Comparative studies spanning multiple species are integrating data from various studies and blurring distinctions between phylogenetics, phylogeography and population genetics (Cutter, 2013). This opens the exciting prospect of addressing questions about evolutionary convergence (Stern, 2013), the importance of repeatability of evolution (Gaut, 2015; Weinreich, 2006) and eventually their relationship to speciation (Cutter, 2015; Mandeville, Parchman, McDonald, & Buerkle, 2015). A predictable model of evolution provides a useful tool in the search for genetic differences causing phenotypic changes. Covariation of amino-acid residues in distantly related species that share phenotypically similar traits has been interpreted as convergent evolution and used to link genes to phenotypic traits (Stewart, Schilling, & Wilson, 1987; Wierer, Schrey, Kühne, Ulbrich, & Meyer, 2012). Availability of genome-wide data provides an opportunity to understand the prevalence of convergent amino-acid changes (Foote et al., 2015). However, theoretical models of molecular evolution suggest that covariation need not be caused by adaptive convergence (Talavera, Lovell, & Whelan, 2015) and might instead be the result of slower evolutionary rates of certain sites or differences in amino acid acceptability at a given site due to epistasis (Zou & Zhang, 2015). Development of relevant evolutionary models will be needed to improve phylogenetic inference methods, establish the link to selection at population genetic time-scales, as well as distinguish adaptive and non-adaptive covariation of states of sites.

Long-term comparative studies have largely been limited to nucleotide level changes (Alföldi & Lindblad-Toh, 2013). However, large-scale changes such as genome duplication, translocations, inversions and in-dels are expected to be highly relevant to evolution (adaptation and speciation) at shorter evolutionary time-scales (Faria & Navarro, 2010; Feder, Nosil, & Flaxman, 2014; Kirkpatrick & Barton, 2006). Chromosomal inversions are barriers to gene flow and their importance to speciation and adaptation is well established (Kirkpatrick & Barton, 2006). Improvements in genome

assembly methods has improved our ability to detect structural variants such as inversions, insertions and deletions (Chaisson, Wilson, & Eichler, 2015).

Overall, the genomics era has opened up access to natural populations and as sequencing technology improves, it will enable exploration of a previously inaccessible new axis of haplotypic or structural variation. Increased importance of introgression and its adaptive potential are being understood with the availability of large-scale whole genomic datasets (Hedrick, 2013). Reducing costs of sequencing has also made it possible to study the evolution of recombination rate over different time-scales (Jensen-Seaman et al., 2004; Smukowski & Noor, 2011), chromatin accessibility (Connelly, Wakefield, & Akey, 2014) and methylation (Seymour, Koenig, Hagmann, Becker, & Weigel, 2014). Integrating such diverse datasets and understanding their role in speciation and adaptation will be a worthy challenge for the future.

1.4 Methods

1.4.1 Genome assembly

Sequencing of DNA molecules has made it possible to determine the exact order of consecutive nucleotide bases (Shendure & Ji, 2008). Improvements in the technology over the years has made the process of sequencing easier and cheaper. However, the length of continuous stretches of DNA that can be sequenced has always been a limiting factor. Such continuous stretches of DNA that can be sequenced are called sequencing reads (Eklblom & Wolf, 2014). Genome assembly is the process of ordering and orienting such sequencing reads based on shared sequence to obtain contiguous stretches of the genome. By sequencing a large number of such reads, a large fraction of the genome is covered by overlapping reads that can be assembled into longer regions.

For species that lack a genome assembly, it is possible to assemble the genome de novo based on just the information available in the raw sequencing reads. However, as genomes of different organisms have accumulated it has become possible to assist the assembly of genomes of related species by using the information already present in assembled genomes (Jaebum Kim et al., 2013). Reference assisted assembly of genomes utilize high quality assemblies of the same or closely related species to guide the de novo assembly of sequencing reads (Kolmogorov, Raney, Paten, & Pham, 2014). Hybrid strategies that rely on both de novo and mapping assembly have also been implemented (Utturkar et al., 2014). Availability of a multitude of sequencing technologies such as Sanger, 454, Illumina, Solid, Ion Torrent and Pacific-biosciences has also generated different breeds of genome assemblers fine-tuned for each of these platforms as well as those genome assemblers that are able to use data from multiple platforms (W. Zhang et al., 2011).

Reducing costs of high throughput sequencing technologies has made it possible to sequence whole genomes of various non-model species (Ekblom & Wolf, 2014; Ellegren, 2014). Substantial progress has been made in the sequencing technologies as well as genome assembly methods. However, most genome assemblies are far from perfect (Denton et al., 2014). Comparisons of different genome assembly programs and strategies have been performed to evaluate the improvements in methodology (Bradnam et al., 2013; Salzberg et al., 2012). Quality of genome assemblies have been evaluated based on different statistics such as those that measure the contiguity of the assembly, short & long range accuracy, number of miss-assemblies, ability to assemble genes, repeats or other regions that are difficult to assemble. Despite such advances draft genome assemblies are still plagued by various drawbacks (Nagarajan & Pop, 2013). For example, missing data caused due to gaps in genomes as well as hard to assemble complex genetic variation have been shown to impede discovery of disease causing mutations. Certain newly assembled regions of the genome have also been crucial to enhance our understanding of natural variation and evolution (Chaisson et al., 2015). Hence, it is advisable to be cautious while performing downstream analysis using draft genome assemblies as the backbone.

Improvement of draft assemblies using optical maps (Nagarajan, Read, & Pop, 2008), admixture mapping (Genovese, Handsaker, Li, Kenny, & McCarroll, 2013; Genovese, Handsaker, Li, Altemose, et al., 2013) and linkage maps have shown promise (Fierst, 2015). It is hoped that newer technologies such as Pacific-biosciences and Nanopore sequencing with their longer read lengths will improve the quality of genome assemblies considerably. Alternative approaches to genome assembly have not grown much beyond novelties such as the use of reduced representation libraries for genome assembly (Young et al., 2010), sequencing of haploid embryos to overcome heterozygosity (Langley, Crepeau, Cardeno, Corbett-Detig, & Stevens, 2011) and creation of divergent mapping populations (Hahn, Zhang, & Moyle, 2014). Future challenges for genome assembly methods would be to generate high quality assembly of hard to assemble regions such as centromeres and telomeres, identify structural variants longer than a few base pairs, resolving the phase of the genotype calls, be computationally tractable and fast and handle heterogeneity in data quality (Chaisson et al., 2015; Schatz, 2015).

1.4.2 Transcriptomics

The transcriptome, the entirety of transcribed genomic regions, has various levels of complexity such as alternative splice-forms (Matlin, Clark, & Smith, 2005), tissue specificity of expression, developmental stage specific differences (Domazet-Lošo & Tautz, 2010), and influence by environmental condition. Such multifarious complexity makes it harder to comprehend

changes in the transcriptome than changes at the level of DNA. Despite such challenges, differences at the level of expression could be used to characterize the genetic basis of speciation relevant phenotypes (J. W. Poelstra, Vijay, Hoepfner, & Wolf, 2015) and alternative splicing for species specific splicing patterns (Q. Zhang, Hill, Edwards, & Backström, 2014). Linking changes in DNA sequence to changes in transcriptional level of specific genes is challenging, but constitutes the decisive link between variation at the level of the heritable information and its translation into phenotypic variation (Grabherr et al., 2011).

Technical challenges in the study of transcriptomes are far from resolved. Errors in the underlying genome assembly and annotation can have a pronounced impact on the quality of conclusions that can be drawn from such large datasets (Vijay, Poelstra, Künstner, & Wolf, 2013). Biological complexity at the transcriptional level complicates the measurement of gene expression in all its myriad forms. Phenotypic plasticity of traits also manifests itself at the level of gene expression. Environment can strongly influence gene expression. Hence, common garden experiments that try to control for extrinsic variation are extremely important to arrive at conclusions that can be replicated.

1.4.3 Populations genomics

Genomic variation evolves through the action of several different factors such as mutation, selection, genetic drift and migration. Hence, the genome provides a historical record of the events that have shaped its diversity landscape. For example, positive selection for a certain trait leads to a strong reduction in diversity at the locus that has been selected. Only the allele that is under positive selection is retained, while almost all the other alleles are lost. Similarly, migration between populations leads to a greater than expected proportion of allele sharing between populations. Early studies that looked at population level variation looked at polymorphism in allozyme data to study population dynamics. Other markers such as micro-satellite markers, mitochondrial markers and Y-chromosome markers have provided additional information regarding sex-specific patterns of gene flow. Access to whole genome re-sequencing data has made it possible to ask other interesting questions regarding introgression from Neanderthals to humans (Green et al., 2010; Sankararaman et al., 2014), identification of population specific adaptation (Bersaglieri et al., 2004) and role of introgression in speciation (Baack & Rieseberg, 2007).

The genomic revolution now enables us to scan entire genomes of numerous individuals from multiple populations. Genome scans exploring various aspects of genetic variation within and among populations are a central component of speciation genomic approaches. Under the assumption that demographic processes affect all parts of the genome (with the same ploidy) alike,

scanning genetic variation within and between populations along the genome can provide insight into signatures of selection (Lewontin & Krakauer, 1973). Certain regions of the genome are thought to have undergone population specific selection events. Identification of such selected regions is performed using selection scans. This involves calculating various intra and inter population statistics that are capable of identifying signatures of selection. Methods that use haplotype information have also seen wide-spread adoption due to the greater power offered by using information available at the haplotype level (Sabeti et al., 2002). However, the diversity landscape of the genome is highly dependant on the genetic recombination rates. Linked selection along with recombination rate differences is a major determinant of diversity and differentiation patterns across the genome. Theoretical models of linked selection (Brian Charlesworth, 2013) as well as empirical data (Cruickshank & Hahn, 2014) from multiple systems has established the pervasive role of background selection. Hence, newer methods that incorporate the recombination rates and background selection maps are being proposed to identify selective sweeps (Huber, DeGiorgio, Hellmann, & Nielsen, 2015). In the context of speciation genomics, genome scans generally start by population differentiation metrics quantifying the genetic variation between a pre-defined set of populations, F_{ST} being arguably the most prominent (Beaumont, 2005). Closer scrutiny of outlier regions with additional summary statistics may then further the understanding of the underlying processes. Candidate regions can then be analysed in detail to understand the functional importance of the genetic changes between populations.

1.5 Study system

Traditionally, the study of speciation genetics had been largely restricted to model species amenable to traditional genetic approaches such as *Drosophila* (see above). With the recent sequencing revolution it has been possible to extend the study to various other systems such as fungi, plants, butterflies, fish, birds, reptiles and mammals spanning the entire continuum of life (Seehausen et al., 2014). Comparative analysis of such different study systems provides an opportunity to identify similarities that are shared and novelties that make each system unique. While some study systems are easy to cultivate, cross and genetically modify others are not, yet have been studied for decades and have relevant ecological information available. In this thesis, my colleagues and I studied a corvid species complex (in chapter I & II) with a very broad palaeartic geographical distribution along with a marine mammal system (in chapter III) with an almost pan-oceanic distribution. We finally used a comparative approach (in chapter IV) to investigate genomic differentiation across multiple bird species complexes.

1.5.1 Avian system – Corvids

The genus *Corvus* consists of ~40 of the more than 120 bird species described in the family Corvidae. These birds range in size from the relatively small jackdaws (34-39 cm long) to the larger ravens (58-69 cm long) and can be found on all continents except South America and Antarctica (del Hoyo, J., Elliott, A., Sargatal, J., Christie, D.A. & de Juana, 2014; Knud A Jønsson, Fabre, Ricklefs, & Fjeldså, 2011). Corvids are considered to be extremely intelligent animals, with folklore replete with tales of their brilliance. Research has shown that corvids have independently evolved many of the complex cognitive abilities similar to that of apes (Emery & Clayton, 2004; Emery, 2006). The impressive ability of New Caledonian crows (*Corvus moneduloides*) to manufacture and use tools (Hunt, 1996; Taylor, Hunt, Holzhaider, & Gray, 2007), solving of problems cooperatively by rooks (Bird & Emery, 2009; Seed, Clayton, & Emery, 2008) as well as social learning in American crows (Cornell, Marzluff, & Pecoraro, 2012) makes these birds well suited models for the study of evolution of cognitive abilities.

