The evolutionary fate of a hybrid lineage

The potential for genomic differentiation in an admixed species

Angélica María Cuevas Pulido

Dissertation presented for the degree of
Philosophiae Doctor (*PhD*)
2022

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

ACKNOWLEDGMENTS

A mi Mamita y Carito, no creo que me alcancen las palabras para agradecerles su apoyo y amor incondicional, sin el cual nunca hubiese tenido la fuerza para embarcarme en este viaje tan lejos de casa, que empezó mucho antes de este PhD y que ya nosotras sabemos bien, ha durado mucho mas de lo esperado. Mil gracias por siempre creer en mi, especialmente cuando me fuese difícil creer en mi misma. Gracias por sus sacrificios y por siempre estar dispuestas a escucharme, las incontables horas de charlas virtuales que nos hacen sentir cerca aun cuando un océano nos separa, son invaluables. Tenerlas a ustedes es lo que me mantiene a flote. Las amo hasta mas allá del sol!

László, I have been very lucky to have you by my side in this last stretch of my PhD. Many thanks for your patience and kindness, for lifting my spirit when I most needed it, for cheering me up with homemade meals and ice cream sandwiches. Köszönöm szépen! You have definitely made this journey a lot easier. Te quiero!

To my favourite Norwegians Thea and Kaja, who made Norway such a welcoming place, you made me feel at home when I was missing family and thinking I could not be further away. To my great friends Hanneke, Sabrina, Mari, Iulia and Luis, thanks for making me laugh, for all the brunches, long existential conversations, summer swims, wine + quiz nights, contemporary dance, you name it. Having friends like you is like having family in a foreign land, your support kept me going and made my time in Oslo much more enjoyable.

TABLE OF CONTENTS

LIST OF PAPERS

Paper I

Intraspecific genomic variation and local adaptation in a young hybrid species.

Angélica Cuevas, Mark Ravinet, Glenn-Peter Sætre and Fabrice Eroukhmanoff
Molecular Ecology
https://doi.org/10.1111/mec.15760

Paper II

Predictors of genomic differentiation within a hybrid taxon.

Angélica Cuevas, Fabrice Eroukhmanoff, Mark Ravinet, Glenn-Peter Sætre, Anna Runemark
PLOS Genetics
https://doi.org/10.1371/journal.pgen.1010027

Paper III

Rapid polygenic response to secondary contact in a hybrid species.

Glenn-Peter Sætre*, **Angélica Cuevas*,** Jo Skeie Hermansen*, Tore Oldeide Elgvin, Laura Piñeiro
Fernández, Sæther, Stein A., Camila Lo Cascio Sætre, and Fabrice Eroukhmanoff
*These authors contributed equally to this work
Proceedings of the Royal Society B.
https://doi.org/10.1098/rspb.2017.0365

SUMMARY

Hybridization plays an important role in the evolution, adaptation and diversity of species, and ultimately in the speciation process. Mechanisms like adaptive introgression have been found to be more common than previously thought. Moreover, hybridization events have the potential to trigger species radiation. Homoploid hybrid speciation has been a controversial topic regarding both the genomic integrity of the hybrid lineage and the evolution of reproductive isolation to the parental species. It is well-known that hybridization events can lead to the generation of novel admixed populations that might have the potential to prevail and become their own individual lineage. However, it is also interesting to ask what are the challenges and opportunities a hybrid species faces after the development of some reproductive barriers against the parental species. Whether the potential for divergence and rapid adaptation, conferred by novel rearrangements of parental genomes, could be sufficient for a hybrid species to thrive. Alternatively, whether genomic incompatibilities can drastically limit the evolution of a hybrid lineage. Moreover, how external environmental pressures play a role on the evolution of an admixed genome and whether a hybrid species is capable of rapidly adapt to novel challenges, like secondary contact with its parents. In light of these questions, my dissertation aims to study the evolutionary potential and limitations of a hybrid lineage, the Italian sparrow, a hybrid species resulting from hybridization between the house and the Spanish sparrow. Investigating the genomic differentiation of the Italian sparrow along the Italian peninsula (**paper I**) we found that climatic variation has an effect on genomic variation, suggesting ongoing processes of local adaptation in this hybrid species. Nonetheless, results from **paper I** and **II** suggest that incompatibilities are a limiting factor to the evolution of the Italian sparrow. Looking to determine the most important factors affecting genomic variation (**paper II**) we revealed shared patterns of differentiation among populations inhabiting the islands of Corsica, Crete and Sicily, which could be jointly explained by constraints due to incompatibilities and selection. Moreover, less steep relationships between recombination and genomic differentiation found in the islands' lineages, in comparison to parental species, may suggest an effect of unbalanced purging of minor-parent alleles in regions of low recombination, reducing the variation available for divergence (**paper II**). Finally, after secondary contact with one of its parents (**paper III**), the Spanish sparrow, the hybrid Italian sparrow exhibits a pattern of polygenic character displacement accompanied by habitat segregation, poorer body condition and a significant drop in population size following the recent invasion of the Spanish sparrow. Overall, my work sheds light on the factors that shape genomic variation in a hybrid species, specifically focusing on the potential and challenges that a hybrid lineage could encounter.

INTRODUCTION

INTRODUCTION

Speciation – A historic background

The changing nature of organisms has long fascinated intellectuals and philosophers. In the 4th century BC, Aristotle considered nature to be dynamic and as such, he suggested that it presents qualitative traits of change like formation, destruction and transformation. In _De Generatione Animalium_ he put forward the idea that complex beings could originate from much simpler organisms (Andrade, 2009). Later, in the 18th and 19th centuries, philosophers and naturalists from the German _Naturphilosophie_ like Immanuel Kant and Georg W F Hegel proposed a line of thought explaining the diversity in nature as a consequence of a guided change towards higher forms, a conceptual hierarchical transformation of organisms. French naturalists, like George-Louis Leclerc de Buffon and Jean-Baptiste Lamarck, among others, influenced by Newtonianism, also presented the idea of change through time, including components like gradualism and continuity, where simpler organisms would transform towards higher complexity (Andrade, 2009). These early lines of thoughts and their exponents did not have a robust scientific framework, and lacked an empirical foundation, thus presenting critical flaws, such as a deterministic concept of hierarchical transformation towards higher levels of complexity. However, they represent the early stages of a non–static notion of nature and the possibility of transformation through time, paving the way for the development of the evolutionary theory later established by Darwin.

Darwin's reflections that led to the basis of the evolutionary theory were greatly influenced by the geologist Charles Lyell. The uniformitarian geological change proposed by Lyell gave Darwin the temporal framework needed to suggest that transformation in biological organisms could be gradual (Gould, 1987). Darwin also explored ideas on embryology, morphology and phylogeny developed by Ernst Haeckel and Karl von Baer, guiding his thought to the possibility of a common ancestor shared by different species. Thomas Malthus's essay on population growth and the scarcity of resources was also crucial for the development of Darwin's concept of natural selection. It helped him develop the idea of natural selection where environmental pressures act as the external driving forces that lead species to adapt and diversify (Vorzimmer, 1969; Depew & Weber, 1995; Andrade, 2009). After his travels on the _Beagle_ Darwin integrated his views on embryology, geology and phylogeny together with population dynamics and the patterns and diversity in nature that he observed during his journey, landing on a more robust theory of evolution based on natural selection, famously published on 1859 with his book '_On the origin of species by means of natural selection_' (hereafter "_The Origin_"). The development of his ideas on evolution leading to the

publication of "*The Origin*" can also be found in a series of notebooks known as the Transmutation Notebooks (Kottler, 1978; Mallet, 2008b). Darwin's theory of evolution by natural selection was shared by Alfred Russel Wallace, which was jointly presented to the Linnean Society in 1858 and resulted in a publication titled "*On the Tendency of Species to form Varieties; and on the Perpetuation of Varieties and Species by Natural Means of Selection*". Darwin's, and Wallace's, work is the foundation of the evolutionary thinking, bringing attention to the processes and mechanisms of what we now call speciation, where nature not only shapes populations and selects for advantageous traits but can also lead to the divergence of a lineage into two completely independent species. Some prominent biologists, like Ernst Mayr and Theodosius Dobzhansky, among others, considered that Darwin never properly addressed the process of speciation, the very same dilemma posed in the title of his book, mainly because, according to Mayr, Darwin lacked an accurate definition of species based on interbreeding barriers (Mallet 2008; Mallet 2010). However, Darwin presented the idea of continuity of the evolutionary change, and stated that it could occur in varieties (variation within a population) as well as in higher hierarchical levels as species. According to Darwin, species present much bigger morphological gaps while varieties (or subspecies) can present intermediate forms. These gaps were explained by the principle of extinction, a consequence of selective pressures eliminating the intermediates. In his early notes, *The Transmutation* notebooks, Darwin presents one of the most known drawings (Fig. 1A), and later modified in *The Origin*, where he exposed the idea of origin of differential lineages across time as a result of divergent evolution of populations sharing a common ancestor (Kottler, 1978), implying that evolutionary laws that apply to varieties (populations) can also transform species (Mallet, 2008b, 2010). Although Darwin never explicitly developed a species concept in terms of reproductive barriers, he was well aware of hybrid sterility (explored in the "*Hybridims*" Chapter VIII on *The Origin*) and mentions geographic isolation and sexual selection as important factors for speciation (Mallet, 2008b). However, it seems that Darwin considered that reproductive isolation was not an adequate criterion to delimitate species. Thus, Darwin posed the idea of continuity, highlighting the difficulty of defining sharp boundaries between gradual variation, and in some way getting closer to the concept of a speciation continuum.

