The Genomics of Hybrid Speciation

Cassandra Nicole Trier

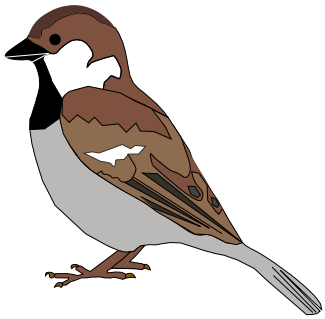
Dissertation presented for the degree of
Philosophiae Doctor (PhD)
2018

Centre for Ecological and Evolutionary Synthesis
Department of Biosciences
Faculty of Mathematics and Natural Sciences
University of Oslo

Series of dissertations submitted to the
Faculty of Mathematics and Natural Sciences, University of Oslo
No. 2020

ISSN 1501-7710

Cover: Hanne Baadsgaard Utigard.
Print production: Reprosentralen, University of Oslo.



Science is magic that works.

-Kurt Vonnegut

Acknowledgements

This PhD has been a crazy journey and the truth is, I would have never made it through without support from so many wonderful people along the way. First of all, I would like to thank my supervisors Glenn-Peter Sætre and Kjetill S. Jakobsen. Glenn, you didn't know what you were signing up for when I knocked on your door as a new masters student looking for a research group! Now, eight years later, I can say that it has been a pleasure being in your research group and I will always be grateful for the guidance and support you have given me through the years. Kjetill, you have always been there if I need advice or just someone to talk to about my project. Your up-beat personality and great sense of humor make you easy to talk to and I always enjoy our chats. Thank you so much for your support.

I would also like to thank my co-authors and friends in the sparrow group for so many great discussions that left me inspired in my own work. In particular, thank you Angelica, Anna R., Camilla, Caroline, Fabrice, Jo, Mark and Melissah for being a constant source of encouragement through the years. You have always made me feel like my opinion is valuable even when I have doubted so myself; it has meant more to me than you know.

The Centre for Ecological and Evolutionary Synthesis (CEES) has been such a fantastic work environment where I have made life-long friends. I feel so fortunate to have worked with such an amazing group of people. Thank you to all my co-workers for the ridiculous lunchtime conversations and making CEES a fun place to work. A special thanks to Anders, Anna M., Bastiaan, Boris, Eric, Heidi, Helle, Inger Maren, Katie, Kjetil, Luis, Ole Kristian, Olja, Pernille, Ryan, Sanne and Unni for the fun that also extended outside of work at barbecues, parties, hikes, kids play-dates, you name it.

To my knitting girls, your love and constant support has meant the world to me. Life is so much easier with friends like you. Thank you for keeping me sane and bringing so much joy and laughter into my life.

I have never stopped missing my friends from California. To Alex, Anna H., Jack, Jiro, Kevin, Nicole, Matt and Tarek, thank you for being life-long friends that keep me grounded and have put up with me all these years!

I would also like to thank all my extended family that held back from teasing me too much about "taking care of the birds" all this time. Thank you for your continuous love and support. A special thank you to Grandpa Tom who always believed in me and I know would be proud.

To Mom and Dad, it has only been with your unconditional love that I had the courage to come to Norway in the first place. You constantly sacrifice so I can have a good life, even when it means your only daughter moves across the world. Thank you for always being there for me, believing in me and pushing me to follow my passion.

To my dear husband, Tore, I'm not sure any words suffice. Not only have you stuck by me as my number one fan throughout the PhD, but you have been my officemate, co-author and best friend. People wonder how I could share an office with my husband and the truth is, it was easy because you and our sons are always the highlight of my day. Thank you for quite simply everything along the way, I couldn't have done it without you.

And last but not least, my sons. Oliver and Eirik, thank you for everyday reminding me of what is really important in life.

Cassandra Trier

Oslo, June 2018

Table of Contents

List of Papers

Paper I.

C. N. Trier*, J. S. Hermansen*, G.-P. Sætre, R. I. Bailey. (2014) Evidence for mito-nuclear and sex-linked reproductive barriers between the hybrid Italian sparrow and its parent species. *PLoS Genetics* **10**:e1004075.

Paper II.

T.O. Elgvin*, **C. N. Trier***, O.K. Tørresen, I. Hagen, S. Lien, M. Ravinet, H. Jensen, G.-P. Sætre. (2017) The genetic mosaicism of hybrid speciation. *Science Advances* **3:**e1602996.

*These authors contributed equally to the paper.

Paper III.

A. Runemark, **C. N. Trier**, F. Eroukhmanoff, J.S. Hermansen, M. Matschiner, M. Ravinet, T.O. Elgvin, G.-P. Sætre. (2018) Variation and constraints in hybrid genome formation. *Nature Ecology and Evolution* **2**:549-556.

Summary

The central question of this thesis is how hybridization can lead to the formation of new species. To approach this question, I used the hybrid Italian sparrow (*Passer italiae*) to explore the consequences of hybridization on genomic architecture and reproductive isolation at the genomic level. With the use of transcriptomic data and cline analyses in **Paper I**, we investigated if there is evidence for reproductive barriers between the hybrid species and its parents, the house sparrow (*P. domesticus*) and Spanish sparrow (*P. hispaniolensis*); and if so, which genes or genomic regions may be involved? We found that there is evidence of reproductive isolation between the Italian sparrow and its parent species and that Z-linked genes and mito-nuclear gene complexes play an integral role.

In **Paper II**, we sought to examine the genomic architecture of the Italian sparrow in comparison to its parents. This first consisted of *de novo* assembling a high-quality reference genome of the house sparrow. We then mapped whole-genome sequencing data from populations of the parent house and Spanish sparrows, as well as the Italian sparrow to the reference genome. By using comparative genomics, we were able to characterize patterns of admixture and differentiation in the Italian sparrow genome in relation to both parent species. We show that the genomic landscape of the Italian sparrow is highly heterogenous with regions inherited alternately from either parent across the genome in a mosaic pattern. High divergence regions between the Italian sparrow and either of its parents were found to be disproportionately located on the Z chromosome and genes involved in body patterning, beak morphology and the immune system were over-represented in these regions. We also found regions where the Italian sparrow is divergent from both parents that may represent areas of novel divergence in this homoploid hybrid lineage.

In **Paper III**, we utilized genomic data from multiple geographically isolated Italian sparrow populations from different Mediterranean islands to explore the extent to which the hybrid genome can vary. We find that there is

variation in the genomic combinations that compose a functional hybrid species, yet there are some areas that are invariably inherited from one parent. These regions of genomic constraint are over-represented on the Z chromosome and hold candidate incompatibility loci involved in DNA repair and mito-nuclear function.

Overall, this dissertation helps demonstrate how a new species can arise via hybridization by painting a picture of how admixture can shape the genomes of differentiated populations and lead to the formation of reproductive barriers.

Introduction

Background

Hybridization is widespread in nature and can have various impacts on species diversification. Traditionally, interbreeding between distinct populations has been viewed as a detriment to species divergence and a 'biological mistake' (Mayr 1932; Fisher 1930). Yet, around 10% of animal and 25% of plant species are known to hybridize (Mallet 2007), making hybridization a prominent feature in nature. Hybridization has even been shown to have shaped the genome of our own species (Sankararaman et al. 2014). In recent years, it has become apparent that hybridization can also serve as a source of genetic variation promoting diversification (Abbott et al. 2013; Seehausen et al. 2014). Since new genetic variation can have adaptive potential, species divergence can alternatively be facilitated through hybridization rather than impeded (Grant and Grant 1994; Heliconius Genome Consortium 2012). In fact, hybridization is particularly common among rapidly radiating groups (Mallet 2007; Grant et al. 2015) as it can allow for rapid local adaption through the introgression of new variation subject to selection (Lamichhaney et al. 2018). In some instances, hybridization can even give rise to new species through the recombination of parental

genomes, leading to a third population of mixed ancestry that remains distinct from both its parents (Rieseberg 1997; Mallet 2007; Mavarez and Linares 2008). Thus, the creative role of hybridization in evolution spans a continuum from adaptive introgression to hybrid speciation. This thesis focuses on the latter and aims at gaining a better understanding of how hybridization can play a primary role in the origin of new species.

Hybrid speciation

When two differentiated populations mate, hybrid offspring possessing novel, mixed genotypes can arise (Rieseberg 1997; Buerkle et al. 2000; Coyne and Orr 2004). These hybrid offspring may have reduced fitness relative to their parent taxa (Mayr 1963; Coyne and Orr 2004) due to genetic incompatibilities leading to inviability or infertility in first generation (F1) hybrids (Arnold and Hodges 1995) or ecological intermediacy (Schluter 1993; 1995; Coyne and Orr 2004). However, empirical studies have shown that hybrid genotypes possess a wide range of fitnesses and that hybrids can have equivalent or higher fitness than their parents (Arnold and Hodges 1995; Arnegard et al. 2014).

Hybridization provides the means in which new genetic variation from multiple loci with adaptive potential are transferred simultaneously. Therefore, adaptive evolution may proceed more rapidly following hybridization than would be expected from mutations alone (Grant and Grant 1994; Heliconius Genome Consortium 2012; Abbott et al. 2016). F1 hybrids can also experience an increase in growth rate, size and reproductive success, known as heterosis or hybrid vigor (Arnold and Hodges 1995), which may constitute an adaptive advantage relative to its parents. Though the genetic basis for hybrid vigor is subject to much debate, recent hypotheses suggest it may be attributed to complementary interactions of alleles at multiple loci (epistasis) (Baack and Rieseberg 2007). Hybrid vigor is often broken down in subsequent generations as recessive alleles become

exposed and parental gene combinations are broken up via recombination (Dobzhansky 1948; Templeton 1981; Felsenstein 1981). Yet, if the hybrid lineage is able to persist and develop reproductive barriers against both its parents, it has the potential to become its own species. This process in which interspecific hybridization gives rise to novel species is known as hybrid speciation.

An important aspect of hybrid speciation is the hybrid karyotype in relation to its parents. Allopolyploid hybrids maintain a different number of chromosome complements than their parents and consequently develop immediate reproductive isolation. This, combined with potential heterosis, may lead to the establishment of new lineages and has frequently been observed in plant taxa (Rieseberg 1997; Mallet 2007; Rieseberg and Willis 2007; Hegarty and Hiscock 2008; Soltis and Soltis 2009). In contrast, homoploid hybrids, which share the same ploidy level as their parents, may struggle to develop reproductive barriers strong enough to remain distinct from their parents (Baack & Rieseberg 2007). For this reason, homoploid hybrid speciation has historically been considered to be a rare outcome of hybridization (Mallet 2007; Schumer et al. 2014), particularly in animals (Mavarez and Linares 2008). The hybrid must first escape genetic incompatibilities and fitness loss, and then the homogenizing effect of gene flow from its parents; despite complementary ploidy levels and the fact that their parent's reproductive barriers were sufficiently weak that they interbred in the first place (Buerkle et al. 2000; Coyne and Orr 2004). In some instances however, the recombination of parental alleles from initial hybridization could trigger the formation of reproductive barriers between the parents via ecological divergence (Gross and Rieseberg 2005), assortative mating (Mavarez et al. 2006; Melo et al. 2009) or genetic incompatibilities (Rieseberg 1997; Schumer et al. 2014; Abbott et al. 2016). In particular, one mechanism predicted to allow for homoploid hybrid speciation is transgressive segregation - the production of hybrid traits outside the range of its parents - that enable the hybrid to colonize new ecological niches (Rieseberg et al. 1999; Mallet 2007). While there are few well-documented cases of homoploid hybrid animal species, new methods have led to a growing

number of empirical examples in the past decade in flies (Schwarz et al. 2005), bats (Larsen et al. 2010), butterflies (Mavarez and Gonzalez 2006; Kunte et al. 2011; Heliconius Genome Consortium 2012), birds (Elgvin et al. 2011; Hermansen et al. 2011) and fishes (Salzburger et al. 2002; Meyer et al. 2006; Keller et al. 2013) suggesting that homoploid hybrid speciation may be more common than initially thought (Mavarez and Linares 2008; Abbott et al. 2013).

Detecting hybrid origin

For many years, conflicting phylogenetic trees from nuclear and organellar markers have been used to test for hybrid ancestry in proposed hybrids (Bullini 1994; Dowling and Secor 1997; Soltis and Soltis 2009). In the past, this has proven problematic when few markers are used because incomplete lineage sorting can also produce discordant phylogenies (reviewed in Ballard and Rand 2005). Genetic mosaicism, where there is evidence of alternating inheritance from two parental lineages, has been considered strong evidence for hybrid speciation, but it is also difficult to demonstrate with few markers since hybridization upon secondary contact can give the same signal (vonHoldt et al. 2011; Schumer et al. 2014). As data sets have gotten bigger and genome-wide markers or whole genome sequencing are becoming commonplace, this has allowed for more robustly testing of hybrid ancestry hypotheses. However, it has also raised new questions in regards to what criteria are needed to determine if a taxon is derived from hybrid speciation.

As the scale of genetic data has gotten larger, it has become increasingly apparent that hybridization is a prominent feature in nature (Mallet 2007; Abbott et al. 2013) and the number of cases of proposed hybrid species has increased dramatically (Schumer et al. 2013; 2014). This has led to some confusion in the literature as to what constitutes 'hybrid speciation' (Mallet 2007; Mavarez and Linares 2008; Jiggins et al. 2008;

Abbott et al. 2013; 2016). Schumer *et al.* 2014, argued that many purported examples of hybrid speciation do not have strong enough evidence and proposed criteria required in order to conclude that a species is product of hybrid speciation. They suggest there needs to be evidence of i) reproductive isolation between the hybrid lineage from its parents, ii) signatures of hybridization in the genome, and iii) reproductive isolation being a direct product of hybridization. Since speciation is a continuum and defining the point at which two divergent populations become a species has proven difficult (Mallet 1995), it should come as no surprise that there is a debate on what evidence is required for a species to be considered of hybrid origin. In particular, Feliner *et al.* 2017 have criticized Schumer *et al.* 2014, specifically criterion iii, arguing that if a hybridization led to an established, ecologically and morphologically distinct hybrid lineage, this should be considered hybrid speciation regardless of whether or not hybridization directly led to reproductive isolation.

There does however appear to be some agreement that support for hybrid speciation should include evidence of both hybrid ancestry and reproductive isolation from the parental lineages. Genomic data provides great opportunities to investigate these aspects of proposed hybrid species. For instance, since many more genomic regions can be sampled, the chances of identifying the areas involved in reproductive isolation are higher (Schumer et al. 2014). Also, large-scale signatures of hybridization such as genomic mosaicism can help distinguish between patterns of hybridization from those driven by incomplete lineage sorting, genetic drift or selection (Nice et al. 2013; Schumer et al. 2014). While there is certainly a gray area in what is considered hybrid speciation, what is perhaps more fruitful than squabbling over its definition is using newly available genomic data to better understand the processes involved in hybridization leading to the formation of new species. In this thesis, the goal was to do just that; use genomic data to explore how hybridization can result in a new, distinct and reproductively isolated lineage.

Hybrid zones

Hybrid zones are regions where "genetically distinct groups of individuals meet and mate, resulting in at least some offspring of mixed ancestry" (Harrison 1993). They provide an excellent opportunity to study reproductive barriers as they are in effect natural laboratories where parental ancestry blocks are broken down and new combinations of genes are exposed to selection (Barton and Hewitt 1989; Buerkle et al. 2000; Payseur and Nachman 2005; Harrison and Larson 2014).

Postzygotic reproductive barriers between a hybrid and its parents are often the result of Bateson-Dobzhansky-Muller (BDM) incompatibilities (Bateson 1909, Dobzhansky 1936; Muller 1940; Orr 1996; Turelli and Orr 1995). Under a BDM model, hybrids may receive alleles from two populations that have diverged at different loci without suffering a loss in fitness in either population, but when these divergent alleles introgress into a new genomic background, they may interact poorly resulting in a fitness loss (Figure 1). This could be due to selection against heterozygotes, selection against certain alleles in a foreign genetic background or a combination of the two (Barton 2001; Buerkle and Rieseberg 2001; Baack and Rieseberg 2007). The rate of gene flow of an allelic variant in a hybrid zone is consequently determined by its effect on fitness and linkage disequilibrium to other genomic regions (Gompert and Buerkle 2011b). This is true for all loci experiencing divergent selection in the hybrid zone and is expected to be especially strong for BDM incompatibilities as they often have large fitness effects. Therefore, hybrid zone analyses provide an opportunity for studying genomic regions that may drive speciation.

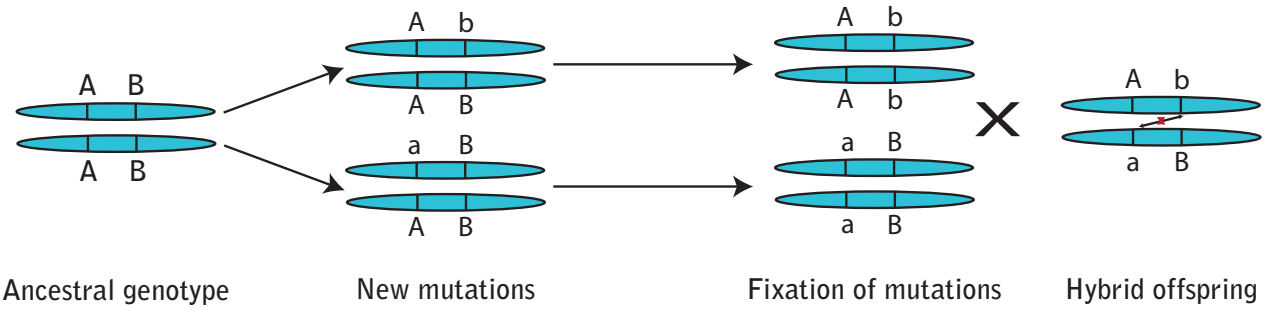


Figure 1. BDM model of genetic incompatibilities. After populations of the ancestral genotype split, new mutations arise and reach fixation without a fitness loss. When the divergent populations meet again and mate, new combinations of incompatible alleles result in a fitness loss in the hybrid offspring.

By examining patterns of differentiation of loci across hybrid zones, it is possible to identify loci that exhibit reduced introgression and consequently are candidates for being involved in reproductive barriers between species (Barton and Hewitt 1985; Gompert and Buerkle 2011a; Gompert et al. 2012). One manner of quantifying differential rates of introgression among loci in hybrid zones is through the use of cline theory. Cline theory measures the gradient in a trait or allele frequency across a geographic or genomic range. In terms of gene flow, cline theory can be used to measure rates of gene flow across a geographic range (geographic clines) (Szymura and Barton 1986; Carling and Brumfield 2008; Teeter et al. 2010; Taylor et al. 2012) or into different genomic backgrounds (genomic clines) (Szymura and Barton 1986; Gompert and Buerkle 2011b; Taylor et al. 2014). Cline theory predicts that the width of a cline is dependent on a balance between selection and dispersal (Slatkin 1973; Barton and Hewitt 1985). Neutral alleles are predicted to introgress proportionally to dispersal distance, i.e. the gene flow out of the hybrid zone. Meanwhile, alleles that reduce fitness or contribute to assortative mating are expected to introgress less due to strong selection creating narrow clines (Figure 2) (Barton and Hewitt 1989; Harrison 1993; Buerkle and Lexer 2008; Gompert and Buerkle 2011b). If partial reproductive isolation exists between two taxa at a given locus, steep clines in allele frequencies indicative of reduced introgression would be expected in hybrid zones (Slatkin 1973; Nagylaki 1975; May et al. 2015).

In the context of hybrid speciation, advantageous alleles are expected to spread rapidly and be driven to fixation in the hybrid lineage while BDM incompatibilities are expected to be purged via recombination (Buerkle and Rieseberg 2008). Once the hybrid lineage diverges from one or both of the parents in areas of the genome, novel incompatibility factors may then arise if the hybrid backcrosses with its parent taxa (Buerkle and Rieseberg 2008). Only a few generations of recombination may be sufficient in creating a high fitness hybrid lineage with reproductive isolation from its parents (Buerkle et al. 2000; Buerkle and Rieseberg 2008). Detection of loci experiencing reduced introgression between a hybrid and its parental lineages in hybrid zones can therefore provide insight into how reproductive barriers form in the hybrid speciation process.

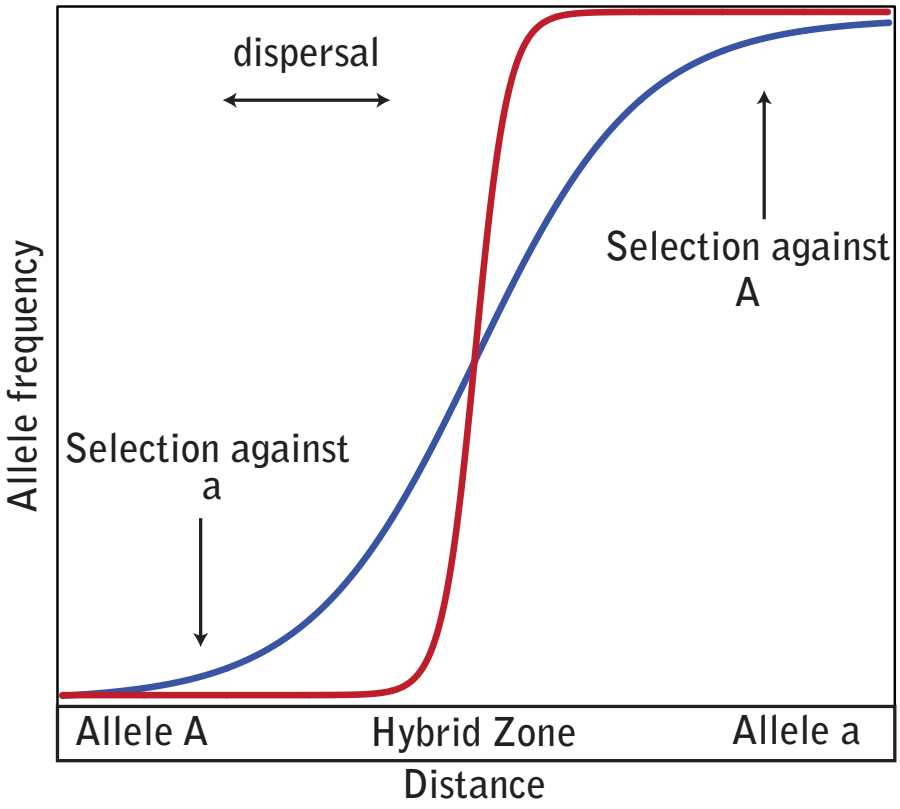


Figure 2. Example of geographic clines. Two populations fixed for different alleles hybridize creating steep (red) and shallow (blue) clines in allele frequencies depending on the strength of selection in relation to dispersal.

The genomic architecture of hybrid speciation

Genomic analyses have consistently shown that there is variation in the permeability of foreign alleles across the genomic landscape (Payseur and Nachman 2005; Baack and Rieseberg 2007; Harrison and Larson 2014; Payseur and Rieseberg 2016). The term 'genomic islands' has become commonly used to refer to regions where gene flow is restricted and thus divergence between lineages is relatively high (Turner et al. 2005). Genomic islands are suggested to develop through selective sweeps of favorable variants along with physically linked surrounding neutral variation (Via and West 2008) while the homogenizing effect of gene flow reduces differentiation in the rest of the genome (Burri et al. 2015). The sizes of these regions is expected to be influenced by selection and the rate of recombination (Nachman and Payseur 2012; Samuk et al. 2017; Ravinet et al. 2017) and they vary largely between hybridizing species pairs. For example, some hybridizing species have only small regions of divergence (Good et al. 2015; Toews et al. 2016) while others have substantial areas of genome-wide divergence (Ellegren et al. 2012; Parchman 2013). Identifying and characterizing high divergence areas, or 'islands' where introgression is reduced can therefore aid in understanding which genome regions may be involved in the maintenance of species barriers.

In particular, sex chromosomes have been considered 'hot spots' of species divergence as numerous studies have found higher differentiation and reduced introgression on sex-linked loci (Macholán et al. 2007; Teeter et al. 2010; Ellegren et al. 2012; Carneiro et al. 2013; Taylor et al. 2014). This pattern is likely attributed to i) a smaller effective population size, ii) lower recombination rates, iii) exposure of recessive alleles to selection in the heterogametic sex, and iv) non-random accumulation of genes involved in sex and reproduction (Charlesworth et al. 1987; Qvarnström and Bailey 2008; Mank et al. 2010). For these reasons, sex chromosomes are often viewed as farther along in the speciation continuum than the rest of the genome. An example in hybridizing birds are collared (*Ficedula albicollis*) and pied

(*Ficedula hypoleuca*) flycatchers. They exhibit higher species differentiation on sex chromosomes than autosomes (Ellegren et al. 2012) and evidence has shown that both male traits under sexual selection and female preference are sex-linked, thus accelerating the speciation process on sex chromosomes (Sæther et al. 2007).

As there are few well-documented cases of homoploid hybrid species, little is known about the genomics of this form of speciation including the nature of reproductive barriers and how genomic mosaicism evolves. One proposed mechanism for driving hybrid incompatibilities and species divergence is through mito-nuclear incompatibilities (Bar-Yaacov et al. 2015; Hill 2017). The mitochondrial genome is maternally inherited and responsible for energy production via oxidative phosphorylation in animals. It is also non-recombining with a mutation rate an order of magnitude higher than the nuclear genome, yet both nuclear and mitochondrial genomes are responsible for mitochondrial function (Rand et al. 2004). Mito-nuclear dysfunction is expected to have strong effects on hybrid fitness as it has been shown to affect (among other things) aging and fertility (Camus et al. 2015; Patel et al. 2016). Therefore, strongly selected upon co-adapted mito-nuclear gene complexes are strong candidates for BDM incompatibilities and may act as potent reproductive barriers during hybridization and speciation (Hill 2017).

With whole genome sequencing becoming increasingly more affordable, new opportunities have arisen to study hybrid speciation through patterns of introgression on a genomic scale. The examination of differentiation and admixture between populations across the genome can shed light on the evolutionary processes involved in hybrid speciation (Payseur and Rieseberg 2016). In this thesis, I use genomic data on a proposed hybrid species to examine its genetic composition in relation to its parents and help elucidate how a new species was formed as a product of hybridization.

The *Passer* sparrow system

The Italian sparrow represents one of the few examples of vertebrate species shown to be of hybrid origin (Elgvin et al. 2011; Hermansen et al. 2011). Found on the Italian peninsula and some Mediterranean islands (Figure 3), the Italian sparrow was for many years suggested to have arisen through hybridization between the Spanish sparrow and the house sparrow (Summers-Smith 1988; Töpfer 2006) based on its intermediate male plumage coloration (Figure 3). More recently, genetic studies have provided support for this hypothesis by demonstrating that the Italian sparrow is genetically mosaic with genes on the Z chromosome nearly fixed for alternate parent's alleles as well as mitochondrial DNA (mtDNA) nearly fixed for house sparrow inheritance (Elgvin et al. 2011; Hermansen et al. 2011). However, these studies were preformed with small sets of genetic markers with limited power in detecting hybrid ancestry.

One proposed scenario for the origin of the Italian sparrow is that as the human commensal house sparrow's range expanded throughout Europe <10,000 years ago alongside the spread of agriculture during the Neolithic revolution, it encountered and hybridized with the Spanish sparrow which is believed to have already been present in the Mediterranean (Hermansen et al. 2011; Sætre et al. 2012; Ravinet et al. 2018). A challenge in studying hybrid species systems can be the lack of geographic overlap between the species, which makes it more difficult to investigate reproductive barriers. Yet, the Italian sparrow's distribution overlaps with those of its parents as it encounters and occasionally hybridizes with the house sparrow in narrow contact zones in the Alps (Hermansen et al. 2011) and lives in sympatry with the Spanish sparrow in a small, recently established contact zone on the Gargano peninsula of southeast Italy (Figure 3). Furthermore, morphologically divergent Italian sparrow populations are also present on the islands of Corsica, Crete, Sicily and Malta (Figure 3). This unique sparrow system, combined with modern genomic tools, therefore provides an excellent opportunity to study how reproductive barriers have evolved in a

hybrid species and this is where my PhD research begins.