Crows have a relatively constant shape and are mostly all-black. However, grey or white plumage occurs in at least seven species: Daurian jackdaw, collared crow, house crow, piping crow, grey crow, pied crow and hooded crow. Although these species show such a distinct plumage, they do not form a mono-phyletic group compared to the all-black forms. The repeated independent occurrence of the plumage coloration suggests a simple genetic switch (Knud Andreas Jønsson et al., 2015). The Western and Daurian jackdaws have contrasting colouration, but are known to hybridize occasionally (Madge & Burn, 1994). Similarly, the taxon pairs with contrasting plumage coloration, dwarf ravens and pied crows as well as carrion and hooded crows hybridize in well defined hybrid zones (Haas & Brodin, 2005; Londei, 2008).

The all black Carrion and grey coated Hooded crows from Europe form a narrow hybrid zone that has been studied for over a century (Meise, 1928). Despite the striking difference in the colour phenotype, most of the genome shows almost no differentiation. Apart from the difference in plumage colouration, differences in behaviour has also been observed. The stable maintenance of the hybrid zone is thought to be the result of strong assortative mating. The European crow hybrid zone has been the focus of studies along the entire span of the hybrid zone (Haas & Brodin, 2005).

Species complex and replicate Hybrid zones

Apart from the hybrid zone running through Germany, another hybrid zone between the all black carrion crow (*C.c. orientalis*) and the grey coated hooded crow is found in Russia. A further switch in colour phenotype occurs

from the Russian carrion crows to the collared crows found in China. Such a leap frog pattern seen in the colour phenotype suggests that similar genetic changes might be involved in causing these changes. The European crow super-species complex is spread across Europe all the way upto Russia and extends into China. Further switches in the plumage colour occur within the corvid phylogeny such as the Indian house crow morphs and Jackdaws. Hence, the genus *corvus* is very well suited for the study of speciation and genetics of plumage colour.

1.5.2 Marine system – Orca

Killer whales or Orcas (*Orcinus orca*) are an iconic species that have been portrayed as powerful predators that roam the seas and have occupied a prominent position in human culture. Monitoring the behaviour of killer whales has shown that they are highly social creatures with intricate group dynamics (Whitehead, 1998). Dietary habits of site-faithful individuals have been studied for many years through direct observation, molecular and visual identification of prey remains from predation events, faecal samples and stomach contents (Ford et al., 1998; Herman et al., 2005; Krahn et al., 2007; Matthews & Ferguson, 2013; Pitman & Ensor, 2003; Saulitis, Matkin, Barrett-Lennard, Heise, & Ellis, 2000). Morphological characterisation of body length and pigmentation features has been performed using laser metrics and aerial photogrammetry (Baird & Stacey, 1988; Durban & Parsons, 2006; Fearnbach, Durban, Ellifrit, & III, 2011).

Killer whales lack natural predators and feed on fish, reptiles, birds and mammals (Foote, Newton, Piertney, Willerslev, & Gilbert, 2009). Orcas are physically larger than all other species in the dolphin family (Delphinidae) and have a widespread distribution spanning all ocean basins. However, in various locations they have specialised into sympatric ecotypes with different prey capture strategies (Riesch, Barrett-Lennard, Ellis, Ford, & Deecke, 2012). Three distinct ecotypes have been characterised in the North Pacific: a mammal eating ecotype referred to as 'transient', a fish eating ecotype known as 'resident' and the 'offshore' ecotype that is found further offshore and feeds on sharks and other fish. Killer whales found in the waters near Antarctica have diversified into several distinct morphotypes. The individuals that are most similar to the common killer whale in coloration are classified as type A. Types B1 has a large eye patch and feeds on Weddell seals (Pitman & Durban, 2012). Type B2 also has a large eye patch but eats penguins (Matthews & Ferguson, 2013). Type C has a smaller forward-slanted eye patch and feeds on a variety of different fish (Whitehead, 1998).

Extensive morphological and behavioural studies along with the availability of a high quality reference genome as well as genomic resources (Foote et al., 2015; McGowen, Grossman, & Wildman, 2012) for an outgroup species provides an opportune time for population genetic study of different

orca ecotypes. Orcas share many similarities with humans in that both are mammals and have highly social behavioural patterns such as post-reproductive females that help their kin (Brent et al., 2015). Such similarities between orca ecotypes and human populations provides an interesting counter narrative given our anthropocentric view of the world. Studying speciation in a marine mammal provides important insight into the processes involved in marine systems compared to studies in other vertebrates. The transient and resident ecotypes from the North Pacific and the B1, B2 and C ecotypes from the Antarctic occur at different levels of differentiation from each other. Hence, this provides snapshots of different stages of the speciation continuum.

1.5.3 Comparative avian system

Study of individual species provides valuable insight regarding the evolutionary patterns and processes that drive speciation. Yet, the broader applicability of these patterns at macro-evolutionary time scales requires a comparative approach (Seehausen et al., 2014). Do some species show extraordinary patterns due to their unique demographic history? How do life-history, genome organisation and karyotype constrain nucleotide diversity landscapes? Understanding the effect of micro-evolutionary processes across species boundaries promises to provide answers to many of these questions.

Passerines are the most speciose bird lineage with over 5000 species. Such large diversity coupled with the rather stable karyotype of birds (Ellegren, 2013) makes them an ideal study system to perform comparative studies. Such integrative studies that leverage genomics to span multiple time-scales will be crucial for improving our understanding further (Cutter, 2013). Populations of various passerine bird species at different stages of speciation (Burri et al., 2015; Ellegren, Smeds, Burri, Olason, Backström, et al., 2012; Lamichhaney et al., 2015; J W Poelstra et al., 2014), have been sequenced to reveal the heterogeneous differentiation landscape. Availability of this extraordinary dataset that consists of population level sampling for multiple Corvids, Flycatcher and Darwin's Finch species provides us an opportunity to understand how the differentiation landscape changes over much larger time-scales. Evaluating the dynamics of the differentiation landscape will provide an opportunity to test the predictions from different hypothesis and motivate the expansion of the dataset to more species.

2 Research Aims

2.1 General Aims

The general aim of my thesis is to understand the evolutionary processes acting during early stages of population divergence. To reach this goal, I used a combination of transcriptome analyses characterizing divergence at an intermediate level between the DNA blueprint and the phenotype in question, and genome-wide resequencing characterizing genetic variation across the genome in light of population genetic theory. The empirical systems I used shed light on different aspects promoting divergence. Multiple population comparisons of corvids (nucleotide sequence and transcriptome) allowed investigating the genetic basis and impact of a candidate speciation trait (mating relevant colour). The killer whale system lends itself to study the role of rapid niche shifts for genome-wide differentiation and also screen for local signatures of selection. Finally, comparative analyses across the speciation continuum within both systems and in addition between large taxonomic distances in several avian families allowed addressing the role of structural genomic features on heterogeneous differentiation landscapes.

2.2 Specific Aims

Paper I – To generate a high quality reference genome of the European crow by sequencing one individual to a high coverage. To quantify the genome-wide patterns of DNA sequence nucleotide diversity and divergence, as well as gene expression differences between carrion crow and hooded crow populations from both sides of the European crow hybrid zone.

Paper II – To generate whole genome re-sequencing data from multiple populations of crows across Europe, Russia and China to look at patterns of nucleotide diversity and divergence across four switches in phenotype. To reconstruct the demographic history of this species complex. To investigate whether regions of elevated nucleotide differentiation are shared based on phenotype, geography, genomic architecture or whether selection is population-specific.

Paper III – To study nucleotide sequence differentiation across the speciation continuum using whole-genome sequencing data from five killer whale ecotypes. To test the hypothesis that founder effects associated with niche shifts promote speciation. To screen for genomic signatures of adaptation.

Paper IV – To compare differentiation landscapes across speciation continua in several bird species to answer the question in how far general, structural genomic features are associated with elevated differentiation.

3 Summaries of Papers

Paper I – The genomic landscape underlying phenotypic integrity in the face of gene flow in crows

As species diverge they accumulate differences in their genomes. We are only beginning to understand how such differences build up across the genome, are the differences confined to small parts of the genome or do they randomly accumulate across the genome? Do these genomic differences contribute to phenotypic differences and what role do they play in speciation? Populations in early stages of divergence differing in phenotypic characters conducive to reproductive isolation are very well suited to address these questions.

The European crow hybrid zone has long been a textbook example of speciation (Meise, 1928). Black carrion crows (*Corvus (corone) corone*) and grey coated hooded crows (*Corvus (corone) cornix*) hybridize along a narrow zone that has been stably maintained for a century (Meise, 1928), presumably even since secondary contact approximately 8,000 years ago (Mayr, 1942). Extensive morphological, ecological and behavioural work (Nicola Saino, 1992; Saino & Scatizzi, 1991; Saino, 1992) suggests some degree of reproductive isolation related to plumage colour under conditions of ongoing gene flow (backcrossing) within the hybrid zone. Mean genetic differentiation between these populations at numerous candidate loci has been shown to be relatively homogeneous and rather low for phenotypically distinct populations (Haas et al., 2009; J W Poelstra, Ellegren, & Wolf, 2013; Saino, Lorenzini, Fusco, & Randi, 1992; Wolf, Bayer, et al., 2010). The genetic basis for the difference in colouration is not known, but suspected to involve the melanogenesis pathway (Hill & McGraw, 2006; J. W. Poelstra et al., 2015).

In this study, we assembled a high-quality genome of one Swedish hooded crow male individual that was sequenced at ~152x depth of coverage using multiple paired-end and mate-pair libraries with insert sizes from 150bp – 20 kb. This reference genome was used to call SNP's from whole-genome re-sequencing data of 60 crows sequenced to a mean depth of 12.2x. Unrelated male hooded crows were sampled from Sweden and Poland, carrion crows from Germany and Spain. In agreement with previous studies mean genome-wide differentiation was low ($F_{st} = 0.02$) with the carrion crow

population from Spain being the most divergent. The carrion crow population from Germany was on a genome-wide level overall more similar to the hooded crow population from Poland and Sweden than to the carrion crow population from Spain and showed clear signatures of admixture.

Greater genetic similarity between colour morphs than within colour morphs provides an opportunity to understand the genetic basis of the colour difference stably maintained in the face of genome-wide gene flow. Sequence differentiation was calculated across the genome in 50Kb windows to identify candidate regions of divergent selection showing higher differentiation than the genomic background. Five narrow “peaks” of differentiation could be identified by selecting the windows that had strongly elevated levels of differentiation between hooded and carrion crows (top 1% F_{ST}). In all populations these windows also had a nucleotide diversity lower than the genomic background. Moreover, mean differentiation of the sex chromosome Z was higher than that of the autosomes. This is in line with the theoretical expectation that the sex chromosome in birds should be subject to higher levels of genetic drift, as it only has 2/3 of the effective population size N_e of autosomes.

In addition to the rather crude window based approach, we also reconstructed local phylogenies using an HMM-SOM based method that is able to delineate regions with as few as three SNP's and assign it to the appropriate cactus (phylogenetic hypothesis based on SNP variation). This software provides a more sensitive method to identify and assign regions to different phylogenetic hypothesis. The HMM-SOM method implemented in the software Saguaro (Zamani et al., 2013) was used to build ten different “cacti” across the genome that could be grouped into just two classes, those that separated populations based on their colour (selection hypothesis) and those that separated Spain from the other three populations (demographic background). Using this method, we could identify additional candidate regions that were missed by the window based approaches.