With the birth of genetics by Mendel (1866), and its independent rediscovery in the 20th century by Hugo de Vries, Karl Joseph Correns and Erich Tschermak von Seysenegg (Fairbanks & Abbott, 2016), gene theory and the laws of inheritance were introduced to biology. They constitute the missing link that Darwin's evolutionary theory was lacking to explain inheritance of adaptive traits across generations. The gene concept, which can be viewed as a discrete inheritance element, also

permitted the integration of mathematical approaches, allowing a new angle for the study of evolution. Later, in the very early beginning of population genetics, Charles Pierce suggested that the evolution of organisms follows statistical laws (Pierce, 1877 in Andrade, 2009). However, only with the incorporation of probability theory by Ronald Fisher in 1918, explaining evolutionary population dynamics as changes in the frequency of genes, was the science of population genetics formalized (Fisher, 1918). Fisher offered the mathematical framework that allows the understanding of evolution as a gradual process based on the generation of heritable genetic variation, by combination of alleles, giving natural selection the material to act on.
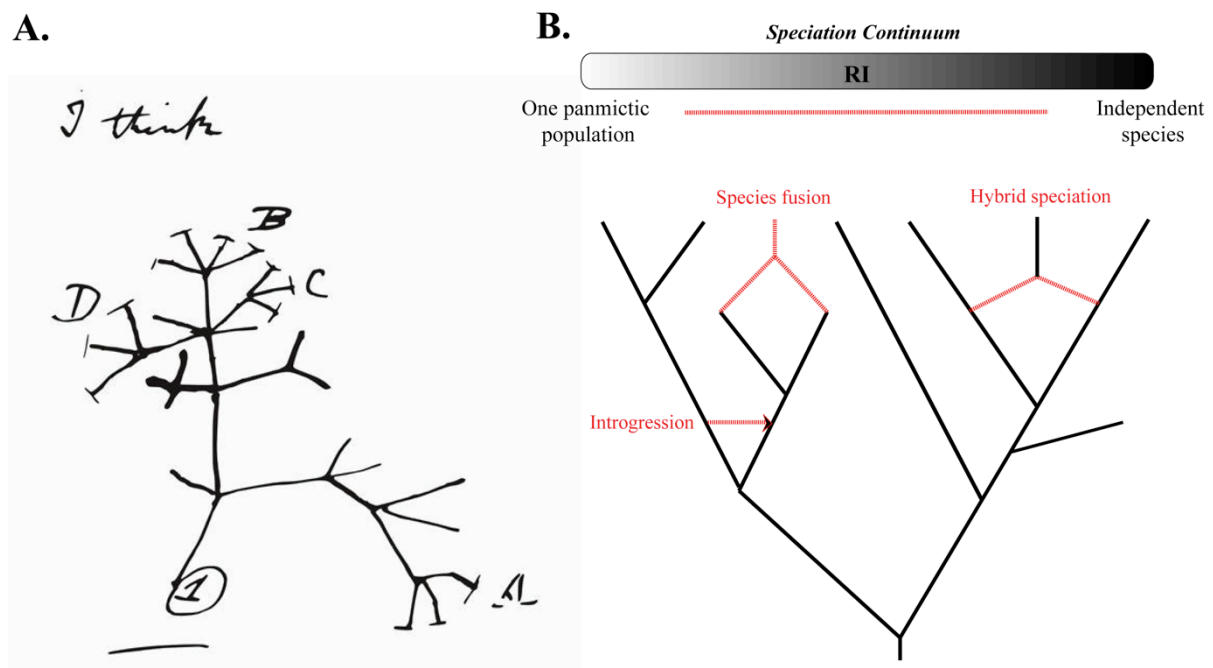
Figure 1. The speciation process. A. Darwin's sketch on the process of speciation, drawn in his Transmutation Notebooks. **B.** The current interpretation of the evolutionary process. The speciation continuum as a range of reproductive isolation (RI) where hybridization can occur at intermediate levels of RI, marked in red. And a phylogeny depicting some of the different evolutionary outcomes, including those triggered by hybridization (in red).

By the 20th century Sewall Wright further developed Fisher's proposal and incorporated other elements that play a role in the evolution of a population. Wright suggested several modes of evolution as potential scenarios in which changes in gene frequencies can lead to a fitness peak in an adaptive landscape. He studied evolution as a statistical process incorporating factors like mutation rate, migration rate, level of selective pressure, population size and epistasis. He also considered the random process of genic drift as a process that could act together with selection and result in a climb towards a fitness peak (Wright, 1931, 1932). Wright's model of population genetics, and his proposed different evolutionary scenarios, inspired two of the most influential

schools of thought in evolution: the neutral theory developed mainly by Motto Kimura and Tomoko Ohta, where genetic drift, neutral mutation and recombination are thought to play a major role in (molecular) evolution; and adaptationism developed by Theodosius Dobzhansky, among others, where natural selection was thought to favour diversity (e.g. through *balancing selection*) and ecological niches would determine the fitness peaks in Wright's adaptive landscape, thus making adaptability a consequential property. The adaptationism of Dobzhansky quickly influenced disciplines like biogeography, systematics and palaeontology with scientists like Ernest Mayr and George Simpson conceptualizing it in what we know as the Modern Synthesis of Evolution (1948) (in Andrade, 2009).

Reproductive Isolation, Sympatry and the Speciation Continuum

During the Modern Synthesis, Ernst Mayr was one of the main proponents of the study of speciation and its mechanisms. Among his most influential contributions is the biological species concept (BSC) and the emphasis on allopatric speciation, which he thought as the only real mechanism for species to diverge (Coyne, 1994; Mallet, 2008b; Wang et al., 2020; Wu et al., 2020). In the allopatric speciation model geographic isolation facilitates the development of reproductive isolation between diverging populations that would otherwise be counteracted by interbreeding and gene flow in non-allopatric scenarios (Mayr, 1942). Similar ideas were initially exposed by Dobzhansky in his discussion on "isolating mechanisms" (Dobzhansky, 1937). For Dobzhansky the ultimate step of speciation was *reinforcement*, meaning that allopatric speciation was only complete when divergent populations come into secondary contact and the accumulated genetic divergence was strengthened (Coyne, 1994). Mayr criticized sympatric speciation and considered it an implausible mode of speciation, arguing that gene flow would hinder the development of reproductive isolation, crucial for populations to evolve into independent new lineages. After the presentation of these concepts in his book *"Systematics and the Origin of Species"* (Mayr, 1942) speciation became understood, as Coyne (1994) phrases it: *"speciation... as the origin of reproductive isolating mechanisms"*; species were thus defined in terms of their reproductive isolation and natural selection became less relevant for speciation (Mallet, 2008b). Even though allopatric speciation and the biological species concept (BSC) were widely accepted by the scientific community, evidence supporting sympatric speciation started an early debate (Mallet, 2008b). Contemporary to Mayr, Huxley (1942) exposed cases of altitudinal differences in birds, species radiation of fish and differential adaptation in *Peromyscus* rodents as examples of the gradual change, an idea similar to the continuity in speciation initially presented by Darwin (Mallet, 2008a). Nonetheless, Darwin

recognized and discussed the importance of geographic isolation on the generation of new lineages and the development of reproductive isolation. He, however, did not define species on the bases of reproductive barriers. For Darwin the key to speciation laid in the '*principle of divergence*', natural selection and extinction (Mallet, 2008b, 2008a).

Thus, the debate on speciation mainly focused on i.) genetic mechanisms upon which barriers to gene flow evolve, facilitating the development of **reproductive isolation (RI)** and ii.) whether variation and selection (or drift) could generate divergent lineages in sympatry or whether geographic isolation is crucial for the generation of reproductive barriers.