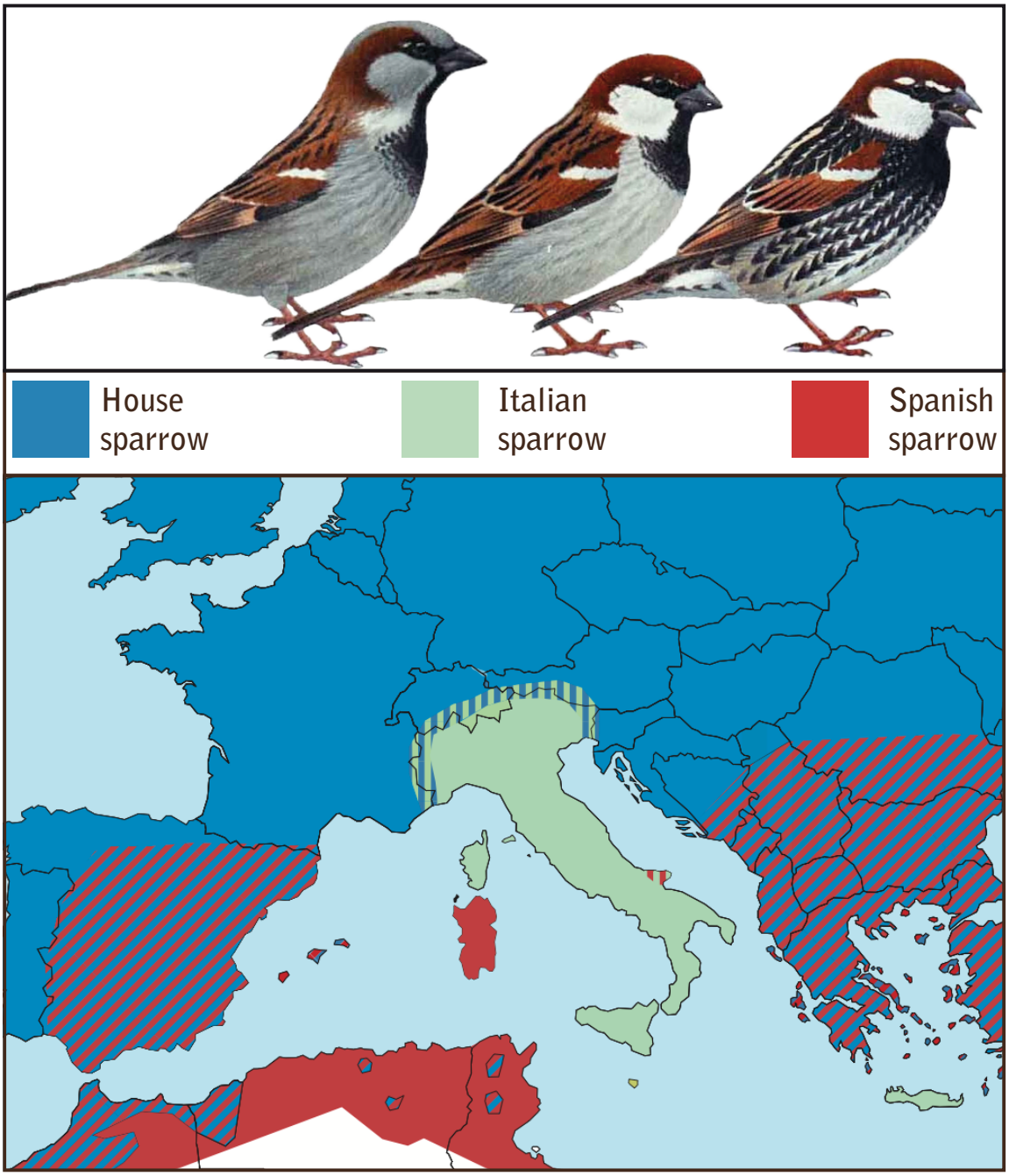


Figure 3. Distribution map and male plumage. Top: Illustrations of male plumage patterns in house (left), Italian (center) and Spanish (right) sparrows modified from Svensson *et al.* 1999. Bottom: Distribution map of house, Italian and Spanish sparrows in Europe and Northern Africa.

Aims

The over-arching aim of my thesis has been to gain a better understanding of the hybrid speciation process at a genomic level. When I began my work, we had just scratched the surface of the Italian sparrow hybrid system. It had been shown that the Italian sparrow is phenotypically and genetically mosaic (Summers-Smith 1988; Hermansen et al. 2011; Elgvin et al. 2011), yet little was known about the nature of reproductive barriers between the hybrid and its parents. While Hermansen *et al.* 2011 suggested there was evidence for partial reproductive isolation, this was inferred from the allele frequencies of micro-satellite markers with limited power in differentiating species. Therefore, the first goal of my PhD research in **Paper I** was to test whether or not there is evidence for reproductive barriers between the Italian sparrow and its parents in zones of contact. This would serve to both solidify the Italian sparrow's status as a hybrid species and provide insight into how the Italian sparrow has maintained as its own distinct, hybrid lineage. Once we confirmed that there was evidence for reproductive isolation between the Italian sparrow and its parents, the next step was to investigate the hybrid system on a larger scale. Understanding which genomic regions were inherited from either parent, where there is evidence of selection and identifying areas of the genome potentially involved in reproductive isolation would provide further insight into how the Italian sparrow became its own species. Thus, in **Paper II**, we sought to characterize the Italian sparrow in relation to its parents by comparing entire genomes of the three focal taxa.

Finally, in **Paper III,** the goal was to extend our genomic understanding of the hybrid system by comparing multiple geographically isolated populations of the Italian sparrow on different Mediterranean islands. Investigating which genomic regions were invariably inherited from one parent or the other, would help in identifying areas that may be constrained for the formation of a functional hybrid, as well as which regions are more free to vary. Together, these papers shed light onto the hybrid

speciation process by providing a detailed investigation into the genome of Italian sparrow and highlighting regions that may have been important in the formation of species barriers that have enabled the Italian sparrow to remain distinct from its parents.

Paper Summaries

Paper I – Evidence for mito-nuclear and sex-linked reproductive barriers between the hybrid Italian sparrow and its parent species

In this paper, we tested if there was evidence of reproductive barriers between the Italian sparrow and its parent species at geographic range boundaries. We first utilized whole-transcriptome sequencing of the house and Spanish sparrows to identify 86 parent species diagnostic SNP markers. We then sequenced these markers in Italian sparrows (n=385) from populations across its range throughout Italy and Sicily including contact zones with both parents at its range boundaries. Spanish sparrows (n=142) from Spain, Sardinia and a Spanish/Italian sympatric zone in southeast Italy, and house sparrows (n=85) from the Czech Republic and Norway were also sequenced for the same SNP set. We employed Bayesian genomic and geographic cline analyses to identify markers exhibiting steep clines at range boundaries, indicative of a decrease in introgression of a parent's alleles into the Italian sparrow's genetic background. We demonstrated that there is evidence for post-zygotic reproductive barriers between the hybrid and its parents and identified seven markers that exhibited the steepest clines as candidate loci involved in reproductive isolation. A disproportionately large number of these candidate loci were found on the Z

chromosome. Also, the mitochondria and nuclear genes with mitochondrial function demonstrated patterns of reduced introgression of Spanish sparrow alleles in the Italian sparrow. We conclude mito-nuclear incompatibilities isolate Italian and Spanish sparrows while sex-linked incompatibilities isolate Italian and house sparrows. We also found no evidence for hybridization between sympatric Italian and Spanish sparrows in the Gargano peninsula indicating pre-zygotic barriers may also exist between the taxa. We suggest habitat dependent assortative mating plays a part as the Italian sparrow more closely resembles the human commensal house sparrow rather than the Spanish sparrow, which occupies more rural habitats. Overall, we conclude that the mechanisms of reproductive isolation in hybrid speciation may be similar to those of non-hybrid speciation with the exception that they are against two parent species rather than one.

Paper II – The genomic mosaicism of hybrid speciation

In Paper II, the goal was to examine hybrid speciation at a whole-genome level by characterizing patterns of parental inheritance and looking for signatures of selection in the Italian sparrow genome. To accomplish this, we first whole-genome sequenced and *de novo* assembled a reference house sparrow genome. We then mapped whole genome data from populations of the three focal taxa to the reference genome. Through population genetic and admixture analyses as well as phylogenetic inference we characterized the composition of the Italian sparrow's genome in relation to its parents. We found balanced yet heterogenous levels of parental contribution in the hybrid genome and identified regions where the Italian sparrow exhibits divergence from both parent species. Areas of novel divergence in the Italian sparrow lineage demonstrated patterns of variation consistent with balancing selection suggesting a heterozygotic advantage in the hybrid and were enriched for genes involved in immune system regulation. We

speculate that the admixed Italian sparrow genome may have had an advantageous effect on fitness thereby facilitating its spread. Furthermore, we identified regions of high divergence between the hybrid and each of its parents. These regions were disproportionately located on the Z chromosome and overrepresented in gene networks likely to be involved in reproductive barriers and species-specific adaptations such as body patterning and beak morphology. Additionally, we found evidence for heteroplasmy with mitochondrial sequences from both house and Spanish sparrows in two Italian sparrows. We conclude that both a mosaic pattern of parental inheritance and novel divergence in the hybrid lineage have contributed to the formation of the Italian sparrow.

Paper III – Variation and constraints in hybrid genome formation

In this paper, we investigated the extent to which the Italian sparrow genome varies between populations and if there is evidence for constraints in hybrid genome formation. We used whole-genome data from four Mediterranean island populations (Crete, Corsica, Sicily and Malta) of Italian sparrows with divergent morphologies to test this by comparing patterns of admixture and differentiation between them. We show that there is variation in the genomic combinations making up the Italian sparrow genome and suggest that these novel hybrid genomic combinations may have arisen independently more than once in the Mediterranean islands. In particular, we find that the islands differentiate in regions holding candidate genes for beak shape and plumage color and suggest that the differing combinations of parental genomes allowed for adaptive differentiation between the isolated island populations. We also find that some areas of the genome are inherited invariably from one of the parent species indicating the presence of genomic constraints. Mito-nuclear genes and genes involved in DNA repair were strongly constrained to house sparrow inheritance indicating the importance of incompatibilities

in reproductive isolation between Spanish and Italian sparrows. There were fewer genes invariably inherited from the Spanish sparrow, but the ones that were identified affected external phenotype rather than genome function. In general, these constrained regions are over-represented on the Z chromosome, consistent with the pattern of reduced introgression on sex chromosomes. This paper demonstrates that while there are varying genomic combinations that can lead to a functional hybrid genome which likely facilitate local adaptation, some regions are constrained in parallel producing strong reproductive barriers against one parent.

Discussion

Patterns of admixture in a hybrid species

To date, the Italian sparrow represents one of the few homoploid hybrid species studied in depth with genomic data. The papers in my dissertation help shed light on the processes at play in this poorly understood mode of speciation. First and foremost, my work helps to further validate the Italian sparrow's status as a hybrid species by demonstrating evidence of reproductive isolation as well as extensive, genome-wide admixture from its parents. The Italian sparrow's mosaic pattern of parental inheritance more closely resembles what has been observed in *Helianthus* sunflowers (Rieseberg et al. 2003) and the tiger swallow tail butterflies (Kunte et al. 2011) than hybrid species shown to differ in only small genomic regions such as the *Heliconius* butterflies (Heliconius Genome Consortium 2012). Patterns of mosaicism across a hybrid's genome are likely a product of both the age of the hybrid species and degree of divergence between the parental lineages.

The Italian sparrow is a hybrid species proposed to have arisen around 8,000 years ago when the house sparrow's range rapidly expanded throughout Europe during the spread of agriculture with the Neolithic

revolution (Hermansen et al. 2011; Sætre et al. 2012). The parental house and Spanish sparrows are inferred to have split from a common ancestor < 1 million years ago (Ravinet et al. 2018). At later stages of speciation, high levels of divergence are expected to be maintained genome-wide while introgression is localized to small genomic regions (Ravinet et al. 2018). Therefore, the largely heterogenous landscape of parental inheritance in the Italian sparrow may reflect the fact that the parent species themselves are not highly divergent. Larger block-like patterns of inheritance are observed on the Z chromosome in the Italian sparrow, which is consistent with recombination rate variation and selection shaping patterns of hybrid ancestry. In areas of low recombination, BDM incompatibilities are expected to be purged along with larger surrounding areas of neutral variation, producing block-like inheritance patterns. In contrast, in areas of high recombination or weak selection, haplotype blocks are expected to be broken apart more easily, allowing for decoupling of neutral variation from incompatibility loci (Schumer et al. 2018). Furthermore, the Z chromosome also has a lower recombination rate since recombination occurs only in males. The larger ancestry blocks on the Z chromosome relative to autosomes in the Italian sparrow supports previous studies (Carling and Brumfield 2008; Ellegren et al. 2012) in finding that sex chromosomes are at a later stage of speciation than the rest of the genome.

While the Italian sparrow maintains large portions of its genome from both parents, the amount from each respective parent is variable reflecting the fact that there is a spectrum of genomic parental contributions within the hybrid lineage as demonstrated in Paper III. The relatively low divergence between the parents may mean that few incompatibilities exist between them and occur in their hybrid offspring. With fewer incompatibilities in need of purging from the hybrid lineage, multiple functional hybrid genomic combinations are possible. The papers in this PhD provide an in-depth look at admixture across the genome of a hybrid species helping to elucidate how hybridization can shape genomes.

Reproductive isolation from the parents

Escape from genetic incompatibilities is believed to be easier when the parent species are more closely related (Orr et al. 1997; Tubaro and Lijtmaer 2002), yet this may in turn make the establishment of reproductive isolation between the hybrid and its parents more difficult. The Italian sparrow may represent a good balance of these conditions as Paper III demonstrates that there are variable genomic combinations that constitute the Italian sparrow, while some areas are genomically constrained to inheritance from one parent. In Paper I, we found evidence for reproductive isolation between the Italian sparrow in zones of contact and identified candidate reproductive isolation loci. Subsequently, (Hermansen et al. 2014) found that these candidate reproductive isolation loci are a subset of loci isolating the parent species. This suggests the Italian sparrow has sorted incompatibilities isolating its parents in order to form reproductive barriers against both.

In line with many previous studies (Macholán et al. 2007; Carling and Brumfield 2008; Teeter et al. 2010; Gompert and Buerkle 2011b; Ellegren et al. 2012; Taylor et al. 2014), the papers in this dissertation find candidate incompatibility loci were disproportionately found on the Z chromosome. The Z chromosome has a reduced effective population size due to female heterogamety, which is expected to increase the rate of fixations via drift or selection. However, selection tests on coding sequences and Tajima's D estimates are consistent with a role for selection on the Z chromosome. Together, the papers in this thesis provide support for sex chromosomes being paramount in establishing reproductive barriers during hybrid speciation.

In Papers I and III, we also find evidence for less well-documented mito-nuclear incompatibilities playing a crucial part in isolating the Spanish and Italian sparrow. There is strong selection for compatibility between the nuclear and mitochondrial genomes (Hill 2016) as mito-nuclear dysfunction can have severe consequences on fitness (Camus et al. 2015; Patel et al. 2016). In fact, other hybridizing bird species have been shown to experience

reduced introgression of mitochondrial loci compared to nuclear loci (Kvist and Rytkoenen 2006; Carling and Brumfield 2008; Taylor et al. 2013). Moreover, fitness costs associated with suboptimal respiration have been found in hybridizing bird species (Olson et al. 2010; McFarlane et al. 2016). The majority of Italian sparrows have inherited solely house sparrow mtDNA and I found evidence that there is also strong selection against Spanish sparrow inheritance in nuclear genes with mitochondrial function. In Paper I, we suggest this pattern is consistent with patterns of genomic conflict in the form of 'mothers curse'. Since mitochondria are maternally inherited, male detrimental mutations will not be subject to selection and female advantageous, yet male detrimental mutations can accumulate (Beekman et al. 2014). This effect can be remedied in males through compensatory mutations which is expected to arise on the Z chromosome where male-specific fitness effects are more likely (Connallon et al. 2018). This is consistent with our findings that candidate reproductive isolation loci were predominantly sex-linked. Paternal leakage is another suggested mechanism for overcoming negative fitness effects of maternally inherited mitochondria. Notably, heteroplasmy has also been observed in hybridizing taxa (Shitara et al. 1998; Kvist et al. 2003; Radojicic et al. 2015; Śmietanka and Burzyński 2017), which we found evidence for in Papers II and III.

While mito-nuclear incompatibilities were shown to be instrumental in isolating the Italian and Spanish sparrows, Paper III found that genes invariably inherited from the Spanish sparrow as candidates for affecting external phenotype and included a candidate gene for plumage color. A role for pre-zygotic barriers in isolating the Italian sparrow from the house sparrow may therefore be important. This has been supported by a cline analysis in the Alps hybrid zone between house and Italian sparrows demonstrating selection on plumage traits (Bailey et al. 2015).

In all, my work provides evidence for the existence of reproductive barriers between the Italian sparrow and its parents and highlights the importance of sex-linked and mito-nuclear incompatibilities in the formation of hybrid species.

Selection within the hybrid lineage

The papers in this dissertation also provide evidence of selection within the hybrid lineage. One proposed mechanism of hybrid speciation is that a hybrid maintains fitness advantages compared to the parents by being able to occupy a new ecological niche (Mallet 2007) and that hybrids may even displace their parents if they experience higher fitness in a given habitat (Buerkle et al. 2000). Both Papers II and III find evidence for balancing selection in areas where the Italian sparrow is divergent from both parent taxa and in Paper II these regions were enriched for genes with functions in the immune system. We postulate that the admixed genome may also have an advantageous effect in traits with an intermediate optimum and that this may have facilitated its spread. Genomic regions where the morphologically divergent populations of Italian sparrows differentiated from each other were also found to harbor candidate loci for beak shape, feather development and melanogenesis underscoring the fact that novel variation combined into admixed genomes can allow for adaptive differentiation between isolated populations of hybrid species.

Future perspectives & concluding remarks

Though my work provides a glimpse of how hybrid speciation can occur at a genomic level, there are still many avenues of further research to help elucidate the hybrid speciation process. One question that remains to be answered is how the timing and magnitude of hybridization events affect speciation potential. Phylogenetic analysis of mitochondrial markers in the system have supported the hypothesis that the Italian sparrow arose as a result of human activity during the Neolithic revolution (Hermansen et al. 2011; Sætre et al. 2012), however, further research is needed to more robustly test this. More specifically, model based estimation of the timing

and magnitude of gene flow would provide further insight into the hybrid speciation process (Payseur and Rieseberg 2016).

Additionally, while my work points to candidate regions and genes involved in reproductive isolation, the functional genetics behind these proposed areas is still unknown. Further research into the phenotypic effects of these regions, and the identification of the specific genes and causative factors is needed. It would also be interesting to investigate if structural variation plays a role in reproductive isolation in the system. Inversions have been shown to rapidly drive species divergence (Noor et al. 2001) and combining two parental genomes into a hybrid genome may result in structural variation potentially forming powerful reproductive barriers. In fact, both pre- and post-zygotic reproductive barriers have been shown to map to inversions in species pairs (Kirkpatrick 2010).

Another question raised from my work and an avenue for further research is the role and extent of heteroplasmy in Italian sparrows. If mito-nuclear incompatibilities are instrumental as reproductive barriers, then how does heteroplasmic mitochondria fit into the picture? The system would benefit from a more detailed investigation into how heteroplasmic individuals differ from the Italian sparrows with solely house sparrow mitochondria in terms of fitness, patterns of parental inheritance in nuclear-encoded mitochondrial genes and degree of isolation from the parents.

Finally, another line of research that needs to be further explored is the role of ecology and geography in shaping hybrid traits in the Italian sparrow. In Paper II, we suggest that the hybrid genome may have experienced fitness advantages facilitating its spread, one of the proposed mechanisms for hybrid success (Buerkle et al. 2000; Mallet 2007). Additionally, Paper III demonstrated there is evidence for adaptive differentiation among island populations of Italian sparrows and beak shape has been shown to be strongly influenced by precipitation regimes in Italy (Eroukhmanoff et al. 2013). This points to a role of ecology and local adaptation in the formation of the Italian sparrow. However, the phylogeographic history of a species is important in understanding how new species can successfully spread, and combining it with phenotypic and

genomic data is needed to understand how species are able to occupy new niches (Trucchi et al. 2016). Thus, investigating the phylogeographic history of the Italian sparrow and pairing it with the morphological, ecological and genomic data would be powerful way to move forward and investigate how the Italian sparrow has been successful in occupying the Italian peninsula and surrounding islands.

Overall, throughout my PhD research, I have used genomic data to paint a picture as to how hybridization between two lineages has the potential to create a novel species. My work serves as a detailed investigation of an example of homoploid hybrid speciation in animals and provides a jumping off point for future hybrid speciation research. It is an exciting time as fast developing technology has enabled new approaches for studying hybrid speciation and as more information becomes available, it is becoming clear that hybridization is a powerful force of variation and can be a catalyst for speciation.

Acknowledgements

Many thanks to Tore Elgvin, Fabrice Eroukhmanoff and Mark Ravinet for taking the time to read my thesis and provide helpful comments. Also a big thank you to Anna Mazzarella for advice with layout editing.

References

Abbott, R. J., N. H. Barton, and J. M. Good. 2016. Genomics of hybridization and its evolutionary consequences. *Molecular Ecology* **25**:2325–2332.

Abbott, R.J., D. Albach, S. Ansell, J. W. Arntzen, S. J. E. Baird, N. Bierne, J. Boughman, A. Brelsford, C. A. Buerkle, R. Buggs, R. K. Butlin, U. Dieckmann, F. Eroukhmanoff, A. Grill, S. H. Cahan, J. S. Hermansen, G. Hewitt, A. G. Hudson, C. Jiggins, J. Jones, B. Keller, T. Marczewski, J. Mallet, P. Martinez-Rodriguez, M. Möst, S. Mullen, R. Nichols, A. W. Nolte, C. Parisod, K. Pfennig, A. M. Rice, M. G. Ritchie, B. Seifert, C. M. Smadja, R. Stelkens, J. M. Szymura, R. Vainola, J. B. W. Wolf, and D.

Zinner. 2013. Hybridization and speciation. *Journal of Evolutionary Biology* **26:**229–246.
Arnegard, M. E., M. D. McGee, B. Matthews, K. B. Marchinko, G. L. Conte, S. Kabir, N. Bedford, S. Bergek, Y. F. Chan, F. C. Jones, D. M. Kingsley, C. L. Peichel, and D. Schluter. 2014. Genetics of ecological divergence during speciation. *Nature* **511**:307–311.
Arnold, M. L., and S. A. Hodges. 1995. Are natural hybrids fit or unfit relative to their parents? *Trends in Ecology and Evolution* **10**:67–71.
Baack, E. J., and L. H. Rieseberg. 2007. A genomic view of introgression and hybrid speciation. *Current Opinion in Genetics and Development* **17**:513–518.
Bailey, R. I., M. R. Tesaker, C. N. Trier, and G. P. Saetre. 2015. Strong selection on male plumage in a hybrid zone between a hybrid bird species and one of its parents. *Journal of Evolutionary Biology* **28**:1257–1269.
Ballard, J., and D. M. Rand. 2005. The population biology of mitochondrial DNA and its phylogenetic implications. *Annual Review of Ecology Evolution and Systematics* **36**:621–642.
Bar-Yaacov, D., Z. Hadjivasiliou, L. Levin, G. Barshad, R. Zarivach, A. Bouskila, and D. Mishmar. 2015. Mitochondrial involvement in vertebrate speciation? The Case of mito-nuclear genetic divergence in chameleons. *Genome Biology and Evolution* **7**:3322–3336.
Barton, N. H. 2001. The role of hybridization in evolution. *Molecular Ecology* **10**:551–568.
Barton, N. H., and G. M. Hewitt. 1989. Adaptation, speciation and hybrid zones. *Nature* **341**:497–503.
Barton, N. H., and G. M. Hewitt. 1985. Analysis of hybrid zones. *Annual Review of Ecology and Systematics* **16**:113–148.
Bateson, W. 1909. Heredity and variation in modern lights. *Darwin and Modern Science:* 85-101.
Beekman, M., D. K. Dowling, and D. K. Aanen. 2014. The costs of being male: are there sex-specific effects of uniparental mitochondrial inheritance? *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **369**:1–7.
Buerkle, C. A., and C. Lexer. 2008. Admixture as the basis for genetic mapping. *Trends in Ecology and Evolution* **23**:686–694.
Buerkle, C. A., and L. H. Rieseberg. 2001. Low intraspecific variation for genomic isolation between hybridizing sunflower species. *Evolution* **55**:684–691.
Buerkle, C. A., and L. H. Rieseberg. 2008. The rate of genome stabilization in homoploid hybrid species. *Evolution* **62**:266–275.
Buerkle, C. A., R. J. Morris, M. A. Asmussen, and L. H. Rieseberg. 2000. The likelihood of homoploid hybrid speciation. Heredity **84**:441–451.
Bullini, L. 1994. Origin and evolution of animal hybrid species. *Trends in Ecology and Evolution* **9**:422–426.
Burri, R., A. Nater, T. Kawakami, C. F. Mugal, P. I. Olason, L. Smeds, A. Suh, L. Dutoit, S. Bureš, L. Z. Garamszegi, S. Hogner, J. Moreno, A.

Qvarnström, M. Ružić, S.-A. Sæther, G.-P. Sætre, J. Toeroek, and H. Ellegren. 2015. Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of Ficedula flycatchers. *Genome Research* **25**:1656–1665.

Camus, M. F., J. B. W. Wolf, E. H. Morrow, and D. K. Dowling. 2015. Single nucleotides in the mtDNA sequence modify mitochondrial molecular function and are Associated with sex-specific effects on fertility and aging. *Current Biology* **25**:2717–2722.

Carling, M. D., and R. T. Brumfield. 2008. Haldane's rule in an avian system: Using cline theory and divergence population genetics to test for differential introgression of mitochondrial, autosomal, and sex-linked loci across the passerinabunting hybrid zone. *Evolution* **62**:2600–2615.

Carneiro, M., S. J. E. Baird, S. Afonso, E. Ramirez, P. Tarroso, H. Teotónio, R. Villafuerte, M. W. Nachman, and N. Ferrand. 2013. Steep clines within a highly permeable genome across a hybrid zone between two subspecies of the European rabbit. *Molecular Ecology* **22**:2511–2525.

Charlesworth, B., J. A. Coyne, and N. H. Barton. 1987. The relative rates of evolution of sex chromosomes and autosomes. *The American Naturalist* **130**:113–146.

Connallon, T., M. F. Camus, E. H. Morrow, and D. K. Dowling. 2018. Coadaptation of mitochondrial and nuclear genes, and the cost of mother's curse. *Proceedings of Biological Science*. **285**:e20172257.

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

Dobzhansky, T. 1948. Genetics of Natural Populations. *Genetics* **33**:588–602.

Dobzhansky, T. 1936. Studies on hybrid sterility. II. Localization of sterility factors in Drosophila pseudoobscura hybrids. *Genetics* **21**:113–135.

Dowling, T. E., and A. C. L. Secor. 1997. The role of hybridization and introgression in the diversification of animals. *Annual Review of Ecology and Systematics* **28**:593–619.

Elgvin, T. O., J. S. Hermansen, A. Fijarczyk, T. Bonnet, T. Borge, S. A. Sæther, K. L. Voje, and G.-P. Sætre. 2011. Hybrid speciation in sparrows II: a role for sex chromosomes? *Molecular Ecology* **20**:3823–3837.

Ellegren, H., L. Smeds, R. Burri, P. I. Olason, N. Backström, T. Kawakami, A. Künstner, H. Mäkinen, K. Nadachowska-Brzyska, A. Qvarnström, S. Uebbing, and J. B. W. Wolf. 2012. The genomic landscape of species divergence in Ficedula flycatchers. *Nature* **491**:756–760.

Eroukhmanoff, F., J. S. Hermansen, R. I. Bailey, S. A. Sæther, and G.-P. Sætre. 2013. Local adaptation within a hybrid species. *Heredity* **111**:286–292.

Feliner, G.N., M. Álvarez, J. Fuertes-Aguilar, M. Heuertz, I. Marques, F. Moharrek, R Piñeiro, R Riina, J.A.Rosselló, P.S.Soltis, I. Villa-Machío 2017. Is homoploid hybrid speciation that rare? An empiricist's view. *Heredity* **118**: 513-516.

Felsenstein, J. 1981. Skepticism towards Santa Rosalia, or why are there so few kinds of animals? Evolution **35**:124–138.
Fisher, R.A., 1930. *The genetical theory of natural selection*. Clarendon Press, Oxford.
Gompert, Z., and C. A. Buerkle. 2011a. A hierarchical bayesian model for next-generation population genomics. *Genetics* **187**:903–917.
Gompert, Z., and C. A. Buerkle. 2011b. Bayesian estimation of genomic clines. *Molecular Ecology* **20**:2111–2127.
Gompert, Z., L. K. Lucas, C. C. Nice, J. A. Fordyce, M. L. Forister, and C. A. Buerkle. 2012. Genomic regions with a history of divergent selection affect fitness of hybrids between two butterfly species. *Evolution* **66**:2167–2181.
Good, J. M., D. Vanderpool, S. Keeble, and K. Bi. 2015. Negligible nuclear introgression despite complete mitochondrial capture between two species of chipmunks. *Evolution* **69**:1961–1972.
Grant, P. R., and B. R. Grant. 1994. Phenotypic and genetic effects of hybridization in Darwin's finches. *Evolution* **48**:297.
Grant, P. R., B. R. Grant, and K. Petren. 2015. Hybridization in the recent past. *The American Naturalist* **166**:56–67.
Gross, B. L., and L. H. Rieseberg. 2005. The ecological genetics of homoploid hybrid speciation. *Journal of Heredity* **96**:241–252.
Harrison, R. G. 1993. Hybrid zones and the evolutionary process. Oxford University Press, New York, NY.
Harrison, R. G., and E. L. Larson. 2014. Hybridization, introgression, and the nature of species boundaries. *Journal of Heredity* **105**:795–809.
Hegarty, M. J., and S. J. Hiscock. 2008. Genomic clues to the evolutionary success of review polyploid plants. *Current Biology* **18**:R435–R444.
Heliconius Genome Consortium. 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* **487**:94–98.
Hermansen, J. S., F. Haas, C. N. Trier, R. I. Bailey, A. J. Nederbragt, A. Marzal, and G.-P. Sætre. 2014. Hybrid speciation through sorting of parental incompatibilities in Italian sparrows. *Molecular Ecology* **23**:5831–5842.
Hermansen, J. S., S. A. Saether, T. O. Elgvin, T. Borge, E. Hjelle, and G.-P. Sætre. 2011. Hybrid speciation in sparrows I: phenotypic intermediacy, genetic admixture and barriers to gene flow. *Molecular Ecology* **20**:3812–3822.
Hill, G. E. 2016. Mitonuclear coevolution as the genesis of speciation and the mitochondrial DNA barcode gap. *Ecology and Evolution* **6**:5831–5842.
Hill, G. E. 2017. The mitonuclear compatibility species concept. *The Auk* **134**:393–409.
Jiggins, C. D., C. Salazar, M. Linares, and J. Mavarez. 2008. Hybrid trait speciation and *Heliconius* butterflies. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **363**:3047–3054.
Keller, I., C. E. Wagner, L. Greuter, S. Mwaiko, O. M. Selz, A. Sivasundar, S. Wittwer, and O. Seehausen. 2013. Population genomic signatures of

divergent adaptation, gene flow and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes. *Molecular Ecology* **22**:2848–2863.