As a complement to the population genomic approach, we characterized the transcriptome of birds raised under common garden conditions from all four populations. RNA-seq data were used to characterise gene expression divergence between carrion and hooded crows in forebrain, gonads, liver and regrowing feather follicles. While most genes had similar expression levels in both carrion and hooded crows, many genes involved in the melanin pigmentation pathway showed a clear difference in their expression patterns. The differentially expressed genes RASGRF1, NDP and HPGDS overlapped with either the F_{ST} peaks or the cacti that could distinguish the colour morphs. Genes showing clear differential expression in the skin between colour morphs are good candidates for further studies. Moreover, genes such as CACNG calcium channels, identified in the genome scans, likely contribute to gene expression differences as primary regulators, though they were not differentially expressed themselves. Additional insight from immuno-

histo-chemical visualization of active melanocytes indeed suggests that pervasive gene expression differences in the melanogenesis pathway are not due to low densities or malfunction in hooded crows, but rather reflect upstream regulatory changes setting off a transcriptional cascade.

This study provides first insight into how nucleotide and transcriptome level divergence builds up across the genome at early stages of phenotypic divergence. The candidate loci showing localised divergence against a low genome wide background identified in this study provide a good set-up to test the hypothesis regarding “speciation islands” and their role in the speciation process.

Paper II – Patterns and processes of genomic differentiation in replicate contact zones

As natural populations diverge they inevitably accumulate genetic differences in their genomes. These changes tend to be non-randomly distributed across the genome in what is often referred to as heterogeneous landscapes of genetic differentiation. However, it has been difficult to understand the relationship between this heterogeneous landscape, the accompanying phenotypic changes, and the build-up of reproductive isolation. Studying populations at different stages of the speciation continuum within species complexes sharing genome architectures and life history traits allows to gain insight into the underlying evolutionary forces. Particularly revealing are comparisons contrasting population pair comparisons that have most likely not been affected by gene flow to population pairs with a history of introgression (Burri et al., 2015; M. Kronforst et al., 2013; Nadeau et al., 2014; Renaut et al., 2013). We here make use of a crow species complex allowing to compare the differentiation landscapes between allopatric populations and populations connected by gene flow. Special to this system is the repeated occurrence of contact zones separating populations with divergent colour phenotypes known to be involved in mate choice, where all black morphs hybridize with pied forms. This setup is well suited to distinguish differentiation peaks caused by processes common to all populations compared to those caused by population specific processes, particularly those involved in maintaining the marked colour contrast against introgression.

In this study, we sequenced 128 genomes from across the speciation continuum of the crow species complex *Corvus (corone) spec.* stretching across Europe, Russia and China. Samples were obtained from both sides of two well characterised hybrid zones between different colour morphs, one in Europe and another in Russia as well as another steep phenotypic transition zone of all-black *C. (c.) orientalis* and *C. (c.) pectoralis* between Russia and China. We see a moderate isolation-by-distance pattern across Europe. The Spanish all-black carrion crow *C. (c.) corone* as well as the eastern Russian

C. (c.) orientalis populations showed the deepest divergence hence not supporting a basal split by phenotype. Analyses of zygosity disequilibrium indicated coalescing demographic histories at approximately 300,000 years for the Eurasian complex consistent with a radiation during climatological perturbations at the onset of the 'Riss' Pleistocene glacial. A single recent origin for all hooded crow populations spread over a vast geographic region was supported by much lower mean genome-wide differentiation between hooded crow populations compared to the mean genome-wide differentiation between all other populations and a series of other population genetic analyses. Gene flow across both the hybrid zones as well as east-Asian zone of contact was supported by significant Patterson's D statistics, elevated linkage disequilibrium and admixture analyses.

The genomic landscape of differentiation measured by z-standardized F_{ST} in 50Kb windows was highly heterogeneous. In all the 54 pairwise comparisons, the windows with elevated levels of differentiation (>99th percentile) were significantly clustered into distinct peaks. As predicted by theoretical expectations this broad-scale heterogeneity resulted in large part from linked selection profiles that were shared between populations. F_{ST} landscapes were highly correlated among comparisons with a large fraction of the peaks being at the same positions in different pairwise population comparisons. As the genome-wide mean level of differentiation increased, the population differentiation peaks crystallized and the shared landscapes became increasingly visible. Pairwise population comparisons between allopatric populations also shared the same peaks, demonstrating that the shared differentiation landscape was largely determined by genomic features, not (exclusively) as a result of selection against gene flow.

Next, we focused our attention on the three target contact zones addressing the question of parallelism in their evolutionary dynamics. The differentiation landscapes were highly dissimilar among the three contact zones and none of the outlier peaks were shared across all three zones. The European part of the corone-cornix hybrid zone had a prominent peak showing a pattern of divergent selection stretching across a 2.8Mb region on chromosome 18. This peak had been identified by us in a previous study (see Paper-I), and contains melanogenesis-related genes such as *PRKCA*, *SLC16A6*, *AXIN2*, *CACNG1*, *CACNG4*, *CACNG5* and the *RGS9* gene which is involved in visual perception.

In the cornix-orientalis hybrid zone the peaks had a much lower amplitude and were more evenly distributed across the genome. Two moderately sized peaks showed on chromosome 21 and the Z chromosome. Genes from the melanogenesis pathway such as *CLCN6*, *MFN2* and *MTOR* are located in these peaks. A fixed difference in the *LRP5* gene which interacts with the WNT pathway was found across both the cornix-orientalis and orientalis-pectoralis contact zones. The orientalis-pectoralis contact zone showed a prominent peak spanning 16 consecutive windows on chromosome 23.

In this study we used multiple pairs of population comparisons to understand the patterns of differentiation for populations across multiple spatial and temporal scales. We present a strategy to utilise this information to disentangle processes that are shared by all populations and those that are specific to a population. Results suggest population specific processes across contact zones acting on top of background processes of linked selection affecting all populations alike. Stable maintenance of the steep colour contrast at zones of hybridization seems to be realized by different, though partially overlapping genes, suggesting a multigenic architecture of the trait.

Paper III– Genome-culture coevolution promotes rapid divergence in the killer whale

How does ecological adaptation and behavioural innovation influence speciation? How does the build-up of genomic differentiation proceed as species diverge? Do the same genomic regions show higher differentiation at different stages along the speciation continuum? Regions of the genome that harbour genes involved in causing phenotypic differences between ecotypes are thought to exhibit signatures of selection. We try to answer these questions and find population specific signatures of selection in killer whale ecotypes.

Killer whales are apex predators that consume a diverse variety of prey such as fish, reptiles, birds and mammals (Foote et al., 2009). Although all ocean basins are inhabited by killer whales, in certain locations killer whales have diversified into ecotypes that have developed specialised prey capture strategies through behavioural innovations (Pitman & Durban, 2012). Various ecotypes of killer whale (*Orcinus orca*) characterized by their feeding ecology (Riesch et al., 2012) are at different levels of differentiation from each other and can be used to study the build up of genomic differentiation at different stages of speciation (Seehausen et al., 2014). Prevalence of matrilineal group structure and transfer of knowledge from matriarchs to their kin is thought to play an important role in killer whale speciation (Whitehead, 1998).

We generated low coverage genome sequences suitable for population genomic analysis for 48 individuals, consisting of 10 individuals each of a mammal-eating ('transient') ecotype and fish-eating ('resident') ecotype that are sympatric in the North Pacific; and 7 individuals of a large mammal-eating form (type B1), 11 individuals of a partially sympatric smaller form which feeds on penguins (*type B2*), and 10 individuals each of the smallest form of killer whale which feeds upon fish (*type C*), all three from Antarctic waters. Reconstruction of ancestral demographic history revealed bottlenecks during founder events, promoting the basis for ecological divergence and genetic drift resulting in genome-wide differentiation across the speciation continuum. The genome-wide mean differentiation as estimated by F_{ST}

ranged from 0.09 to 0.57 providing a snapshot of different stages of the speciation process. The sympatric populations from the Pacific showed a clear separation from the ecotypes of the Antarctic, with all the sampled Antarctic ecotypes being closely related.

The sequence differentiation landscape was extremely heterogeneous across the genome, but very similar between comparisons. The presence of shared genomic features was reflected in the nucleotide diversity being correlated across all five ecotypes. Despite this shared background of differentiation, certain genic regions showed differentiation that could be functionally relevant to specific ecotypes. In comparisons between mammal-eating and fish-eating ecotypes, genes involved in the formation of the three primary germ layers of the digestive system during embryonic development showed the most significant levels of differentiation. The comparisons between ecotypes inhabiting the extreme cold of the Antarctic pack ice with ecotypes from the more temperate North Pacific showed the most significant enrichment in genes involved in adipose tissue development.

Even though the evolutionary processes driving divergence cannot be determined with certainty, fixed allelic variants within genes could potentially produce phenotypic differences. Genes encoding proteins primarily expressed in the gastrointestinal tract contained two fixed non-synonymous amino acid substitutions. Fixed non-synonymous amino acid substitutions were further found in the exons of several genes that encode proteins associated with reproductive function, including testes development, regulation of spermatogenesis, spermatocyte development and survival and initiating the acrosome reaction of the sperm.

Humans are unique in having a rapid evolution due to the interaction of culture and genes as exemplified by the relationship of cattle domestication and lactase persistence alleles (Laland, Odling-Smee, & Myles, 2010). The killer whale system presented here provides an interesting non-human example with a potential role for gene-culture interaction. Studies that analyze the speciation continuum for genome-wide differentiation also provide useful insight into the patterns and processes that shape the process of speciation.

Paper IV – Genomic signatures of species diversification – a comparative perspective

The availability of high-throughput sequencing technology at a relatively lower cost has made it possible to study the molecular and population genetics of speciation in numerous non-model taxa (Ellegren, 2014; Wolf, Lindell, et al., 2010). Genomic approaches to speciation have helped shape our understanding of the evolutionary processes involved in speciation in diverse taxa such as sunflower plants (Renaut et al., 2013), aspen trees (J. Wang, Street, Scofield, & Ingvarsson, 2015), insects (Simon H. Martin et al., 2013;

Soria-Carrasco et al., 2014, p. -), fish (McGee, Neches, & Seehausen, 2015) and birds (Ellegren, Smeds, Burri, Olason, Backström, et al., 2012; J.W. Poelstra & Vijay, 2014)(for review see (Seehausen et al., 2014). In papers I to III we use such genome scan approaches to identify regions of the genome that show signatures of selection. Regions identified by such scans are often considered to be candidate speciation genes.

Despite substantial progress, it is still hard to link the genomic patterns of population differentiation that are observed to the underlying process of species diversification (Berner & Salzburger, 2015; Toews et al., 2015). Genomic regions with elevated levels of nucleotide sequence differentiation ('speciation islands') could either be the result of divergent selection against homogenizing gene flow promoting reproductive barriers (Feder, Egan, et al., 2012; Nosil & Feder, 2013) or could simply be a consequence of genomic features exposing effects of linked selection in regions of low recombination regions (Cruickshank & Hahn, 2014; Cutter & Payseur, 2013; M A F Noor & Bennett, 2009). Distinguishing between these scenarios is complicated by the fact that numerous factors are known to perturb patterns of genetic variation along the genome.

A range of strategies such as the functional validation of candidate regions in non-model species (M. R. Kronforst & Papa, 2015), theoretical approaches (Bank, Ewing, Ferrer-Admettla, Foll, & Jensen, 2014), experimental evolution (Dettman, Sirjusingh, Kohn, & Anderson, 2007) and manipulative experiments (Soria-Carrasco et al., 2014) under different selective regimes as well as comparative population genetic approaches (Nadeau et al., 2014; Vijay et al., 2015) that utilise natural spatio-temporal distribution of populations are being used to disentangle the effects of background selection from those thought to contribute to speciation. In this paper, we propose a *macro-level comparative approach* to assess the contribution of shared linked selection versus population specific processes associated with divergent selection in generating heterogeneity in genomic differentiation profiles.

We utilise population genomic datasets from three avian species complexes to disentangle processes responsible for differentiation islands. Patterns of nucleotide diversity, differentiation and divergence across the genome were conserved on a broad scale within and across all three clades. Although the correlations within clades was high, correlation coefficients between species were lower, yet consistently positive reaching up to 0.36 between flycatcher and crow populations. Similarly, genetic differentiation, divergence as well as population-specific estimates of PBS were correlated both within and across clades. Next, we quantified the overlap of outlier windows with centromeres and sub-telomeric regions predicted in zebra finch. In all three clades, overlap was non-zero (FC: 58.53% & 60.98%, DF: 14.63% & 29.27%, CR: 21.95% & 31.7%) and significantly greater than expected by chance alone.