Initially proposed by Bateson (1909) and later developed by Dobzhansky (1937) and Muller (1942) (in Futuyma, 2013), a theory on how **incompatibilities** can arise in the genome, without presenting a detrimental effect on the carrier of the mutation that have a negative effect to interbreeding, aimed to explain the development of **RI**. They proposed that epistatic interactions between divergent alleles that initially developed in different populations would present genetic incompatibilities if they were to meet in the same genome (i.e. a hybrid individual), mainly because such combination of alleles have never been tested by natural selection (Fig. 2). In this scenario carriers of new alleles that harbour potential reproductive barriers to interbreeding will not suffer negative effects to their reproductive success when interbreeding with individuals within the same population, and such novel variations may have fixed genomic differences involved in reproductive isolation.

RI can be developed by several types of reproductive barriers (including incompatibilities through the epistatic effects between derived loci): i) Premating barriers, those that prevent interbreeding (e.g ecological or sexual isolation), ii) Postmating, prezygotic barriers, which hinder the production of a zygote even if members of different species mate (e.g. mechanical isolation of gametes) and iii) Postzygotic barriers, where hybrid individuals are formed but their survival and maintenance is problematic (e.g. low fitness, hybrid inviability or sterility). *Haldane's rule* is one of the most known and well-studied examples of postzygotic barriers, more specifically of hybrid sterility, where the heterogametic sex is infertile (Haldane, 1922). One of the major mechanisms of Haldane's rule is thought to be through **Bateson–Dobzhansky–Muller incompatibilities (BDMI)**; given that incompatibilities are often recessive, they will be more easily exposed in the heterogametic sex (Kulmuni & Butlin, 2021).
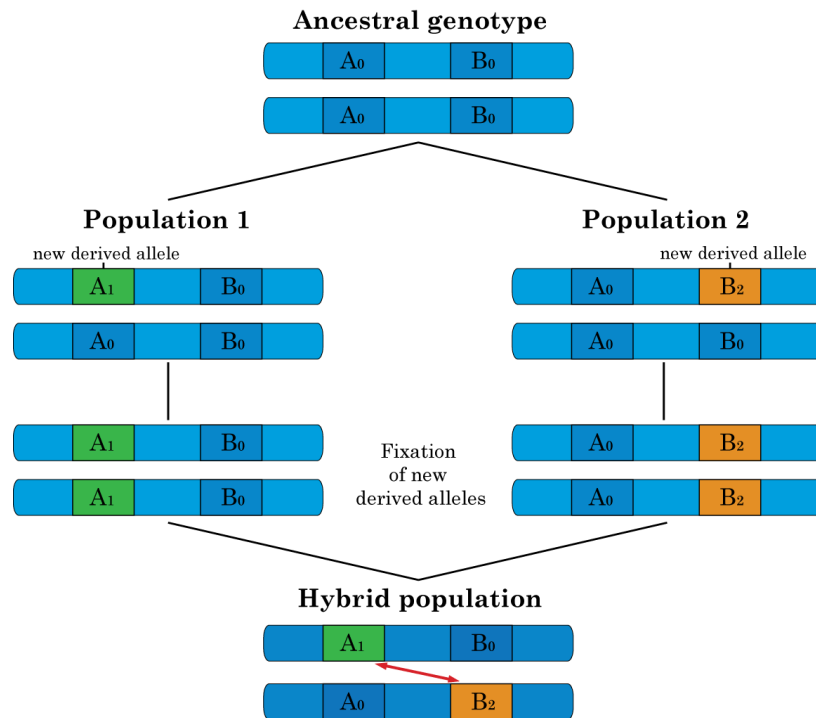
Figure 2. Bateson–Dobzhansky–Muller (BDM) model of incompatibilities. If for a pair of loci, with ancestral genotype A_0A_0–B_0B_0 divergent new mutations in either allele in two allopatric populations were developed; population 1 (A_1A_1–B_0B_0) developing a new derived allele A and population 2 (A_0A_0–B_2B_2) a new allele in the B locus; a hybrid, as a result of a cross between individuals of population 1 and population 2, would present a new combination of alleles (with genotype A_0A_1–B_0B_2) for which epistatic interactions could lead to genetic incompatibilities (among new derived alleles marked by a red arrow), with the potential of generating reproductive barriers against interbreeding. In the case of epistatic interactions between loci A_0 and B_0 having complementary effect in fitness, the newly acquired alleles A_1 and B_2, in populations 1 and 2 respectively, could however increase in frequency if they present a greater fitness than the ancestral combination (e.g. via natural selection), in their respective genomic background (A_1 in B_0 genomic background or B_2 in A_0 genomic background).

One of the most common criteria to classify the different modes of speciation is based on the degree of geographic isolation among divergent populations, which would also inform us on the level of gene flow, facilitating or hampering the accumulation of divergence that can lead to **RI**. Allopatric and sympatric speciation are in fact the geographical extremes, the former occurs between populations that are completely isolated by a geographic barrier while the latter happens in the presence of gene flow, with reproductive barriers developing in a panmictic population (Coyne & Orr, 2004; Butlin et al., 2008; Fitzpatrick et al., 2009; Futuyma, 2013). **Sympatric speciation** has been the most debated mode of speciation mainly because high levels of gene flow would hinder the accumulation of genomic differences that could contribute to the development of RI. Models supporting sympatric speciation suggest that genomic divergence should be strongly favoured to contrast the homogenising effect of gene flow (Bolnick & Fitzpatrick, 2007). For example,

assortative mating can occurs in the light of gene flow, via e.g. ecological divergent selection (Bolnick & Fitzpatrick, 2007; Papadopulos et al., 2011; Egan et al., 2015; Jiggins et al., 2006; Malinsky et al., 2015). Physical linkage, structural genomic variation (e.g. inversion polymorphisms) and linkage disequilibrium (LD), can also play a role on the development of RI.

Defining mode of speciation based on geographic patterns could be informative, as certain evolutionary mechanisms can only occur while in sympatry (e.g. introgression, reinforcement, hybridization) or in allopatry (e.g. drift, accumulation of genomic variation that could lead to reproductive barriers, without gene flow and recombination hampering genomic differentiation), but others could occur in both (e.g. natural or sexual selection). However, focusing only on geographic distance might be restrictive (Butlin et al., 2008). In the last decades there have been calls for rethinking how we study speciation, advocating for an emphasis towards the genomic mechanisms that drive speciation, and approaching geographical patterns only based on how they can either facilitate or hinder those mechanisms (Bolnick & Fitzpatrick, 2007; Butlin et al., 2008; Fitzpatrick et al., 2009; Nosil & Feder, 2012; Stankowski & Ravinet, 2021). The debate has extended onto the species concept (Mallet, 1995; Wu & Ting, 2004), on how crucial **RI** is for the definition of species, challenging the relevance of the widely accepted **biological species concept (BSC)** (Wang et al., 2020), and whether gene flow is allowed throughout the speciation process or a time of complete, or almost complete, allopatry is necessary; as it could be expected there are advocates for all alternatives (Wu, 2001; Butlin & Stankowski, 2020; Mallet, 2020; Wang et al., 2020; Wu et al., 2020).

With the overwhelming development of genome sequencing and genome-wide analysis tools, we have been able to disentangle in more detail the various and complex genomic patterns generated by the speciation process giving us a better understanding of the evolutionary mechanisms that trigger RI. We now know that the development of **RI** is a continuous process that evolves through time and varies across the genome (Kulmuni et al., 2020; Kulmuni & Butlin, 2021; Stankowski & Ravinet, 2021). While certain regions in the genome, coding for traits under divergent selection or involved in incompatibilities in hybrids, could generate some degree of assortative mating between coexisting ecotypes, other parts of the genome can still recombine freely, an idea tackled in part by the 'genic view of speciation' by Wu (2001) and the 'species as genomic clusters' of Mallet (1995, 2020). **Reproductive isolation** can be based on a small number of loci, or accumulated on genomic islands of speciation, and in both scenarios a lack of genome-wide divergence between already reproductively isolated populations can occur (Stankowski & Ravinet, 2021). Thus, as mentioned

previously, ecological divergence has the potential to generate barriers to reproduction in the presence of gene flow (Funk et al., 2006; Wang et al., 2020). There is also evidence of introgression occurring between well established species that present significant reproductive isolation, like in *Heliantus* sunflowers (Sambatti et al., 2012), and even between species that are not considered sister species (Turner et al., 2005; Edelman et al., 2019). In fact, it has been argued that complete **RI** is not needed for speciation, it can occur well after divergent lineages have been established (Mallet, 2008a; Kulmuni et al., 2020). All this leads us to think that perhaps complete RI is not the most relevant factor in speciation. Moreover, **RI** cannot be determined as a specific on/off point in speciation. Gene flow reduction and accumulation of genomic differentiation that favours reproductive barriers may be continuous and sometimes simultaneous processes. To better understand the speciation process it is crucial to study the mechanisms that play a role and their consequences along such continuum. Focusing solely on a sharp limit to when **RI** is completed or when genomic differentiation is sufficient, limits the understanding of what is a far more complex process.