Kirkpatrick, M. 2010. How and why chromosome inversions evolve. *PLoS Biology* **8**:e1000501

Kunte, K., C. Shea, M. L. Aardema, J. M. Scriber, T. E. Juenger, L. E. Gilbert, and M. R. Kronforst. 2011. Sex chromosome mosaicism and hybrid speciation among tiger swallowtail butterflies. *PLoS Genetics* **7**:e1002274–14.

Kvist, L., and S. Rytkoenen. 2006. Characterization of a secondary contact zone of the great tit *Parus major* and the Japanese tit p-minor (*Aves : Passeriformes*) in far eastern Siberia with DNA markers. *Zootaxa* 55–73.

Kvist, L., J. Martens, A. A. Nazarenko, and M. Orell. 2003. Paternal leakage of mitochondrial DNA in the great tit (*Parus major*). *Molecular Biology and Evolution*. **20**:243–247.

Lamichhaney, S., F. Han, M. T. Webster, L. Andersson, B. R. Grant, and P. R. Grant. 2018. Rapid hybrid speciation in Darwin's finches. *Science* **359**:224–228.

Larsen, P. A., M. R. Marchan-Rivadeneira, and R. J. Baker. 2010. Natural hybridization generates mammalian lineage with species characteristics. *Proceedings of the National Academy of Science U.S.A.* **107**:11447–11452.

Macholán, M., P. Munclinger, M. Šugerková, P. Dufková, B. Bímová, E. Božíková, J. Zima, and J. Piálek. 2007. Genetic analysis of autosomal and X-linked markers across a mouse hybrid zone. *Evolution* **61**:746–771.

Mallet, J. 1995. A species definition for the modern synthesis. *Trends in Ecology and Evolution* **10**:294–299.

Mallet, J. 2007. Hybrid speciation. *Nature* **446**:279–283.

Mallet, J. 2005. Hybridization as an invasion of the genome. *Trends in Ecology and Evolution* **20**:229–237.

Mank, J. E., B. Vicoso, S. Berlin, and B. Charlesworth. 2010. Effective population size and the faster-X effect: Empirical results and their interpretation. *Evolution* **64**:663–674.

Mavarez, J., and M. Gonzalez. 2006. A set of microsatellite markers for *Heliconius melpomene* and closely related species. *Molecular Ecology Notes* **6**:20–23.

Mavarez, J., and M. Linares. 2008. Homoploid hybrid speciation in animals. *Molecular Ecology* **17:**4181–4185.

Mavarez, J., C. A. Salazar, E. Bermingham, C. Salcedo, C. D. Jiggins, and M. Linares. 2006. Speciation by hybridization in *Heliconius* butterflies. *Nature* **441**:868–871.

May, R. M., J. A. Endler, and R. E. McMurtrie. 2015. Gene frequency clines in the presence of selection opposed by gene flow. *The American Naturalist* **109**:659–676.

Mayr, E. 1963. *Animal species and evolution*. Harvard University Press, Cambridge.
Mayr, E. 1932. Birds collected during the Whitney South Sea Expedition. 21, Notes on thickheads (Pachycephala) from Polynesia. 1–23. New York City : The American Museum of Natural History, New York City.
McFarlane, S. E., P. M. Sirkiä, M. Ålund, and A. Qvarnström. 2016. hybrid dysfunction expressed as elevated metabolic rate in male *Ficedula* flycatchers. *PLoS ONE* **11:**e0161547.
Melo, M. C., C. Salazar, C. D. Jiggins, and M. Linares. 2009. Assortative mating preferences among hybrids offers a route to hybrid speciation. *Evolution* **63**:1660–1665.
Meyer, A., W. Salzburger, and M. Schartl. 2006. Hybrid origin of a swordtail species (*Teleostei : Xiphophorus clemenciae*) driven by sexual selection. *Molecular Ecology* **15**:721–730.
Muller, H. J. 1940. An analysis of the process of structural change in chromosomes of *Drosophila*. *Journal of Genetics* **40**:1–66.
Nachman, M. W., and B. A. Payseur. 2012. Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. *Philosophical Transactions of the Royal Society London:* Biological Science **367**:409–421.
Nagylaki, T. 1975. Conditions for the existence of clines. *Genetics* **80**:595–615.
Nice, C. C., Z. Gompert, J. A. Fordyce, M. L. Forister, L. K. Lucas, and C. A. Buerkle. 2013. Hybrid speciation and independent evolution in lineages of alpine butterflies. *Evolution* **67**:1055–1068.
Noor, M., K. L. Grams, L. A. Bertucci, and J. Reiland. 2001. Chromosomal inversions and the reproductive isolation of species. *Proceedings of the National Academy of Sciences U.S.A.* **98**:12084–12088.
Olson, J. R., S. J. Cooper, D. L. Swanson, M. J. Braun, and J. B. Williams. 2010. The relationship of metabolic performance and distribution in black-capped and Carolina chickadees. *Physiological and Biochemical Zoology* **83**:263–275.
Orr, H. A. 1996. Dobzhansky, Bateson, and the genetics of speciation. *Genetics* **144**:1331–1335.
Orr, H. A., L. D. Madden, J. A. Coyne, R. Goodwin, and R. S. Hawley. 1997. The developmental genetics of hybrid inviability: A mitotic defect in *Drosophila* hybrids. *Genetics* **145**:1031–1040.
Parchman, T.L., Z. Gompert., M.J. Braun, R.T. Brumfield, D.B. McDonald, J.A.C. Uy, G. Zhang, E.D. Jarvis, B.A. C.A. Schlinger, Buerkle 2013. The genomic consequences of adaptive divergence and reproductive isolation between species of manakins. *Molecular Ecology* **22**:3304-3317.
Patel, M. R., G. K. Miriyala, A. J. Littleton, H. Yang, K. Trinh, J. M. Young, S. R. Kennedy, Y. M. Yamashita, L. J. Pallanck, and H. S. Malik. 2016. A mitochondrial DNA hypomorph of cytochrome oxidase specifically impairs male fertility in *Drosophila melanogaster. Elife* **5**:e16923.

Payseur, B. A., and L. H. Rieseberg. 2016. A genomic perspective on hybridization and speciation. *Molecular Ecology*. (Special Issue).

Payseur, B. A., and M. W. Nachman. 2005. The genomics of speciation: investigating the molecular correlates of X chromosome introgression across the hybrid zone between *Mus domesticus* and *Mus musculus*. *Biological Journal of the Linnean Society* **84**:523–534.

Qvarnström, A., and R. I. Bailey. 2008. Speciation through evolution of sex-linked genes. *Heredity* **102**:4–15.

Radojicic, J. M., I. Krizmanic, P. Kasapidis, and E. Zouros. 2015. Extensive mitochondrial heteroplasmy in hybrid water frog (*Pelophylax spp*.) populations from Southeast Europe. *Ecology and Evolution* **5**:4529–4541.

Rand, D. M., R. A. Haney, and A. J. Fry. 2004. Cytonuclear coevolution: the genomics of cooperation. *Trends in Ecology and Evolution* **19**:645–653.

Ravinet, M., T.O. Elgvin, C Trier, M. Aliabadian, A. Gavrilov, G-P Sætre. 2018. Signatures of human-commensalism in the house sparrow genome. *Proceedings of a Royal Society B* **287**:1884.

Ravinet, M., K. Yoshida, S. Shigenobu, A. Toyoda, A. Fujiyama, and J. Kitano. 2018. The genomic landscape at a late stage of stickleback speciation: High genomic divergence interspersed by small localized regions of introgression. *PLoS Genetics* **14**:e1007358.

Ravinet, M., R. Faria, R. K. Butlin, J. Galindo, N. Bierne, M. Rafajlović, M.A.F. Noor, B. Mehlig, and A. M. Westram. 2017. Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *Journal of Evolutionary Biology* **30**:1450–1477.

Rieseberg, L. H. 1997. Hybrid origins of plant species. *Annual Review of Ecology Evolution and Systematics* **28**:359–389.

Rieseberg, L. H., and J. H. Willis. 2007. Plant speciation. *Science* **317**:910–914.

Rieseberg, L. H., M. A. Archer, and R. K. Wayne. 1999. Transgressive segregation, adaptation and speciation. *Heredity* **83**:363–372.

Rieseberg, L. H., O. Raymond, D. M. Rosenthal, Z. Lai, K. Livingstone, T. Nakazato, J. L. Durphy, A. E. Schwarzbach, L. A. Donovan, and C. Lexer. 2003. Major ecological transitions in wild sunflowers facilitated by hybridization. *Nature Chemistry* **301**:1211–1216.

Salzburger, W., S. Baric, and C. Sturmbauer. 2002. Speciation via introgressive hybridization in East African cichlids? *Molecular Ecology* **11**:619–625.

Samuk, K., G. L. Owens, K. E. Delmore, S. E. Miller, D. J. Rennison, and D. Schluter. 2017. Gene flow and selection interact to promote adaptive divergence in regions of low recombination. *Molecular Ecology* **26**:4378–4390.

Sankararaman, S., S. Mallick, M. Dannemann, K. Prüfer, J. Kelso, S. Pääbo, N. Patterson, and D. Reich. 2014. The genomic landscape of Neanderthal ancestry in present-day humans. *Nature* **507**:354–357.

Schluter, D. 1993. Adaptive Radiation in Sticklebacks: Size, Shape, and Habitat Use Efficiency. *Ecology* **74**:699–709.
Schluter, D. 1995. Adaptive Radiation in Sticklebacks: Trade-Offs in Feeding Performance and Growth. *Ecology* **76**:82–90.
Schumer, M., C. Xu, D. L. Powell, A. Durvasula, L. Skov, C. Holland, J. C. Blazier, S. Sankararaman, P. Andolfatto, G. G. Rosenthal, and M. Przeworski. 2018. Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science* **360**:3684–660.
Schumer, M., R. Cui, B. Boussau, R. Walter, G. Rosenthal, and P. Andolfatto. 2013. An evaluation of the hybrid speciation hypothesis for *Xiphophorus clemenciae* based on whole genome sequences. *Evolution* **67**:1155–1168.
Schumer, M., G. G. Rosenthal, and P. Andolfatto. 2014. How common is homoploid hybrid speciation? *Evolution* **68**:1553–1560.
Schwarz, D., B. M. Matta, N. L. Shakir-Botteri, and B. A. McPheron. 2005. Host shift to an invasive plant triggers rapid animal hybrid speciation. *Nature* **436**:546–549.
Seehausen, O., R. K. Butlin, I. Keller, C. E. Wagner, J. W. Boughman, P. A. Hohenlohe, C. L. Peichel, G.-P. Sætre, C. Bank, Å. Brännström, A. Brelsford, C. S. Clarkson, F. Eroukhmanoff, J. L. Feder, M. C. Fischer, A. D. Foote, P. Franchini, C. D. Jiggins, F. C. Jones, A. K. Lindholm, K. Lucek, M. E. Maan, D. A. Marques, S. H. Martin, B. Matthews, J. I. Meier, M. Möst, M. W. Nachman, E. Nonaka, D. J. Rennison, J. Schwarzer, E. T. Watson, A. M. Westram, and A. Widmer. 2014. Genomics and the origin of species. *Nature Reviews Genetics* **15:**176–192.
Shitara, H., J.-I. Hayashi, S. Takahama, H. Kaneda, and H. Yonekawa. 1998. Maternal inheritance of mouse mtDNA in interspecific hybrids: segregation of the leaked paternal mtDNA followed by the prevention of subsequent paternal leakage. *Genetics* **148**:851–857.
Slatkin, M. 1973. Gene flow and selection in a cline. *Genetics* **75**:733–756.
Soltis, P. S., and D. E. Soltis. 2009. The Role of Hybridization in Plant Speciation. Annual Review of Plant Biology **60**:561–588.
Summers-Smith, J. D. 1988. *The sparrows: A study of the genus Passer*. T & AD Poyser, Calton.
Svensson, L., Grant, P.J, Mullarney, K., Zetterstroem, D., 1999. *Gyldendals store fugleguide: Europas og middelhavsområdets fugler i felt.* Gyldendal, Oslo.
Szymura, J. M., and N. H. Barton. 1986. Genetic analysis of a hybrid zone between the fire-bellied toads, *Bombina bombina* and *B. variegata*, near Cracow in southern Poland. *Evolution* **40**:1141–1159.
Sæther, S. A., G.-P. Sætre, T. Borge, C. Wiley, N. Svedin, G. Andersson, T. Veen, J. Haavie, M. R. Servedio, S. Bureš, M. Král, M. B. Hjernquist, L. Gustafsson, J. Träff, and A. Qvarnström. 2007. Sex chromosome-linked species recognition and evolution of reproductive isolation in flycatchers. *Science* **318**:95–97.

Sætre, G.-P., S. Riyahi, M. Aliabadian, J. S. Hermansen, S. Hogner, U. Olsson, M. F. Gonzalez Rojas, S. A. Sæther, C. N. Trier, and T. O. Elgvin. 2012. Single origin of human commensalism in the house sparrow. *Journal of Evolutionary Biology* **25**:788–796.

Śmietanka, B., and A. Burzyński. 2017. Disruption of doubly uniparental inheritance of mitochondrial DNA associated with hybridization area of European *Mytilus edulis* and *Mytilus trossulus* in Norway. *Marine Biology* **164**:743.

Taylor, S. A., D. J. Anderson, and V. L. Friesen. 2013. Evidence for asymmetrical divergence-gene flow of nuclear loci, but not mitochondrial loci, between seabird sister species: Blue-footed (*Sula nebouxii*) and Peruvian (*S. variegata*) boobies. *PLoS ONE* **8**:e62256.

Taylor, S. A., D. J. Anderson, C. B. Zavalaga, and V. L. Friesen. 2012. Evidence for strong assortative mating, limited gene flow, and strong differentiation across the blue-footed/Peruvian booby hybrid zone in northern Peru. *Journal of Avian Biology* **43**:311–324.

Taylor, S. A., R. L. Curry, T. A. White, V. Ferretti, and I. Lovette. 2014. Spatiotemporally consistent genomic signatures of reproductive isolation in a moving hybrid zone. *Evolution* **68**:3066–3081.

Teeter, K. C., L. M. Thibodeau, Z. Gompert, C. A. Buerkle, M. W. Nachman, and P. K. Tucker. 2010. The variable genomic architecture of isolation between hybridizing species of house mouse. *Evolution* **64**:472–485.

Templeton, A. R. 1981. Mechanisms of Speciation - a Population Genetic Approach. Annual Review of Ecology and Systematics **12**:23–48.

Töpfer, T. 2006. The taxonomic status of the Italian Sparrow - *Passer italiae* (Vieillot 1817): Speciation by stabilised hybridisation? A critical analysis. *Zootaxa* 117–145.

Toews, D. P. L., S. A. Taylor, R. Vallender, A. Brelsford, B. G. Butcher, P. W. Messer, and I. J. Lovette. 2016. Plumage genes and little else distinguish the genomes of hybridizing warblers. *Current Biology* **26**:1–7.

Trucchi, E., B. Facon, P. Gratton, E. Mori, N. C. Stenseth, and S. Jentoft. 2016. Long live the alien: is high genetic diversity a pivotal aspect of crested porcupine (*Hystrix cristata*) long-lasting and successful invasion? *Molecular Ecology* **25**:3527–3539.

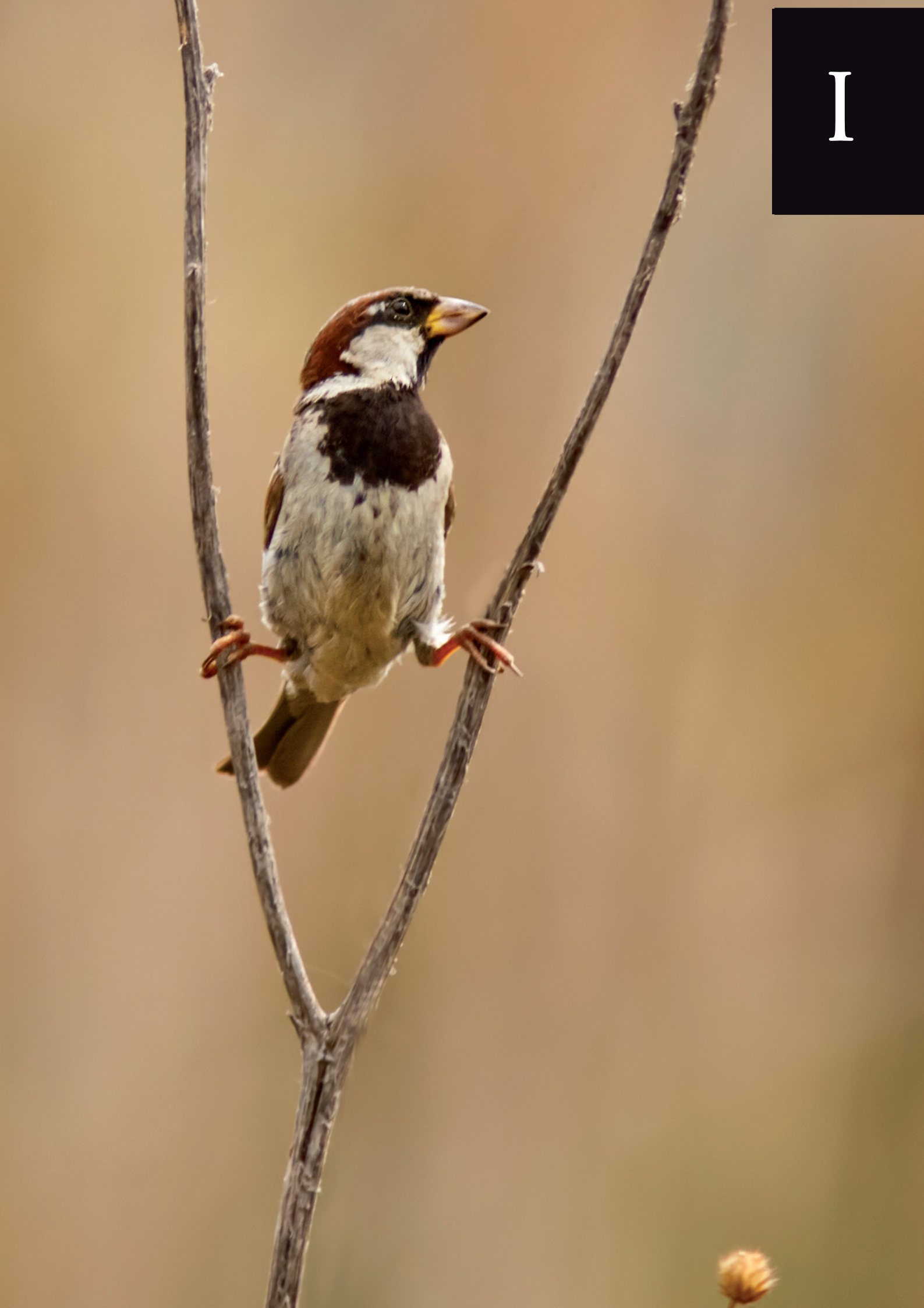
Tubaro, P. L., and D. A. Lijtmaer. 2002. Hybridization patterns and the evolution of reproductive isolation in ducks. *Biological Journal of the Linnean Society* **77**:193–200.

Turelli, M., and Orr, H. A. 1995. The dominance theory of Haldane's rule. *Genetics* **140**:389–402.

Turner, T. L., M. W. Hahn, and S. V. Nuzhdin. 2005. Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biology* 3:1572–1578.

Via, S., and J. West. 2008. The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular Ecology* **17**:4334–4345.

vonHoldt, B. M., J. P. Pollinger, D. A. Earl, J. C. Knowles, A. R. Boyko, H. Parker, E. Geffen, M. Pilot, W. Jedrzejewski, B. Jedrzejewska, V. Sidorovich, C. Greco, E. Randi, M. Musiani, R. Kays, C. D. Bustamante, E. A. Ostrander, J. Novembre, and R. K. Wayne. 2011. A genome-wide perspective on the evolutionary history of enigmatic wolf-like canids. *Genome Research* **21**:1294–1305.



I

OPEN ACCESS Freely available online

PLOS | GENETICS

Evidence for Mito-Nuclear and Sex-Linked Reproductive Barriers between the Hybrid Italian Sparrow and Its Parent Species

Cassandra N. Trier[ϑ], **Jo S. Hermansen**[ϑ], **Glenn-Peter Sætre***, **Richard I. Bailey**

Centre for Ecological and Evolutionary Synthesis, Department of Biology, University of Oslo, Oslo, Norway

Abstract

Studies of reproductive isolation between homoploid hybrid species and their parent species have rarely been carried out. Here we investigate reproductive barriers between a recently recognized hybrid bird species, the Italian sparrow *Passer italiae* and its parent species, the house sparrow *P. domesticus* and Spanish sparrow *P. hispaniolensis*. Reproductive barriers can be difficult to study in hybrid species due to lack of geographical contact between taxa. However, the Italian sparrow lives parapatrically with the house sparrow and both sympatrically and parapatrically with the Spanish sparrow. Through whole-transcriptome sequencing of six individuals of each of the two parent species we identified a set of putatively parent species-diagnostic single nucleotide polymorphism (SNP) markers. After filtering for coverage, genotyping success (>97%) and multiple SNPs per gene, we retained 86 species-informative, genic, nuclear and mitochondrial SNP markers from 84 genes for analysis of 612 male individuals. We show that a disproportionately large number of sex-linked genes, as well as the mitochondria and nuclear genes with mitochondrial function, exhibit sharp clines at the boundaries between the hybrid and the parent species, suggesting a role for mito-nuclear and sex-linked incompatibilities in forming reproductive barriers. We suggest that genomic conflict via interactions between mitochondria and sex-linked genes with mitochondrial function (''mother's curse'') at one boundary and centromeric drive at the other may best explain our findings. Hybrid speciation in the Italian sparrow may therefore be influenced by mechanisms similar to those involved in non-hybrid speciation, but with the formation of two geographically separated species boundaries instead of one. Spanish sparrow alleles at some loci have spread north to form reproductive barriers with house sparrows, while house sparrow alleles at different loci, including some on the same chromosome, have spread in the opposite direction to form barriers against Spanish sparrows.

Citation: Trier CN, Hermansen JS, Sætre G-P, Bailey RI (2014) Evidence for Mito-Nuclear and Sex-Linked Reproductive Barriers between the Hybrid Italian Sparrow and Its Parent Species. PLoS Genet 10(1): e1004075. doi:10.1371/journal.pgen.1004075

Editor: Chris D. Jiggins, University of Cambridge, United Kingdom

Received April 5, 2013; **Accepted** November 18, 2013; **Published** January 9, 2014

Funding: We have received funding from the Research Council of Norway, grant number: 204523, Molecular Life Science (MLS), University of Oslo, Norway, grant number: NA, and Centre for Ecological and Evolutionary Synthesis (CEES), University of Oslo, Norway, grant number: NA. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: g.p.satre@bio.uio.no

ϑ These authors contributed equally to this work.

Introduction

Hybridization between divergent populations has diverse impacts on evolution [1–3], including the rapid formation of hybrid species [1–6]. Homoploid hybrid speciation (HHS) is the process through which hybridization between two taxa results in a third, novel taxon that remains distinct by means of reproductive barriers against both parent taxa, without a change in number of chromosome sets. This mode of speciation is thought to be rare in nature as hybridization must be initiated by gene exchange between two taxa, but this gene exchange also subsequently reduces the likelihood of hybrid speciation occurring. Gene flow from the parents must be countered or reduced after initial contact despite complementary ploidy levels and weak initial isolation [2,5].

In non-hybrid speciation involving species with chromosomal sex-determination, sex-linked genes have repeatedly been found to strongly influence reproductive isolation (RI) [7–10]. The prominent role of sex chromosomes as reproductive barriers between non-hybrid species is attributed to a fast rate of genetic divergence, exposure of recessive alleles to selection in the heterogametic sex, a predominance of genes with sexual functions, and high average linkage between the disproportionately many sex-linked genes involved in RI [7,8]. However, an equally important role for sex chromosomes in HHS is far from settled. One of the most likely mechanisms for HHS is thought to be through transgressive segregation: the production of trait values outside the range of both parent taxa in the hybrids, allowing adaptation to ecological niches unavailable to those parent taxa [5,11]. Whereas divergent selection on reproductive traits is expected to be heavily involved in non-hybrid speciation, increasing the influence of sex chromosomes, transgressive ecological adaptations are more likely to be autosomal and evolve under stabilizing selection in parents [8,11–13]. Divergent selection produces extreme phenotypes when it leads to parent taxa being fixed for alleles with opposite effects at each locus. Hence in this situation, additive genetic variation leads to intermediate hybrid trait values [11,13]. In contrast, divergence under stabilizing selection occurs through weakly selected turnover of alleles contributing to a trait with an intermediate optimum. This promotes divergence on autosomes [12], and leads to a

Author Summary

Hybridization between two species has the potential to create a third, hybrid species. However this process, known as hybrid speciation, is thought to be unlikely because it requires reproductive barriers against both parent species to develop despite the barriers between parents being weak enough to allow for the formation of viable, fertile hybrids. The Italian sparrow, which occupies the entire Italian peninsula and some Mediterranean islands, is the product of past hybridization between house and Spanish sparrows and therefore represents one of the few documented cases of vertebrate hybrid speciation in nature. We show that reproductive barriers between Italian sparrows and their parent species exist and that genes on the sex (Z) chromosome and mitochondria are heavily involved. We suggest that speciation in this system may have been driven by dissociation of the sex (Z) chromosome into blocks of different parent species-specific genes, which have shifted alongside mitochondrial genes to form reproductive barriers where the hybrid now meets each of its parent species.

mixture of loci fixed for alleles with both positive and negative effects on trait values in each taxon. Thus, through additive effects alone, hybrids with more positive- or negative-effect alleles than in either parent taxon will often be produced, leading to transgressive phenotypes [11].

The study of RI in hybrid species systems can be complicated by a lack of geographical overlap between the hybrid and one or both of its parent species [5]. In *Passer* sparrows, however, the distribution of the hybrid Italian sparrow *Passer italiae* [14,15] overlaps with those of both its parent species, the Spanish sparrow *P. hispaniolensis* and the house sparrow *P. domesticus* (Figure 1) [14] allowing for the study of reproductive barriers. The Italian sparrow is in contact with the house sparrow in a stable, narrow hybrid zone in the Alps [14,16] and with the Spanish sparrow in a recently established sympatric zone in southeast Italy. In Sardinia, off the west coast of Italy, Spanish sparrows occur allopatrically (Figure 1). House and Spanish sparrows are themselves broadly, and often locally, sympatric across the entire Spanish sparrow range, remaining phenotypically distinct in all but a few locations [16]. Hence reproductive barriers exist and are typically effective in maintaining isolation between the parent species, but can be broken down to form viable hybrid populations and species.

Previous studies have indicated that Italian sparrows are almost fixed for house sparrow mitochondrial DNA [14,15]. Moreover, two markers on the Z chromosome (birds are female-heterogametic with a ZZ/ZW sex chromosome system) were found to be fixed for the Spanish sparrow allele in one case (*CHD1Z*) and nearly fixed for the house sparrow allele in the other (*PLAA*), indicating strong mosaicism on the Z chromosome not paralleled by autosomal markers [15].

The evidence of Z chromosome mosaicism from existing studies [15] may indicate that Italian sparrow isolating mechanisms are more similar to those involved in non-hybrid speciation than would be expected given a strong influence of transgressive segregation. Furthermore, *CHD1Z* has been shown to be under divergent selection and associated with RI between several non-hybrid bird species pairs [15 and references therein]. For example, *CHD1Z* shows evidence of divergent selection in sympatric but not allopatric population comparisons of pied and collared flycatchers (*Ficedula hypoleuca* and *F. albicollis*) [17]. Should *CHD1Z* prove to be an informative marker in Italian sparrows, it may be particularly likely to represent a functional variant directly involved in RI (or at least a linked marker within the same gene) rather than a linked marker within a neutral gene.