Using this comparative approach we get few clues about the prevalence and stability of linked selection across large evolutionary timescales. These results are consistent with background selection being a key driver of genomic heterogeneity. A role of recurrent positive selection in the form of meiotic drive could explain the association we see with respect to telomeres and centromeres. The increased use of comparative, phylogenetic approaches provides a novel strategy to understand the processes driving genomic landscape of selection.

3.1 Concluding Remarks and Future Prospects

Understanding evolutionary processes at the interface between genotype and phenotype has the potential to provide a clearer picture of the principles underlying evolutionary change. Access to genome-wide data in natural populations at different stages of the speciation process provides a rather novel opportunity for studying this link using a diverse set of phenotypes (Seehausen et al., 2014). Population genomic approaches, as exemplified in this thesis, can contribute to our understanding of how selective and neutral processes shape the phenotypic and genetic landscape of speciation as well as adaptation. When applied to many different system, this will provide crucial information for creating a more comprehensive and predictable picture of the way evolution works (Stern & Orgogozo, 2009). Catalogues that document inference of positive selection in humans (Cheng, Chen, Richards, Deng, & Zeng, 2009; Li et al., 2014) and other species (Proux, Studer, Moretti, & Robinson-Rechavi, 2009) have been created. Such datasets allow the comparison of regions of the genome that show signatures of selection across species (Enard, Depaulis, & Roest Crolius, 2010). Detailed annotation and analysis of such catalogues of loci will provide new insight (A. Martin & Orgogozo, 2013) and reveal the limitations of methods used to detect selection.

Extrapolating from the avian and mammalian systems studied here, suggests that broad-scale patterns of genome-wide sequence differentiation are primarily driven by selective processes common to all populations. Few major peaks per chromosomes, informative correlations between differentiation, divergence and diversity statistics and potential association with centromeres (in birds) suggested that differences may preferentially accumulate in regions of low recombination as suggested by background selection (Brian Charlesworth, 2013; Cruickshank & Hahn, 2014). An alternative explanation for the differentiation peaks may be given by recurrent positive selection, which in case of centromeres would imply centromeric drive dynamics (Malik, 2009). Species differ in the magnitude of the influence of background selection vs other forms of selection. Efforts to quantify the relative contributions of different forms of selection across the tree of life have been

attempted using numerous approaches (Corbett-Detig, Hartl, & Sackton, 2015; Cutter & Payseur, 2013). Future studies attempting to quantify selection will benefit from a high resolution map of the positions of centromeres and recombination rates.

Despite striking commonalities in differentiation landscapes, chapters I-III of this thesis also provided evidence for idiosyncratic evolutionary dynamics of specific populations pairs. Functional genomic characterization of these candidate regions are necessary to validate these population genomic results – an enormous challenge for non-model organisms not accessible to traditional genetics. Detailed gene expression analyses as illustrated in chapter I may be seen as an attempt in this direction. Such large scale expression datasets provide input for methods that infer regulatory networks (Nachman, Regev, & Friedman, 2004). Evolution of regulatory networks (Erwin & Davidson, 2009) and its relevance for speciation have been explored in theoretical models (Lindtke & Buerkle, 2015; ten Tusscher & Hogeweg, 2009) and promise a comprehensive explanation for linking genotype and phenotypes.

A deeper understanding of mechanisms involved in speciation and adaptation will help us understand the link between genotype and phenotype. Our ability to identify population structure at increasing resolution and research into the genetic basis of phenotypic traits has stimulated discussion in the field of conservation genetics (Aaron B.A. Shafer et al., 2014). The field of reproductive medicine could also benefit from a better understanding of the evolution of reproductive barriers. Speciation research has also inspired the creation of “synthetic species” based on the idea of reproductive barriers (Moreno, 2012).

To more generally address the relative importance of different processes acting during population divergence, evidence from many systems differing in ecological contrast, life history, effective population size and generation time will be needed (Leffler et al., 2012). Moreover, further developments in theoretical models are necessary to integrate features such as demographic history, genome architecture, role of karyotype evolution, recombination rate and divergence in gene expression into a single model. The role of less studied factors such as epigenetics, phenotypic plasticity and environmental interaction deserve special attention in order to develop an overarching theory of evolution (Laland et al., 2015).

4 Svensk sammanfattning

4.1 Bakgrund

Den enorma biologiska mångfalden på jorden har delats in i olika hierarkiska grupper såsom arter. Den svenska vetenskapsmannen Carl von Linné var pionjär inom sådan klassificering av organismer som grundar sig på fysiska egenskaper. Nästan hundra år senare var Charles Darwin (1859) en av de första att inse att hierarkin i en sådan klassificering i de flesta fall var en följd av gemensam härkomst. Naturforskare har omsorgsfullt samlat prover och klassificerat dem som arter baserat på någon form av fenotypisk eller genetisk särprägel. Emellertid har man endast börjat förstå den genetiska grunden för bildningen av arter. I denna avhandling använder vi kråka och späckhuggare för att få en första bild av hur genomet ändras under artbildning.

4.2 Matrial och metoder

4.2.1 Studiesystem

Släktet *Corvus* har ett intressant mönster av upprepade fenotypförändringar från en helsvart till en svartgrå. Den upprepade förekomsten av detta mönster tyder på att en enkel genetisk brytare av genflöde kan orsaka denna förändring i fenotypen. I denna avhandling har vi studerat hybridzonen för den europeiska kråkan och jämfört den med den ryska hybridzonen samt den östasiatiska kontaktzonen. Följdaktligen så använder vi helgenomsprover från kråkans artkomplex för att studera artbildning.

Späckhuggare (*Orcinus orca*) är en ikonisk art som har framställts som kraftfulla rovdjur som driver omkring i haven och har en framträdande plats i den mänskliga kulturen. Späckhuggarna är större än alla andra arter i delfinfamiljen (Delphinidae) och har en bred spridning som spänner över alla världens oceaner. Dock har de på olika platser specialiserat sig till delvis sympatriskt förekommande ekotyper med olika fångststrategier. Späckhuggare delar många likheter med människor i och med att båda är däggdjur med utpräglade sociala beteendemönster. Sådana likheter mellan späckhuggarekotyper och befolkningsgrupper ger en intressant kontrast till vår antropocentriska syn på världen.

4.2.2 Metoder

Sequensering av DNA-molekyler har gjort det möjligt att bestämma den exakta ordningsföljden av nukleotidbaser. Förbättringar i tekniken under årens lopp har gjort sekvenseringsprocessen enklare och billigare. De minskade kostnaderna för storskaliga sekvenseringsteknologier har gjort det möjligt att sekvensera hela genom av olika icke-modellarter, mäta genuttryck med hjälp av RNA-sekvensering samt att studera populationsgenomik med re-sekvenseringsdata. I denna avhandling, utnyttjar vi kraften i dessa metoder för att studera hur genuttryck och DNA-sekvenser förändras under artbildning.

4.3 Resultat

Vi utförde *de novo*-assembly av kråkgenomet och använde sekvensdata från 30 kråkor från varje sida av den europeiska hybridzonen för att förstå differentieringsmönstren över hela genomet. Vi mätte också genuttryck med RNA-sekvensering för att identifiera gener som uttrycks olika mellan olika vävnader. (Artikel I)

Genom en mer omfattande datainsamling av kråkpopulationer från hela världen, jämför vi differentieringslandskapet i europeiska och ryska hybridzoner. En kontroll med allopatriska jämförelser användes för att identifiera områden där differentieringsmönstren är gemensamma för alla populationer. Dessa gemensamma regioner togs bort från var och en av hybridzonjämförelserna för att identifiera hybridzon-specifika differentieringsstopp. (Artikel III)

Vi använde sekvensdata på populationsnivå från fem olika ekotyper av späckhuggare för att karakterisera uppbyggnaden av genetiska skillnader över genomet. Genomiska regioner som innehåller ekotypspecifika genetiska förändringar identifierades också. Dessa regioner visade sig innehålla gener som potentiellt kan vara involverade i ekotypspecifik anpassning. (Artikel III)

I den sista artikeln jämförde vi differentieringslandskapen i kråk-, flugsnappar- och Darwinfinkpopulationer. Vi visar att nukleotiddiversitetslandskapet till stor del är konserverad mellan olika arter. Nukleotiddiversiteten minskar nära centromerer och telomerer medan differentiering ökar. (Artikel IV)

4.4 Diskussion

Denna avhandling ger en första uppfattning om differentieringslandskapet som ligger bakom de tidiga stadierna av artbildning och hur det utvecklar sig

över tiden. Vi hoppas att de mönster som vi har upptäckt i kråk- och späckhuggarsystemen kommer att ha en bred tillämpbarhet på alla ryggradsdjursystem i allmänhet.

Det genomiska differentieringslandskapet är mycket heterogent och verkar drivas av ett flertal processer. Analys av differentieringslandskapet över den Europeiska kråkans hybridzon har identifierat regioner i genomet som visar större genetiska förändringar än förväntat och visa gener i dessa regioner kan därför ses som kandidatgener som kan vara inblandade i artbildning. Utökning av studien till två oberoende kråkhybridzoner och jämförelsen med allopatriska populationer har gett en unik möjlighet att särskilja mönster som är gemensamma för alla populationer mot de som är specifika för hybridzoner. Mätning av genuttryck från många vävnader från de två färgtyperna ger en tydlig bild av hur genuttrycket skiljer sig medan artbildningen fortskrider. Följaktligen tillhandahåller kråksystemet en ögonblicksbild av de tidiga stadierna av artbildning.

Både kråksystemet och späckhuggarekotyperna ger ögonblicksbilder av differentieringslandskapet i olika rumsliga konfigurationer och tidsmässiga ögonblicksbilder av kontinuerlig artbildning. De fem olika ekotyperna av späckhuggare avviker olika mycket från varandra och faller längs ett spektrum av artbildning. Var och en av ekotyperna har också utvecklat habitat-specifika kostanpassningar. Genom att mäta skillnader i sekvensen, kan genomiska regioner som visar ekotypens specifika förändringar identifieras. Dessa ekotypspecifika regioner tillhandahåller kandidatgener för ekotypspecifik anpassning.

5 Acknowledgements

I am very grateful to my main supervisor Jochen Wolf for accepting me as a PhD-student and all the patient instruction and inspiration. It has been an exciting, exhilarating, adventurous and awesome learning experience. I have enjoyed every moment of working with you and hope that there will be opportunities for future work together. Thanks also to Manfred Grabherr, my co-supervisor for enabling me to participate in bioinformatics projects and providing me an opportunity to work with him. Manfred's extensive and broad knowledge always helped us find light when it got too dark.

Thanks to Hans Ellegren and his group for sharing lab meetings and friday journal clubs with Jochens group which provided a very instructive and stimulating atmosphere. Thank you Robert Ekblom and Jochen for organising the NGS club and all the participants over the years. The whole Evolutionary biology program for being a great place to work. Evolution discussion groups were a great place to hear about the work being done by other students in the department and also practice talks. Thank you for organising this Kim.

Thanks to my lab mates over the years: Jelmer (with whom i worked on numerous projects), Christen (thank you so much for your help with various things, especially with figures just before submitting:), Matthias (it was fun working on the “core” crow team), Julia, Kristaps, Reza & Lars Dronasson, Andy (Orca, Convergent Evolution, The Neanderthal dream), Glib (co-expression clusters), Aaron (for providing a unique perspective), Özge, Anca (Inversion detection PCR), Melanie, Claire, Chi-Chih, Ramon, Verena (thank you for feedback on my thesis drafts), Axel. Thank you Henrik & Mark for discussions on crow genome annotation. For discussions and insights about all things bioinformatics, thanks to Linnea, Pall, Douglas, Axel, Robert, Per and others. Special thanks to Linnea and Josefine for help with the Swedish summary. Thanks to Talla Venkat and Chi-Chih for help with cover design.