A concept that lately has become more adopted in the scientific literature is that of the "**speciation continuum**", first quoted by Drès and Mallet (2002), where genomic divergence from populations to species is understood as a continuous process of reduction of gene flow (i.e. reproductive isolation). At one extreme of the continuum populations are panmictic with complete random mating, while in the other, two lineages present total, or almost total, RI. It has been revisited several times in the last ~20 years (Hendry, 2009; Nosil et al., 2009; Nosil, 2012; Seehausen et al., 2014; Shaw & Mullen, 2014; Stankowski & Ravinet, 2021), and it has been used as a framework in empirical studies to asses the accumulation of divergence along the continuum (Hendry et al., 2009a; Burri et al., 2015). The continuum is also bidirectional, divergence between lineages is as likely as the reduction of differentiation or even fusion of previously divergent species (Taylor et al., 2006; Stankowski & Ravinet, 2021). This flexibility in directionality can be in part explained by the genic view of speciation by Wu (Figure 1 in Wu, 2001) where divergence among species varies across the genome, thus species reversal can occur when differentiation is not widespread across large genomic blocks but rather present in scattered loci.

But where can one draw the line where populations become species in this continuous and flexible view of speciation? Examples of what we call well-defined species vary across a range of levels of **RI** and gene flow. In fact, Roux et al., (2016), using simulations and incorporating genomic data of species pairs into an approximate Bayesian computation framework (ABC), showed that

reproductive isolation represents a blurred, wide state. Roux et al., (2016) show that the "*grey zone of speciation —the intermediate state in which species definition is controversial— spans from 0.5% to 2%*" of genomic divergence. Where reproductive barriers have been detected as low as at 0.075% of divergence in populations presenting gene flow. Thus, the delimitation of species — identification of sufficient **RI**— is, as Roux et al., (2016) put it, *"difficult by nature"* and *"the commonness of semi-isolated entities, between which gene flow can be demonstrated but only concerns a fraction of loci, further challenge the species concept"*. Depending on which stage in the process of speciation the lineages we are studying are placed, we will have a different snapshot of the speciation process (Butlin & Stankowski, 2020; Stankowski & Ravinet, 2021). Like Darwin already hinted in the 1850's, there is a continuous variation that makes varieties and species the same thing but just in a different point on the timeline i.e. there are different levels of **RI**. Indeed, as Mallet mentioned, maybe we are simply coming back to Darwin (Mallet, 2008a, 2008b, 2010).

The continuous nature of speciation does not mean that all species follow the same path along it; there are many different paths towards **RI**. Speciation is complex and does not necessarily always represent a bifurcation of lineages, the evolution of species can be reticulated (introgression, hybridization). Species are flexible entities, where two independent species coming into secondary contact (e.g. after an expansion of their geographic distribution) could either lead to the collapse of species, reinforcement or hybridization. One of the most interesting evolutionary mechanisms, in my opinion, is hybridization, a mechanism that has even led to adaptive radiations, or further to hybrid speciation (Rieseberg, 1991; Rieseberg et al, 1996; Meier et al., 2017; Lamichhaney et al., 2018; Svardal et al., 2020). Thus, the map inside the continuum of RI and gene flow can sometimes look like a web (Fig. 1B).

Hybridization – A Source of Genomic Variation and its Role in Speciation

Hybridization was defined by Barton and Hewitt (1985) as the interbreeding between genomically differentiated lineages generating hybrid offspring of mixed ancestry. However the study of hybridization dates back to Linnaeus (1742) and his experiment on hybrids of *Tragopogon* flowers. Linnaeus even argued that "*it is impossible to doubt that there are new species produced by hybridisation generation*" (Baack & Rieseberg, 2007; Abbott & Rieseberg, 2012). Darwin also addressed hybrids widely in his work and was very aware of hybrid sterility, he even dedicated a whole chapter in *The Origin* entitled '*Hybridism*'. Around the 1950's, during the time of the Modern Synthesis, Edgar Anderson was one of the main advocates of the importance of

hybridization in evolutionary processes. He argued that varieties or subspecies could be formed by positive selection on novel traits formed in hybrid individuals (Anderson, 1949).

The potential role of hybridization in evolution was for long a controversial issue among biologists, especially among zoologists who initially considered it as an evolutionary dead end (Abbott et al. 2013). Mayr (1963) and other supporters of the BSC considered hybridization as the means from which isolation mechanisms can breakdown, which as such threatens the existence of species, as viewed by the BSC perspective (Mallet, 2005). However, in botany, hybridization has been historically recognized as a source of novelty and adaptation (Anderson, 1949; Anderson & Stebbins, 1954). The main issue for a lineage to overcome during hybridization is the potential for genomic incompatibilities generated by divergent loci involved in e.g. hybrid sterility or heterozygote disadvantage. Yet, with the very fast development of evolutionary genomics and whole-genome sequencing techniques, we have now recognized that hybridization is a widespread process and an important source of genomic variation also in animals. Results of several studies suggest that hybridization can trigger diversification by facilitating adaptation via novel combinations (reviewed in Marques et al., 2019). Across taxa, between 10% of animal species interbreed at least with one other species (Mallet, 2005; Abbott et al., 2013; Mallet et al., 2016). In fact, taking into consideration the continuous nature of the speciation process, perhaps hybridization is more common than previously imagined. The evolution of RI is a continuous process that can take several hundreds or thousands of generations and during that process, species are dynamic, their geographic distribution, adaptation, population size and gene exchange, constantly change throughout their history, thus, as Abbott *et al.*, (2013) phrased it, hybridization '*could occur at different stages of a complex and continuous interaction*' of species.

Expansion of geographic distribution has the potential to generate overlaps of species previously in isolation, transforming what initially was an allopatric distribution into species in parapatry, for instance, species expansion after the last glaciers receded have generated approximately 37% of contact zones (i.e. hybrid zone) between species that were previously isolated across the northern hemisphere (Barton & Hewitt, 1985). Depending on where in the speciation continuum newly overlapping species are located, and the strength of their RI, the outcome of such encounters could vary (e.g. species fusion, reinforcement, interbreeding, hybrid speciation). In the case of non-complete RI, lineages can potentially hybridize generating offspring with mixed ancestry as a result of interbreeding between two genetically distinct species. Contact between hybridizing lineages often produces a *hybrid zone* where hybrid individuals are usually maintained by the continuous

influx of genetic material from each parental species, when hybrids are not capable of expanding outside of the hybrid zone. For example, heterozygote disadvantage would generate lower fitness in hybrids individuals occurring in the environment occupied by either of parental lineage, or alternatively, intrinsic epistatic incompatibilities could also hamper hybrid stability confining admixed individuals to the hybrid zone. Environmental divergence between the parents' geographic distribution can generate strong differential selection on adaptive loci, especially in alleles differentially fixed between parents, affecting allele frequency along the hybrid zone. Allele frequency can vary quite drastically when transitioning from an area predominately occupied by one parental species to the other, similar patterns occur in loci involved in strong reproductive isolation. Such patterns of change in parental allele frequency along a hybrid zone are called *'clines'*. As their width and steepness are strongly determined by the strength of selection, reproductive isolation and linkage of neutral regions, among other evolutionary factors, clines have been widely used to study regions involved in e.g. reproductive isolation across the genome (Barton & Hewitt, 1985; Payseur, 2010; Gompert & Buerkle, 2011; Taylor et al., 2014; Trier et al., 2014).

In hybrid zones, initial reproductive isolation may have not been as drastic as for hampering hybridization. Yet, in time, if circumstances change, stable hybrid zones can have different fates. For instance, **reinforcement** of RI can be mediated by epistasis of environmental incompatibilities. While loci involved in incompatibilities can maintain species differences, they account for small genomic regions of differentiation that can easily be broken down by recombination in hybrid populations. Therefore, it has been proposed that if associations of RI loci with adaptive, neutral or other RI loci develop, these associations will facilitate reinforcement by increasing the divergent portion of the genome. Such coupling can be triggered in hybrid zones if clines of loci involved in RI coincide with those of adaptive loci, for instance loci involved in adaptation to environmental differentiation across the geographic range of the hybrid zone (Bierne et al., 2011). Alternatively, hybridization can also be a mechanism that favours diversity, like in the case of adaptive introgression, hybrid speciation and even species radiations. In **hybrid speciation**, hybridization results in the formation of a taxon that is reproductively isolated from its parent species. If, for instance, a novel genomic combination in the hybrid has some adaptive value, increasing hybrid fitness, such allele combination will increase in frequency favouring differentiation of the hybrid simultaneously to both parental species. Lineages resulting from hybridization may even outcompete the parental species in certain environments and colonize new niches as documented in *Helianthus* sunflowers where hybrid taxa colonize extreme environments (Schwarzbach et al., 2001; Rieseberg et al., 2003). Importantly, hybrid individuals can simultaneously develop RI

against both parental species. RI can be acquired instantly by chromosomal doubling, like in the case of allopolyploidization (much more frequent in plants) or by homoploid hybrid speciation (HHS), where rearrangements of loci involved in parental reproductive barriers can generate simultaneous isolation from both parents, becoming thus a third independent lineage (Mallet, 2007; Abbott & Rieseberg, 2012). However, it has been argued that HHS may not be very common in animals (Schumer et al., 2014).