Evidence is accumulating that mito-nuclear interactions cause postzygotic isolation and are influential in speciation [18,19]. Mitochondrial DNA also commonly shows strongly shifted clines relative to nuclear markers, and this is often attributed to adaptive introgression [20]. Such introgression can occur in the face of detrimental effects on males, a phenomenon known as "mother's curse" [21,22]: selection in males has no direct effect on mitochondrial fitness due to maternal inheritance, causing a selective sieve allowing the excessive build-up of male-detrimental mutations. Hence, if "mother's curse" is acting in the Italian sparrow we would expect male-compensatory alleles to track the spread of house sparrow mitochondria and show concordant clines.

Here, we analyze species-informative single nucleotide polymorphism (SNP) markers, located within functional transcribed genes, across the breeding range of the Italian sparrow, as well as the contact zones with its parent species. We use a cline analysis approach [23] to identify candidate hybrid-parent RI genes, and hence to elucidate the mechanisms involved in HHS. In particular, we look for evidence of coincidence between nuclear and mitochondrial clines and discuss the possibility that they represent the outcome of mother's curse in a hybrid species, and whether this mechanism may be influential in hybrid speciation. We test whether markers with clines falling on current hybrid-parent species boundaries are disproportionately (i) Z-linked, thus showing similarities with non-hybrid speciation, or (ii) autosomal, suggesting differences in mechanisms of hybrid speciation relative to non-hybrid speciation and a greater influence of stabilizing selection and transgressive segregation.

Results

We identified putatively parent species-diagnostic SNP markers through transcriptome sequencing of six individuals of each of the two parent species. After filtering for coverage, sufficient flanking sequence, genotyping success (>97%) and multiple SNPs per gene, we retained 86 species-informative, genic, nuclear and mitochondrial SNP markers from 84 genes for analysis. Using this marker set, we found the Italian sparrow to exhibit high levels of genomic admixture over the entire study area (Figure 1 and Table S1). We also found evidence for on-going but restricted gene exchange between Italian sparrows and house sparrows in the contact zone in the Alps (Figure S1), though no evidence for gene exchange between Italian and Spanish sparrows in the sympatric zone in southeast Italy (see below). However, STRUCTURE [24] analysis revealed evidence of migration between Italian sparrow populations on mainland Italy and Spanish sparrow populations on Sardinia. Early generation migrants were present in both locations indicating ongoing gene flow through dispersal events (Figure 1).

As gene flow was observed between the Italian sparrow and both parent species, we implemented a cline analysis framework to look for genes exhibiting steep clines, and therefore decreased gene flow at the species boundaries. The SNPs are within functional coding genes and hence any such clines may indicate a direct influence of the gene on RI. They may also, however, be neutral but closely linked to loci under selection. These genes nevertheless represent the most likely candidates to be involved in RI at this inferential stage of analysis.

Cline analysis is a method used to measure the steepness, shape and location of changes in allele frequency or locus-specific

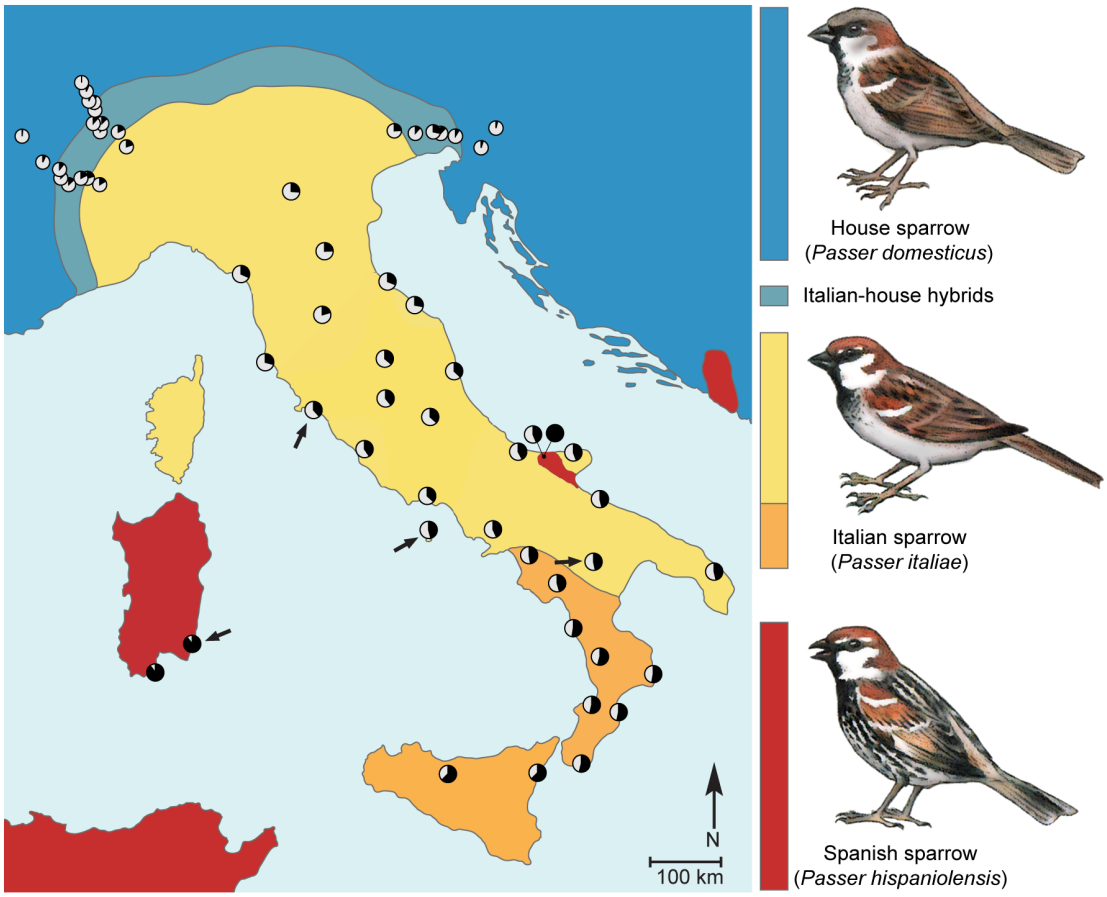


Figure 1. Phenotypic and genetic makeup of the hybrid Italian sparrow. Coloration of the map denotes phenotypic distribution as indicated by the bird drawings to the right of the map (blue: house sparrow, turquoise: Italian-house hybrids, yellow: typical Italian sparrow, orange: Italian sparrows with plumage intermediate between typical Italian and Spanish sparrows, red: Spanish sparrow). Bird drawings indicate species-specific male plumage characteristics of the three taxa [16]. Pie charts denote mean hybrid index at sampling localities where white and black color indicate house and Spanish sparrow genetic contribution, respectively. Locations with evidence of recent gene exchange between Spanish and Italian sparrows are indicated by arrows.
doi:10.1371/journal.pgen.1004075.g001

ancestry, as well as in quantitative traits [23,25]. It is typically used to examine geographic clines where, given the assumption that the cline is maintained by a balance between dispersal into a hybrid zone and selection against hybrids [25–27], various parameters including the strength of selection acting on traits or loci can be estimated. Cline analysis is therefore useful for identifying loci involved in RI in hybrid zones [23]. However, many contact zones do not conform to the assumption of a dispersal/selection balance and show a more complex pattern of contact and changes in locus-specific ancestry or allele frequency. This has led to the emergence of genomic cline analysis, in which geographic distance is replaced by a 'hybrid index', and cline width and location represent the amount and bias of introgression at a locus into the foreign genomic background [23,28–30]. With the caveat that use of genomic clines does not fully remove the influences of genetic drift and geographic structure alongside selection on introgression, these analyses can be employed on any geographic pattern of contact and so are amenable for use in studies of hybrid speciation.

Bayesian genomic cline analysis (*BGC*) [29,31] fits the Barton cline and estimates the parameters α (excess of house or Spanish alleles) and β (rate or steepness of cline), analogous to geographic cline center and width respectively [28,30]. As there was some variation between runs, our *BGC* analysis revealed 31–35 genes to have excess house sparrow ancestry while 18–22 genes exhibited excess Spanish sparrow ancestry (Figure 2, Figure 3, and Figure S2). Furthermore, 25–27 genes exhibited steeper clines than neutral expectations (Figure 3, Figure S2) while 14–16 genes exhibited clines shallower than neutral expectations. Of the 25–27 genes with steeper clines than neutral expectations, 10 exhibited excess house sparrow ancestry, and another 5–8 exhibited excess Spanish sparrow ancestry. The remaining 8–10 genes exhibiting steep clines were not significantly shifted in either parental direction (Figure S2).

Combined *BGC* and geographical analysis using Geneland [32,33] further revealed seven genes to exhibit abrupt allele frequency shifts and thus steep clines at the hybrid-parent range boundaries (Figure 2 and Figure 3, Table S2). Of these seven genes, five (i.e. 71.4%) were Z-linked (Figure 2 and Figure 3). Three of these Z-linked genes shifted at the Italian-Spanish boundary (Figure 2 and Figure S2) alongside mitochondrial *ND2*, whereas clines in two Z-linked and one autosomal gene were located at the Italian-house boundary (Figure 2, Figure 3 and Figure S3). These results indicate a mosaic pattern of introgression along the Z chromosome, as predicted by previous results [15].

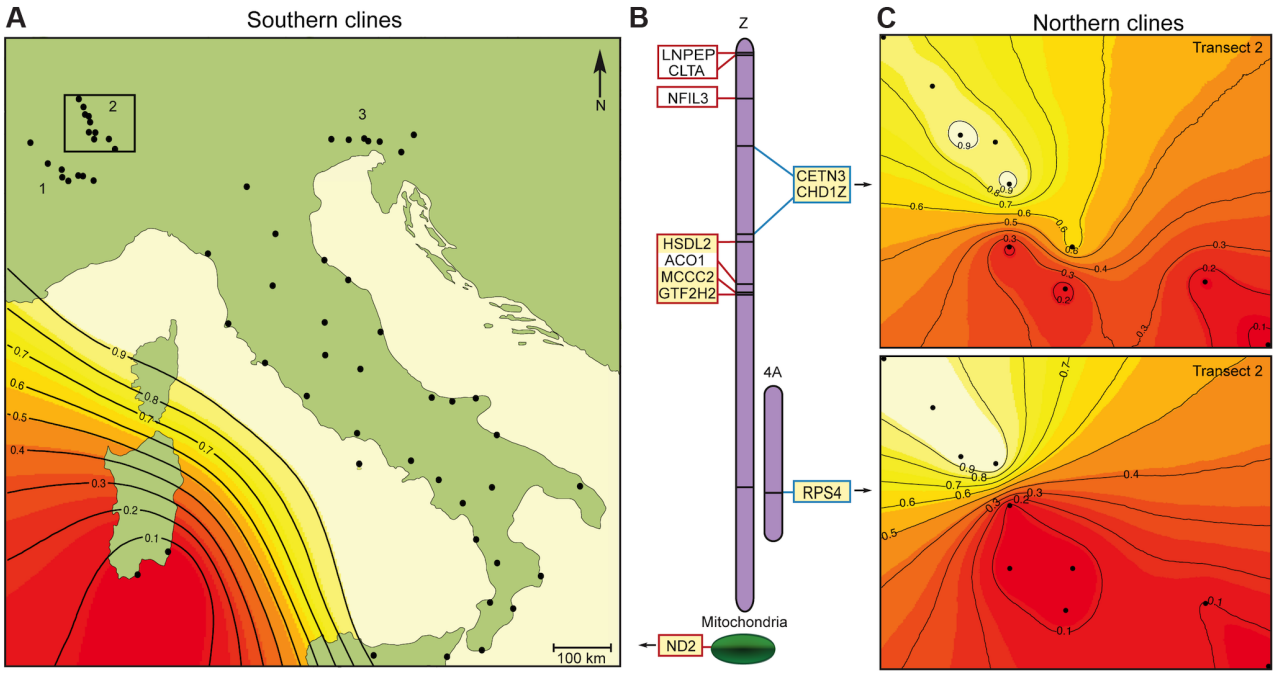


Figure 2. Genetic incompatibilities between the hybrid Italian sparrow and its parent species. (A) Representative geographic cline (*ND2*) for the mitochondrion and Z-linked genes shifting significantly between the Italian and Spanish sparrows of mainland Italy/Sicily and Sardinia. Colors refer to posterior likelihood of belonging to group corresponding to the Italian sparrow (>0.9, no color) relative to the Spanish sparrow (<0.1, red). The numbers refer to three transects through the Italian-house sparrow hybrid zone in the Alps. Black dots denote sampling locations. (B) Genomic location (in zebra finch) of genes inferred to be involved in hybrid-parent reproductive isolation. Blue outlines denote genes shifting significantly between the Italian and house sparrow, and red outlines denote genes shifting significantly between the Italian and Spanish sparrow. Markers highlighted in yellow have significantly steeper clines (significant β) than the neutral expectation according to a *BGC*-analysis (see main text) in addition to being significantly skewed towards either hybrid-parent species boundary (significant α), and hence represent the strongest candidate RI genes. Markers in white have significant α only. Chromosomal location for the Z-linked and Chr. 4A genes are indicated. (C) Geographic clines along transect 2 for the three genes shifting significantly in the Italian-house sparrow hybrid zone in the Alps. Upper panel shows results from the Z-linked genes *CHD1Z*/*CETN3*, lower panel shows results for the autosomal gene *RPS4*. Colors refer to posterior likelihood of belonging to group corresponding to the house sparrow (>0.9, white) relative to the Italian sparrow (<0.1, red). Black dots denote sampling locations.
doi:10.1371/journal.pgen.1004075.g002

There was a significant overrepresentation of Z-linkage among the genes exhibiting steep clines at the species boundaries considering that the Z chromosome holds about 3–7% of the genome of birds [9,34,35] (One-tailed binomial test: null probability based on flycatcher genome = 0.066, successes = 5, trials = 7, $P = 2.35 \times 10^{-5}$).

As observed in previous studies [14,15], we found Italian sparrows to be nearly fixed for house sparrow mitochondrial haplotypes (Figure 2 and Figure 3; Table S2). Moreover, two of the three Z-linked genes that exhibit steep clines at the Italian-Spanish boundary are classified as nuclear-encoded mitochondrial proteins (*HSDL2* and *MCCC2*) [36,37]. This is a significant overrepresentation of mitochondrial function compared to 8.3% in chickens [34,37] (One-tailed binomial test: null probability = 0.083, successes = 2, trials = 3, $P = 0.02$; null probability data from The Gene Ontology Project's Gene Association file for *Gallus gallus*, GOC validation date: 4 December 2012, gaf-version 2.0 and The Gene Ontology Project's Gene Ontology file, date: 4 December 2012, cvs revision version 4708) among the nuclear genes shifting at this boundary. The Z-linked gene *HSDL2* exhibited a near-identical pattern of fixation for house sparrow alleles in the Italian sparrow as the mitochondrial marker *ND2* (Figure 3; Table S2).

In three transects through the hybrid zone in the Alps, the Z-linked markers *CETN3* and *CHD1Z* and autosomal marker *RPS4* exhibited the steepest clines (Figure 2 and Figure 3). Unlike in other bird species [9,34,35], *CHD1Z* and *CETN3* appear to be tightly linked in sparrows (Figure S4). Outlier analyses indicated that *CHD1Z* but not *CETN3* is a candidate for being under divergent selection (Figure S5).

The three markers *MCCC2*, *GTF2H2* and *HSDL2* also show evidence of statistical association, although *HSDL2* is predicted to be a long physical distance from the other two based on the zebra finch genome (Figure S4). A conservative estimate can therefore be made that two out of four sets of markers (*ND2*, *RPS4*, one from *CHD1Z*/*CETN3* and one from *HSDL2*/*MCCC2*/*GTF2H2*) with steep clines on range boundaries are Z-linked. This remains a significant overrepresentation of Z-linked markers (One-tailed binomial test: null probability = 0.066, successes = 2, trials = 4, $P = 0.02$).

While the steepest genomic clines were found for markers with major geographic clines at the Italian sparrow range boundaries (Figure 2 and Figure 3; Figure S2, Table S2), clines steeper than neutral expectations were also found within the Italian sparrow's range in a number of both autosomal and Z-linked genes (Figure 3c and Figure S2), some of which were also significantly shifted (significant α; Figure 3 a–c, Figure S2). Seven other markers were strongly shifted towards an excess of Spanish sparrow alleles, but had clines much shallower than the rest (Figure 3 a–c). Not all of the α and β estimates for these markers were significant however, and the parental allele frequency difference was <0.5 in every case. With such a low parental allele frequency difference and

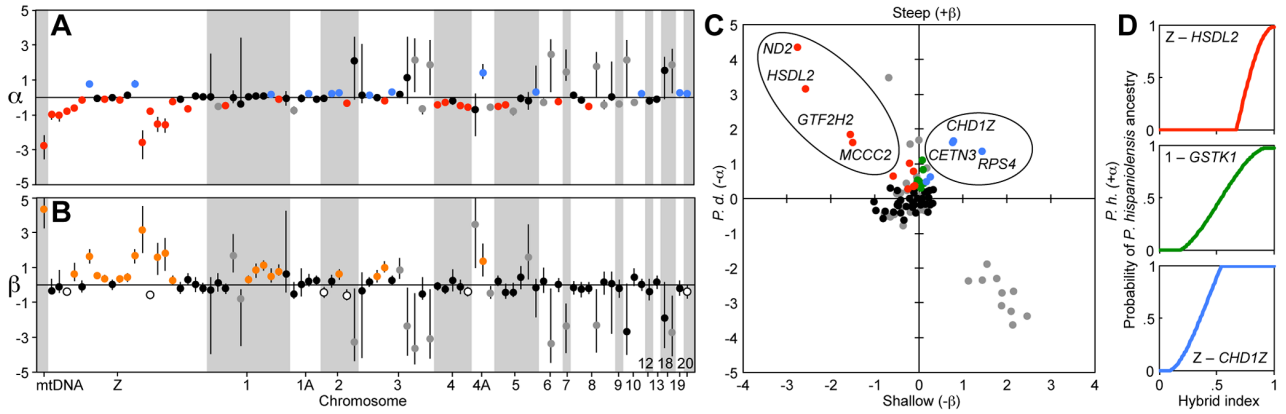


Figure 3. Candidate RI genes revealed by genomic cline analysis. (A–B) Markers are ordered along chromosomes as in Table S3. (A) *BGC* run 2 estimates of genomic cline center (α) with 95% credibility intervals for all 86 SNP markers. Red dots indicate markers with significant excess house sparrow ancestry, blue dots indicate markers with significant excess Spanish sparrow ancestry, grey dots indicate markers with either significant house or Spanish excess ancestry but where the difference in allele frequency between the parent species are below 0.5, and black dots indicate markers that do not differ from neutral expectations. (B) *BGC* run 2 estimates of genomic cline rate (β) with 95% credibility intervals for all 86 SNP markers. Orange dots indicate markers with clines significantly steeper than neutral expectations, white dots indicate markers with significantly shallower clines than neutral expectations, grey dots indicate markers with either significantly steeper or shallower clines than neutral expectations but where the difference in allele frequency between the parent species is below 0.5, and black dots indicate markers that do not differ from neutral expectations. (C) Genomic cline center (α) plotted against genomic cline rate (β). Red dots indicate markers that exhibit significant excess house sparrow (*P. d.*) ancestry, clines steeper than neutral expectations and where the difference in allele frequency between the parent species is greater than 0.5. The red dots that are encircled and named shift at the Italian-Spanish boundary and are hence candidate RI genes. Blue dots indicate markers that exhibit significant excess Spanish sparrow (*P. h.*) ancestry, clines steeper than neutral expectations and where the difference in allele frequency between the parent species is greater than 0.5. The blue dots that are encircled and named shift at the Italian-house boundary and are hence candidate RI genes. Green dots indicate markers that have allele frequency differences between the parent species greater than 0.5 and clines steeper than neutral expectations but do not exhibit excess house or Spanish sparrow ancestry. These markers are candidates for being incompatibilities within the Italian sparrow. Grey dots indicate markers in which the allele frequency difference between the parent species is less than 0.5. (D). Examples of *BGC* genomic clines representative of the marker categories described in panel (C). For illustrations of all clines, see Figure S2.

doi:10.1371/journal.pgen.1004075.g003

significantly shallow rather than steep clines, we do not consider these as potential RI genes. However, this does not rule out other forms of selection on these loci within the hybrid species.

Whereas Sardinian Spanish sparrows show evidence of ongoing introgression from Italian sparrows, Spanish sparrows in the recently established sympatric population in southeast Italy appear to be genetically pure (Table S1, 'Lesina (Spanish)'). We found no difference in F_{ST} between these Spanish sparrows and sympatric Italian sparrows versus Spanish sparrows and nearby allopatric Italian sparrows (Table 1). Thus, there was no sign of gene flow in sympatry between Italian and Spanish sparrows.

Discussion

Reproductive barriers between the hybrid Italian sparrow and its parent taxa

In this study, we utilized the Italian sparrow's gene exchange and geographical overlap with both parent species, house and Spanish sparrows, to investigate hybrid-parent reproductive barriers using a cline analysis framework. Results of geographic and genomic cline analyses reveal that several markers with an excess of Spanish sparrow alleles within Italian sparrows are associated with RI at the house-Italian range boundary, and that several markers with house sparrow excess are associated with RI at the Spanish-Italian range boundary. A disproportionately high number of the markers showing steep clines at one or the other hybrid-parent species boundary are Z-linked, including when potential physical linkage is accounted for. Of these SNPs, which are all within functioning transcribed genes, the strongest candidates to be within genes directly involved in RI are those that exhibit the most extreme cline parameters. This includes *CHD1Z* (Z chromosome) and *RPS4* (chromosome 4A) at the Italian-house sparrow boundary in the Alps, and *HSDL2* (Z chromosome) and *ND2* (mitochondria) at the Italian-Spanish sparrow boundary between the Italian mainland and Sardinia. Furthermore, two of the three nuclear markers with steep clines coincident with that of the mitochondria at the Italian-Spanish boundary (*HSDL2* and *MCCC2*) are classified as nuclear-encoded mitochondrial proteins [36,37]; a statistical overrepresentation. As outlier loci revealed by cline analysis may result from genetic drift rather than from selection on introgression [29], we note that a significant overrepresentation of sex-linkage and mitochondrial function among candidate RI genes as reported here is not expected to arise through drift.

Excessive sex linkage of RI supports the hypothesis that HHS in Italian sparrows is facilitated by divergent selection between the

Table 1. Test for gene exchange in sympatry.

	Lesina Spanish	Guglionesi	Lesina Italian
Guglionesi	**0.4163*****		
Lesina Italian	**0.4105*****	0.0025	
Mass Montanari	**0.3926*****	0.0073	0.0041

FST values between Gargano peninsula populations of Spanish sparrows and sympatric (Lesina) and nearby allopatric Italian sparrows.
***indicates highly significant genotypic divergence, $P<0.0001$.

doi:10.1371/journal.pgen.1004075.t001

parent species. HHS differs from non-hybrid speciation in that an isolating mechanism is required against each parent. This may be aided by transgressive segregation leading to extreme hybrid phenotypes [5,11]; a process promoted more by stabilizing selection in the parent species than by divergent selection [11,13]. Stabilizing selection is more likely to produce transgressive hybrid phenotypes through purely additive effects due to the greater likelihood of complementary gene action [11]. While traits under divergent selection may also produce transgressive phenotypes in some circumstances, for example with epistasis, the emphasis on transgression in HHS represents a contrast to theories of non-hybrid adaptive speciation, in which genes under divergent selection are thought more likely to contribute to isolation [38,39].The mechanisms promoting Italian sparrow HHS may therefore more closely resemble those involved in typical non-hybrid speciation. While transgression cannot be ruled out, Italian sparrows appear phenotypically intermediate between the parents and share the house sparrow's human-commensal niche.

We recognize that the genes with steep clines may represent neutral markers in linkage disequilibrium with the RI genes under selection (this is likely to be the case for the mitochondrial *ND2*, which is in linkage disequilibrium with all other mitochondrial genes), rather than being directly involved in RI. In the case of *RPS4*, there is no *a priori* expectation that it should be involved in RI. However this gene, the sole autosomal representative associated with hybrid-parent isolation, is found on chromosome 4A in zebra finch [35]. This chromosome is orthologous to the mammalian X chromosome and *RPS4* is in fact X-linked in Eutherian mammals [40]. Chromosome 4A appears to be enriched for genes with properties similar to sex-linked genes, including the avian homolog to the human sex-determining *SRY*, and about one third of this chromosome has even translocated to the sex chromosomes in the whole avian superorder Sylvioidea [40]. Hence, the sole autosomal gene with a significantly steep cline on a species boundary may be linked to genes with properties more typical of sex-linked genes.

We hypothesized that *CHD1Z* may directly influence RI because it has been previously highlighted as a candidate speciation gene in other bird systems [15 and references therein]. This therefore represents a stronger RI gene candidate than *RPS4* or *CETN3* on the Italian-house boundary. Though *CHD1Z* and *CETN3* appear be closely linked to the same RI locus, our outlier analysis indicates *CHD1Z* is under divergent selection while *CETN3* is not, and is therefore a more likely candidate to be the RI locus. *CHD1* (the generic name for this gene in all organisms, represented by divergent Z-linked and W-linked copies in birds) is a chromatin-remodeling factor potentially affecting the expression of many genes [41]. Of particular interest is its essential role in chromosome centromere localization [42]. "Centromeric drive" is a proposed mechanism of intra-genomic conflict potentially causing rapid evolution of incompatibilities and speciation [43]. In this process, male-detrimental centromere drivers causing biased meiosis lead to selection for compensatory mutations in centromeric proteins, potentially including *CHD1*. Centromeric drive has been proposed as an important mechanism in the non-hybrid speciation of pied and collared flycatchers [9], in which *CHD1Z* also shows evidence of involvement in RI [17]. This adds weight to the argument that *CHD1Z* represents a candidate RI gene, maintaining parapatric differentiation between hybrid Italian sparrows and the house sparrow.

We also hypothesized a role for mito-nuclear interactions, and in particular the tracking of spreading mitochondrial variants by nuclear male 'restorer' genes, compensating for 'mother's curse'. The overrepresentation of nuclear genes with a mitochondrial function on the Italian-Spanish sparrow boundary provides some support for this hypothesis. However, only one of these genes, *HSDL2*, shows a cline almost as steep as mitochondrial *ND2*. While this may be a linked neutral marker, the fact that *HSDL2*'s protein product is located within mitochondria [37,44] supports its candidacy as an RI gene. *HSDL2* is thought to be involved in fatty acid metabolism, although its exact functions are unknown [44].

Our results thus appear consistent with the influence of mito-sex chromosome conflict acting as a reproductive barrier at the Spanish-Italian sparrow boundary. Furthermore, due to the lack of global dosage compensation in birds, Z-linked genes typically have higher expression and – combined with the fact that Z chromosomes spend two thirds of their time in males - stronger fitness effects in males than females. Consequently, genes with male-specific fitness effects are overrepresented on the Z chromosome [8,45]. We thus postulate that nuclear male-compensatory 'restorer' genes are most likely to occur on the Z chromosome, leading to reduced fitness in hybrids with mito-sex chromosome mismatches. *HSDL2* in fact has been shown to have higher expression levels in male than female chickens [46]. As an alternative to "mother's curse", isolation through co-adaptation between nuclear and mitochondrial genes is also possible. Because natural selection can only act on such bi-directional co-evolution through female fitness effects, we propose that mito-nuclear co-adaptation should involve disproportionately many autosomal genes, as they spend equal time in both sexes and show no overall sex bias in gene expression [45]. Hence our results are more consistent with mito-nuclear conflict.

There is no evidence for hybridization between Spanish and Italian sparrows in the sympatric zone of southeast Italy, supporting previous results [14]. This suggests a role for prezygotic barriers between the two taxa. Spanish sparrows are much less associated with humans than house and Italian sparrows and occupy a different habitat, providing some habitat-dependent assortative mating [16]. We suggest that evolution of Italian sparrows towards the house sparrow human-commensal niche may have contributed to rapid development of prezygotic isolation with Spanish sparrows alongside, or even reinforced by, the aforementioned mito-nuclear postzygotic barrier.

Reproductive barriers within the hybrid Italian sparrow

In addition to the steep clines at the species range boundaries, some genes exhibited clines steeper than the neutral expectation within the Italian sparrow's range. One possible interpretation of this result is that moving clines of incompatibility genes may have become trapped by environmental transitions or population density troughs before reaching the current hybrid-parent boundaries. In this way, intraspecific incompatibilities within a hybrid species may increase future diversification relative to non-hybrid species, in particular through the effect of divergence hitchhiking in promoting the build-up of novel isolating mechanisms surrounding pre-existing incompatibilities [47,48]. Isolation by adaptation may be occurring in Italian sparrows [49], so moving clines of incompatibility genes may also have become trapped by association with niche differentiation [50]. Nevertheless, neutral processes cannot be ruled out. The spatial spread of a partially reproductively isolated taxon into the range of the other taxon may lead to neutral allele frequency clines at historical invasion wave fronts [51], although these would become more diffuse over time since the spread.