My masters in bioinformatics went a long way in preparing me for a Phd. Thanks to Margareta Krabbe, Lars-Göran Josefsson for being great masters program coordinators. Many thanks to Thorsten Heidorn for organising the iGEM team which was a lot of fun while being a very useful experience.

Thanks to Matthew Webster for inspiring me about vertebrate genomics and GC-content evolution.

Thanks to all my NRI friends for all the great times. Many other people over the years have made my stay here memorable, worthy and right. Thanks to all these people who are too numerous to mention by name. I wish to thank my mother Shantala Priyadarshini for her encouragement and support.

During my time here i have thoroughly enjoyed the high quality of science, socialism, Swedes, snow (walking on the frozen river), sweets, sunless days (i will sorely miss the darkness forest rally) and short mid-summer nights. I am extremely grateful to the Swedish people for providing me an opportunity to do study here.

The computations were performed on resources provided by SNIC through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under the projects b2010059, b2010060, b2012209, b2013182, b2014050 and b2014099. I would like to thank the UPPMAX support team for excellent support in utilizing kalkyl, tintin, milou and halvan nodes.

Although i was fortunate enough to escape becoming a physician, unlike a resolute few (Choudhury, 2011) not only did i become an engineer but also an IT professional! While my 'bug-period' died in the sixth grade, it was thankfully resurrected from the dead, but inside computer code.

Suddenly, i feel like the leaf on a tree from that children's story that arrogantly thought that it was a self-made, independent leaf. A fierce wind has finally blown me of my branch to the dirt on the ground. As my life slowly ebbs away, i look at the magnificent old tree that has been my home for the past few years and realise that i have never been on my own. My entire life has been part of something bigger and more beautiful than anything i could have imagined. As i try to awaken from this delusion of self, i fear that an arrogant, self-centred kid will rake me up and set me on fire. Hopefully, even as the fire burns the leaf will not be consumed as if it were part of the burning bush from Exodus.

Crows ahvayante kråkor natu bikuskaru yachakan and other wisdom would have never reached me without my mothers belief in me and encouragement.

One thing is clear, we don't have all the answers just a few questions we have been lucky enough to pose. We have to live with these questions in the hope of someday finding that one question whose answer is 42.

6 References

- Alcaide, M., Scordato, E. S. C., Price, T. D., & Irwin, D. E. (2014). Genomic divergence in a ring species complex. *Nature*, *511*(7507), 83–5.
- Alföldi, J., & Lindblad-Toh, K. (2013). Comparative genomics as a tool to understand evolution and disease. *Genome Research*, *23*(7), 1063–8.
- Andersson, M. B. (1994). *Sexual selection*. Princeton University Press.
- Andolfatto, P., Depaulis, F., & Navarro, A. (2001). Inversion polymorphisms and nucleotide variability in *Drosophila*. *Genetics Research*, *77*(01), 1–8.
- Andrew, R. L., & Rieseberg, L. H. (2013). Divergence is focused on few genomic regions early in speciation: incipient speciation of sunflower ecotypes. *Evolution; International Journal of Organic Evolution*, *67*(9), 2468–82.
- Baack, E. J., & Rieseberg, L. H. (2007). A genomic view of introgression and hybrid speciation. *Current Opinion in Genetics & Development*, *17*(6), 513–8.
- Baird, R. W., & Stacey, P. J. (1988). Variation in saddle patch pigmentation in populations of killer whales (*Orcinus orca*) from British Columbia, Alaska, and Washington State. *Canadian Journal of Zoology*, *66*(11), 2582–2585.
- Bank, C., Ewing, G. B., Ferrer-Admetlla, A., Foll, M., & Jensen, J. D. (2014). Thinking too positive? Revisiting current methods of population genetic selection inference. *Trends in Genetics*, *30*(12), 540–546.
- Barton, N. H., & Hewitt, G. M. (1989). Adaptation, speciation and hybrid zones. *Nature*, *341*(6242), 497–503.
- Bateson W. (1909). Heredity and variation in modern lights. In *Darwin and modern science* (ed. Seward A. C.) (pp. 85–101).
- Beaumont, M. A. (2005). Adaptation and speciation: what can Fst tell us? *Trends in Ecology & Evolution*, *20*(8), 435–440.
- Berg, J. J., & Coop, G. (2014). A population genetic signal of polygenic adaptation. *PLoS Genetics*, *10*(8), e1004412.
- Berner, D., & Salzburger, W. (2015). The genomics of organismal diversification illuminated by adaptive radiations. *Trends in Genetics*.
- Bersaglieri, T., Sabeti, P. C., Patterson, N., Vanderploeg, T., Schaffner, S. F., Drake, J. A., ... Hirschhorn, J. N. (2004). Genetic signatures of strong recent positive selection at the lactase gene. *American Journal of Human Genetics*, *74*(6), 1111–20.
- Bird, C. D., & Emery, N. J. (2009). Insightful problem solving and creative tool modification by captive nontool-using rooks. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(25), 10370–5.
- Bradnam, K. R., Fass, J. N., Alexandrov, A., Baranay, P., Bechner, M., Birol, I., ... Korf, I. F. (2013). Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. *GigaScience*, *2*(1), 10.
- Brawand, D., Soumillon, M., Neacsulea, A., Julien, P., Csárdi, G., Harrigan, P., ... Kaessmann, H. (2011). The evolution of gene expression levels in mammalian organs. *Nature*, *478*(7369), 343–8.

- Brent, L. J. N., Franks, D. W., Foster, E. A., Balcomb, K. C., Cant, M. A., & Croft, D. P. (2015). Ecological knowledge, leadership, and the evolution of menopause in killer whales. *Current Biology: CB*, *25*(6), 746–50.
- Burri, R., Nater, A., Kawakami, T., Mugal, C. F., Olason, P. I., Smeds, L., ... 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*, gr.196485.115.
- Chaisson, M. J. P., Wilson, R. K., & Eichler, E. E. (2015). Genetic variation and the de novo assembly of human genomes. *Nature Reviews Genetics*, *16*(11), 627–640.
- Chan, Y. F., Marks, M. E., Jones, F. C., Villarreal, G., Shapiro, M. D., Brady, S. D., ... Kingsley, D. M. (2010). Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. *Science (New York, N.Y.)*, *327*(5963), 302–5.
- Charlesworth, B. (2013). Background selection 20 years on: the Wilhelmine E. Key 2012 invitational lecture. *The Journal of Heredity*, *104*(2), 161–71.
- Charlesworth, B., Morgan, M. T., & Charlesworth, D. (1993). The effect of deleterious mutations on neutral molecular variation. *Genetics*, *134*(4), 1289–303.
- Cheng, F., Chen, W., Richards, E., Deng, L., & Zeng, C. (2009). SNP@Evolution: a hierarchical database of positive selection on the human genome. *BMC Evolutionary Biology*, *9*(1), 221.
- Choudhury, R. (2011). The Evolutionary genetics of the parasitic wasp *Nasonia*. Retrieved from <https://urresearch.rochester.edu/institutionalPublicationPublicView.action?institutionalItemId=13610&versionNumber=1>
- Colosimo, P. F., Hosemann, K. E., Balabhadra, S., Villarreal, G., Dickson, M., Grimwood, J., ... Kingsley, D. M. (2005). Widespread parallel evolution in sticklebacks by repeated fixation of *Ectodysplasin* alleles. *Science (New York, N.Y.)*, *307*(5717), 1928–33.
- Connelly, C. F., Wakefield, J., & Akey, J. M. (2014). Evolution and genetic architecture of chromatin accessibility and function in yeast. *PLoS Genetics*, *10*(7), e1004427.
- 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.
- Cornell, H. N., Marzluff, J. M., & Pecoraro, S. (2012). Social learning spreads knowledge about dangerous humans among American crows. *Proceedings. Biological Sciences / The Royal Society*, *279*(1728), 499–508.
- Coyne, J. A., & Orr, H. A. (1989). Patterns of speciation in *Drosophila*. *Evolution*, *43*(2), 362–381.
- Coyne, J. A., & Orr, H. A. (1997). Patterns of speciation in *Drosophila* revisited. *Evolution*, *51*(1), 295–303.
- 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.
- Cutter, A. D. (2013). Integrating phylogenetics, phylogeography and population genetics through genomes and evolutionary theory. *Molecular Phylogenetics and Evolution*, *69*(3), 1172–85.
- Cutter, A. D. (2015). Repeatability, ephemerality and inconvenient truths in the speciation process. *Molecular Ecology*, *24*(8), 1643–4.

- Cutter, A. D., & Payseur, B. A. (2013). Genomic signatures of selection at linked sites: unifying the disparity among species. *Nature Reviews Genetics*, *14*(4), 262–274.
- Darwin, C. R. (1859). *On the Origin of Species*. London: John Murray.
- Darwin, C., & Wallace, A. (1858). On the tendency of species to form varieties; and on the perpetuation of varieties and species by natural means of selection. *Journal of the Proceedings of the Linnean Society of London. Zoology*, *3*(9), 45–62.
- De Bie, T., Cristianini, N., Demuth, J. P., & Hahn, M. W. (2006). CAFE: a computational tool for the study of gene family evolution. *Bioinformatics (Oxford, England)*, *22*(10), 1269–71.
- De Queiroz, K. (2007). Species concepts and species delimitation. *Systematic Biology*, *56*(6), 879–86.
- del Hoyo, J., Elliott, A., Sargatal, J., Christie, D.A. & de Juana, E. (2014). *Handbook of the Birds of the World Alive*. Lynx Edicions, Barcelona.
- Denton, J. F., Lugo-Martinez, J., Tucker, A. E., Schridder, D. R., Warren, W. C., & Hahn, M. W. (2014). Extensive Error in the Number of Genes Inferred from Draft Genome Assemblies. *PLoS Computational Biology*, *10*(12), e1003998.
- Dettman, J. R., Sirjusingh, C., Kohn, L. M., & Anderson, J. B. (2007). Incipient speciation by divergent adaptation and antagonistic epistasis in yeast. *Nature*, *447*(7144), 585–+.
- Dobzhansky, T. G. (1935). *A Critique of the Species Concept in Biology*
- Dobzhansky, T. G. (1937). *Genetics and the Origin of Species*. Columbia University Press.
- Domazet-Lošo, T., & Tautz, D. (2010). A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature*, *468*(7325), 815–8.
- Durban, J. W., & Parsons, K. M. (2006). Laser-metrics of free-ranging killer whales. *Marine Mammal Science*, *22*(3), 735–743.
- Eichler, E. E., Flint, J., Gibson, G., Kong, A., Leal, S. M., Moore, J. H., & Nadeau, J. H. (2010). Missing heritability and strategies for finding the underlying causes of complex disease. *Nature Reviews. Genetics*, *11*(6), 446–50.
- Ekblom, R., & Wolf, J. B. W. (2014). A field guide to whole-genome sequencing, assembly and annotation. *Evolutionary Applications*, *7*(9)
- Ellegren, H. (2013). The Evolutionary Genomics of Birds. *Annual Review of Ecology, Evolution, and Systematics*, *44*(1), 239–259.
- Ellegren, H. (2014). Genome sequencing and population genomics in non-model organisms. *Trends in Ecology & Evolution*, *29*(1), 51–63.
- Ellegren, H., Smeds, L., Burri, R., Olason, P. I., Backström, N., Kawakami, T., ... Wolf, J. B. W. (2012). The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature*, *491*(7426), 756–760.
- Ellegren, H., Smeds, L., Burri, R., Olason, P. I., Backström, N., Kawakami, T., ... Wolf, J. B. W. (2012). The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature*, *491*(7426), 756–60.
- Emery, N. J. (2006). Cognitive ornithology: the evolution of avian intelligence. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *361*(1465), 23–43.
- Emery, N. J., & Clayton, N. S. (2004). The mentality of crows: convergent evolution of intelligence in corvids and apes. *Science (New York, N.Y.)*, *306*(5703), 1903–7.
- Enard, D., Depaulis, F., & Roest Crolius, H. (2010). Human and non-human primate genomes share hotspots of positive selection. *PLoS Genetics*, *6*(2), e1000840.