Even when hybridization is no longer taking place, certain genomic patterns can reveal past hybridization events; one of the best-documented genomic consequence of hybridization is adaptive introgression. **Introgression** is the process in which genetic material of a divergent lineage is incorporated into the genome of another species and maintained in the population by e.g. selection (Arnold & Kunte, 2017; Taylor & Larson, 2019). This occurs when F1 or F2 hybrid individuals backcross with one of the parental species and, either through selection or drift, the foreign genomic region starts segregating among populations of the host species. Introgression can occur even if hybrid individuals suffer from low fitness, as long as recombination breaks linkage, dissociating adaptive (or neutral) loci from disadvantageous alleles, novel adaptive combinations can potentially be generated. Several cases of adaptive introgression have been reported in natural systems (Arnold & Kunte, 2017; Taylor & Larson, 2019), for instance the wing-pattern mimicry in *Heliconious* butterflies is known to cross species boundaries through introgression (The Heliconius Genome Consortium et al., 2012). Adaptive introgression has been documented in a wide range of contexts, like insecticide resistance in mosquitoes malaria vectors (Norris et al., 2015), or adaptive fur colour in hares (Jones et al., 2018). It has even been suggested that adaptive introgression from archaic humans, like the Denisovan and Neandertal into modern humans occurred, potentially increasing diversity of the MHC immune defence (Abi-Rached et al., 2011) and acquiring altitude adaptations (Huerta-Sanchez et al., 2014).

Permeability to introgression varies quite drastically across the genome and it is highly dependent on factors like recombination rate and linkage to adaptive or multilocus reproductive barriers loci. Theory predicts that selection against deleterious loci within introgressed genomic blocks is less forgiving in regions with low recombination rate, generating thus a positive correlation between introgression and recombination rate (Veller et al., 2019). Areas of high recombination can rapidly break the linkage of neutral loci from deleterious sites, thus, purging deleterious introgressed DNA without removing the entire introgressed region. Recent studies have indeed found reduced introgression in low recombination regions in swordtail fish, sticklebacks and *Heliconius* butterflies

(Ravinet et al., 2018b; Schumer et al., 2018; Edelman et al., 2019; Martin et al., 2019). Introgressed regions are also less commonly found near regions of high gene density like in the case of Nearthendal introgression in modern humans (Sankararaman et al., 2014; Juric et al., 2016). Sex chromosomes are also typically less prone to introgression than autosomes (partially due to the lack of recombination), in species where one sex is heterogametic, consistent with selection against infertility (Sætre et al., 2003; Storchová et al., 2010; Patrick Gainey et al., 2018).

Several methods have been developed to identify traces of hybridization in the genome, for instance discordance between phylogenetic trees vs. gene trees are often interpreted as an indication of foreign genomic regions introgressed into the host genetic background. However other processes, like incomplete lineage sorting, can also produce conflicting patterns during phylogenetic inference (Baack & Rieseberg, 2007). Multilocus comparison among species focusing on a decrease of interspecific genomic differences if introgression has occurred, as well as coalescent-based Bayesian modelling, have been used in *Heliconious* butterflies to identify regions of admixed ancestry (Bull et al., 2006; Kronforst et al., 2006). Among recent approaches for detecting introgression and putative hybrid taxa, genomic clustering methods like STRUCTURE or ADMIXTURE (Pritchard et al., 2000; Alexander & Novembre, 2009) where each individual is assigned to a particular genomic cluster based on similarity, are widely used. In this case, this framework assumes that a hybrid population would present a mixed contribution from other defined cluster. However, variation in drift among the groups and uneven sampling can cause erroneous results. Test of imbalance in allele sharing, like ABBA–BABA statistics (Patterson's D and the f4-admixture ratio) (Patterson et al., 2012; Martin et al., 2015; Peter, 2016; Zheng & Janke, 2018; Malinsky et al., 2021) have recently been developed and widely used to detect genetic exchange after lineage divergence, however false positives may occur when ancestral populations presented subdivisions or when there is variation in substitution rates (Svardal et al., 2020). Demographic modelling has also been used to fit the best demographic scenario on the evolutionary history of the lineages involved; timing, amounts and direction of gene flow as well as population size can also be inferred from these models (Pinho & Hey, 2010; Excoffier et al., 2013, 2021). Finally, reconstruction of complex phylogenies –fitting best model of reticulated speciation– using graph construction methods, has been used to estimate hybridization in phylogenies with large number of lineages (Lipson et al., 2013; Mathur & Adlakha, 2016).

An extreme example of how hybridization and introgression of adaptive traits can generate novel combinations, facilitating genetic variation and even triggering diversification, is the case of

adaptive radiations. Hybridization between divergent lineages can considerably increase genomic variation with the potential of generating novel phenotypes, allowing hybrids to drastically diverge if the ecological opportunity exists (Seehausen, 2004; Marques et al., 2019). This is the case of the well–known Cichlids radiation where these fish have colonized a wide variety of environments. Cichlids harbour an important variation in their phenotypic traits, which has certainty allowed them to occupy a wide range of habitats across the African and American continents. Roughly half of this diversity is thought to have originated from adaptive radiations in the Great Lakes of Africa (Victoria, Tanganyika and Malawi) (Svardal et al., 2021), with examples of diversifications occurring in sympatry, without clear geographic barriers in some cases (Malinsky et al., 2015). Studies on the radiations in Lake Malawi (Svardal et al., 2020) and Lake Victoria (Meier et al., 2017) have shown that individuals of hybrid origin were the early founders of these populations, later radiating into communities of ~850 and ~700 species, respectively. Interestingly, despite the likely high genomic variation of the founder populations, current lineages present relatively low within-species genetic diversity, low mutation rate and a widespread gene flow among species, generating quite similar genomes (Svardal et al., 2021). Such an extraordinary explosion of species has thus been explained as a consequence of novel genomic combinations of an already admixed genome, occurring in conjunction with a large supply of environmental diversity. The initial hybrid founder population was the result of an ancient hybridization event and subsequent rearrangements into several different new admixed genomes, providing an overwhelming amount of genetic variation and thus facilitating species radiations (Meier et al., 2017, 2019).

Ancient hybridization events can trigger subsequent species radiation, not only due to the access to ecological opportunity, but the admixed nature of a hybrid genome itself can predispose the genome for the development of multiple species reproductively isolated. Two interesting mechanisms proposed by Seehausen (2013) explain how a hybrid genome can generate hybrid swarms that later could evolve into multiple independent species. Firstly, novel rearrangements of parental genomic blocks will increase the standing genetic variation in the hybrid population, freeing regions to form novel combinations underlying adaptation. For instance, large fixed genomic islands of adaptation, present in a parental genome, can constrain flanking regions to freely diverge, while in the hybrid genome, via recombination, novel rearrangements of genes belonging to such selective sweeps can be formed, boosting the possibility for novel adaptations. Second, coupling of loci involved in reproductive isolation with other loci responding to ecological divergent selection, becomes more likely in a hybrid population than in, for example, parapatric populations that are under ecological divergent selection alone. The genomic pool in a hybrid

population will present a large variability of adaptive loci as well as of regions presenting BDM incompatibilities, independently developed in the parental lineages (through recombination), and if adaptive loci are undergoing strong disruptive ecological selection, the opportunity for coupling reproductive isolation loci and ecological loci increase drastically, facilitating thus the evolution of several different isolated species. Interestingly, these mechanisms can, to a certain extent, be extrapolated to the origin of **homoploid hybrid species**.