The Italian sparrow genome represents a mosaic, particularly on the Z chromosome, in which some Spanish sparrow alleles

have spread to form reproductive barriers against house sparrows, while house sparrow alleles at different loci on the same chromosome have spread in the opposite direction to form barriers against Spanish sparrows. We envisage that HHS in the Italian sparrow may match the 'mosaic genome hybrid speciation' model [52] with the addition of secondary spatial spread of the hybrid genotype. Such discordant spread would most likely occur through selective sweeps, if one parental allele had a fitness advantage over the other in the mosaic genomic background. However, if strong selective sweeps occurred, some mechanism would be needed to cause them to stop at the current hybrid-parent boundaries. We note that these boundaries lie on major barriers to dispersal in the case of the *Passer* sparrows, and that beneficial alleles often do not sweep across the whole range of a species for a variety of reasons including geographical barriers.

Conclusions

Using a cline analysis framework we have identified sets of candidate RI genes and genomic regions between the hybrid Italian sparrow and its parent species. These results support our predictions that mito-nuclear interactions and loci on the Z chromosome strongly influence RI. In this regard, we suggest that HHS in the Italian sparrow resembles non-hybrid speciation, and we would therefore predict that the same loci would be involved in RI between the parent taxa; house and Spanish sparrows. An important next step is therefore to replicate this study in a region of parental sympatry and hybridization [53].

Materials and Methods

Sample collection

Only males, which are diploid for the Z chromosome, were analyzed to avoid issues related to haplodiploidy of the Z chromosome. Blood samples from the three taxa ($n = 612$) were taken from 64 locations between 2007 and 2011 (Figure 1 and Table S1): Spanish sparrows ($n = 142$) from Badajoz, Spain, Sardinia, Italy and a Spanish/Italian sympatric zone in southeast Italy; allopatric house sparrows ($n = 85$) from Hradec Králové, Czech Republic and Oslo, Norway; Italian sparrows, Italian-house hybrids and parapatric house sparrows from the Italian peninsula and the Alps ($n = 385$). The sparrows were caught using mist nets. About 25 µl of blood was extracted by venipuncture of a brachial vein and stored in 1 ml of Queens lysis buffer. Appropriate catching and sampling permits were obtained for all sampling locations from the relevant authorities. DNA was isolated using Qiagen DNeasy 96 Blood and Tissue Kits (Qiagen N.V., Venlo, Netherlands) according to the manufacturer's instructions. For transcriptome sequencing, three house (Oslo, Norway) and three Spanish (Badajoz, Spain) sparrows of each sex were sampled in October 2010. Liver, heart and brain tissue samples were taken and stored on RNAlater (100 mg tissue in 1400 µl buffer) according to the manufacturer's protocol.

Transcriptomic cDNA library preparation and sequencing

Total RNA isolation from pooled liver, heart and brain samples followed by normalized cDNA library preparation was performed by Vertis Biotechnologie AG, Freising, Germany. Total RNA was isolated from the cell powders using the mirVana RNA kit (Ambion) including an on-column DNase treatment. From the total RNA samples, poly(A)+ RNA was prepared and fragmented with ultrasound (1 pulse of 30 sec at 4°C). First-strand cDNA was synthesized from the fragmented RNA using a N6 randomized primer and M-MLV RNaseH-reverse transcriptase. 454 adapters A and B were ligated to the 5′ and 3′ ends of the cDNA. The cDNA was amplified with PCR using a proof reading enzyme. Normalization was carried out by one cycle of denaturation and reassociation of the cDNA, resulting in N1-cDNA. Reassociated ds-cDNA was separated from the remaining ss-cDNA (normalized cDNA) by passing the mixture over a hydroxylapatite column. After hydroxylapatite chromatography, the ss-cDNA was PCR amplified. For GS FLX Titanium sequencing, the cDNA in the size range of 450–700 bp was eluted from preparative agarose gels. The resulting cDNA was double stranded, and had a size of about 450–700 bp. The six samples from each species were pooled in equal amounts and pyrosequenced on a Roche GS FLX Titanium sequencer at the Norwegian Sequencing Center using the manufacturer's protocol.

Sequence alignment and SNP mining

The house sparrow reads were aligned and mapped against the zebra finch genome, and reads of both species were then mapped against the resulting contigs, providing a list of potential species-informative SNPs. Species-diagnostic SNPs from the twelve samples were chosen and subsequently filtered for those without sufficient flanking sequence for PCR-primer design. Genes were annotated by blasting against the zebra finch and chicken genomes. Two exceptions were SNPs within *CHD1Z* and *ND2* genes, which were genotyped using existing primers [15]. The two *CHD1Z* SNPs in this initial set are within an intron [15]. The genomes of *Passer* sparrows have so far not been mapped. Genomic locations of the various markers were therefore inferred based on the Zebra finch *Taeniopygia guttata* genome [35].

Genotyping

Multiplex sets of PCR primers were designed and all individuals genotyped at each SNP locus using the Sequenom MassARRAY system at CIGENE, Norwegian University of Life Sciences, Ås, Norway. A total of 124 putatively diagnostic SNP markers from 107 different genes were genotyped successfully. Statistical analyses were carried out on a subset of 86 species informative SNPs after further filtering (Table S3). This involved removing SNPs with <97% genotyping success over all samples, plus removing all but one SNP from each gene (except in *APC* and *A2ML1*, in which two markers were included due to marked differences in parental allele frequencies).

Detection of gene exchange

Recent migration between Spanish sparrows on Sardinia and Italian sparrows was identified using the USEPOPINFO model in STRUCTURE [22], with 100,000 iterations, a burnin of 50,000, GENSBACK set to three generations and the rest of the settings as default. A hybrid index value [54] was calculated for each individual, based on the 86 SNPs. In the Alps transects, the presence of many individuals with intermediate hybrid index indicated hybridization with house sparrows (Figure S1).

Detection of SNPs associated with reproductive isolation

If genotype frequencies for individual SNPs change more sharply with changing hybrid index than a neutral expectation, this indicates a potential association with reproductive isolation. Examination of individual SNP cline width with respect to hybrid index was carried out using Bayesian genomic clines method as implemented in *BGC* [31]. Spanish sparrows from Badajoz and Gargano, and house sparrows from Oslo and Hradec Králové,

were used to indicate parental genotype frequencies (Table S3). For the genomic clines analyses, all individuals from the Alps, the Italian peninsula (excluding Spanish sparrows from the southeast Italian sympatric zone), Sicily and Sardinia were pooled into one admixed population. Genotypes at the mitochondrial *ND2* locus were coded as diploid homozygote as *BGC* failed to run with a haploid marker included. Three independent runs with 100,000 iterations each were run with the first 25,000 iterations discarded as burnin, MCMC samples thinned by recording every fifth value, while the rest of the *BGC* settings were as default. SNPs were identified as significantly deviating from null expectations when the 95% credibility intervals of the cline parameters α and β did not cross zero. Once candidate SNPs were chosen using the genomic clines approach, geographic locations of sharp changes in allele frequency for each SNP were identified in GENELAND [32,33] using the uncorrelated allele frequency model and allowing 1–10 clusters. Each SNP was run three times at 3 million iterations with a burn-in of 200. For some SNPs Geneland identified more than two geographic clusters, indicating multiple rapid changes in allele frequency. In these cases the main cline was determined to be on a species boundary if Geneland identified a cline on that boundary and *BGC* indicated a significantly shifted cline center (significant α) (Figure 2 and Figure S2). We further narrow our focus primarily to markers that also have significantly steep clines (significant positive β) in all three *BGC* runs (Figure 2 and Figure S2).

Linkage disequilibrium

On top of using the zebra finch genome to identify gene location, genetic linkage was estimated using GENEPOP [55] to calculate a *P* value for genotypic disequilibrium between every SNP pair and by employing a Fisher test to combine probabilities across all populations in the Italian peninsula, the Alps and Sardinia (Figure S4).

Population divergence in sympatric Italian and Spanish sparrows

To assess if there is introgression between sympatric Italian and Spanish sparrows, F_{ST} values and genotypic differentiation were calculated between four populations in southeast Italy (Lesina, which is a sympatric population of Italian and Spanish sparrows, and the nearby allopatric Italian sparrow populations of Mass. Montanari and Guglionesi) in GENEPOP [55] using males and the chosen 86 SNPs. F_{ST} estimates were calculated between i) the Italian sparrows from each population and ii) the Spanish sparrows in Lesina and the Italian sparrows in each of the three populations. A shift in F_{ST} towards or away from Spanish sparrows in Italian sparrows from Lesina, relative to the two allopatric populations, would indicate introgression or displacement respectively.

Outlier tests for divergent selection

BAYESCAN [56] and LOSITAN [57] are softwares that implement methods to test for evidence of both divergent and balancing selection through F_{ST} outlier analysis of molecular markers. We used both softwares to determine which of *CHD1Z* and *CETN3*, two highly linked markers, showed stronger evidence for selection in the Alps house-Italian sparrow hybrid zone. Data from all 85 nuclear SNPs (haploid mtDNA markers cannot be run alongside diploid markers, and *ND2* is invariant in the Alps) were used and the three transects were pooled, excluding sites with just a single individual sampled. For LOSITAN the options 'neutral mean F_{ST}' and 'force mean F_{ST}' were chosen, along with the infinite alleles mutation model and 50 k simulations. For BAYESCAN, default settings were used.

Ethics statement

Handling of birds were conducted according to guidelines approved by the relevant authorities in the respective countries (Museum National d'Histoire Naturelle, Centre de Recherches sur la Biologie de Populations d'Oiseaux, Paris (France), Institute for Environmental Protection and Research – ISPRA (Italy), Consejería de Industria, Energía y Medio Ambiente (Spain), Norwegian Food Safety Authority (Norway), Ministrstvo za okolje in proctor, Agencija Republike Slovenije za okolje (Slovenia) and Bundesamt für Umwelt BAFU, Abteilung Artenmanagement (Switzerland)).

Supporting Information

Figure S1 Individuals with intermediate crown color have already been reported in the Alps contact zone [14,16]. Further evidence of hybridization comes from a large proportion of individuals with hybrid index intermediate between house and Italian sparrows in the contact zone. Hybrid index for each individual is plotted against distance from the house sparrow end of each transect, along a) transect 1 (see Fig. 2), b) transect 2 and c) transect 3. d) Histogram of hybrid indices for populations at the house sparrow end of the three transects ('House'), the rest of the transect collections ('Hybrid zone'), and northern Italian populations away from the transects ('Italian').
(DOC)

Figure S2 Bayesian genomic clines for all 86 SNP markers. The legend in each panel indicates genomic location (chromosome number), SNP name and significance of cline center (α) and cline rate (β) parameters for three independent *BGC* runs. Significant excess house sparrow ancestry is indicated by H's, significant excess Spanish sparrow ancestry is indicated by S's, significantly steep clines are indicated by +'s and significantly narrow clines are indicated by −'s. The number of H's, S's, +'s and −'s indicate for how many runs the parameter estimates for a given marker differed from neutral expectations. When none of the runs differed from neutral expectations, this is indicated by ns. Red panels indicate markers that exhibit significant excess house sparrow ancestry and have clines steeper than neutral expectations, where the difference in allele frequency between the parent species is greater than 0.5 and for which Geneland revealed a shift at the Italian-Spanish sparrow boundary. These are hence candidate RI genes. Blue panels indicate markers that exhibit significant excess Spanish sparrow ancestry, clines steeper than neutral expectations, a difference in allele frequency between the parent species greater than 0.5 and with a shift at the Italian-house sparrow boundary. These are hence candidate RI genes. Green panels indicate markers that have allele frequency differences between the parent species greater than 0.5 and where the clines are steeper than neutral expectations but do not exhibit a major allele frequency cline on either hybrid-parent species boundary. These markers are candidates for being incompatibilities within the Italian sparrow. Only markers with significant '+' on more than one *BGC* run are highlighted.
(DOC)

Figure S3 GENELAND geographic clines for Alps transects 1–3 (Fig. 2a), for the three loci exhibiting rapid changes in allele frequency coinciding with the hybrid zone between Italian and house sparrows (*CHD1Z*, *CETN3* and *RPS4*; see main text). Axes represent longitude (x axis) and latitude (y axis) in decimal degrees.

Colors refer to posterior likelihood of belonging to the group corresponding to the house sparrow (>0.9, white) relative to the Italian sparrow (<0.1, red). Black dots denote sampling locations. *CHD1Z* and *RPS4* results for transect 2 are also represented in Fig. 2c.
(DOC)

Figure S4 Heat map of genotypic disequilibrium between Z-linked markers with significantly steep genomic clines. Marker positions in kbp are based on the zebra finch genome. Internal cell values indicate distances between genes in kbp. Cell colors represent genotypic disequilibrium, combining *P* values across all Italian peninsula and Sardinian populations. Dark blue $P = 0$; medium blue $P < 0.05$; pale blue $P < 0.1$.
(DOC)

Figure S5 F_{ST}-outlier analyses. (A) LOSITAN graphic of test for evidence of divergent selection in the Alps, using the 85 nuclear SNPs and combining data from all three transects (not including populations represented by a single individual). SNPs in the red area are candidates for positive divergent selection. At a false discovery rate of 0.05 only *CHD1Z* and *RPS4* were significant outliers for divergent selection. (B) BAYESCAN graphic of test for evidence of divergent selection in the Alps, using the 85 nuclear SNPs and combining data from all three transects (not including populations represented by a single individual). The estimated alpha coefficient indicates the strength and direction of selection. A positive value of alpha suggests diversifying selection, whereas negative values suggest balancing or purifying selection. *CHD1Z* and *RPS4* (highlighted in red and named) were the only markers that were significant at a false discovery rate of 0.1.
(DOC)

Table S1 Sample population details.
(DOC)

Table S2 F_{ST} between Italian sparrow and its parent species where a steep cline exists on the hybrid-parent boundary, plus estimates of cline shift (α) and steepness (β). Values in shaded boxes are not significant (95% credibility intervals overlap with zero).
(DOC)

Table S3 Details of SNPs used in analyses.
(DOC)

Acknowledgments

We thank S. A. Sæther, T. Borge, T. O. Elgvin, F. Haas, A. Marzala-Reynolds and P. Munclinger as well as numerous field assistants for help during fieldwork, A. J. Nederbragt, O.K. Tørresen and S. Henriksen for bioinformatics help, H. Hegdal for artwork and T. O. Elgvin, F. Eroukhmanoff, Ø. H. Holen and S. A. Sæther for helpful discussions.

Author Contributions

Conceived and designed the experiments: GPS RIB. Performed the experiments: CNT JSH GPS RIB. Analyzed the data: CNT JSH RIB. Contributed reagents/materials/analysis tools: Wrote the paper: CNT JSH GPS RIB.

References

1. The *Heliconius* Genome Consortium (2012) Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature 487: 94–98.
2. Abbot R, Albach D, Ansell S, Arntzen JW, Baird SJE, et al. (2013) Hybridization and speciation. J Evol Biol 26: 229–246.
3. Sætre GP (2013) Hybridization is important in evolution, but is speciation? J Evol Biol 26: 256–258.
4. Buerkle CA, Morris RJ, Asmussen MA, Rieseberg LH (2000) The likelihood of homoploid hybrid speciation. Heredity 84: 441–451.
5. Mallet J (2007) Hybrid speciation. Nature 446: 279–283.
6. Buerkle CA, Rieseberg LH (2008) The rate of genome stabilization in homoploid hybrid species. Evolution 62: 266–275.
7. Sæther SA, Sætre G-P, Borge T, Wiley C, Svedin N, et al. (2007) Sex chromosome–linked species recognition and evolution of reproductive isolation in flycatchers. Science 318: 95–97.
8. Qvarnström A, Bailey RI (2009) Speciation through evolution of sex-linked genes. Heredity 102: 4–15.
9. Ellegren H, Smeds L, Burri R, Olason PI, Backström N, et al. (2012) The genomic landscape of species divergence in *Ficedula* flycatchers. Nature 491: 756–760.
10. Meyer M, Kircher M, Gansauge M-T, Li H, Racimo F, et al. (2012) A high-coverage genome sequence from an archaic Denisovan individual. Science 338: 222–226.
11. Rieseberg LH, Archer MA, Wayne RK (1999) Transgressive segregation, adaptation and speciation. Heredity 83: 363–372.
12. Charlesworth B, Coyne JA, Barton NH (1987) The relative rates of evolution of sex chromosomes and autosomes. Am Nat 130: 113–146.
13. Bailey RI, Eroukhmanoff F, Sætre G-P (2013) Hybridization and genome evolution II: Mechanisms of species divergence and their effects on evolution in hybrids. Curr Zool 59: 675–685.
14. Hermansen JS, Sæther SA, Elgvin TO, Borge T, Hjelle E, et al. (2011) Hybrid speciation in sparrows I: phenotypic intermediacy, genetic admixture and barriers to gene flow. Mol Ecol 20: 3812–3822.
15. Elgvin TO, Hermansen JS, Fijarczyk A, Bonnet T, Borge T, et al. (2011) Hybrid speciation in sparrows II: a role for sex chromosomes? Mol Ecol 20: 3823–3837.
16. Summers-Smith JD (1988) The Sparrows: A study of the genus *Passer*. Calton: T & AD Poyser. 342 p.
17. Backström N, Lindell J, Zhang Y, Palkopoulou E, Qvarnström A, et al. (2010) A high-density scan of the Z chromosome in *Ficedula* flycatchers reveals candidate loci for diversifying selection. Evolution 64: 3461–3475.
18. Burton RS, Barreto FS (2012) A disproportionate role for mtDNA in Dobzhansky– Muller incompatibilities? Mol Ecol 21: 4942–4957.
19. Pritchard VL, Edmands S (2013) The genomic trajectory of hybrid swarms: outcomes of repeated crosses between populations of *Tigriopus californicus*. Evolution 67: 774–791.
20. Toews DPL, Brelsford A (2012) The biogeography of mitochondrial and nuclear discordance in animals. Mol Ecol 21: 3907–3930.
21. Frank, SA Hurst LD (1996) Mitochondria and male disease. Nature 383: 224.
22. Gemmel NJ, Metcalf VJ, Allendorf FW (2004) Mother's curse: the effect of mtDNA on individual fitness and population viability. Trends Ecol Evol 19: 238–244.
23. Payseur BA (2010) Using differential introgression in hybrid zones to identify genomic regions involved in speciation. Mol Ecol Resour 10: 806–820.
24. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155: 945–959.
25. Barton NH, Hewitt GM (1985) Analysis of hybrid zones. Annu Rev Ecol Syst 16: 113–148.
26. Barton NH, Hewitt GM (1989) Adaptation, speciation and hybrid zones. Nature: 341: 497–503.
27. Barton NH, Gale KS (1993) Genetic analysis of hybrid zones. In Harrison RG, editor. Hybrid zones and the evolutionary process. Oxford: Oxford Univ. Press. pp. 13–45.
28. Szymura JM, Barton NH (1986) Genetic analysis of a hybrid zone between the fire-bellied toads, *Bombina bombina* and *B. variegate*, near Cracow in southern Poland. Evolution 40: 1141–1159.
29. Gompert Z, Buerkle CA (2011) Bayesian estimation of genomic clines. Mol Ecol 20: 2111–2127.
30. Fitzpatrick BM (2013) Alternative forms for genomic clines. Ecol Evol 3: 1951–1966.
31. Gompert Z, Buerkle CA (2012) bgc: Software for Bayesian estimation of genomic clines. Mol Ecol Resour 12: 1168–1176.
32. Guillot G, Estoup A, Mortier F, Cosson JF (2005) A spatial statistical model for landscape genetics. Genetics 170: 1261–1280.
33. Guillot G, Mortier F, Estoup A (2005) GENELAND: a computer package for landscape genetics. Mol Ecol Notes 5: 712–715.
34. International Chicken Genome Sequencing Consortium (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432: 695–716.
35. Warren WC, Clayton DF, Ellegren H, Arnold AP, Hillier LW, et al. (2010) The genome of a songbird. Nature 464: 757–762.
36. Pagliarini DJ, Calvo SE, Chang B, Sheth SA Vafai SB, et al. (2008) A mitochondrial protein compendium elucidates complex I disease biology. Cell 134: 112–123.
37. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. (2000) Gene Ontology: tool for the unification of biology. Nat Genet 25: 25–29.
38. Rundle HD, Nosil P (2005) Ecological speciation. Ecol Lett 8: 336–352.

39. Schluter D (2009) Evidence for ecological speciation and its alternative. Science 323: 737–741.
40. Pala I, Hasselquist D, Bensch S, Hansson B (2012) Patterns of molecular evolution of an avian neo-sex chromosome. Mol Biol Evol 29: 3741–3754.
41. Agate RJ, Choe M, Arnold AP (2004) Sex differences in structure and expression of the sex chromosome genes *CHD1Z* and *CHD1W* in zebra finches. Mol Biol Evol 21: 384–396.
42. Okada M, Okawa K, Isobe T, Fukagawa T (2009) CENP-H–containing complex facilitates centromere deposition of CENP-A in cooperation with FACT and CHD1. Mol Biol Cell 20: 3986–3995.
43. Henikoff S, Kami A, Malik SM (2001) The centromere paradox: stable inheritance with rapidly evolving DNA. Science 293: 1098–1102.
44. Kowalik D, Haller F, Adamski J, Moeller H (2009) In search for function of two human orphan SDR enzymes: Hydroxysteroid dehydrogenase like 2 (HSDL2) and short-chain dehydrogenase/reductase-orphan (SDR-O). J Steroid Biochem Mol Biol 117: 117–124.
45. Ellegren H, Parsch J (2007) The evolution of sex-biased genes and sex-biased gene expression. Nat Rev Genet 8: 689–698.
46. Goerlich VC, Nätt D, Elfwing M, Macdonald B, Jensen P (2012) Transgenerational effects of early experience on behavioral, hormonal and gene expression responses to acute stress in the precocial chicken. Horm Behav 61: 711–718.
47. Seehausen O (2004) Hybridization and adaptive radiation. Trends Ecol Evol 19: 198–207.
48. Via S, West J (2008) The genetic mosaic suggests a new role for hitchhiking in ecological speciation. Mol Ecol 17: 4334–4345.
49. Eroukhmanoff F, Hermansen JS, Bailey RI, Sæther SA, Sætre G-P (2013). Local adaptation within a hybrid species. Heredity 111: 286–292.
50. Barton NH, de Cara MAR (2009) The evolution of strong reproductive isolation. Evolution 63: 1171–1190.
51. Currat M, Ruedi M, Petit RJ, Excoffier L (2008) The hidden side of invasions: massive introgression by local genes. Evolution 62: 1908–1920.
52. Jiggins CD, Salazar C, Linares M, Mavarez J (2008) Hybrid trait speciation and *Heliconius* butterflies. Phil Trans R Soc B 363: 3047–3054.
53. Rieseberg LH (1997) Hybrid origins of plant species. Annu Rev Ecol Syst 28: 359–389.
54. Gompert Z, Buerkle CA (2010) INTROGRESS: a software package for mapping components of isolation in hybrids. Mol Ecol Resour 10: 378–384.
55. Rousset F (2008) GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. Mol Ecol Resour 8: 103–106.
56. Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics 180: 977–993.
57. Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: A workbench to detect molecular adaptation based on a F_{st}-outlier method. BMC Bioinformatics 9: 323.

II

GENETICS

The genomic mosaicism of hybrid speciation

Tore O. Elgvin,[1]* Cassandra N. Trier,[1]* Ole K. Tørresen,[1] Ingerid J. Hagen,[2] Sigbjørn Lien,[3] Alexander J. Nederbragt,[1] Mark Ravinet,[1] Henrik Jensen,[2] Glenn-Peter Sætre[1†]

Hybridization is widespread in nature and, in some instances, can result in the formation of a new hybrid species. We investigate the genetic foundation of this poorly understood process through whole-genome analysis of the hybrid Italian sparrow and its progenitors. We find overall balanced yet heterogeneous levels of contribution from each parent species throughout the hybrid genome and identify areas of novel divergence in the hybrid species exhibiting signals consistent with balancing selection. High-divergence areas are disproportionately located on the Z chromosome and overrepresented in gene networks relating to key traits separating the focal species, which are likely involved in reproductive barriers and/or species-specific adaptations. Of special interest are genes and functional groups known to affect body patterning, beak morphology, and the immune system, which are important features of diversification and fitness. We show that a combination of mosaic parental inheritance and novel divergence within the hybrid lineage has facilitated the origin and maintenance of an avian hybrid species.

INTRODUCTION

Hybridization is increasingly recognized as a potentially creative force contributing to adaptation and species diversification (*1*, *2*). New species may arise as a direct consequence of interbreeding between diverged taxa, in which the hybrid lineage comprises a recombinant genome with the same ploidy level and is reproductively isolated from its parent species (*3*, *4*). This process—known as homoploid hybrid speciation—may take many forms with respect to genomic makeup, ranging from introgression of a single or few genes into a foreign genomic background (*5*) to balanced genomic contributions from both parent lineages (*6*, *7*). At both ends of the spectrum, novel allelic combinations and subsequent evolution in the hybrid lineage may facilitate its escape from inferior fitness and aid in the evolution of reproductive barriers toward both parents, which is essential for maintaining its isolation. Characterizing these combinations and their genomic distribution is therefore of paramount importance for understanding the hybrid speciation process and enriching our knowledge of the role of hybridization in shaping biodiversity. However, there are few well-documented cases of homoploid hybrid species in nature, and, consequently, the genomics of this mode of speciation are poorly understood.

The Italian sparrow (*Passer italiae*) is a homoploid hybrid species found in mainland Italy and a few Mediterranean islands (Fig. 1A) that has arisen from hybridization between the house sparrow (*Passer domesticus*) and the Spanish sparrow (*Passer hispaniolensis*). Although ecologically more similar to the house sparrow, male Italian sparrows have plumage patterns that comprise a mosaic of male Spanish and house sparrow traits (Fig. 1A) (*8*). The taxonomic status of the Italian sparrow has been subject to much debate, and several hypotheses have previously been proposed for its evolutionary relationship to the other *Passer* sparrows (*8*, *9*). Its intermediate appearance, in particular, led to the proposition that it is of hybrid origin (*10*) [see also Anderson (*8*)]. Only recently have genetic studies given support for this idea, demonstrating the Italian sparrow's genetic and phenotypic mosaicism (*11*–*13*). A proposed scenario for its origin is that the Italian sparrow arose <10,000 years ago as the human commensal house sparrow expanded throughout Europe where it encountered and hybridized with the Spanish sparrow (*12*).

The system is unique in that the hybrid Italian sparrow remains in geographic contact with both its parent species. To the north of its range, the Italian sparrow meets the house sparrow in a narrow hybrid zone in proximity to the Alps. In addition, it lives in sympatry with the Spanish sparrow on the southeastern Italian peninsula of Gargano, with little evidence of gene flow (*13*). The lack of introgression in Gargano may be due to premating barriers such as habitat differentiation and timing of breeding, or to postzygotic isolating factors. Evidence for postzygotic reproductive barriers has been shown in areas of contact (*13*). Furthermore, hybridization is likely to have directly contributed to the development of reproductive barriers between the Italian sparrow and its parents as preexisting parental incompatibility alleles are sorted within the Italian sparrow lineage (*14*).

The Italian sparrow's hybrid ancestry has thus far been characterized by a limited set of genetic markers, and the genomic architecture and regions underlying reproductive barriers in the system have yet to be identified. Here, we investigate the genetic composition of an avian hybrid species by using a whole-genome analysis of the Italian sparrow and its parent species, the house sparrow and the Spanish sparrow. Genome data provide powerful prospects for understanding hybrid speciation. Although conflicting signatures of ancestry represent a hallmark of hybrid speciation, the traces of hybridization may be obscured by processes such as backcrossing to either parent lineage, sorting of ancestral polymorphisms, genetic drift, and selection. Hence, the degree of admixture may vary widely throughout the hybrid genome (*15*, *16*), and studying only a subset of the genome can give confounding results. Whole-genome data also offer the alluring potential for elucidating candidate regions responsible for the formation and maintenance of the hybrid species.