- Erwin, D. H., & Davidson, E. H. (2009). The evolution of hierarchical gene regulatory networks. *Nature Reviews. Genetics*, 10(2), 141–8.
- Faria, R., & Navarro, A. (2010). Chromosomal speciation revisited: rearranging theory with pieces of evidence. *Trends in Ecology & Evolution*, 25(11), 660–9.
- Fearnbach, H., Durban, J. W., Ellifrit, D. K., & III, K. C. B. (2011). Size and long-term growth trends of Endangered fish-eating killer whales.
- Feder, J. L., Egan, S. P., & Nosil, P. (2012). The genomics of speciation-with-gene-flow. *Trends in Genetics : TIG*, 28(7), 342–50.
- Feder, J. L., Gejji, R., Yeaman, S., & Nosil, P. (2012). Establishment of new mutations under divergence and genome hitchhiking. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 367(1587), 461–74.
- Feder, J. L., Nosil, P., & Flaxman, S. M. (2014). Assessing when chromosomal rearrangements affect the dynamics of speciation: implications from computer simulations. *Frontiers in Genetics*, 5, 295.
- Fierst, J. L. (2015). Using linkage maps to correct and scaffold de novo genome assemblies: methods, challenges, and computational tools. *Frontiers in Genetics*, 6, 220.
- Flaxman, S. M., Feder, J. L., & Nosil, P. (2013). Genetic hitchhiking and the dynamic buildup of genomic divergence during speciation with gene flow. *Evolution; International Journal of Organic Evolution*, 67(9), 2577–91.
- Foote, A. D., Liu, Y., Thomas, G. W. C., Vinař, T., Alföldi, J., Deng, J., ... Gibbs, R. A. (2015). Convergent evolution of the genomes of marine mammals. *Nature Genetics*, 47(3), 272–275.
- Foote, A. D., Newton, J., Piertney, S. B., Willerslev, E., & Gilbert, M. T. P. (2009). Ecological, morphological and genetic divergence of sympatric North Atlantic killer whale populations. *Molecular Ecology*, 18(24), 5207–5217.
- Ford, J. K., Ellis, G. M., Barrett-Lennard, L. G., Morton, A. B., Palm, R. S., & Balcomb III, K. C. (1998). Dietary specialization in two sympatric populations of killer whales (*Orcinus orca*) in coastal British Columbia and adjacent waters. *Canadian Journal of Zoology*, 76(8), 1456–1471.
- Funk, D. J., Nosil, P., & Etges, W. J. (2006). Ecological divergence exhibits consistently positive associations with reproductive isolation across disparate taxa. *Proceedings of the National Academy of Sciences of the United States of America*, 103(9), 3209–13.
- Gaut, B. S. (2015). Evolution Is an Experiment: Assessing Parallelism in Crop Domestication and Experimental Evolution: (Nei Lecture, SMBE 2014, Puerto Rico). *Molecular Biology and Evolution*, 32(7), 1661–1671.
- Gavrilets, S. (2004). *Fitness landscapes and the origin of species 2004* Princeton. NJ: Princeton University Press.
- Genovese, G., Handsaker, R. E., Li, H., Altemose, N., Lindgren, A. M., Chambert, K., ... McCarroll, S. A. (2013). Using population admixture to help complete maps of the human genome. *Nature Genetics*, 45(4), 406–14, 414e1–2.
- Genovese, G., Handsaker, R. E., Li, H., Kenny, E. E., & McCarroll, S. A. (2013). Mapping the human reference genome's missing sequence by three-way admixture in Latino genomes. *American Journal of Human Genetics*, 93(3), 411–21.
- Gompert, Z., Lucas, L. K., Nice, C. C., Fordyce, J. A., Alex Buerkle, C., & Forister, M. L. (2013). Geographically multifarious phenotypic divergence during speciation. *Ecology and Evolution*, 3(3), 595–613.
- Grabherr, M. G., Pontiller, J., Mauceli, E., Ernst, W., Baumann, M., Biagi, T., ... Grabherr, R. M. (2011). Exploiting nucleotide composition to engineer promoters. *PLoS One*, 6(5), e20136.

- Grant, B. R., & Grant, P. R. (2003). What Darwin's Finches Can Teach Us about the Evolutionary Origin and Regulation of Biodiversity. *BioScience*, 53(10), 965.
- Graur, D., Zheng, Y., Price, N., Azevedo, R. B. R., Zufall, R. A., & Elhaik, E. (2013). On the immortality of television sets: "function" in the human genome according to the evolution-free gospel of ENCODE. *Genome Biology and Evolution*, 5(3), 578–90.
- Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., ... Pääbo, S. (2010). A draft sequence of the Neandertal genome. *Science (New York, N.Y.)*, 328(5979), 710–22.
- Haas, F., & Brodin, A. (2005). The Crow *Corvus corone* hybrid zone in southern Denmark and northern Germany. *Ibis*, 147(4), 649–656.
- Haas, F., Pointer, M. A., Saino, N., Brodin, A., Mundy, N. I., & Hansson, B. (2009). An analysis of population genetic differentiation and genotype-phenotype association across the hybrid zone of carrion and hooded crows using microsatellites and MC1R. *Molecular Ecology*, 18(2), 294–305.
- Hahn, M. W., Zhang, S. V., & Moyle, L. C. (2014). Sequencing, assembling, and correcting draft genomes using recombinant populations. *G3 (Bethesda, Md.)*, 4(4), 669–79.
- Haldane, J. B. S. (1922). Sex ratio and unisexual sterility in hybrid animals. *Journal of Genetics*, 12(2), 101–109.
- Hedges, S. B., Marin, J., Suleski, M., Paymer, M., & Kumar, S. (2015). Tree of life reveals clock-like speciation and diversification. *Molecular Biology and Evolution*, 32(4), 835–845.
- Hedrick, P. W. (2013). Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. *Molecular Ecology*, 22(18), 4606–4618.
- Herman, D. P., Burrows, D. G., Wade, P. R., Durban, J. W., Matkin, C. O., Leduc, R. G., ... Krahn, M. M. (2005). Feeding ecology of eastern North Pacific killer whales *Orcinus orca* from fatty acid, stable isotope, and organochlorine analyses of blubber biopsies. *Marine Ecology. Progress Series*, 302, 275–291.
- Hewitt, G. M. (2001). Speciation, hybrid zones and phylogeography - or seeing genes in space and time. *Molecular Ecology*, 10(3), 537–49.
- Hill, G. E., & McGraw, K. J. (2006). *Bird Coloration: Function and evolution*.
- Hillis, D. M. (1987). Molecular versus morphological approaches to systematics. *Annual Review of Ecology and Systematics*, 18, 23–42.
- Huang, H., & Rabosky, D. L. (2014). Sexual Selection and Diversification: Reexamining the Correlation between Dichromatism and Speciation Rate in Birds. *The American Naturalist*, 184(5), E000–E000.
- Huber, C. D., DeGiorgio, M., Hellmann, I., & Nielsen, R. (2015). Detecting recent selective sweeps while controlling for mutation rate and background selection. *Molecular Ecology*.
- Hunt, G. R. (1996). Manufacture and use of hook-tools by New Caledonian crows. *Nature*, 379(6562), 249–251.
- Irwin, D. E., Bensch, S., & Price, T. D. (2001). Speciation in a ring. *Nature*, 409(6818), 333–7.
- Jensen-Seaman, M. I., Furey, T. S., Payseur, B. A., Lu, Y., Roskin, K. M., Chen, C.-F., ... Jacob, H. J. (2004). Comparative recombination rates in the rat, mouse, and human genomes. *Genome Research*, 14(4), 528–38.
- Jönsson, K. A., Fabre, P.-H., Kennedy, J. D., Holt, B. G., Borregaard, M. K., Rahbek, C., & Fjeldså, J. (2015). A supermatrix phylogeny of corvid passerine birds (Aves: Corvides). *Molecular Phylogenetics and Evolution*, 94(Pt A), 87–94.

- Jönsson, K. A., Fabre, P.-H., Ricklefs, R. E., & Fjeldså, J. (2011). Major global radiation of corvid birds originated in the proto-Papuan archipelago. *Proceedings of the National Academy of Sciences of the United States of America*, 108(6), 2328–33.
- Kim, J., Kim, I., Han, S. K., Bowie, J. U., & Kim, S. (2012). Network rewiring is an important mechanism of gene essentiality change. *Scientific Reports*, 2, 900.
- Kim, J., Larkin, D. M., Cai, Q., Asan, Zhang, Y., Ge, R.-L., ... Ma, J. (2013). Reference-assisted chromosome assembly. *Proceedings of the National Academy of Sciences of the United States of America*, 110(5), 1785–90.
- Kirkpatrick, M., & Barton, N. (2006). Chromosome inversions, local adaptation and speciation. *Genetics*, 173(1), 419–34.
- Kocher, T. D. (2004). Adaptive evolution and explosive speciation: the cichlid fish model. *Nature Reviews. Genetics*, 5(4), 288–98.
- Kolmogorov, M., Raney, B., Paten, B., & Pham, S. (2014). Ragout-a reference-assisted assembly tool for bacterial genomes. *Bioinformatics (Oxford, England)*, 30(12), i302–9.
- Kondrashov, F. A. (2012). Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proceedings. Biological Sciences / The Royal Society*, 279(1749), 5048–57.
- Korte, A., & Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods*, 9, 29.
- Krahn, M. M., Herman, D. P., Matkin, C. O., Durban, J. W., Barrett-Lennard, L., Burrows, D. G., ... Wade, P. R. (2007). Use of chemical tracers in assessing the diet and foraging regions of eastern North Pacific killer whales. *Marine Environmental Research*, 63(2), 91–114.
- Kronforst, M., Hansen, M. B., Crawford, N., Gallant, J., Zhang, W., Kulathinal, R., ... Mullen, S. (2013). Hybridization Reveals the Evolving Genomic Architecture of Speciation. *Cell Reports*, 5(3), 666–677.
- Kronforst, M. R., & Papa, R. (2015). The functional basis of wing patterning in *Heliconius* butterflies: the molecules behind mimicry. *Genetics*, 200(1), 1–19.
- Laland, K. N., Odling-Smee, J., & Myles, S. (2010). How culture shaped the human genome: bringing genetics and the human sciences together. *Nature Reviews Genetics*, 11(2), 137–148.
- Laland, K. N., Uller, T., Feldman, M. W., Sterelny, K., Müller, G. B., Moczek, A., ... Odling-Smee, J. (2015). The extended evolutionary synthesis: its structure, assumptions and predictions. *Proceedings. Biological Sciences / The Royal Society*, 282(1813), 20151019–.
- Lamichhaney, S., Berglund, J., Almén, M. S., Maqbool, K., Grabherr, M., Martinez-Barrio, A., ... Andersson, L. (2015). Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*, 518(7539), 371–375.
- Langley, C. H., Crepeau, M., Cardeno, C., Corbett-Detig, R., & Stevens, K. (2011). Circumventing heterozygosity: sequencing the amplified genome of a single haploid *Drosophila melanogaster* embryo. *Genetics*, 188(2), 239–46.
- Leffler, E. M., Bullaughey, K., Matute, D. R., Meyer, W. K., Ségurel, L., Venkat, A., ... Przeworski, M. (2012). Revisiting an old riddle: what determines genetic diversity levels within species? *PLoS Biology*, 10(9), e1001388.
- Leigh Van Valen. (1976). Ecological Species, Multispecies, and Oaks
- Lewontin, R. C. (1965). *The genetics of colonizing species: proceedings.*
- Lewontin, R. C., & Krakauer, J. (1973). Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics*, 74(1), 175–95.