Homoploid hybrid speciation – Hybridization as a creative process

The famous drawing on Darwin's notebook (Fig. 1A), and later modified in *The Origin*, was a phylogenetic tree reflecting how he thought the evolutionary process of speciation occurs in nature. Darwin proposed bifurcation from a common ancestor as an explanation for the large biodiversity he encountered during his journey on *The Beagle*, however, he was well aware of the occurrence of hybrids in nature. More than 150 years later, we now know that speciation is a continuous process where species delimitation has become blurred. Hence, bifurcation from a single common ancestor is an oversimplification of the speciation process. Hybridization among well–defined species also occurs in nature, generating novelty within the species involved or sometimes even resulting in the formation of a new independent hybrid taxa (Fig. 1B). **Homoploid hybrid speciation (HHS)**, a hypothesis initially proposed by G. Ledyard Stebbins in 1950's and later named **recombinational speciation** by Verne Grant (1971), is the process from which hybridization directly lead to the development of a new, genetically stable and reproductive isolated species. Some hybrid genotypes could be fertile and simultaneously present reproductive barriers to both parental species. If these hybrid individuals increase in frequency they eventually can generate a distinct, independent lineage. In contrast to allopolyploidization, HHS does not involve any change in the number of chromosome sets that confer instant RI against the parents; sorting of parental incompatibilities in the hybrid genome can result in the acquisition of reproductive barriers against both parental species simultaneously. In the hybrid genome, parental incompatibilities can be fixed independently, where one set of loci might generate reproductive barriers against one parent while a second set, in another region of the genome, will isolate the hybrid against the second parent.

By definition, for hybridization to take place, sympatry and incomplete RI between species are crucial factors, since gene flow is needed. Time of divergence between the parental species could roughly determine the level of RI; closely related species most likely present weaker RI increasing the likelihood of hybridization (Mallet, 2007; Abbott & Rieseberg, 2012). Secondary contact of

previously allopatric species can occur by a range expansion after certain ecological changes (e.g. retreat of ice after a glaciation period). However, as described previously, hybridization events do not guarantee the development of a hybrid species. The success of the hybrid lineage is dependant on the rapid development of RI against both parental species, genome stabilization by sorting of incompatibilities and maintaining compatible combinations, and eventually the adaptation to novel divergent habitats (if the ecological opportunity exists), to avoid competition with the parents (Abbott & Rieseberg, 2012). Emerging hybrid populations need to be either reproductively or ecological isolated from their parental species to avoid the loss of incipient differentiation. Before the hybrid genome gains a level of stability that ensures the maintenance of the genomic differences and isolation from the parental lineages, it is under threat of homogenisation by gene flow.

Unlike in allopolyploidy, HHS F1 hybrids do not instantly acquire RI from the parent lineages, thus such reproductive barriers usually evolve through the sorting of parental incompatibilities during several generations of viable hybrid genotypes. The genetic mechanisms of simultaneous isolation to both parents have been proposed via reciprocal genomic translocations (Grant, 1971; Rieseberg, 1997; Gross & Rieseberg, 2005). Parental species can present complementary translocations such that combinations in F1 hybrids would generate mostly unbalance gametes (75%) and only a small percentage of balanced novel gametes (12,5%), thus lowering the hybrid fertility (Fig. 3). F2 homozygotes from the combination of balanced gametes will present complementary genetic factors generating fertile individuals although isolated from both parental species as interbreeding will generate unbalanced chromosomes leading to sterility (Futuyma, 2013).

This mechanism could ensure some level of reproductive isolation between the hybrid and the parental species, however further chromosomal evolution and ecological isolation across several generations (of individuals with the suitable novel genotypes) would ensure the **genome stabilization** needed for the hybrid taxa to establish. In general, genomic composition is affected by factors like mutation, drift and recombination within a framework of selection and demographic changes shaping the genome over time. However, unique to admixed genomes, mechanisms like purging of incompatibilities and fixation of compatible genomic parental combinations are needed for the formation of a stable and functional genome (Buerkle & Rieseberg, 2008). **Stabilization of admixed genomes** can take several generations; mathematical modelling suggests that the number of generations required for the stabilization of an admixed genome varies across species
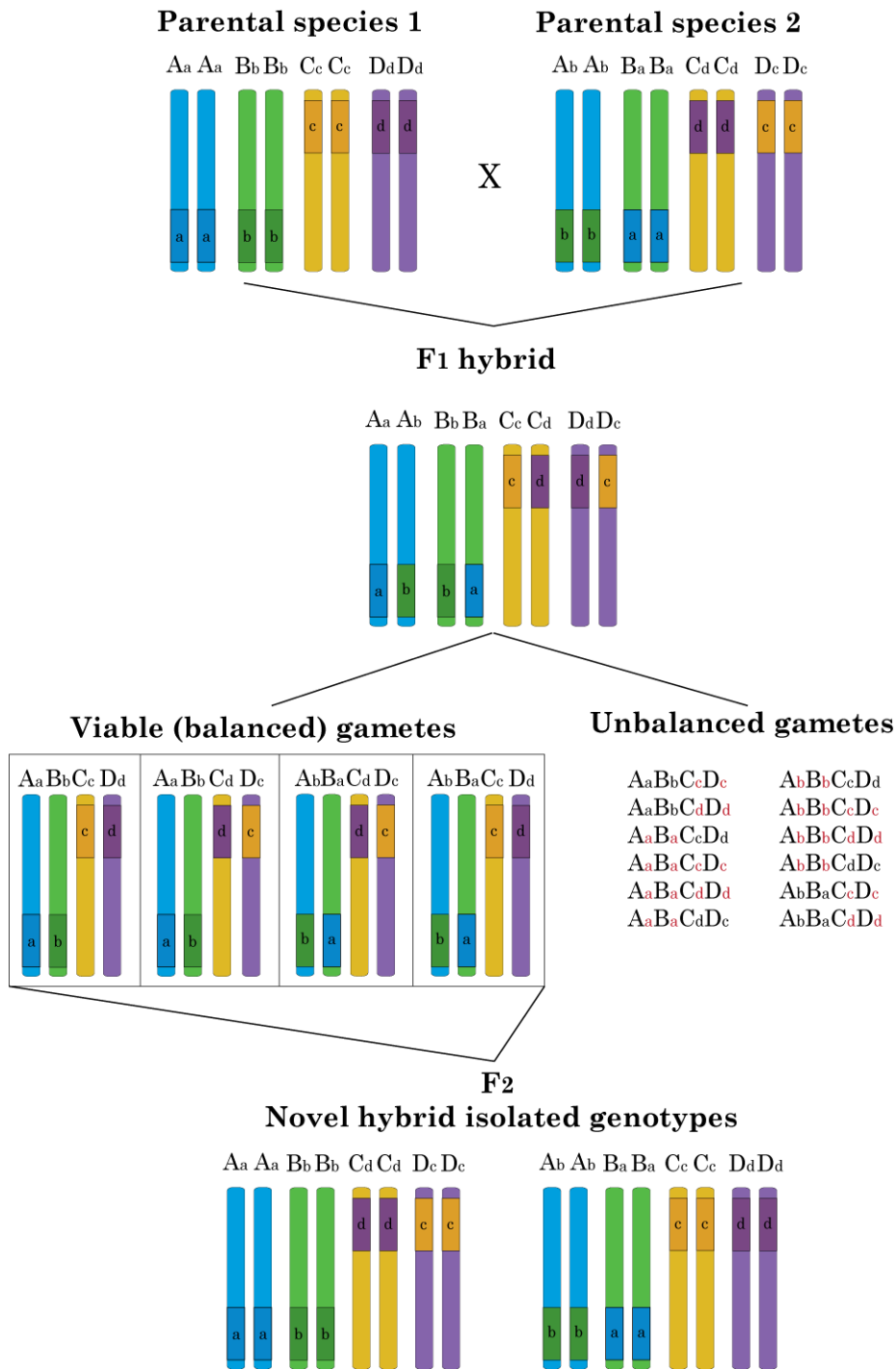
Figure 3. Genetic mechanism for the development of simultaneous reproductive isolation to parental species (adapted from Rieseberg, 1997 and Futuyma 2013). Two independent translocations in the parental species 2, where reciprocal regions of chromosomes A and B are translocated. Similarly, these occurs for a second locus on chromosomes C and D. Heterozygote F1 hybrids thus present 75% of unbalanced gametes and 25% of viable–balanced gametes (without deficiencies or duplications). Half of those balanced gametes would recover the parental genotypes while the other half present novel, recombinant genotypes. If such gametes are combined, a F2 generation will produce some viable, novel, balanced individuals that are isolated to both parental species, as their offspring (largely unfertile) will only present unbalanced combinations.

(Buerkle & Rieseberg, 2008). Some studies have shown that one determining factor is the number of available hybrid individuals with the ideal hybrid genotype (see Fig. 3) to reproduce and increase the hybrid population size (McCarthy et al., 1995; Abbott & Rieseberg, 2012). **Genomic stabilization** will also occur at different rates in different parts of the genome, varying drastically between neutral loci and functionally important regions (Buerkle & Rieseberg, 2008; Schumer et al., 2016; Moran et al., 2021). Thus, genome stabilization has a determining effect on the composition of admixed genomes, which in turn biases the genomic diversity of hybrid lineages.