With the aid of a high-quality de novo reference genome created for the house sparrow, we mapped and analyzed genome data from key populations of the three focal taxa. Population genetic parameters, admixture analyses, and phylogenetic inference were used to characterize the genetic composition throughout the hybrid genome in relation to its parents. This approach also allowed us to identify areas in which the Italian sparrow segregates for alternative parental alleles and genes

[1]Department of Biosciences, Centre for Ecological and Evolutionary Synthesis, University of Oslo, P.O. Box 1066, N-0316 Oslo, Norway. [2]Department of Biology, Centre for Biodiversity Dynamics, Norwegian University of Science and Technology, N-7491 Trondheim, Norway. [3]Department of Animal and Aquacultural Sciences, Faculty for Biosciences, Centre for Integrative Genetics, Norwegian University of Life Sciences, P.O. Box 5003, Ås, Norway.
*These authors contributed equally to this work.
†Corresponding author. Email: g.p.satre@ibv.uio.no

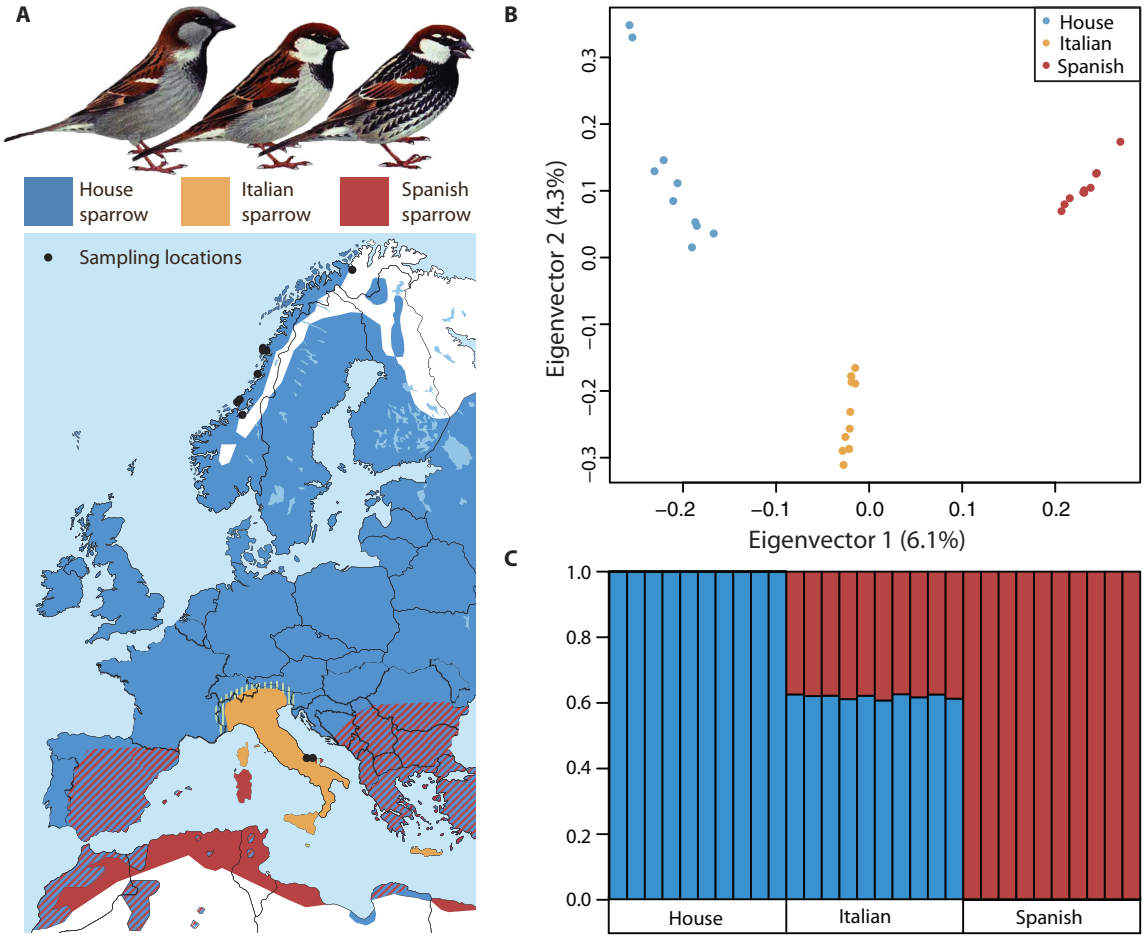


Fig. 1. Population structuring and species information. (**A**) Top: Illustrations of male plumage patterns in house, Italian, and Spanish sparrows modified from Svensson *et al.* (*83*). Bottom: A distribution map of house, Italian, and Spanish sparrows throughout Europe and northern Africa (*9*). (**B**) PCA of the LD-pruned high-quality SNP set. (**C**) Population structuring based on admixture analysis for house, Italian, and Spanish populations.

where the hybrid differs from both parents, which may play instrumental roles in the reproductive barriers involved in the hybrid system.

RESULTS

House sparrow reference genome

We created a high-quality de novo reference genome by sequencing and assembling the genome of a house sparrow (~130× coverage; table S1). The final assembly encompasses 1.04 Gb (gigabases), and with the aid of a medium-density linkage map, we ordered and oriented 88% of the assembly scaffolds into chromosomes, resulting in an N50 sequence size of 68.7 Mb (megabases) (table S2). We then sequenced whole genomes of 10 males of each of the focal species and one tree sparrow (*Passer montanus*; outgroup) at ~10× coverage per individual (table S3) and mapped them to the reference genome for downstream analysis.

Population structuring

Population genetic parameter estimates reveal genome-wide intermediacy in the Italian sparrow compared to its parents, although it is overall closer to the house sparrow (Table 1). The global average of differentiation (F_{ST}) was, as expected under a hybrid speciation model, higher between the parents, the house sparrow and the Spanish sparrow (hereinafter HS; F_{ST} = 0.33), compared to differentiation between the Italian sparrow and either the house sparrow (HI; F_{ST} = 0.18) or the Spanish sparrow (SI; F_{ST} = 0.25).

A principal components analysis (PCA) of high-quality single-nucleotide polymorphisms (SNPs), obtained by variant calling the three focal taxa against the house sparrow reference genome and pruning for linkage disequilibrium (LD), demonstrated striking structuring between the three focal taxa. The house and Spanish sparrows separate along the first eigenvector with the Italian sparrow positioned in between with no overlap among the clusters (Fig. 1B). Furthermore, separation along eigenvector 2 indicates novel differentiation of the Italian sparrow against both parents, which may be due to extensive segregation of alternative alleles from both parents in the hybrid, differences from both parents due to selection, and/or de novo mutations in the hybrid lineage subsequent to its formation.

Parental contributions to the hybrid lineage

Whole-genome admixture analysis revealed that the Italian sparrows, on average, assign 61.9% to the house sparrow ancestry and 38.1% to

the Spanish sparrow ancestry (Fig. 1C). In addition, admixture analysis across the Italian sparrow genome [100-kb (kilobase) nonoverlapping windows] revealed a large variation (σ = 19% combined for all windows) in the assignment probability to either parent (Fig. 2 and fig. S1).

Maximum likelihood phylogenies created using RAxML (*17*) demonstrated a discordant evolutionary history throughout the hybrid genome. Individual trees were created for 100-kb nonoverlapping windows and classified according to whether the Italian sparrows grouped monophyletically with the house sparrow, Spanish sparrow, or in its own clade. The vast majority of trees (76%) remained unresolved, that is, the Italian sparrows did not form a monophyletic group alone or with either parent. This is expected because any window may harbor alleles derived from both parents as well as novel mutations in the hybrid lineage. Nonetheless, in 14% of the genomic windows, Italian sparrows grouped with house sparrows, in 9% with Spanish sparrows, and in less than 1%, they formed their own clade (Fig. 2, fig. S1, and table S4).

Table 1. Mean values of population genomic statistics for 100-kb sliding windows with 25-kb steps across the genome. d_f, density of fixed differences.

Parameter	Species	Autosomes	Z chromosome	Whole genome
F_{ST}	HS	0.32 ± 0.10	0.45 ± 0.09	0.33 ± 0.10
	HI	0.17 ± 0.07	0.29 ± 0.16	0.18 ± 0.08
	SI	0.23 ± 0.12	0.41 ± 0.17	0.25 ± 0.14
d_f (10^{-3})	HS	0.048 ± 0.28	1.440 ± 1.90	0.146 ± 0.67
	HI	0.003 ± 0.01	0.005 ± 0.01	0.003 ± 0.01
	SI	0.003 ± 0.02	0.211 ± 0.70	0.018 ± 0.19
π	House	0.0073 ± 0.0021	0.0042 ± 0.0025	0.0071 ± 0.0023
	Italian	0.0070 ± 0.0021	0.0048 ± 0.0025	0.0068 ± 0.0023
	Spanish	0.0051 ± 0.0022	0.0025 ± 0.0023	0.0049 ± 0.0023
Θ*	House	0.0075 ± 0.002	0.0043 ± 0.002	0.0073 ± 0.002
	Italian	0.0074 ± 0.002	0.0046 ± 0.002	0.0072 ± 0.002
	Spanish	0.0055 ± 0.002	0.0027 ± 0.002	0.0053 ± 0.002

*Watterson estimator of Θ based on the number of segregating sites.

A 100-kb window size was chosen for sliding window analyses because LD decays within this distance (fig. S2). To test whether a smaller window size improved the proportion of resolved phylogenies by limiting the number of windows where the Italian sparrows segregate for haplotype blocks from both parents, RAxML analyses were also performed for 50- and 10-kb window sizes. Despite the reduction in window size, the proportion of resolved phylogenies remained largely similar to that of the 100-kb window analyses (table S5), and the latter was therefore kept for downstream analysis.

Incomplete lineage sorting can leave similar genomic footprints as those of hybridization. However, we find significant introgression between both parents and the Italian sparrow on the whole-genome level, as well as when the autosomes and Z chromosome are considered individually (Table 2). When ABBA BABA tests are performed in sliding windows across the genome, the f_d estimator (*18*) showed very similar admixture proportions between the Italian sparrow and either parent, with more gene flow between the Italian and Spanish sparrows on autosomes and more introgression between the house and Italian sparrows on the Z chromosome (Table 3). In addition, the SDs of f_d estimates were large, lending further support to the Italian sparrow's mosaic composition of alternating genomic ancestry from either parent, as demonstrated in the admixture and RAxML analyses.

To investigate the composition of mitochondrial DNA (mtDNA) in the Italian sparrow, we created a haplotype network and ran a fastSTRUCTURE (*19*) analysis of complete mtDNA sequences. Although the Italian sparrow has been shown to be nearly fixed for house sparrow mtDNA (*12*), surprisingly, two Italian sparrows demonstrate evidence of admixture from both parental mitochondrial haplotypes (Fig. 3, A and C). Further, these "admixed" individuals display intermediate estimates of sequence divergence (d_{XY}) from the parent species along the mitochondria (Fig. 3B and fig. S3). To test for heterozygosity in mtDNA sequences, we also variant-called each individual's mtDNA as diploid and used VCFtools (*20*) to calculate per individual inbreeding coefficient (F) estimates. VCFtools calculates F on the basis of the number of observed homozygotes, number of sequenced sites, and the expected number of homozygotes under the Hardy-Weinberg equilibrium. F estimates for the admixed individuals indicated an excess of heterozygosity (F = −0.25 and −0.36) compared to the other Italian individuals, which were largely homozygous (average, F = 0.67). Furthermore, the admixed individuals do not differ from the other Italian sparrows in sequence coverage (Mann-Whitney U test, P = 0.4) nor in their nuclear sequences (Mann-Whitney U test on nuclear F estimates, P = 0.09; Fig. 1A). Thus, there is no evidence for contamination of the

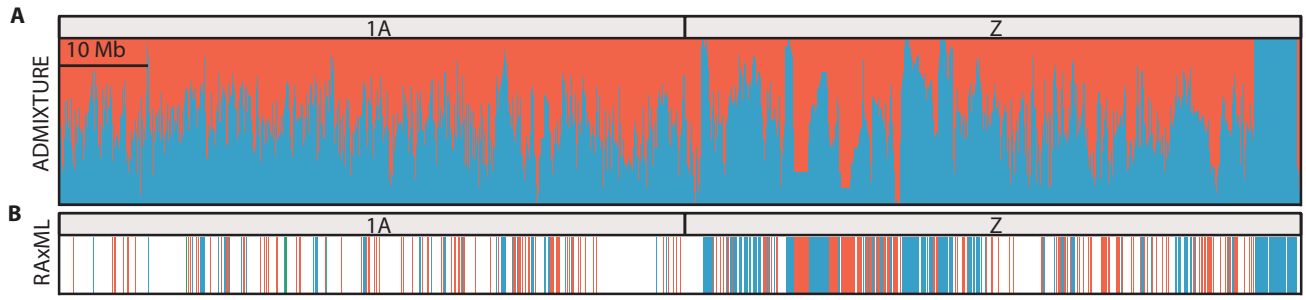


Fig. 2. Phylogenetic inference and admixture analysis of the Italian sparrow. (**A**) ADMIXTURE analysis of 100-kb nonoverlapping windows across the Italian sparrows' chromosomes 1A and Z, with the house sparrow ancestry shown in blue and Spanish sparrow ancestry shown in red. (**B**) RAxML tree assignment results for 100-kb nonoverlapping windows across the genome for chromosomes 1A and Z. Windows depict whether the Italian sparrows grouped monophyletically with house sparrows (blue), Spanish sparrows (red), in its own clade (green), or were unresolved (white).

Table 2. Results from Patterson's *D* test for introgression between lineages with block jack-knifed SE estimates and significance values.

	P1	P2	P3	Patterson's *D*	Jack-knifed SE	*Z* score	*P*
Autosomes	House	Italian	Spanish	0.195	0.001	27.77	<0.00001
	Spanish	Italian	House	0.035	<0.001	4.98	<0.00001
Z	House	Italian	Spanish	0.552	0.004	8.34	<0.00001
	Spanish	Italian	House	0.575	0.006	7.62	<0.00001
Whole genome	House	Italian	Spanish	0.211	<0.001	24.23	<0.00001
	Spanish	Italian	House	0.065	<0.001	5.39	<0.00001

Table 3. Mean f_d values for 100-kb sliding windows with 25-kb steps across the genome.

P1	P2	P3	f_d (%) Autosomes	f_d (%) Z chromosome	f_d (%) Whole genome
House	Spanish	Italian	31.7 ± 15.4	40.3 ± 23.7	32.1 ± 16.0
Spanish	House	Italian	27.5 ± 21.7	48.7 ± 27.8	32.4 ± 24.9

samples. Together, all the results from the mitochondrial analyses suggest the presence of heteroplasmy, in which individuals retain mtDNA haplotypes from both parental species.

Signatures of selection

Genomic regions exhibiting high interspecific divergence may have been targets of species-specific selection and are often considered candidates for barriers to gene flow. For hybrid systems, this extrapolation is not straightforward because the hybrid lineage inherits evolutionary histories from two sources, thereby distorting such signals. We used disparities in F_{ST} values between lineages to identify genomic regions where the Italian sparrow displays elevated divergence from either or both of its parents. For each comparison between the Italian sparrow and either parent, we selected the top 1% 100-kb windows where the Italian sparrow showed the largest difference in F_{ST} values between one parent and the other (Eqs. 1 and 2; Fig. 4A and fig. S4)

(1) House versus Italian sparrow divergence (HI) = HI F_{ST} − SI F_{ST}

(2) Spanish versus Italian sparrow divergence (SI) = SI F_{ST} − HI F_{ST}.

Concern has been raised that relative measures of divergence, such as F_{ST}, could be elevated because of reduced intraspecific variation from processes unrelated to speciation, such as background selection in low-recombining regions (*21*–*23*). Our three-taxa system enables us to account for this confounding factor. If a high F_{ST} value is driven by low nucleotide diversity due to reduced variation in the ancestral population or regions of low recombination, then the F_{ST} values would be expected to be elevated in all species comparisons. However, if one parent/hybrid comparison has a high F_{ST}, whereas the other has low F_{ST}, then this indicates a differentiated region separating the hybrid and the former parent that is not attributed to low intraspecific nucleotide diversity from background selection. Similarly, regions of higher divergence between the Italian sparrow and both parents, compared to the divergence between the parents, indicate potential regions privately isolating the Italian sparrow from both its parents. An advantage of using F_{ST} estimates in this system is that it is sensitive to recent selection events, allowing for the identification of areas of hybrid-specific evolution. Regions of novel divergence in the Italian sparrow—that is, windows in which the Italian sparrow displays high divergence against both parents (hereafter referred to as PI)—were targeted using a similar method as the HI and SI windows. For both of the hybrid/parent comparisons, parental divergence in the same window was subtracted from the Italian sparrow divergence against the respective parent. The 1% windows exhibiting the highest F_{ST} disparities, common to both hybrid/parent comparisons, were kept as PI windows (Eq. 3; Fig. 4D and fig. S4).

(3) Parents versus Italian sparrow divergence (PI) = SI F_{ST} − HS F_{ST} ∪ HI F_{ST} − HS F_{ST}.

Overall, we find higher F_{ST} values, lower nucleotide diversity, more extreme Tajima's *D* (TD) values, and elevated levels of LD within the outlier windows relative to the nonoutlier windows (Fig. 4 and Table 4). These patterns bear signatures of selection, suggesting that the outlier windows may harbor or are located adjacent to genes with a role in species divergence. Particularly striking are the strongly positive TD values in windows of novel divergence in the Italian sparrow, whereas both parents exhibit negative TD values in the corresponding windows (Fig. 4B). High TD values result from an excess of medium frequency alleles, consistent with balancing selection or a population bottleneck, whereas low values indicate an excess of low-frequency polymorphisms following a selective sweep or population expansion (*24*). These contrasting values suggest that the Italian sparrow segregates for alleles that have undergone selection in the parents and subsequently have been subject to balancing selection within the hybrid lineage. Because the nonoutlier windows exhibit an overall negative tendency in TD, indicating a genome-wide demographic signal of expansion and/or background selection in the three species, the strongly positive TD density distribution in PI divergence windows is particularly compelling.

The direction of selection (DoS) statistic was also estimated for all 100-kb genomic windows to test for selection in protein-coding sequence between the three species comparisons. The DoS statistic is conceptually similar to the McDonald-Kreitman test and uses protein-coding SNPs to measure the direction and extent of selection on the basis of nonsynonymous and synonymous fixed differences and polymorphisms between lineages (*25*). DoS estimates revealed more extreme signals of selection within the outlier windows compared to the genomic background. Furthermore, the extent and DoS differ between the species comparisons within the outlier windows, whereas genomic background values are very similar between the species comparisons.

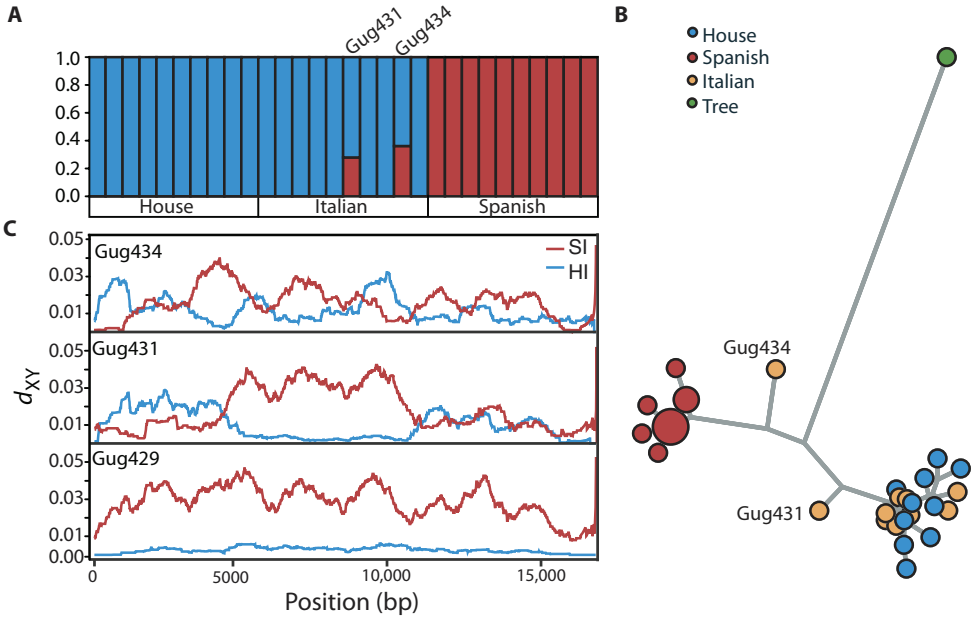


Fig. 3. Mitochondrial ancestry in the Italian sparrow. (**A**) fastSTRUCTURE analysis of complete mitochondrial sequences for the house, Italian, and Spanish sparrow individuals. (**B**) Mitochondrial haplotype network of all sparrow individuals. (**C**) Plots of sliding window (1-kb window and 100-bp step) sequence divergence (d_{XY}) along the mitochondria between the house and Italian (blue) and Spanish and Italian (red) sparrows for three Italian individuals. The top and middle panels depict the mixed mitochondrial ancestry of two Italian individuals, whereas the bottom panel shows the sole house sparrow ancestry of a single Italian sparrow mitochondria that is representable for the eight remaining Italian sparrows (see fig. S3).

Stark differences were revealed in the DoS statistic between the parents and the HI and SI comparisons in the PI regions (Fig. 4C). Although the parents exhibit the full range of possible DoS values with most of the genes near neutrality, both HI and SI comparisons are largely negative, indicating an excess of nonsynonymous polymorphisms. This is expected under balancing or weak purifying selection (*26*). Together, the excess of nonsynonymous polymorphism and high TD values suggest the presence of balancing selection in the PI windows. Parental differentiation is, on average, only slightly higher than the genomic background in PI windows (Table 4); however, there is a higher density of genes under positive divergent selection in the parents within these windows compared to all other genomic regions (Fig. 4C). These DoS patterns may be driven by low numbers of nonsynonymous fixed sites that do not necessarily elevate the F_{ST} values across the entire genomic window between the parent taxa. Hence, it appears that some genes within PI windows are under divergent selection in the parents.

Within the SI windows, the TD value distributions are shifted to the left in the Spanish sparrow and to the right in the house sparrow and Italian sparrow (Fig. 4B). DoS estimates are variable in the SI windows but reveal a higher density of genes with an excess of nonsynonymous fixed differences between the Spanish and Italian sparrows compared to all other genomic regions (Fig. 4C), indicating that some genes are under positive divergent selection. Similarly, there is an excess of nonsynonymous fixed differences between the parent taxa within the SI windows. This suggests that these regions, in which the Italian sparrow may be assumed to have inherited from the house sparrow, harbor genes under divergent selection in the parent taxa. Moreover, the TD values indicate that this selection has mainly occurred in the Spanish sparrow lineage.

The HI windows exhibit strongly negative TD values in the Spanish and Italian sparrows, with more neutral values in the house sparrow (Fig. 4B). Although the strongly negative values could be driven by selective sweeps or purifying selection, DoS estimates indicate that there are no genes with an excess of nonsynonymous fixed differences in any of the species comparisons (Fig. 4C). This suggests that the high divergence within HI windows is attributed to the Italian sparrow's inheritance of regions, which are under background selection or have undergone a selective sweep in the Spanish sparrow but are near neutrality in the house sparrow.

The Z chromosome versus autosomes

For all comparisons, we find overall higher divergence, lower nucleotide diversity (Table 1), and a larger proportion of resolved phylogenies on the Z chromosome relative to autosomes (table S4, Fig. 2, and fig. S1). Contrasting patterns of divergence and polymorphisms are consistent with a faster rate of evolution and/or reduced gene flow on the Z chromosome. Furthermore, although we observe a mosaic pattern of parental inheritance across the Italian sparrow genome, this pattern is strikingly more pronounced on the Z chromosome with large genomic regions alternating in inheritance from one parent or the other (Fig. 2 and fig. S1) in a block-like fashion.

Among the outlier windows, there is a significant overrepresentation on the Z chromosome (SI, $\chi^2 = 87.064$, $P < 0.0001$; HI, $\chi^2 = 79.794$, $P < 0.0001$; PI, $\chi^2 = 787.461$, $P < 0.0001$). Because sex chromosomes have a lower effective population size than autosomes, it can be difficult to parse out signals of selection from increased rates of genetic drift. However, DoS estimates on all genes throughout the genome for the three species comparisons revealed more extreme estimates of selection on Z-linked genes compared to autosomes, as well as a lower density of genes evolving neutrally, although the density distributions were only significantly different in the HI and SI comparisons (Fig. 5, two-sided permutation test: HS, $P = 0.0649$; HI, $P < 0.00001$; SI, $P = 0.0013$).

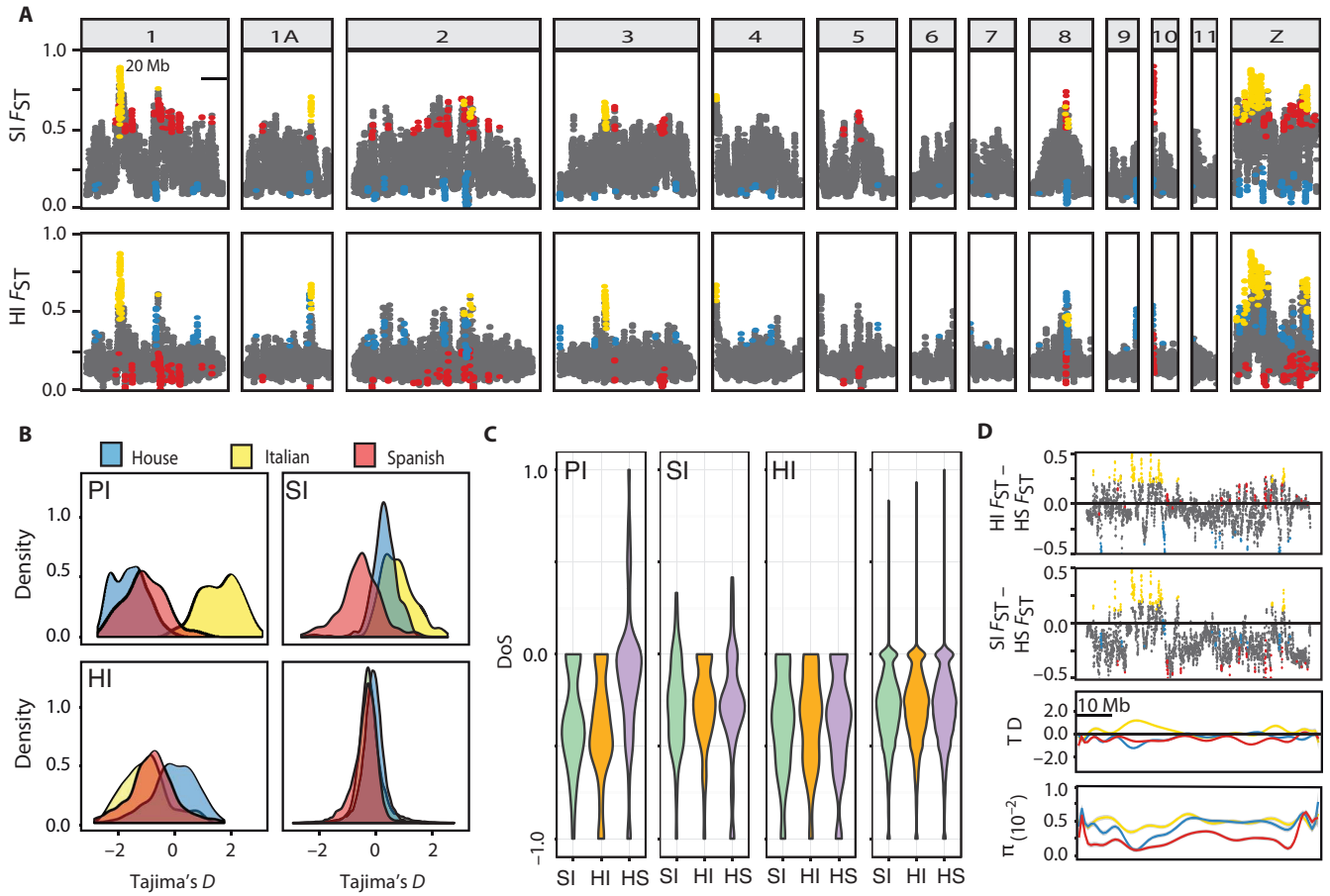


Fig. 4. Divergence landscape and selection tests for the *Passer* taxa. (**A**) F_{ST} estimates for 100-kb overlapping windows with 25-kb steps across the largest chromosomes in the genome. Windows are identified as HI divergence peaks in blue, SI in red, and PI in yellow. Microchromosomes are plotted in fig. S4. (**B**) Density plots of TD estimates for PI, SI, and HI outlier windows, as well as all nonoutlier windows (bottom right) for each focal species. (**C**) Density of DoS values within PI, SI, and HI windows and the genomic background for the three species comparisons. (**D**) Plots of disparities in F_{ST} values between HI/parent species (top) and SI/parent species (second from top) for all genomic windows on the Z chromosome, highlighting windows in PI peaks (yellow), SI peaks (red), and HI peaks (blue). The two bottom panels depict smoothed plots of TD and π in all 100-kb overlapping genomic windows along the Z chromosome for each focal sparrow species (house, blue; Spanish, red; Italian, yellow).

Gene ontology analysis of high-divergence regions

Annotated genes residing within the outlier windows were extracted for ontology analyses, resulting in a total of 83 PI, 159 HI, and 76 SI outlier genes (table S6). Gene enrichment analysis revealed ontology classes significantly overrepresented for each of the three comparisons (Table 5 and tables S7 and S9), several of which have bearing on key traits separating the species in the system. For the PI genes, one of the significantly enriched classes was "dorsoventral pattern formation," which encompasses a wide range of anatomical features, from body plan to color patterning. Plumage coloration constitutes a strong mating barrier in many bird species and is the most conspicuous phenotypic trait separating males of the focal species (*27*). Furthermore, the PI windows included four genes known to be associated with melanogenesis in vertebrates (table S10) (*28*), and a total of 13 such genes were identified in the other species comparisons, significantly more than expected by chance (one-sided permutation test; $P < 0.044$). Also enriched in PI windows were genes involved in the "negative regulation of immune response."

Two of the eight significantly enriched functional groups in SI divergence regions include "palate development" and "regulation of bone morphogenetic protein (BMP) signaling pathway" (Table 5). *BMP* proteins have been shown to play a key role in beak morphology and diversification among Darwin's finches (*29*). Another SI gene, *PTCH1*, is a craniofacial signaling gene involved in adaptive variation in lower jaw shape in cichlids (*30*). These findings suggest that evolution of craniofacial structures has been an important component driving the Italian-Spanish divergence. The functional group with the lowest corrected *P* values and the highest number of assigned HI genes was the "regulation of G protein–coupled receptor signaling," which has been shown to directly modify behavioral and morphological variation in birds (Table 5 and table S7) (*31*).