- Li, M. J., Wang, L. Y., Xia, Z., Wong, M. P., Sham, P. C., & Wang, J. (2014). dbPSHP: a database of recent positive selection across human populations. *Nucleic Acids Research*, 42(Database issue), D910–6.
- Lindblad-Toh, K., Garber, M., Zuk, O., Lin, M. F., Parker, B. J., Washietl, S., ... Kellis, M. (2011). A high-resolution map of human evolutionary constraint using 29 mammals. *Nature*, 478(7370), 476–82.
- Lindtke, D., & Buerkle, C. A. (2015). The genetic architecture of hybrid incompatibilities and their effect on barriers to introgression in secondary contact. *Evolution; International Journal of Organic Evolution*, 69(8), 1987–2004.
- Linnaeus, C. (1758). *Systema Naturae per regna tria naturae, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis*.
- Londei, T. (2008). The pied crow, *Corvus albus* and Somali crow, *corvus edithae* do not hybridise as soon as they meet. *Scopus*, 27, 1–5.
- Madge, S., & Burn, H. (1994). *Crows and Jays: A Guide to the Crows, Jays and Magpies of the World*. Christopher Helm.
- Maher, B. (2008). Personal genomes: The case of the missing heritability. *Nature*, 456(7218), 18–21.
- Malik, H. S. (2009). The centromere-drive hypothesis: a simple basis for centromere complexity. *Progress in Molecular and Subcellular Biology*, 48, 33–52.
- Mallet, J. (2007). Species, concepts of. In S. A. Levin (Ed.), *Encyclopedia of Biodiversity* (pp. 1–15). Elsevier.
- Mandeville, E. G., Parchman, T. L., McDonald, D. B., & Buerkle, C. A. (2015). Highly variable reproductive isolation among pairs of *Catostomus* species. *Molecular Ecology*, 24(8), 1856–72.
- Martin, A., & Orgogozo, V. (2013). The loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. *Evolution*, 67(5).
- Martin, S. H., Dasmahapatra, K. K., Nadeau, N. J., Salazar, C., Walters, J. R., Simpson, F., ... Jiggins, C. D. (2013). Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Research*, 23(11), 1817–1828.
- Martin, S. H., Dasmahapatra, K. K., Nadeau, N. J., Salazar, C., Walters, J. R., Simpson, F., ... Jiggins, C. D. (2013). Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Research*.
- Matlin, A. J., Clark, F., & Smith, C. W. J. (2005). Understanding alternative splicing: towards a cellular code. *Nature Reviews. Molecular Cell Biology*, 6(5), 386–98.
- Matthews, C. J. D., & Ferguson, S. H. (2013). Spatial segregation and similar trophic-level diet among eastern Canadian Arctic/north-west Atlantic killer whales inferred from bulk and compound specific isotopic analysis. *Journal of the Marine Biological Association of the United Kingdom*, 94(06), 1343–1355.
- Mayr, E. (1942). *Systematics and the Origin of Species, from the Viewpoint of a Zoologist*. Harvard University Press.
- Mayr, E. (1954). *Change of genetic environment and evolution*. In *Evolution as a Process*. (F. E. Huxley, J. Hardy, Ed.). Princeton, NJ: Princeton University Press.
- McGee, M. D., Neches, R. Y., & Seehausen, O. (2015). Evaluating genomic divergence and parallelism in replicate ecomorphs from young and old cichlid adaptive radiations. *Molecular Ecology*.
- McGowen, M. R., Grossman, L. I., & Wildman, D. E. (2012). Dolphin genome provides evidence for adaptive evolution of nervous system genes and a molecular rate slowdown. *Proceedings. Biological Sciences / The Royal Society*, 279(1743), 3643–51.

- Meierjohann, S., Schartl, M., & Volff, J.-N. (2004). Genetic, biochemical and evolutionary facets of Xmrk-induced melanoma formation in the fish *Xiphophorus*. *Comparative Biochemistry and Physiology. Toxicology & Pharmacology: CBP*, 138(3), 281–9.
- Meise, W. (1928). *Die Verbreitung der Aaskr{ä}he (Formenkreis Corvus corone L.)*. Friedlaender in Komm.
- Miyata, T., Yasunaga, T., & Nishida, T. (1980). Nucleotide sequence divergence and functional constraint in mRNA evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 77(12), 7328–32.
- Moebius, W., Murray, A. W., & Nelson, D. R. (2015). How obstacles perturb population fronts and alter their genetic structure. *bioRxiv*, 021964.
- Moreno, E. (2012). Design and construction of “synthetic species”. *PLoS One*, 7(7), e39054.
- Moyle, L. C., Muir, C. D., Han, M. V., & Hahn, M. W. (2010). The contribution of gene movement to the “two rules of speciation”. *Evolution; International Journal of Organic Evolution*, 64(6), 1541–57.
- Muller, H. J. (1942). Isolating mechanisms, evolution and temperature. In *Biol. Symp* (Vol. 6, pp. 71–125).
- N. H. Barton, B. C. (1984). Genetic Revolutions, Founder Effects, and Speciation. *Annual Review of Ecology and Systematics*, 15, 133–164.
- Nachman, I., Regev, A., & Friedman, N. (2004). Inferring quantitative models of regulatory networks from expression data. *Bioinformatics (Oxford, England)*, 20 Suppl 1(suppl_1), i248–56.
- Nadeau, N. J., Ruiz, M., Salazar, P., Counterman, B., Medina, J. A., Ortiz-Zuazaga, H., ... Papa, R. (2014). Population genomics of parallel hybrid zones in the mimetic butterflies, *H. melpomene* and *H. erato*. *Genome Research*, 24(8), 1316–33.
- Nagarajan, N., & Pop, M. (2013). Sequence assembly demystified. *Nature Reviews. Genetics*, 14(3), 157–67.
- Nagarajan, N., Read, T. D., & Pop, M. (2008). Scaffolding and validation of bacterial genome assemblies using optical restriction maps. *Bioinformatics (Oxford, England)*, 24(10), 1229–35.
- Nei, M., & Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.*, 3(5), 418–426.
- Nicola Saino, S. V. (1992). Pair Composition and Reproductive Success across a Hybrid Zone of Carrion Crows and Hooded Crows. *The Auk*, 109(3), 543–555.
- Noor, M. A. F., & Bennett, S. M. (2009). Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity*, 103(6), 439–44.
- Noor, M. A. F., & Feder, J. L. (2006). Speciation genetics: evolving approaches. *Nature Reviews. Genetics*, 7(11), 851–61.
- Nosil, P., & Feder, J. L. (2013). Genome Evolution and Speciation: Toward Quantitative Descriptions of Pattern and Process. *Evolution*, 67(9), 2461–2467.
- Nosil, P., Funk, D. J., & Ortiz-Barrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, 18(3), 375–402.
- Nosil, P., & Schluter, D. (2011). The genes underlying the process of speciation. *Trends in Ecology & Evolution*, 26(4), 160–7.
- Parker, J., Tsagkogeorga, G., Cotton, J. A., Liu, Y., Provero, P., Stupka, E., & Rosstiter, S. J. (2013). Genome-wide signatures of convergent evolution in echolocating mammals. *Nature*, 502(7470), 228–31.

- Pease, J. B., & Hahn, M. W. (2015). Detection and Polarization of Introgression in a Five-Taxon Phylogeny. *Systematic Biology*, 64(4), 651–62.
- Pitman, R. L., & Durban, J. W. (2012). Cooperative hunting behavior, prey selectivity and prey handling by pack ice killer whales (*Orcinus orca*), type B, in Antarctic Peninsula waters. *Marine Mammal Science*, 28(1), 16–36.
- Pitman, R. L., & Ensor, P. (2003). Three forms of killer whales (*Orcinus orca*) in Antarctic waters. *J. CETACEAN RES. MANAGE*, 5(2), 131–139.
- Poelstra, J. W., Ellegren, H., & Wolf, J. B. W. (2013). An extensive candidate gene approach to speciation: diversity, divergence and linkage disequilibrium in candidate pigmentation genes across the European crow hybrid zone. *Heredity*, 111(6), 467–73.
- Poelstra, J. W., & Vijay, N. (2014). The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science*, 344(6190), 1410–1414.
- Poelstra, J. W., Vijay, N., Bossu, C. M., Lantz, H., Ryll, B., Müller, I., ... Wolf, J. B. W. (2014). The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science (New York, N.Y.)*, 344(6190), 1410–4.
- Poelstra, J. W., Vijay, N., Hoepfner, M. P., & Wolf, J. B. W. (2015). Transcriptomics of colour patterning and colouration shifts in crows. *Molecular Ecology*.
- Presgraves, D. C. (2003). A fine-scale genetic analysis of hybrid incompatibilities in *Drosophila*. *Genetics*, 163(3), 955–72.
- Presgraves, D. C. (2010). The molecular evolutionary basis of species formation. *Nature Reviews. Genetics*, 11(3), 175–80.
- Proux, E., Studer, R. A., Moretti, S., & Robinson-Rechavi, M. (2009). Selectome: a database of positive selection. *Nucleic Acids Research*, 37(Database issue), D404–7.
- Provine, W. B. (2004). Ernst Mayr: Genetics and Speciation. *Genetics*, 167(3), 1041–1046.
- Rabosky, D. L., & Matute, D. R. (2013). Macroevolutionary speciation rates are decoupled from the evolution of intrinsic reproductive isolation in *Drosophila* and birds. *Proceedings of the National Academy of Sciences of the United States of America*, 110(38), 15354–9.
- Renaut, S., Grassa, C. J., Yeaman, S., Moyers, B. T., Lai, Z., Kane, N. C., ... Rieseberg, L. H. (2013). Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nature Communications*, 4, 1827.
- Riesch, R., Barrett-Lennard, L. G., Ellis, G. M., Ford, J. K. B., & Deecke, V. B. (2012). Cultural traditions and the evolution of reproductive isolation: ecological speciation in killer whales? *Biological Journal of the Linnean Society*, 106(1), 1–17.
- Ritchie, M. G. (2007). Sexual Selection and Speciation. *Annual Review of Ecology, Evolution, and Systematics*, 38(1), 79–102.
- Robin, V. V., Sinha, A., & Ramakrishnan, U. (2010). Ancient geographical gaps and paleo-climate shape the phylogeography of an endemic bird in the sky islands of southern India. *PLoS One*, 5(10), e13321.
- Rockman, M. V. (2012). The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution; International Journal of Organic Evolution*, 66(1), 1–17.
- Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z. P., Richter, D. J., Schaffner, S. F., ... Lander, E. S. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature*, 419(6909), 832–7.
- Saino, N. (1992). Selection of Foraging Habitat and Flocking by Crow *Corvus corone* Phenotypes in a Hybrid Zone. *Ornis Scandinavica*, 23(2), 111–120.