Not surprisingly, among the first cases of HHS described are plant species. The *Heliantus* sunflowers have been the classic example of HHS and the acquisition of transgressive traits via hybridization. Three wild species, the desert sunflowers *Helianthus anomalus*, *Helianthus deserticola* and *Helianthus paradoxus*, have been identified as homoploid hybrids independent lineages, as a result of hybridization events between *Helianthus annuus* and *Helianthus petiolaris* (Rieseberg, 1991; Rieseberg et al, 1996). These hybrids are adapted to extreme environments that neither of the parental species are capable of occupying (Rieseberg et al., 2003). Cases of HHS with transgressive segregation that facilitates isolation from parental species, are also found in animals, this is the case of alpine hybrid populations of *Lycaeides* butterflies originated by hybridization events between *L. melissa* and *L. idas* (Gompert et al., 2006, 2012; Chaturvedi et al., 2020). Other documented HHS cases in insects include the Terphritid fruit flies (Schwarz et al., 2005), where host selection determines mate choice. One of the most studied HHS cases in animals is that of the *Heliconious* butterflies in Latin America, where *Heliconius heurippa* is identified as a hybrid species result of the interbreeding between *Heliconius melpomene* and *Heliconius cydno* (Salazar et al., 2005; Mavárez et al., 2006). Interestingly, introgression of a relatively small genomic region that control wing colour patterns generates the intermediate coloration phenotype of the hybrid that may have an impact on mating behaviour (The Heliconius Genome Consortium et al., 2012), however, whether this causes reproductive isolation via assortative matting remains unresolved (Mavárez et al., 2021). Another case where sexual selection seems to play an important role in hybrid speciation is the case of Swordtail fishes (*Xiphophorus*), presenting at least two species of hybrid origin (Meyer et al., 2006; Jones et al., 2013; Schumer et al., 2016). Sexual selection on a trait derived by introgression appears to have favoured the speciation of the hybrid *Xiphophorus clemenciae* (Meyer et al., 2006). In birds, Darwin's finches in the Galapagos islands have shown to be classic examples of speciation where drastic morphological variation in beak shape have generated adaptive variation leading to species radiation (Grant & Grant, 1999, 2009; Petren et al., 2005; Huber et al., 2007; Hendry et al., 2009b; Lamichhaney et al., 2015) and a few years ago Lamichhaney et al., (2018) also

reported a very recent case of HHS in this clade. At the Galapagos island of Daphne Major a hybrid population has been developing since 1981 resulting from a single hybridization event between an immigrant male of *Geospiza conirostris* and a female resident of *Geospiza fortis*. Hybrid individuals mate specifically with other hybrids potentially as the result of selection on the transgressive traits in bill morphology. Interestingly, RI from the sympatric *Geospiza fortis* parent seems to have originated specifically from hybridization and completed after only a couple of generations. Reproductive barriers to the other parent species (*G. conirostris*) remains unknown, as they do not occur in sympatry. The Italian sparrow and the golden-crowded manikin are also bird species identified as homoploid hybrids (Hermansen et al., 2011; Sætre et al., 2017; Barrera-Guzmán et al., 2018), where reproductive isolation against the parents have been documented at least against one of the parents, among other putative young hybrid species (reviewed in Ottenburghs, 2018).

Despite the latest efforts to identify and study HHS, one question in evolutionary biology that remains unresolved is how easily hybrid speciation occurs in nature (Buerkle et al., 2000), how fast can hybrid populations establish and whether the stability of an admixed genome occurs with certain frequency. Mallet (2007) suggested that perhaps homoploid hybrid species represent a larger fraction of species in nature than initially thought. He argued that since genetic variation is expected to be triggered by hybridization and hybridization occurs frequently between species, one could expect large amount of genetic variation generated by hybridization events. Besides, heritable genetic variation being the substrate on which natural selection acts and knowing that natural selection is a common force in speciation, it would not be surprising if homoploid hybrid speciation occurs relatively frequently. However, Schumer et al., (2014) argued that even if hybridization is common in nature, criteria to define whether HHS occurs should be more stringent. More specifically they suggest that HHS is only applicable where it is possible to demonstrate that RI to both parent species is directly generated by hybridization. As in species of non-hybrid origin, RI against another species (a parental species in the case of HHS) is not possible to demonstrate if the species do not overlap. Moreover, hybridization can enhance mechanisms that later, after some geographic or ecological barriers, lead to RI, without being easily traceable to the hybridization event itself, like in the case of *Senecio squalidus* and *Pinus densata* (reviewed in Feliner et al., 2017). Nevertheless, it is important to consider that other processes, like repeated hybrid introgression, can also generate similar patterns of admixed ancestry (Ottenburghs, 2018), meaning that basing the HHS concept on the causative role of hybridization on RI is rather limiting. Perhaps rather than focusing in the strict criteria to determine which cases deserve to be

called HHS, a more relevant question is how hybridization generates novel variation that could aid the establishment of a hybrid lineage, or what is the fate of new hybrid lineages and how easily can an admixed genome stabilize. Furthermore, much remains to be learned on whether novel combinations ensure the success of a hybrid by facilitating adaptive divergence, or whether incompatibilities (e.g. BDMI) present stronger constraints to the establishment and evolution of a hybrid genome.

The evolutionary fate of admixed genomes – The potential and limitations of a hybrid species

Admixed genomes being mosaics of parental genomic blocks, harbour variation in the form of novel genetic combinations, some with the potential to facilitate adaptation. Novel adaptive variation generated by hybridization can trigger large bursts of diversity like in the case of adaptive radiations (Meier et al., 2017; Svardal et al., 2020), thus it is possible to imagine that hybrid species could already have a 'promised' evolutionary path. However, genetic divergence between parents can generate incompatibilities limiting possible combinations and constraining the evolution of a hybrid genome. Thus, genome stabilization balancing the limiting factors (e.g. sorting of incompatibilities) and the potential advantageous genomic variation (e.g. fixation of compatible combinations) is needed for the hybrid to have a chance of succeeding.

Besides the ability to purge incompatibilities and retain advantageous combinations to adapt to its environment, the fate of a hybrid species also lies in the evolution of isolation mechanisms (i.e. intrinsic reproductive barriers) against the parent species to escape species fusion if the hybrid and the parents come into secondary contact (Abbott & Rieseberg, 2012). As in species of non-hybrid origin, secondary contact between a hybrid and the parents, can also have phenotypic and genomic effects on the encountering species, generating character convergence or displacement. Even after the development of reproductive barriers of a newly formed hybrid species, if novel variation confers habitat adaptations and the ecological opportunity exist, admixed lineages could capitalize from exploiting new ecosystems, ecologically isolating to the parents to avoid competition.

Positive consequences of admixture – Novelty and Adaptive Potential

In principle, hybridization has the potential to considerably increase genomic diversity in the admixed lineage. The possibility of combining two divergent genomes, both with their own standing genetic variation, into novel rearrangements could enhance the genomic variation of a

hybrid lineage (Stelkens & Seehausen, 2009), beyond that of non-admixed individuals. Since heritable genomic variation is the substrate which natural selection can act upon, hybridization has been increasingly recognized as a process that can facilitate diversification by enriching standing genetic variation and providing novel adaptive genomic diversity (Marques et al., 2019).

Hybrid speciation could lead to elevated variability at ecological relevant loci increasing the adaptive potential of the hybrid taxon and in some cases generating transgressive segregation. Hybrid species sometimes develop transgressive traits that allow them to adapt to new environments, occupying niches that neither of the parents are capable of. Such traits are thought to be the result of novel combinations of parental genomic blocks conferring new adaptations to the hybrid population (Schwarzbach et al., 2001; Rieseberg et al., 2003). Habitat disturbance can also generate ecological opportunity for novel hybrid phenotypes (Abbott & Rieseberg, 2012). Finally, ecological divergence of the hybrid, in comparison to that of the parents, may enhance the development of RI via micro-allopatry. Such ecological divergence may reduce gene flow and competition between the hybrid lineage and its parents, allowing the maintenance of the hybrid through isolation by adaptation (Mallet, 2007; Abbott & Rieseberg, 2012).

Another mechanism that has a positive effect on the adaptive potential of hybrid species is the release of evolutionary constraints to adaptation that may be present in the parents, due to the sorting of genomic blocks in admixed genomes. During hybridization, recombination of parental genomic blocks in the hybrid can break apart linked loci, thereby releasing genetic correlations that otherwise may constrain genome divergence and phenotypic evolution (Seehausen, 2013).

However the stabilization of an admixed genome not only involves sorting genomic blocks and generating novel adaptive combinations, purging of incompatibilities and overcoming constraints are major processes in the evolution of hybrid genomes.

Negative consequences of admixture – Genomic Incompatibilities and Constraints

Although hybridization may have positive effects on the adaptive potential of admixed species, it can also constrain the adaptability of the hybrid taxon due to, e.g. genetic incompatibilities, physical linkage or epistatic and pleotropic effects originally occurring in the parental genomes.