To test whether the significant gene ontologies found within the outlier regions are likely to be observed by chance when sampling windows from the genome, gene ontology (GO) permutations were also run. For each outlier window category (SI, PI, and HI), 50 permutations were performed by randomly sampling the same number of windows as the category being tested and extracting genes from the resulting windows. No overlap was found between the permutations and the outlier windows in significant GOs. Thus, we find that the outlier windows differ from the genomic background and that the significant GO groups are unlikely to be identified randomly.

Table 4. Average population genomic statistics for the high-divergence windows and the genomic background.

Parameter	Species	PI	HI	SI	Background
F_{ST}	Parents	0.380	0.468	0.504	0.329
	HI	0.651	0.389	0.109	0.178
	SI	0.703	0.125	0.574	0.241
TD	House	−1.516	0.010	0.292	−0.132
	Italian	1.625	−1.007	0.750	−0.206
	Spanish	−0.979	−0.785	−0.576	−0.374
π	House	0.0014	0.0046	0.0057	0.0072
	Italian	0.0038	0.0030	0.0057	0.0069
	Spanish	0.0008	0.0023	0.0021	0.0050
r^2*	House	0.4415	0.4901	0.4427	0.3155
	Italian	0.5527	0.4309	0.3957	0.2658
	Spanish	0.3071	0.3896	0.3395	0.2736
$\Theta^{\dagger}$	House	0.0021	0.0046	0.0054	0.0073
	Italian	0.0029	0.0037	0.0050	0.0073
	Spanish	0.0010	0.0027	0.0024	0.0054

*Mean estimates of LD in the form of pairwise r^2 estimates for all SNPs within 1 kb of each other in the specified genomic windows.
†Watterson's estimator of Θ based on the number of segregating sites.

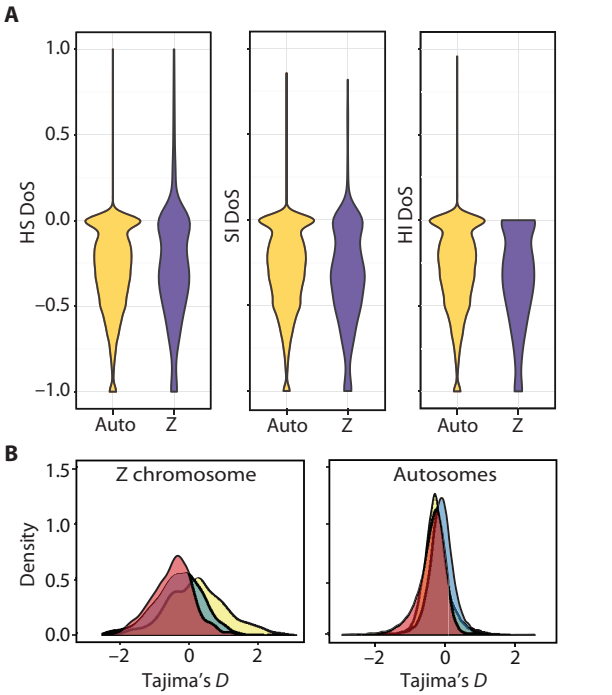


Fig. 5. Signatures of selection on the Z chromosome versus autosomes. (**A**) Density of DoS values for all autosomal and Z-linked genes for the HS, SI, and SI species comparisons. (**B**) Density of TD values for all autosomal and Z-linked genes in the Spanish (red), house (blue), and Italian (yellow) sparrows.

Among the genes within outlier windows, 13 have been previously identified as candidate reproductive barrier genes between the focal species (table S10), a significant overrepresentation (one-sided permutation test; $P < 0.0001$). Two of these genes (*RPS4* and *HSDL2*) exhibited steep genomic clines over the Italian sparrow's range boundaries, indicating involvement in reproductive barriers against the parents (*13*). In addition, five PI genes (*REEP5*, *A2M*, *A2ML1*, *APC*, and *MIA3*) have been shown to exhibit steep clines within the Italian sparrow range (*13*), making them strong candidates for genes under selection in the hybrid species.

DISCUSSION

Genomic mosaicism in a hybrid species

We demonstrate extensive genomic admixture in an avian homoploid hybrid species, with significant contributions from both parent species. The genetic intermediacy of the Italian sparrow is evident through population structure analyses, estimates of population genetic parameters, and phylogenetic inference. Together with tests confirming introgression, our data suggest that hybridization has been the main process behind the evolution of the Italian sparrow.

The Italian sparrow exhibits an overall closer affinity to the house sparrow. Although hybrid speciation at the outset involves an equal mixing of the two genomes, the contributions from the progenitors will rarely be balanced in the hybrid species if subsequent backcrossing is involved (*4*), causing a genetic bias toward one of the parents. Furthermore, local levels of admixture throughout the genome will vary considerably and largely depend on the factors that rendered allele combinations compatible yet also allowed for barriers to gene flow against the parents. Hence, the house sparrow bias in the hybrid genome may result from a range of processes. House sparrows may simply have outnumbered the Spanish sparrow during the initial hybridization events, resulting in an overrepresentation of the former species' genomic background. In addition, a bias may be explained by the hybrid overall experiencing selection that favored alleles from one parent more than the other, as a result of either adaptive evolution or purging of incompatible allele combinations. Intriguingly, the Italian sparrow shares an almost identical ecology to the human commensal house sparrow, whereas the Spanish sparrow occupies more mesic habitats. One could therefore speculate that the Italian sparrow has been exposed to a selective regime more similar to that experienced by the house sparrow, thereby causing a bias toward house sparrow alleles at many genes. Moreover, the extent of intermediacy throughout the Italian sparrow genome may indicate that it originated through bursts of parental interbreeding, which would retain admixture despite the hybrid being in contact with either parent, as has been argued for the tiger swallowtail hybrid system (*7*).

The mitochondrial analyses of the Italian sparrow revealed surprising evidence of heteroplasmy in two individuals. Although heteroplasmy is uncommon in animals, it has been detected in a range of species (*32*), particularly in interspecific hybrids (*33*, *34*), including birds (*35*). It has been proposed that the mechanisms destroying paternal mtDNA in eggs may break down in hybrids, leading to paternal leakage during heterospecific crosses (*33*, *36*). Heteroplasmy can be difficult to detect because nuclear mitochondrial pseudogenes (numts) can map to mitochondrial sequences (*32*), distorting the analysis of true mtDNA.

Table 5. Significantly enriched functional groups within outlier windows.

Window type	Functional group	Bonferroni step-down corrected *P*
HI	Amino acid transport	9.0995×10^{-4}
HI	Negative regulation of G protein–coupled receptor protein signaling pathway	1.6878×10^{-4}
HI	Positive regulation of translation	0.0492
HI	Regulation of guanosine triphosphatase activity	1.5561×10^{-5}
HI	Regulation of autophagy	0.0012
HI	Regulation of muscle system process	1.4758×10^{-5}
SI	Behavioral response to nicotine	1.1522×10^{-4}
SI	Cell differentiation in the spinal cord	0.0026
SI	Cell differentiation involved in kidney development	0.0016
SI	Erythrocyte homeostasis	0.0205
SI	Palate development	0.0201
SI	Regulation of BMP signaling pathway	0.0257
SI	Regulation of mRNA processing	0.0069
SI	Regulation of organ growth	0.0116
PI	Cellular response to retinoic acid	0.0042
PI	Dorsal/ventral pattern formation	0.0027
PI	Negative regulation of innate immune response	0.0081
PI	Negative regulation of stress-activated mitogen-activated protein kinase cascade	0.0027
PI	Regulation of anion transmembrane transport	0.0017

However, we would have expected sequence divergence patterns to be more similar between the two admixed individuals if this was the case, and that sequence coverage then would be higher compared to the other Italian sparrows. In addition, such large mitochondrial regions would not be expected to alternate in higher sequence divergence from either parent throughout the entire mitochondrial genome, as we have observed. Hence, we find that heteroplasmy best explains the mitochondrial patterns seen in the two Italian sparrow individuals. Further analysis is required to determine its prevalence and potential fitness effects in the hybrid lineage.

The role of the Z chromosome in hybrid speciation

Sex chromosomes are known to play a prominent role in the evolution of reproductive isolation (RI) and speciation and have been suggested to be where genomic incompatibilities, such as hybrid inviability or sterility, first develop (*3*, *37*). This has been attributed to faster rate of adaptive evolution (faster X/Z), reduced recombination, and overrepresentation of genes related to sex and reproduction (*38*–*40*). Its role in hybrid speciation has been discussed in previous work [see also Kunte *et al.* (*7*) and Elgvin *et al.* (*11*)] but has not yet been extensively investigated. The current observation of contrasting patterns of divergence and polymorphism on Z chromosomes versus autosomes mirrors results for many other bird taxa, supporting the Z chromosome as a hotspot in avian speciation. We also found more conspicuous patterns of mosaicism and stronger selection signals on the Z chromosome relative to autosomes. An important consideration of Z chromosome evolution is its reduced effective population size relative to that of autosomes (N_e at Z is three-fourth that of autosomes) due to female heterogamety. This may significantly affect the sorting of parental alleles through increased rates of fixation via drift or selection, especially in the initial stages of speciation when the population size is expected to have been small. However, our DoS estimates on protein-coding substitutions revealed fewer genes under neutrality and more genes subject to both positive and purifying selection on the Z chromosome. Overall, and in line with earlier work on the system (*11*, *13*), our data further support the hypothesis that the Z chromosome has an important role also in hybrid speciation.

Selection within high-divergence windows

Homoploid hybrid speciation is thought to require rapid development of isolating barriers because the process is sympatric and the hybrid lineage thereby risks getting swamped by the homogenizing effect of gene flow from either parent species (*4*). Potential mechanisms of escaping such swamping include the emergence of trait combinations that instantly yield incompatibilities toward the parents via deleterious epistatic effects (*41*), assortative mating (*42*, *43*), or transgressive effects that allow for adaptation to novel ecological conditions in the hybrid (*6*, *44*).

The SI and HI genomic windows are areas of the genome where the Italian sparrow has strongly sorted for one parent's genetic variation

and are consequently candidate regions for barriers against the other parent. The strongest patterns of selection were observed in the SI windows where there is evidence for positive divergent selection between the parent taxa, as well as between the Italian and Spanish sparrows in coding regions. This is particularly interesting because two of the significant gene ontologies within this region relate to craniofacial development and include genes known to affect beak morphology and diversification among Darwin's finches (*29*). Because beaks are the main food processing tool in birds, their morphology is often related to individual fitness (*45*) and subject to divergent selection to suit different foraging ecologies (*46*), as has been shown in Italian sparrows (*47*). Because the Italian sparrow ecologically resembles the house sparrow, whereas the Spanish sparrow occupies more mesic habitats, the SI comparison may be expected to show increased divergence in genes controlling beak size and shape, assuming that habitat preferences affect diet and that the hybrid followed a similar genetic trajectory to its human commensal parent. Among the HI windows, the lack of nonsynonymous fixed differences in both HI and HS comparisons is perhaps unsurprising because these regions appear to be largely neutral in house sparrows and subject to selection in the Italian and Spanish lineages. The GO analyses within HI windows revealed significant enrichment of genes involved in the regulation of G protein–coupled receptor signaling that are known to directly modify behavioral and morphologic variation in birds (*31*). However, this group of genes encompasses a wide range of potential functions, and further investigation is needed to determine the phenotypic effects of these genes in sparrows.

Regions where the hybrid differentiated from both parents are candidate areas of novel divergence in the hybrid lineage. Within these areas, we find evidence for balancing selection in the Italian sparrow, which suggests heterozygote advantage (overdominance) in the hybrid. However, there are multiple processes that have confounding effects on sequence variation in a hybrid species and thereby the interpretation of selection signatures, including demography and recombination. It may be the case that background selection, potentially in areas of low recombination, is occurring within these regions, resulting in negative TD values in the parents. Reduced recombination is characterized by decreased nucleotide diversity in surrounding areas, because positive selection and purifying selection are expected to affect larger genomic regions due to stronger linkage among sites (*48*, *49*). In a hybrid, the recombinational landscape may be altered, breaking up haplotype blocks that are largely conserved in the parental lineages. Under this scenario, TD values are expected to increase and contrast the values seen in the source lineages, consistent with the positive TD values in the Italian sparrow. Furthermore, reduced recombination in the parents may also have led to an accumulation of slightly deleterious alleles, which cannot easily be purged from the population. With the release of recombinational blocks in the hybrid, heterosis may occur from the masking of these deleterious alleles.

Although there are a variety processes that may drive the observed TD patterns within the outlier windows, the combination of DoS estimates and TD values suggests a role for balancing selection in at least a portion of the genes within the PI regions. Hybridization can boost genetic variance (*6*), and through complementary gene action of additive alleles in parental lineages, transgressive phenotypes outside both parents' ranges can arise in hybrids (*50*). This is one proposed mechanism for how speciation can occur via hybridization as transgressive traits may allow hybrids to occupy novel ecological niches, in turn, impeding gene flow from its parents (*4*). Hybrids may even displace their parents ecologically if they experience higher fitness in a given habitat (*51*). Furthermore, traits with a history of balancing selection are expected to be more likely to result in hybrid transgression because traits with an intermediate optimum maintain alleles with effects in opposing directions in the parental lineages, allowing for the additive effects of complementary genes in the hybrid (*50*). We found a significant enrichment of genes involved in the regulation of the immune system within PI genomic regions. The immune system may have a large impact on individual fitness, and several of its genetic components are expected to be subject of balancing selection (*52*). It is possible that the combination of such alleles in an admixed genome, such as the Italian sparrow, could have an advantageous effect on the general fitness of the hybrids, thereby facilitating their spread.

We acknowledge that caution should be taken in drawing conclusions about the functional effects of candidate genes from genome comparisons alone (*53*). However, the study highlights candidate regions that harbor genes with known associations to traits that are likely to have a bearing on reproduction in the *Passer* system. In addition, the concordance of genes recognized in the current study and the genes exhibiting steep clines in the study by Trier *et al.* make them—and/or their adjacent locations—strong candidates for the involvement in the hybrid speciation process.

CONCLUSION

To our knowledge, our study represents the first detailed investigation of the genomic admixture in a hybrid species in relation to its parents. We demonstrate substantial parental contributions throughout an avian hybrid species that maintains its integrity despite contact with its parent species. Our study also highlights candidate regions that potentially affect key traits in the system and thereby may have had instrumental roles in the formation of the hybrid species. Overall, we argue that the Italian sparrow serves as a well-documented case of the striking potential of hybridization as a creative force contributing to species diversity.

MATERIALS AND METHODS

Experimental design

The main objective of this study comprised a comparative genomic analysis of the homoploid hybrid species the Italian sparrow and its parents, the house sparrow and Spanish sparrow. The analytical framework included whole-genome sequencing of key populations of the three focal taxa and mapping these to the closely related house sparrow reference genome assembly. The individual chosen for the reference genome assembly was an inbred (pedigree $F = 0.3125$) female house sparrow (individual ID 8887266) sampled in 2002 on the small and inbred island of Aldra (*54*) in northern Norway (66°24′N, 13°6′E; Lurøy kommune, Nordland). A previous study of genome-wide SNP-chip genotyping of this population showed that this individual has a low level of heterozygosity (0.161) compared to the population mean, which is advantageous in the reference assembly process.

We used male house sparrows from island populations in northern Norway ($n = 10$), Italian sparrows from Guglionesi ($n = 10$), and Spanish sparrows from Lesina ($n = 10$), both of the latter locations in central eastern Italy (Fig. 1 and table S3). In addition, one tree sparrow from Giardini Naxos in Sicily was added to the sampling scheme to serve as an outgroup. The sparrows were caught with mist nets, and ~25 μl of blood was extracted through venipuncture of the left brachial vein and stored in either 1 ml of standard lysis buffer (*P. italiae*, *P. hispaniolensis*,

and *P. montanus* samples) or 100% ethanol (*P. domesticus* samples). The appropriate catching and sampling permits were obtained from the appropriate authorities for the respective locations.

Reference genome assembly

DNA from the house sparrow chosen for the reference genome assembly was extracted using the protocol described by Hagen *et al.* (*55*). Detailed information on the reference individual has been deposited in the National Center for Biotechnology Information (NCBI) BioSample database under accession number SAMN02929199.

Sequencing for the de novo assembly of the house sparrow reference genome was performed on an Illumina platform using HiSeq 2000 instruments at the Norwegian Sequencing Centre at the University of Oslo (www.sequencing.uio.no) and at Génome Québec at McGill University (www.genomequebec.com/en/home.html). The sequencing strategy, including platform choice, fragment size, and coverage, was chosen following recommendations of the ALLPATHS-LG assembly software (Broad Institute, Cambridge, MA). ALLPATHS-LG has proven to be a robust assembler for larger eukaryotic genomes (*56*), and it uses a combination of short reads from paired-end and various mate pair (MP) libraries. For a complete list of the library construction and sequence yield, see table S1.

All MP library reads were trimmed for adaptor sequences using cutadapt (v. 1.5) (*57*). Read files were trimmed for the specific adaptors used for the various library construction protocols, including external adapters and junction adapters. The adapter trimming resulted in between 17.71 and 20.93% of the total bases being discarded before assembly. All adapter trimmed reads were used as input for ALLPATHS-LG (v. 46923) (Broad Institute, Cambridge, MA). File preparation was conducted according to the manufacturer's recommendations. The main run was performed using the TARGETS=submission option to make a submission prepared assembly version. The resulting assembly comprised a total of 1.04 Gb divided into 2766 scaffolds ≥1 kb (N50 = 66.3 Mb). Each scaffold was blasted against the NCBI nucleotide database using BLAST+ (v 2.2.29) (*58*). All scaffolds with a top hit that was not avian or reptilian were removed from the assembly. Only top hits with an *e* value greater than e^{-5} and an alignment length greater than 100 base pairs (bp) were considered reliable. This resulted in the removal of 195 scaffolds from the assembly.

For gene annotation of the reference assembly, we used the MAKER (v. 2.31.8) pipeline (*59*). This pipeline used RNA sequencing (RNA-seq) data to collect physical evidence for genes and incorporated additional gene predictions from other programs [see Yandell *et al.* (*60*)]. To obtain physical evidence for gene annotations, we used RNA-seq data from a previous project (*13*), and all of these were downloaded and assembled with Newbler (v. 3.0; -cdna option) to aid in the genome annotation. GeneMark-ES (*61*), with the min_contig option set to 10,000, and CEGMA (*62*) were first applied on the genome assembly file. The resulting CEGMA.gff file was then used to train a SNAP .hmm file (*63*). Both SNAP and GeneMark .hmm files, in addition to a transcriptome assembly and UniProt database (www.UniProt.org), were then fed into a first-pass run of MAKER with the following modifications to the maker_opts.ctl file: est2genome=1, protein2genome=1, keep_preds=1, single_exon=1, max_dna_len=300000, and min_contig=10000. AUGUSTUS (v.3.0.2) (*64*) was trained on the transcriptome assembly and a snap model of the MAKER .gff first run predictions. MAKER was finally run a second round, including the trained predictions from previous steps, with the CEGMA SNAP .hmm file replaced by the predictions SNAP model from the MAKER first run. Only the MAKER second-pass predictions were used for further analysis. After quality filtering (annotation edit distance score, ≥0.5), the resulting annotation included 13,685 protein-coding genes.

Linkage mapping and construction of chromosome sequences

Populations of *P. domesticus* on four different islands in the northern part of Norway (Aldra, Hestmannøy, Leka, and Vega) have been extensively monitored on an individual basis since 1993 [Hestmannøy; for example, Jensen *et al.* (*45*)], 1998 [Aldra; Billing *et al.* (*54*)], or 2001 [Leka and Vega; Hagen *et al.* (*55*)]. A total of 2290 house sparrows on these four islands were genotyped on a 10K SNP array developed for *P. domesticus* (*55*), and data were used to construct a complex pedigree using the program Cervus (*65*). A subset of the pedigree, which included genotype data on 6491 SNPs for 862 individuals included in 105 families, was used in the map construction.

A modified version of the CRIMAP 2.4 software (*66*), including added utilities provided by X. Liu and M. Grosz (Monsanto, St. Louis, MO), was used for the map construction. Initially, SNPs were assigned to linkage groups (LGs) on the basis of pairwise linkages and the grouping algorithm implemented in the AUTOGROUP option of the program. The analysis assigned 6491 SNPs to 29 larger autosomal LGs (macrochromosomes), 456 SNPs to an LG corresponding to the Z chromosome, and 37 SNPs to nine smaller LGs. Four of the smaller LGs contained one or more SNPs showing weak linkage to markers on other LGs, suggesting that they could be merged into other linkage groups, whereas the remaining five LGs built groups on their own, suggesting that they represent microchromosomes. Hence, because the house sparrow karyotype was expected to consist of 38 pairs of chromosomes [$2n = 76$; Bulatova *et al.* (*67*)], we identified LGs that probably correspond to 35 of 38 chromosomes.

After the initial grouping of SNPs, markers on the 29 larger autosomal LGs were ordered using the *BUILD* and *FLIPSN* options in CRIMAP. Following this, 120 bp flanking each SNPs were positioned in the house sparrow assembly and used to assign scaffolds to LGs, order and orientate scaffolds within LGs, and build sequences for 29 autosomal chromosomes in *P. domesticus*. If a scaffold contained only one marker in the linkage map, then the scaffold was assigned the same orientation as in the zebra finch genome.

Following the construction of chromosome sequences, the SNP order within each LG was fine-tuned using physical positions of the SNPs in the chromosome. The *CHROMPIC* option in CRIMAP was then used to phase genotypes within LGs, and a script was written to correct or remove erroneous genotypes on the basis of unlikely tight double recombination events. Finally, multipoint linkage maps for the 29 autosomal LGs were constructed using the *FIXED* option of CRIMAP and the Kosambi correction function (*68*).

Chromosome nomenclature in the house sparrow was determined by alignments against the zebra finch, flycatcher, and chicken genomes. To avoid potential confusion by adding new chromosome names for the smaller linkage groups (microchromosomes), these LGs were not included as separate chromosomes in the current assembly. Because no linkage map was constructed for the Z chromosome, scaffolds on this chromosome were ordered and orientated according to the zebra finch genome. Moreover, other avian mitochondrial genomes were used to identify the scaffold containing the house sparrow mitochondrial genome.

Sequencing

DNA was isolated using either Qiagen DNeasy 96 Blood and Tissue Kits (Qiagen N.V.), according to the manufacturer's instructions, with the exception of eluting the isolate in EB buffer instead of AE buffer (*P. italiae*, *P. hispaniolensis*, and *P. montanus* samples), or the ReliaPrep Large Volume HT gDNA Isolation System (Promega) automated on a Biomek NXp robot (Beckman Coulter), as described by Hagen *et al.* (*P. domesticus* samples).

For each sample, an Illumina TruSeq gDNA 180-bp library was created and quality-controlled for high-throughput massive parallel sequencing. The libraries were then sequenced on the Illumina HiSeq 2000 platform with 100 bp read length and three individuals per lane. All sequencing was performed by the Génome Québec at McGill University (Montreal, Canada) (www.genomequebec.com/en/home.html).

Mapping to the reference genome and variant calling

Before population genomic analyses, interspersed repeat elements were masked from the assembly with RepeatMasker (v 4.0.5) (A. F. A. Smit, R. Hubley, and P. Green; RepeatMasker at http://repeatmasker.org). The default settings were used with the exception of adding the "Do not mask simple" option, which conservatively masked only interspersed repeats and left regions of low diversity unmasked. The whole-genome sequences were mapped to the house sparrow assembly with BWA-MEM (v 0.7.5a-r405) (*69*) using default settings with the exception of adding read group identifiers and a -M parameter to enable Picard (http://picard.sourceforge.net) compatibility for downstream analyses. With the mapped reads, we then removed polymerase chain reaction duplicates with MarkDuplicates in Picard tools (v 1.72) (http://picard.sourceforge.net) with the default settings for every parameter except validation stringency, which was set to lenient. We then realigned the reads mapped to the house sparrow reference genome using GATK's IndelRealigner tool with the default settings.

An SNP set was required for some analyses. Thus, to create an SNP set from the wgs reads, we used GATK's Genome Analysis Toolkit (v 3.2.2) (*70*, *71*). The realigned .bam files were run in GATK's HaplotypeCaller to create a genomic variant call format (gVCF) file for each individual using the default settings. Next, the gVCF files were genotyped using GATK's GenotypeGVCFs function. This resulted in a large VCF file of 37,526,610 SNPs and 5,784,223 indels for all individuals across the genome.

Variants from the VCF file were further quality-filtered using VCFtools (v 0.1.12b) (*32*). SNPs with a genotype quality >20, a quality value >20, and a mean depth >5 across all individuals were kept as a set of high-quality variants. In addition, because the downstream analyses focused on the chromosomes and mitochondria, unplaced scaffolds were removed, leaving 35,867,119 SNPs and 5,389,326 indels. For some analyses, further filtering was applied and detailed in the appropriate methods section.

Ancestry estimates and population structuring

We used NgsAdmix to estimate whole-genome admixture of each sparrow individual from the focal populations because it estimated admixture on the basis of genotype likelihoods from the realigned reads (.bam files) rather than genotype calls. We first calculated genotype likelihoods from the .bam files mapped to the reference genome in ANGSD (v 0.911-47-g4705d60) (*72*) with the parameters -doGlf 2, -doMajorMinor 1, -SNP_pval 1e-6, -doMaf 1. We also filtered the reads for quality scores >20. The resulting file was then input into NgsAdmix and run with $K = 2$ and $K = 3$ ancestral populations to calculate genome-wide admixture for each individual ($K = 2$ shown in Fig. 1; $K = 3$ shown in fig. S6).

The PCA was run on the filtered SNP set excluding the tree sparrow using the R package SNPRelate (v 1.6.2) (*73*). Before the PCA, the SNP set was LD-pruned for r^2 values >0.5 in SNPRelate (l.d.threshold = 0.5). The PCA was then run for the full nuclear genome for all biallelic sites (268,962 SNPs) using default settings.

ADMIXTURE (v 1.23) (*74*) was run in 100-kb stepping windows across the genome on the SNP set that was further filtered to include only SNPs with genotypes in at least one individual of each species (34,776,981 SNPs). A window size of 100 kb was chosen as LD tends to decay within this distance in the genome (fig. S2). The SNP set was converted to a BED file format using PLINK (v 1.07) (*75*). The ADMIXTURE analysis was run with $K = 2$ set as the number of ancestral populations for the analysis. To visualize the admixture of the parent taxa's ancestry in the Italian sparrow, a prior was set fixing the parent species as K1 and K2, so that ancestry estimates to either the house or the Spanish cluster were inferred for each window across the Italian sparrow genome. Aside from this, the default settings were used.

In addition, RAxML (v 8.0.26) (*17*) was run for 100-kb stepping windows across the genome for SNPs genotyped in at least one individual in each species. The model was set to "GTRGAMMA," and the tree sparrow was designated as the outgroup. The resulting trees for each window were then categorized on the basis of whether all Italian sparrow individuals grouped monophyletically with one of its parents, in its own clade, or was "unresolved," meaning the Italian sparrow individuals did not form a monophyletic group. The same method was used for 50- and 10-kb stepping windows to test whether window size affected the proportion of resolved phylogenies.

Mitochondrial analyses

A gVCF file of mitochondrial variants was created separately by running GATK (v 3.3.0) HaplotypeCaller with its haploid option and, aside from that, default settings. The gVCF file was then genotyped with GATK's GenotypeGVCFs tool. The resulting VCF file was filtered in the same manner as the main VCF file in VCFtools (v 0.1.12b) by removing SNPs with a minimum genotype quality and/or minimum quality threshold <20 and a minimum depth of 5. The mitochondrial haplotype network was created in Fitchi (*76*) using the filtered mitochrondrial SNP set. A mitochondrial fastSTRUCTURE (*19*) analysis was also performed on the filtered mitochondrial SNP set and was further filtered for SNPs with a maximum depth of 50 and genotypes in at least 30% of individuals using the default settings and with $K = 2$ to show the admixture of the parent taxa's mitochondria in the Italian sparrow.

A separate VCF file, including calls for every site, was made using GATK's GenotypeGVFs (v 3.3.0) under its default setting with the exception of its –includeNonVariantSites option and ploidy=1 to calculate d_{XY} values for the mitochondria. d_{XY} was calculated in sliding 1-kb sliding windows with 100-bp steps for each sparrow individual along the mitochondria using Martin *et al.*'s script (v. August 2014) with –minimumExploitableData set to 0.3 and the minimum SNPs per window set to 3. The Mann-Whitney *U* test to determine whether the apparent heteroplasmic individuals differed in coverage was tested in R on the mean depth per Italian sparrow individual for the mitochondrial sequences. Similarly, a Mann-Whitney *U* test was implemented in R to determine whether nuclear *F* estimates for the Italian sparrows with heteroplasmic mtDNA differed from values of

the Italian sparrows with house mtDNA. Nuclear *F* values and mtDNA coverage were calculated on a per-individual basis using VCFtools (v 0.1.12b).