- Saino, N., Lorenzini, R., Fusco, G., & Randi, E. (1992). Genetic variability in a hybrid zone between carrion and hooded crows (*Corvus corone corone* and *C. c. cornix*, Passeriformes, Aves) in North-Western Italy. *Biochemical Systematics and Ecology*, *20*(7), 605–613.
- Saino, N., & Scatizzi, L. (1991). Selective aggressiveness and dominance among carrion crows, hooded crows and hybrids. *Bolletino Di Zoologia*, *58*(3), 255–260.
- Salzberg, S. L., Phillippy, A. M., Zimin, A., Puiu, D., Magoc, T., Koren, S., ... Yorke, J. A. (2012). GAGE: A critical evaluation of genome assemblies and assembly algorithms. *Genome Research*, *22*(3), 557–67.
- Sankararaman, S., Mallick, S., Dannemann, M., Prüfer, K., Kelso, J., Pääbo, S., ... Reich, D. (2014). The genomic landscape of Neanderthal ancestry in present-day humans. *Nature*, *507*(7492), 354–7.
- Saulitis, E., Matkin, C., Barrett-Lennard, L., Heise, K., & Ellis, G. (2000). Foraging Strategies of Sympatric Killer Whale (*ORCINUS ORCA*) populations in Prince William sound, Alaska. *Marine Mammal Science*, *16*(1), 94–109.
- Schatz, M. (2015). The next 20 years of genome research. *bioRxiv*, 020289.
- Schluter, D., & Conte, G. L. (2009). Genetics and ecological speciation. *Proceedings of the National Academy of Sciences of the United States of America*, *106 Suppl* (Supplement_1), 9955–62.
- Schneider, A., Suvorov, A., Sabath, N., Landan, G., Gonnet, G. H., & Graur, D. (2009). Estimates of positive Darwinian selection are inflated by errors in sequencing, annotation, and alignment. *Genome Biology and Evolution*, *1*(0), 114–8.
- Seddon, N., Botero, C. A., Tobias, J. A., Dunn, P. O., Macgregor, H. E. A., Rubenstein, D. R., ... Safran, R. J. (2013). Sexual selection accelerates signal evolution during speciation in birds. *Proceedings. Biological Sciences / The Royal Society*, *280*(1766), 20131065.
- Seed, A. M., Clayton, N. S., & Emery, N. J. (2008). Cooperative problem solving in rooks (*Corvus frugilegus*). *Proceedings. Biological Sciences / The Royal Society*, *275*(1641), 1421–9.
- Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., ... Widmer, A. (2014). Genomics and the origin of species. *Nature Reviews. Genetics*, *15*(3), 176–92.
- Servedio, M. R., Van Doorn, G. S., Kopp, M., Frame, A. M., & Nosil, P. (2011). Magic traits in speciation: “magic” but not rare? *Trends in Ecology & Evolution*, *26*(8), 389–97.
- Seymour, D. K., Koenig, D., Hagmann, J., Becker, C., & Weigel, D. (2014). Evolution of DNA methylation patterns in the Brassicaceae is driven by differences in genome organization. *PLoS Genetics*, *10*(11), e1004785.
- Shafer, A. B. A., & Wolf, J. B. W. (2013). Widespread evidence for incipient ecological speciation: a meta-analysis of isolation-by-ecology. *Ecology Letters*, *16*(7), 940–50.
- Shafer, A. B. A., Wolf, J. B. W., Alves, P. C., Bergström, L., Bruford, M. W., Brännström, I., ... Zieliński, P. (2014). Genomics and the challenging translation into conservation practice. *Trends in Ecology & Evolution*, *30*(2), 78–87.
- Shaw, K. L., Ellison, C. K., Oh, K. P., & Wiley, C. (2011). Pleiotropy, “sexy” traits, and speciation. *Behavioral Ecology*, *22*(6), 1154–1155.
- Shaw, K. L., & Mullen, S. P. (2014). Speciation continuum. *The Journal of Heredity*, *105 Suppl* (S1), 741–2.
- Shendure, J., & Ji, H. (2008). Next-generation DNA sequencing. *Nature Biotechnology*, *26*(10), 1135–45.

- Smukowski, C. S., & Noor, M. A. F. (2011). Recombination rate variation in closely related species. *Heredity*, *107*(6), 496–508.
- Soria-Carrasco, V., Gompert, Z., Comeault, A. A., Farkas, T. E., Parchman, T. L., Johnston, J. S., ... Nosil, P. (2014). Stick insect genomes reveal natural selection's role in parallel speciation. *Science (New York, N.Y.)*, *344*(6185), 738–42.
- Stern, D. L. (2013). The genetic causes of convergent evolution. *Nature Reviews Genetics*, *14*(11), 751–64.
- Stern, D. L., & Orgogozo, V. (2009). Is genetic evolution predictable? *Science (New York, N.Y.)*, *323*(5915), 746–51.
- Stewart, C. B., Schilling, J. W., & Wilson, A. C. (1987). Adaptive evolution in the stomach lysozymes of foregut fermenters. *Nature*, *330*(6146), 401–4.
- Swanson, W. J., Clark, A. G., Waldrip-Dail, H. M., Wolfner, M. F., & Aquadro, C. F. (2001). Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(13), 7375–9.
- Talavera, D., Lovell, S. C., & Whelan, S. (2015). Covariation is a poor measure of molecular coevolution. *Molecular Biology and Evolution*, msv109–.
- Taylor, A. H., Hunt, G. R., Holzhaider, J. C., & Gray, R. D. (2007). Spontaneous metatool use by New Caledonian crows. *Current Biology : CB*, *17*(17), 1504–7.
- Ten Tusscher, K. H. W. J., & Hogeweg, P. (2009). The role of genome and gene regulatory network canalization in the evolution of multi-trait polymorphisms and sympatric speciation. *BMC Evolutionary Biology*, *9*(1), 159.
- Toews, D. P. L., Campagna, L., Taylor, S. A., Balakrishnan, C. N., Baldassarre, D. T., Deane-Coe, P. E., ... Winger, B. M. (2015). Genomic approaches to understanding population divergence and speciation in birds.
- Tomback, D. F., & Baker, M. C. (1984). Assortative mating by white-crowned sparrows at song dialect boundaries. *Animal Behaviour*, *32*(2), 465–469.
- Turelli, M., & Orr, H. A. (1995). The dominance theory of Haldane's rule. *Genetics*, *140*(1), 389–402.
- Turelli, M., & Orr, H. A. (2000). Dominance, Epistasis and the Genetics of Postzygotic Isolation. *Genetics*, *154*(4), 1663–1679.
- Turner, L. M., & Harr, B. (2014). Genome-wide mapping in a house mouse hybrid zone reveals hybrid sterility loci and Dobzhansky-Muller interactions. *eLife*, *3*.
- Turner, T. L., Hahn, M. W., & Nuzhdin, S. V. (2005). Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biology*, *3*(9), e285.
- Utturkar, S. M., Klingeman, D. M., Land, M. L., Schadt, C. W., Doktycz, M. J., Pelletier, D. A., & Brown, S. D. (2014). Evaluation and validation of de novo and hybrid assembly techniques to derive high-quality genome sequences. *Bioinformatics (Oxford, England)*, *30*(19), 2709–16.
- van Doorn, G. S., Edelaar, P., & Weissing, F. J. (2009). On the origin of species by natural and sexual selection. *Science (New York, N.Y.)*, *326*(5960), 1704–7.
- Vernot, B., & Akey, J. M. (2014). Resurrecting surviving Neandertal lineages from modern human genomes. *Science (New York, N.Y.)*, *343*(6174), 1017–21.
- Verster, A. J., Ramani, A. K., McKay, S. J., & Fraser, A. G. (2014). Comparative RNAi screens in *C. elegans* and *C. briggsae* reveal the impact of developmental system drift on gene function. *PLoS Genetics*, *10*(2), e1004077.
- Via, S. (2009). Natural selection in action during speciation. *Proceedings of the National Academy of Sciences of the United States of America*, *106* Suppl , 9939–46.
- Via, S. (2012). Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*(1587), 451–60.

- Vijay, N., Bossu, C. M., Poelstra, J. W., Weissensteiner, M., Suh, A., Kryukov, A., & Wolf, J. B. (2015). Idiosyncratic patterns and processes of genomic differentiation in replicate contact zones in crows.
- Vijay, N., Poelstra, J. W., Künstner, A., & Wolf, J. B. W. (2013). Challenges and strategies in transcriptome assembly and differential gene expression quantification. A comprehensive in silico assessment of RNA-seq experiments. In *Molecular Ecology* (Vol. 22, pp. 620–634).
- Wagner, C. E., Harmon, L. J., & Seehausen, O. (2012). Ecological opportunity and sexual selection together predict adaptive radiation. *Nature*, 487(7407), 366–9.
- Wang, I. J., & Bradburd, G. S. (2014). Isolation by Environment. *Molecular Ecology*, 23(23).
- Wang, I. J., & Summers, K. (2010). Genetic structure is correlated with phenotypic divergence rather than geographic isolation in the highly polymorphic strawberry poison-dart frog. *Molecular Ecology*, 19(3), 447–458.
- Wang, J., Street, N. R., Scofield, D. G., & Ingvarsson, P. (2015). *Hard, soft and just right: variations in linked selection and recombination drive genomic divergence during speciation of aspens*. *bioRxiv*. Cold Spring Harbor Labs Journals.
- Warren, D. L., Cardillo, M., Rosauer, D. F., & Bolnick, D. I. (2014). Mistaking geography for biology: inferring processes from species distributions. *Trends in Ecology & Evolution*, 29(10), 572–580.
- Weinreich, D. M. (2006). Darwinian Evolution Can Follow Only Very Few Mutational Paths to Fitter Proteins. *Science*, 312(5770), 111–114.
- White, B. J., Lawniczak, M. K. N., Cheng, C., Coulibaly, M. B., Wilson, M. D., Sagnon, N., ... Besansky, N. J. (2011). Adaptive divergence between incipient species of *Anopheles gambiae* increases resistance to *Plasmodium*. *Proceedings of the National Academy of Sciences of the United States of America*, 108(1), 244–9.
- Whitehead, H. (1998). Cultural Selection and Genetic Diversity in Matrilineal Whales. *Science*, 282(5394), 1708–1711.
- Wiens, J. J., & Graham, C. H. (2005). Niche Conservatism: Integrating Evolution, Ecology, and Conservation Biology. *Annual Review of Ecology, Evolution, and Systematics*, 36(1), 519–539.
- Wierer, M., Schrey, A. K., Kühne, R., Ulbrich, S. E., & Meyer, H. H. D. (2012). A single glycine-alanine exchange directs ligand specificity of the elephant prostin receptor. *PLoS One*, 7(11), e50350.
- Wittbrodt, J., Adam, D., Malitschek, B., Mäueler, W., Raulf, F., Telling, A., ... Schartl, M. (1989). Novel putative receptor tyrosine kinase encoded by the melanoma-inducing Tu locus in *Xiphophorus*. *Nature*, 341(6241), 415–21.
- Wolf, J. B. W., Bayer, T., Haubold, B., Schilhabel, M., Rosenstiel, P., & Tautz, D. (2010). Nucleotide divergence vs. gene expression differentiation: comparative transcriptome sequencing in natural isolates from the carrion crow and its hybrid zone with the hooded crow. *Molecular Ecology*, 19 Suppl 1, 162–75.
- Wolf, J. B. W., Lindell, J., & Backström, N. (2010). Speciation genetics: current status and evolving approaches. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 365(1547), 1717–33.
- Wright, S. (1943). Isolation by Distance. *Genetics*, 28(2), 114–38.
- Wu, C.-I., & Ting, C.-T. (2004). Genes and speciation. *Nature Reviews. Genetics*, 5(2), 114–22.
- Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, 24(8), 1586–91.
- Yeaman, S. (2015). Local Adaptation by Alleles of Small Effect*. *The American Naturalist*, S000–S000.

- Young, A. L., Abaan, H. O., Zerbino, D., Mullikin, J. C., Birney, E., & Margulies, E. H. (2010). A new strategy for genome assembly using short sequence reads and reduced representation libraries. *Genome Research*, *20*(2), 249–56.
- Zamani, N., Russell, P., Lantz, H., Hoepfner, M. P., Meadows, J. R., Vijay, N., ... Grabherr, M. G. (2013). Unsupervised genome-wide recognition of local relationship patterns. *BMC Genomics*, *14*, 347.
- Zhang, Q., Hill, G. E., Edwards, S. V., & Backström, N. (2014). A house finch (*Haemorrhous mexicanus*) spleen transcriptome reveals intra- and interspecific patterns of gene expression, alternative splicing and genetic diversity in passerines. *BMC Genomics*, *15*(1), 305.
- Zhang, W., Chen, J., Yang, Y., Tang, Y., Shang, J., & Shen, B. (2011). A practical comparison of de novo genome assembly software tools for next-generation sequencing technologies. *PloS One*, *6*(3), e17915.
- Zhu, M., & Zhao, S. (2007). Candidate gene identification approach: progress and challenges. *International Journal of Biological Sciences*, *3*(7), 420–7.
- Zou, Z., & Zhang, J. (2015). Are Convergent and Parallel Amino Acid Substitutions in Protein Evolution More Prevalent Than Neutral Expectations? *Molecular Biology and Evolution*, msv091–.

Acta Universitatis Upsaliensis

*Digital Comprehensive Summaries of Uppsala Dissertations
from the Faculty of Science and Technology 1326*

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-265342



ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2016