A mitochrondrial SNP set run as diploid was also created using GATK's (v 3.3.0) GenotypeGVCFs tool under default setting so that the heterozygosity of each individual's mtDNA sequence could be estimated. Overall, *F* values were calculated for each individual using VCFtools (v 0.1.12b) with the diploid SNP set filtered for a minimum depth >3 and a maximum depth of 15.

Population genomic analyses in sliding windows

Estimates for F_{ST}, nucleotide diversity, and TD were calculated in overlapping sliding windows 100 kb in size with 25-kb steps with ANGSD with a minimum quality filter set to >20, minimum mapping quality >20, and genotype likelihood model of 2. F_{ST} values are the weighted mean F_{ST} for each window across the genome. ANGSD calculated nucleotide diversity as θ_D, which was divided by the number of sites in the window to obtain a nucleotide diversity value per site for the window. TD was calculated using a folded site frequency spectrum that treats the ancestral state as unknown.

To calculate the density of fixed differences (d_f) and identify fixed differences in the genome, the SNP set created with GATK's HaplotypeCaller was further filtered so that each population had to have at least three individuals genotyped for each SNP to ensure that SNPs genotyped for very few individuals in a population were not considered fixed differences between species. Fixed differences were found using the R package PopGenome's (v 2.1.6) (*77*) biallelic matrix function to identify sites where a population is monomorphic and another population is monomorphic with a different value.

Incomplete lineage sorting complicates inference of hybridization as the two processes leave similar signatures in the genome and has therefore been emphasized as a critical test when evaluating hybrid species. To distinguish between these processes and quantify the extent of gene flow between the parent and hybrid lineages, we used a four-taxon ABBA BABA test for introgression (D-statistics) and f_d estimator (*18*, *78*, *79*). The analysis uses patterns of ancestral and derived alleles in the ingroups and outgroups to distinguish between incomplete lineage sorting and hybridization and has been shown to be a robust, although conservative, method for identifying introgressed loci (*18*). Whole-genome ABBA BABA (D-statistics) and the jack-knifed SE values were calculated for two topologies {(house, Italian), Spanish [tree]} and {(Spanish, Italian), house [tree]} with the tree sparrow individual used as the outgroup in ANGSD's ABBA BABA multipopulation tool. The tree sparrow fasta file was created directly from the .bam file in ANGSD with the -doFasta 3 option. The f_d estimator used to quantify the gene flow based on ABBA BABA statistics was calculated in 100-kb sliding windows with 25-kb steps with the script from Martin *et al.* (v. August 2014) using the high-quality SNP set with –minimumExploitableData set to 0.3 and the minimum SNPs per window set to 10. Again, the tree sparrow was designated as the outgroup.

The decay of LD was calculated across each chromosome for the three focal taxa by first filtering the SNP set for variants genotyped for at least 80% of the individuals. Pairwise r^2 values were calculated between all SNPs within a 100-kb window in PLINK (v 1.09b) (*75*). Decay plots were created by binning the distance between SNPs in 1-kb increments and averaging the r^2 values within each bin.

To provide a comparison of LD estimates within and outside of outlier divergence windows, pairwise r^2 values were also calculated in PLINK (v 1.09b) (*75*) for all SNPs (genotyped in at least 80% of individuals) within 1 kb of each other and binned into 100-kb genomic windows with 25-kb steps where each bin represents the average r^2 values for all pairwise comparisons within that window.

DoS estimates were calculated separately for all genes on across the genome, as well as for sliding windows across the genome between all three species comparisons (HI, SI, and HS). Nonsynonymous and synonymous fixed differences and polymorphisms were identified in PopGenome (v 2.1.6) (*77*) using the MKT-methods function, and their values were used to calculate DoS according to the formula by Stoletzki and Eyre-Walker (*25*).

Identification of high-divergence windows and genes

To select regions with F_{ST} values high against one parent and low against the other, we subtracted the HI F_{ST} value for every sliding window from the SI F_{ST} value and vice versa. We then took the top 1% windows for each hybrid/parent comparison as high-divergence windows. Regions where the Italian sparrow was divergent against both parent species were identified by subtracting the parent's F_{ST} value for every sliding window from both hybrid/parent comparisons, and the top 1% windows common to both comparisons were identified as private Italian regions of high differentiation (see Eqs. 1 to 3). Genes were then extracted from the sets of HI, SI, and PI high-divergence windows (table S6). A gene was considered to be in the genomic window if a portion of it falls within the region.

Statistical analysis: Gene enrichment analysis and simulations

To investigate whether there were any significantly enriched GO groups or network pathways within the high-divergence windows (SI, HI, and PI), we ran gene enrichment analyses with the ClueGo (v 2.2.5) plugin (*80*) implemented in Cytoscape (v 3.3.0) (*81*). The lists of genes found in each category of our high-divergence windows were input separately and tested with a right-sided hypergeometric enrichment test using both humans (tables S7 to S9) and, next, chickens (tables S11 to S13) as the model organisms. The GO database BiologicalProcess-GOA (09 February 2016, humans; 09 March 2016, chickens) was used along with the network specificity set to medium, Bonferroni step-down method of *P* value correction and a minimum $P < 0.05$ for reporting results. All other settings were set as default. The results between human and chicken analyses were comparable with the chickens being a subset of the human results (tables S11 to S13) because the human genome was a better-annotated genome. Thus, the results based on the analysis with the human reference genome are discussed in the main text.

Gene enrichment analysis permutations were run to test whether the enriched gene ontologies from our outlier windows could also be found when choosing the same number of windows randomly across the genome. To do this, we used BEDTools (v 2.17.0) (*82*) shuffle to randomly select the same number of 100-kb windows as found for the top SI, HI, and PI regions 50 times for each comparison and excluded windows that were already selected as high-divergence windows. We then extracted the genes from all the window selections in the same manner as genes were extracted from our outlier windows. For each permutation, we also randomly sampled genes from each gene list so that the number of genes matched that of genes found in the outlier comparison it was simulating. We then ran ClueGo (v 1.8.0) for each permutation against the human biological process database (09 February 2016, humans) under the default settings with Bonferroni correction and counted the number of times the significant gene ontologies from the

outlier windows also showed up in the randomly selected windows. None of the enriched gene ontologies from our outlier analyses reoccurred during our permutations.

Because important genes for species divergence within our outlier windows may not fall into single enriched gene groups, we also cross-checked our candidate genes with genes known to be involved in pigmentation and SNPs previously identified as being candidate RI genes in the system (table S10). Candidate genes previously identified for the system were the SNP set of species diagnostic markers between the house and Spanish sparrows by Trier *et al.* (*13*). The color gene candidate markers were taken from the list identified by Poelstra *et al.* (*28*). We used a permutation approach to test whether the number of previously identified RI candidate genes and/or candidate color genes among outlier genes was likely to occur by chance. We randomly sampled outlier sets (n = 318 genes) 10,000 times and counted RI/color genes occurring in each sample to generate a null distribution. We then used a one-sided test to examine whether the observed number was greater than the null distribution.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/3/6/e1602996/DC1

fig. S1. Stepping window ADMIXTURE and RAxML analyses across the Italian sparrow genome.
fig. S2. LD decay across all chromosomes.
fig. S3. Mitochondrial sequence divergence between each Italian sparrow and the parent populations.
fig. S4. Divergence peaks between the Italian sparrow and either parent taxa on microchromosomes.
fig. S5. Divergence peaks for PI regions on large chromosomes.
fig. S6. Population structuring analysis of sparrow individuals.
table S1. Sequencing scheme for the house sparrow reference genome assembly.
table S2. Final assembly statistics.
table S3. Sample information.
table S4. Results from genome-wide RAxML analysis.
table S5. Comparison of resolved RAxML phylogenies at variable window sizes.
table S6. ENSEMBLE gene IDs for genes within top divergence windows.
table S7. Significantly enriched GO pathways from ClueGo analysis in HI outlier windows.
table S8. Significantly enriched GO pathways from ClueGo analysis in SI outlier windows.
table S9. Significantly enriched GO pathways from ClueGo analysis in PI outlier windows.
table S10. Genes among outlier windows identified as candidates for the involvement in melanogenesis and RI in sparrows.
table S11. Significantly enriched GO pathways from ClueGo analysis HI outlier windows with chickens as the reference genome.
table S12. Significantly enriched GO pathways from ClueGo analysis in SI outlier windows with chickens as the reference genome.
table S13. Significantly enriched GO pathways from ClueGo analysis in PI outlier windows with chickens as the reference genome.

REFERENCES AND NOTES

1. O. Seehausen, R. K. Butlin, I. Keller, C. E. Wagner, J. W. Boughman, P. A. Hohenlohe, C. L. Peichel, G.-P. Sætre, C. Bank, Å. Brännström, A. Brelsford, C. S. Clarkson, F. Eroukhmanoff, J. L. Feder, M. C. Fischer, A. D. Foote, P. Franchini, C. D. Jiggins, F. C. Jones, A. K. Lindholm, K. Lucek, M. E. Maan, D. A. Marques, S. H. Martin, B. Matthews, J. I. Meier, M. Möst, M. W. Nachman, E. Nonaka, D. J. Rennison, J. Schwarzer, E. T. Watson, A. M. Westram, A. Widmer, Genomics and the origin of species. *Nat. Rev. Genet.* **15**, 176–192 (2014).
2. R. Abbott, D. Albach, S. Ansell, J. W. Arntzen, S. J. E. Baird, N. Bierne, J. Boughman, A. Brelsford, C. A. Buerkle, R. Buggs, R. K. Butlin, U. Dieckmann, F. Eroukhmanoff, A. Grill, S. H. Cahan, J. S. Hermansen, G. Hewitt, A. G. Hudson, C. Jiggins, J. Jones, B. Keller, T. Marczewski, J. Mallet, P. Martinez-Rodriguez, M. Möst, S. Mullen, R. Nichols, A. W. Nolte, C. Parisod, K. Pfennig, A. M. Rice, M. G. Ritchie, B. Seifert, C. M. Smadja, R. Stelkens, J. M. Szymura, R. Väinölä, J. B. W. Wolf, D. Zinner, Hybridization and speciation. *J. Evol. Biol.* **26**, 229–246 (2013).
3. J. A. Coyne, H. A. Orr, *Speciation* (Sinauer Associates Inc., 2004).
4. J. Mallet, Hybrid speciation. *Nature* **446**, 279–283 (2007).
5. Heliconius Genome Consortium, Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* **487**, 94–98 (2012).
6. L. H. Rieseberg, O. Raymond, D. M. Rosenthal, Z. Lai, K. Livingstone, T. Nakazato, J. L. Durphy, A. E. Schwarzbach, L. A. Donovan, C. Lexer, Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* **301**, 1211–1216 (2003).
7. K. Kunte, C. Shea, M. L. Aardema, J. M. Scriber, T. E. Juenger, L. E. Gilbert, M. R. Kronforst, Sex chromosome mosaicism and hybrid speciation among tiger swallowtail butterflies. *PLOS Genet.* **7**, e1002274 (2011).
8. T. R. Anderson, *Biology of the Ubiquitous House Sparrow: From Genes to Populations* (Oxford Univ. Press Inc., 2006).
9. J. D. Summers-Smith, *The Sparrows: A Study of the Genus Passer* (T & AD Poyser, 1988).
10. W. Meise, Zur Systematik und Verbreitungsgeschichte der Haus- und Weidensperlinge, *Passer domesticus* (L.) und *hispaniolensis* (T.). *J. Ornithol.* **84**, 631–672 (1936).
11. T. O. Elgvin, J. S. Hermansen, A. Fijarczyk, T. Bonnet, T. Borge, S. A. Sæther, K. L. Voje, G.-P. Sætre, Hybrid speciation in sparrows II: A role for sex chromosomes? *Mol. Ecol.* **20**, 3823–3837 (2011).
12. J. S. Hermansen, S. A. Sæther, T. O. Elgvin, T. Borge, E. Hjelle, G.-P. Sætre, Hybrid speciation in sparrows I: Phenotypic intermediacy, genetic admixture and barriers to gene flow. *Mol. Ecol.* **20**, 3812–3822 (2011).
13. C. N. Trier, J. S. Hermansen, G.-P. Sætre, R. I. Bailey, Evidence for mito-nuclear and sex-linked reproductive barriers between the hybrid Italian sparrow and its parent species. *PLOS Genet.* **10**, e1004075 (2014).
14. J. S. Hermansen, F. Haas, C. N. Trier, R. I. Bailey, A. J. Nederbragt, A. Marzal, G.-P. Sætre, Hybrid speciation through sorting of parental incompatibilities in Italian sparrows. *Mol. Ecol.* **23**, 5831–5842 (2014).
15. M. Schumer, G. G. Rosenthal, P. Andolfatto, How common is homoploid hybrid speciation? *Evolution* **68**, 1553–1560 (2014).
16. J. Mallet, Hybridization as an invasion of the genome. *Trends Ecol. Evol.* **20**, 229–237 (2005).
17. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
18. S. H. Martin, J. W. Davey, C. D. Jiggins, Evaluating the use of ABBA–BABA statistics to locate introgressed loci. *Mol. Biol. Evol.* **32**, 244–257 (2015).
19. A. Raj, M. Stephens, J. K. Pritchard, fastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics* **197**, 573–589 (2014).
20. P. Danecek, A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin; 1000 Genomes Project Analysis Group, The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
21. T. E. Cruickshank, M. W. Hahn, Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol. Ecol.* **23**, 3133–3157 (2014).
22. M. A. F. Noor, S. M. Bennett, Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity* **103**, 439–444 (2009).
23. B. Charlesworth, Measures of divergence between populations and the effect of forces that reduce variability. *Mol. Biol. Evol.* **15**, 538–543 (1998).
24. F. Tajima, Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595 (1989).
25. N. Stoletzki, A. Eyre-Walker, Estimation of the neutrality index. *Mol. Biol. Evol.* **28**, 63–70 (2011).
26. J. Charlesworth, A. Eyre-Walker, The McDonald–Kreitman test and slightly deleterious mutations. *Mol. Biol. Evol.* **25**, 1007–1015 (2008).
27. R. I. Bailey, M. R. Tesaker, C. N. Trier, G.-P. Sætre, Strong selection on male plumage in a hybrid zone between a hybrid bird species and one of its parents. *J. Evol. Biol.* **28**, 1257–1269 (2015).
28. J. W. Poelstra, N. Vijay, M. P. Hoeppner, J. B. W. Wolf, Transcriptomics of colour patterning and coloration shifts in crows. *Mol. Ecol.* **24**, 4617–4628 (2015).
29. A. Abzhanov, M. Protas, B. R. Grant, P. R. Grant, C. J. Tabin, *Bmp4* and morphological variation of beaks in Darwin's finches. *Science* **305**, 1462–1465 (2004).
30. R. B. Roberts, Y. Hu, R. C. Albertson, T. D. Kocher, Craniofacial divergence and ongoing adaptation via the hedgehog pathway. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 13194–13199 (2011).
31. H. Abe, M. Inoue-Murayama, Structural variation of G protein-coupled receptor in birds. *Recept. Clin. Invest.* **1**, e162 (2014).
32. B. Kmiec, M. Woloszynska, H. Janska, Heteroplasmy as a common state of mitochondrial genetic information in plants and animals. *Curr. Genet.* **50**, 149–159 (2006).
33. J. M. Radojičić, I. Krizmanić, P. Kasapidis, E. Zouros, Extensive mitochondrial heteroplasmy in hybrid water frog (*Pelophylax* spp.) populations from Southeast Europe. *Ecol. Evol.* **5**, 4529–4541 (2015).
34. H. Shitara, J.-I. Hayashi, S. Takahama, H. Kaneda, H. Yonekawa, Maternal inheritance of mouse mtDNA in interspecific hybrids: Segregation of the leaked paternal mtDNA followed by the prevention of subsequent paternal leakage. *Genetics* **148**, 851–857 (1998).
35. L. Kvist, J. Martens, A. A. Nazarenko, M. Orell, Paternal leakage of mitochondrial DNA in the great tit (*Parus major*). *Mol. Biol. Evol.* **20**, 243–247 (2003).
36. L. Bromham, A. Eyre-Walker, N. H. Smith, J. M. Smith, Mitochondrial Steve: Paternal inheritance of mitochondria in humans. *Trends Ecol. Evol.* **18**, 2–4 (2003).

37. A. Qvarnström, R. I. Bailey, Speciation through evolution of sex-linked genes. *Heredity* **102**, 4–15 (2009).
38. B. Charlesworth, J. A. Coyne, N. H. Barton, The relative rates of evolution of sex chromosomes and autosomes. *Am. Natural.* **130**, 113–146 (1987).
39. J. E. Mank, E. Axelsson, H. Ellegren, Fast-X on the Z: Rapid evolution of sex-linked genes in birds. *Genome Res.* **17**, 618–624 (2007).
40. H. Ellegren, The different levels of genetic diversity in sex chromosomes and autosomes. *Trends Genet.* **25**, 278–284 (2009).
41. L. H. Rieseberg, Hybrid origins of plant species. *Annu. Rev. Ecol. Syst.* **28**, 359–389 (1997).
42. C. Salazar, S. W. Baxter, C. Pardo-Diaz, G. Wu, A. Surridge, M. Linares, E. Bermingham, C. D. Jiggins, Genetic evidence for hybrid trait speciation in *Heliconius* butterflies. *PLOS Genet.* **6**, e1000930 (2010).
43. M. C. Melo, C. Salazar, C. D. Jiggins, M. Linares, Assortative mating preferences among hybrids offers a route to hybrid speciation. *Evolution* **63**, 1660–1665 (2009).
44. B. L. Gross, L. H. Rieseberg, The ecological genetics of homoploid hybrid speciation. *J. Hered.* **96**, 241–252 (2005).
45. H. Jensen, I. Steinsland, T. H. Ringsby, B.-E. Sæther, Evolutionary dynamics of a sexual ornament in the house sparrow (*Passer domesticus*): The role of indirect selection within and between sexes. *Evolution* **62**, 1275–1293 (2008).
46. P. R. Grant, B. R. Grant, *How and Why Species Multiply: The Radiation of Darwin's Finches* (Princeton Univ. Press, 2008).
47. F. Eroukhmanoff, J. S. Hermansen, R. I. Bailey, S.-A. Sæther, G.-P. Sætre, Local adaptation within a hybrid species. *Heredity* **111**, 286–292 (2013).
48. B. Charlesworth, M. T. Morgan, D. Charlesworth, The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**, 1289–1303 (1993).
49. R. Burri, A. Nater, T. Kawakami, C. F. Mugal, P. I. Olason, L. Smeds, A. Suh, L. Dutoit, S. Bureš, L. Z. Garamszegi, S. Hogner, J. Moreno, A. Qvarnström, M. Ružić, S.-A. Sæther, G.-P. Sætre, J. Török, H. Ellegren, Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. *Genome Res.* **25**, 1656–1665 (2015).
50. L. H. Rieseberg, M. A. Archer, R. K. Wayne, Transgressive segregation, adaptation and speciation. *Heredity* **83**, 363–372 (1999).
51. C. A. Buerkle, R. J. Morris, M. A. Asmussen, L. H. Rieseberg, The likelihood of homoploid hybrid speciation. *Heredity* **84**, 441–451 (2000).
52. A. Ferrer-Admetlla, E. Bosch, M. Sikora, T. Marqués-Bonet, A. Ramírez-Soriano, A. Muntasell, A. Navarro, R. Lazarus, F. Calafell, J. Bertranpetit, F. Casals, Balancing selection is the main force shaping the evolution of innate immunity genes. *J. Immunol.* **181**, 1315–1322 (2008).
53. B. A. Payseur, L. H. Rieseberg, A genomic perspective on hybridization and speciation. *Mol. Ecol.* **25**, 2337–2360 (2016).
54. A. M. Billing, A. M. Lee, S. Skjelseth, Å. A. Borg, M. C. Hale, J. Slate, H. Pärn, T. H. Ringsby, B.-E. Sæther, H. Jensen, Evidence of inbreeding depression but not inbreeding avoidance in a natural house sparrow population. *Mol. Ecol.* **21**, 1487–1499 (2012).
55. I. J. Hagen, A. M. Billing, B. Rønning, S. A. Pedersen, H. Pärn, J. Slate, H. Jensen, The easy road to genome-wide medium density SNP screening in a non-model species: Development and application of a 10 K SNP-chip for the house sparrow (*Passer domesticus*). *Mol. Ecol. Resour.* **13**, 429–439 (2013).
56. K. R. Bradnam, J. N. Fass, A. Alexandrov, P. Baranay, M. Bechner, I. Birol, S. Boisvert, J. A. Chapman, G. Chapuis, R. Chikhi, H. Chitsaz, W.-C. Chou, J. Corbeil, C. Del Fabbro, T. R. Docking, R. Durbin, D. Earl, S. Emrich, P. Fedotov, N. A. Fonseca, G. Ganapathy, R. A. Gibbs, S. Gnerre, É. Godzaridis, S. Goldstein, M. Haimel, G. Hall, D. Haussler, J. B. Hiatt, I. Y. Ho, J. Howard, M. Hunt, S. D. Jackman, D. B. Jaffe, E. D. Jarvis, H. Jiang, S. Kazakov, P. J. Kersey, J. O. Kitzman, J. R. Knight, S. Koren, T.-W. Lam, D. Lavenier, F. Laviolette, Y. Li, Z. Li, B. Liu, Y. Liu, R. Luo, I. MacCallum, M. D. MacManes, N. Maillet, S. Melnikov, D. Naquin, Z. Ning, T. D. Otto, B. Paten, O. S. Paulo, A. M. Phillippy, F. Pina-Martins, M. Place, D. Przybylski, X. Qin, C. Qu, F. J. Ribeiro, S. Richards, D. S. Rokhsar, J. G. Ruby, S. Scalabrin, M. C. Schatz, D. C. Schwartz, A. Sergushichev, T. Sharpe, T. I. Shaw, J. Shendure, Y. Shi, J. T. Simpson, H. Song, F. Tsarev, F. Vezzi, R. Vicedomini, B. M. Vieira, J. Wang, K. C. Worley, S. Yin, S.-M. Yiu, J. Yuan, G. Zhang, H. Zhang, S. Zhou, I. F. Korf, Assemblathon 2: Evaluating de novo methods of genome assembly in three vertebrate species. *GigaScience* **2**, 1–31 (2013).
57. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12 (2011).
58. C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T. L. Madden, BLAST+: Architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
59. C. Holt, M. Yandell, MAKER2: An annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12**, 491 (2011).
60. M. Yandell, D. Ence, A beginner's guide to eukaryotic genome annotation. *Nat. Rev. Genet.* **13**, 329–342 (2012).
61. A. V. Lukashin, M. Borodovsky, GeneMark.hmm: New solutions for gene finding. *Nucleic Acids Res.* **26**, 1107–1115 (1998).
62. G. Parra, K. Bradnam, I. Korf, CEGMA: A pipeline to accurately annotate core genes in eukaryotic genornes. *Bioinformatics* **23**, 1061–1067 (2007).
63. I. Korf, Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004).
64. M. Stanke, A. Tzvetkova, B. Morgenstern, AUGUSTUS at EGASP: Using EST, protein and genomic alignments for improved gene prediction in the human genome. *Genome Biol.* **7** (suppl. 1), S11.1–S11.8 (2006).
65. S. T. Kalinowski, M. L. Taper, T. C. Marshall, Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol. Ecol.* **16**, 1099–1106 (2007).
66. J.-P. Bidanel, D. Milan, N. Iannuccelli, Y. Amigues, M.-Y. Boscher, F. Bourgeois, J.-C. Caritez, J. Gruand, P. Le Roy, H. Lagant, R. Quintanilla, C. Renard, J. Gellin, L. Ollivier, C. Chevalet, Detection of quantitative trait loci for growth and fatness in pigs. *Genet. Sel. Evol.* **33**, 289–309 (2001).
67. N. S. Bulatova, S. I. Radjabli, E. N. Panov, Karyological description of three species of the genus *Passer*. *Experientia* **28**, 1369–1371 (1972).
68. D. D. Kosambi, The estimation of map distances from recombination values. *Ann. Hum. Genet.* **12**, 172–175 (1943).
69. H. Li, R. Durbin, Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
70. M. A. DePristo, E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, G. del Angel, M. A. Rivas, M. Hanna, A. McKenna, T. J. Fennell, A. M. Kernytsky, A. Y. Sivachenko, K. Cibulskis, S. B. Gabriel, D. Altshuler, M. J. Daly, A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
71. A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, M. A. DePristo, The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
72. T. S. Korneliussen, I. Moltke, A. Albrechtsen, R. Nielsen, Calculation of Tajima's *D* and other neutrality test statistics from low depth next-generation sequencing data. *BMC Bioinformatics* **14**, 289 (2013).
73. X. Zheng, D. Levine, J. Shen, S. M. Gogarten, C. Laurie, B. S. Weir, A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**, 3326–3328 (2012).
74. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
75. S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, P. C. Sham, PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
76. M. Matschiner, Fitchi: Haplotype genealogy graphs based on the Fitch algorithm. *Bioinformatics* **32**, 1250–1252 (2016).
77. B. Pfeifer, U. Wittelsbürger, S. E. Ramos-Onsins, M. J. Lercher, PopGenome: An efficient Swiss army knife for population genomic analyses in R. *Mol. Biol. Evol.* **31**, 1929–1936 (2014).
78. E. Y. Durand, N. Patterson, D. Reich, M. Slatkin, Testing for ancient admixture between closely related populations. *Mol. Biol. Evol.* **28**, 2239–2252 (2011).
79. R. E. Green, J. Krause, A. W. Briggs, T. Maricic, U. Stenzel, M. Kircher, N. Patterson, H. Li, W. Zhai, M. H.-Y. Fritz, N. F. Hansen, E. Y. Durand, A.-S. Malaspinas, J. D. Jensen, T. Marques-Bonet, C. Alkan, K. Prüfer, M. Meyer, H. A. Burbano, J. M. Good, R. Schultz, A. Aximu-Petri, A. Butthof, B. Höber, B. Höffner, M. Siegemund, A. Weihmann, C. Nusbaum, E. S. Lander, C. Russ, N. Novod, J. Affourtit, M. Egholm, C. Verna, P. Rudan, D. Brajkovic, Ž. Kucan, I. Gušic, V. B. Doronichev, L. V. Golovanova, C. Lalueza-Fox, M. de la Rasilla, J. Fortea, A. Rosas, R. W. Schmitz, P. L. F. Johnson, E. E. Eichler, D. Falush, E. Birney, J. C. Mullikin, M. Slatkin, R. Nielsen, J. Kelso, M. Lachmann, D. Reich, S. Pääbo, A draft sequence of the Neandertal genome. *Science* **328**, 710–722 (2010).
80. G. Bindea, B. Mlecnik, H. Hackl, P. Charoentong, M. Tosolini, A. Kirilovsky, W.-H. Fridman, F. Pagès, Z. Trajanoski, J. Galon, ClueGO: A Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25**, 1091–1093 (2009).
81. P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
82. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
83. L. Svensson, P. J. Grant, K. Mullarney, D. Zetterstroem, *Gyldendals store fugleguide: Europas og middelhavsområdets fugler i felt* (Gyldendal, 1999).

Acknowledgments: We thank F. Eroukhmanoff and A. Runemark for helpful comments on the early versions of the manuscript and A. Mazzarella, S. Jentoft, A. Tooming-Klunderud, R. Røsbak, and K. Yttersian Sletta for technical assistance. All computational work was performed on the Abel Supercomputing Cluster [Norwegian Metacenter for High Performance Computing (NOTUR) and the University of Oslo] operated by the Research Computing Services group at The

University Center for Information Technology (www.hpc.uio.no/). Sequencing library creation and high-throughput sequencing were carried out at the NSC, University of Oslo, Norway, and McGill University and Génome Québec Innovation Centre, Canada. This project received support from RCN project 208481, and we especially acknowledge the work carried out by J. K. A. Samy in establishing the genome browser, available at http://cees-genomes.hpc.uio.no/gb2/gbrowse/house_sparrow or just cees-genomes.hpc.uio.no/gb2/gbrowse/house_sparrow. **Funding:** This work was supported by The Research Council of Norway grants 240557, 221956, and 223257.
Author contributions: T.O.E., C.N.T., and G.-P.S. designed the research. T.O.E., C.N.T., O.K.T., and A.J.N. created the house sparrow reference genome. S.L., I.J.H., and H.J. created the house sparrow linkage map. T.O.E., C.N.T., O.K.T., and M.R. analyzed the data. T.O.E. and C.N.T. wrote the first manuscript draft. All authors contributed to the writing of the paper. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors. The raw data produced for this project have been deposited at the NCBI Sequence Read Archive under BioProject PRJNA255814 accession numbers SRR5407744–SRR5407749 (house sparrow reference assembly) and SRR5369936–SRR5369966 (population whole-genome sequencing). The house sparrow reference assembly has been deposited at DDBJ/ENA/GenBank under accession MBAE00000000. The version described in this paper is version MBAE01000000.

Submitted 9 December 2016
Accepted 26 April 2017
Published 14 June 2017
10.1126/sciadv.1602996

Citation: T. O. Elgvin, C. N. Trier, O. K. Tørresen, I. J. Hagen, S. Lien, A. J. Nederbragt, M. Ravinet, H. Jensen, G.-P. Sætre, The genomic mosaicism of hybrid speciation. *Sci. Adv.* **3**, e1602996 (2017).