The very differences acquired between the parental species while in allopatry, including those that confer a certain level of RI, will likely generate genetic incompatibilities (BDMIs) affecting the

hybrid's evolvability. For example, in the case of *Xiphophorus* fishes (Dagilis & Matute, 2020; Powell et al., 2020), genes involved in cancer show to recurrently present incompatibilities in hybrids of several species pairs. Purging of incompatibilities by sorting parental genomic blocks will resolve some of the potential constraints found in hybrids. For instance, because incompatibilities are likely to include novel, derived alleles developed in either of the parental lineages (i.e. alleles not present in the common ancestor of the parent species – Fig. 2), evolution towards ancestral allelic combinations in the admixed genome may remove incompatibilities in the hybrid (Moran et al., 2021).

In most cases, hybrid populations present uneven admixed genomes with an overrepresentation of genomic blocks from one of the parental species. During the early evolution of a hybrid lineage, continuous backcrossing with at least one of the parental species generates this kind of uneven patterns. Loci inherited from the minor contributor are likely to harbour incompatibilities against a major-parent genomic background, making them candidate regions under negative selection. However it has been shown that minor-parental blocks are more likely to be retained in regions presenting high recombination rate due to their increased likelihood of breaking apart from incompatible loci (Ravinet et al., 2018b; Schumer et al., 2018; Edelman et al., 2019; Martin et al., 2019).

Purging of incompatibilities is sometimes restricted by pre-existing complex epistatic interactions and pleiotropic effects originally present in the parental species (Papakostas et al., 2014) even if the persistence of incompatibilities reduces the fitness of hybrid populations.

All these mechanisms point towards a constraining effect over specific genomic regions, limiting how freely they can vary and biasing the overall composition of hybrid genomes. Rieseberg et al., (1996, 2003) found that experimental hybrids (F1) between *Helianthus annuus* and *Helianthus petiolaris* and later backcrosses, replicate chromosomal segments present in the wild hybrid species, and how their phenotypic transgressive traits are also reconstructed. Recent studies of *Lycaeides* butterflies also show a similar pattern, where same genetic combinations found in natural hybrid lineages, re-emerge in younger lineages (Chaturvedi et al., 2020). These might suggest that certain constraints are determining the outcome of hybridization leading to repeated genomic patterns.

As Mallet (2007) pointed, interspecies "*hybrid lineages can be considered hopeful monsters*" (Goldschmidt, 1940). While having to overcome large genetic differences from each parental species into a single genome, purging incompatibilities and usually facing low fertility, admixed

genomes also confer the potential of adaptation to divergent ecological niches due to the large genomic diversity generated by novel genomic combinations.

The Italian sparrow, a hybrid lineage

The Italian sparrow (*Passer italiae*) is a seed-eating passerine bird distributed along the Italian peninsula and the Mediterranean islands of Sicily, Malta, Corsica and Crete (Fig. 4). It is a well-known homoploid hybrid species that originated through past hybridization events between the house sparrow (*P. domesticus*) and the Spanish sparrow (*P. hispaniolensis*) (Hermansen et al., 2011; Elgvin et al., 2017). This hybrid lineage exhibits a mosaic phenotype in males with intermediate traits of those shown by the parental species (Fig. 4); ecologically it is similar to the commensal house sparrow, commonly found living close to human settlements and feeding on crops and human food waste (Summers–Smith, 1988). The intermediate male coloration of the Italian sparrow generated the first hypotheses of hybrid origin for this lineage (Summers–Smith, 1988; Töpfer, 2006). Later, using microsatellites (Hermansen et al., 2011), nuclear genetic marker sequences (Elgvin et al., 2011) and whole-genome sequencing (Elgvin et al., 2017) the admixed, mosaic nature of its genome was confirmed.

It is hypothesised that the parental species came into secondary contact when earlier populations of the house sparrow spread into Europe alongside agriculture after the Neolithic revolution, approximately 6 kyr BP, overlapping with the already resident Spanish sparrow (Elgvin et al., 2017; Hermansen et al., 2011; Ravinet et al., 2018). Although currently the house and Spanish sparrow show signs of nearly complete reproductive isolation in the Eurasian continent, as interbreeding is more or less anecdotal in regions where the two species coexist. The scenario in North Africa is contrasting, where a range of populations presenting genetic admixture between these two lineages have been found (Päckert et al., 2019) (Fig. 4). Thus, it is likely that earlier dispersing populations of the house sparrow in Eurasia did not present such strong reproductive barriers against the Spanish sparrow, allowing them to hybridize, resulting in the hybrid Italian sparrow. Potentially, the hybrid Italian sparrow originated from multiple hybridization events between the parental species (Elgvin et al., 2017; Runemark et al., 2018) in mainland Italy. Evidence suggests that this hybrid has spread across the Italian peninsula and has subsequently undergone local adaptation with somewhat reduced gene flow among divergent populations (Eroukhmanoff et al., 2013). Populations currently occupying the Mediterranean islands are thought to be the result of independent hybridization events or at least long periods of independent evolution, supported by

the low pair-wise correlations of ancestry tracts among islands and significant albeit small differences in ancestry tract sizes (Runemark et al., 2018).
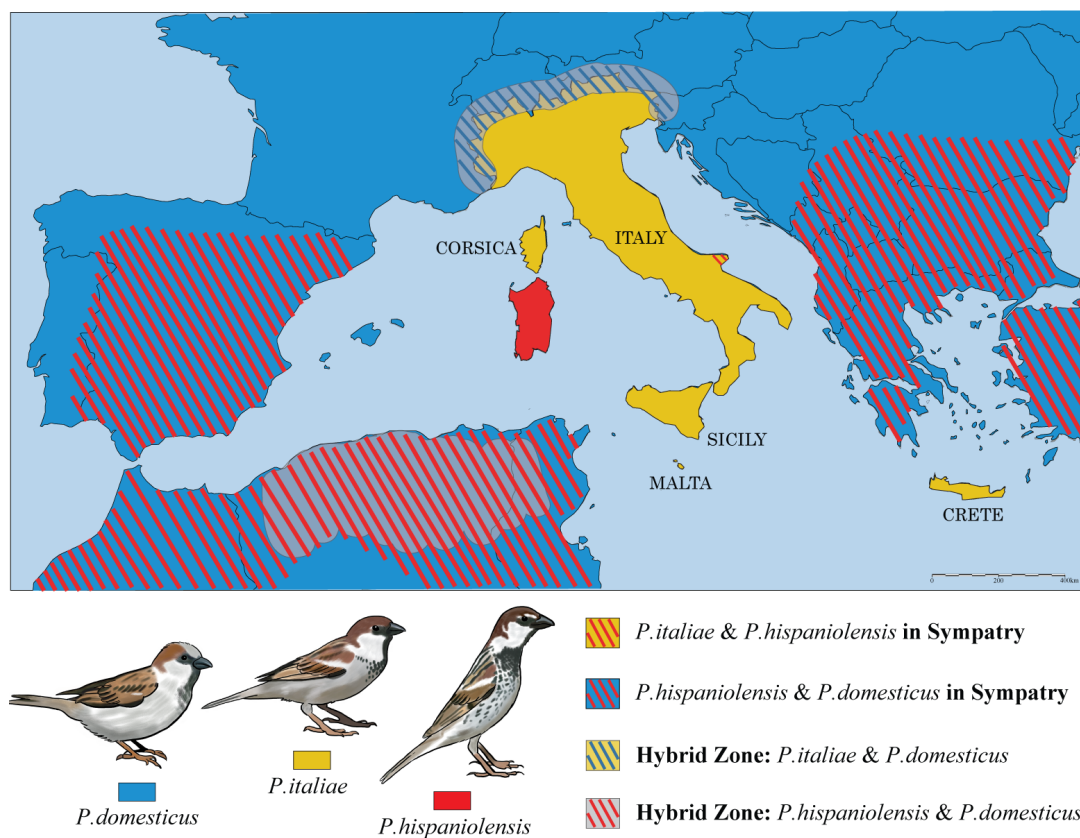
Figure 4. Male plumage colouration patterns and geographic distribution of the Italian Sparrow and its parental species. The distribution map depicts areas of sympatry and hybrid zones. Hybrid zones found in Italy and northern Africa according to Hermansen et al., (2011); Trier et al., (2014); Päckert et al., (2019) and others. Differences in plumage colouration patterns of the three focal species are shown. Sparrows' illustrations by Kaj K. Standal Clausen.

Across its geographic distribution the genomic contribution from each paternal species to the Italian genome varies considerably. In mainland Italy the admixed genome shows a slightly higher contribution (61%) from the house sparrow (Elgvin et al., 2017). While, as suggested by the inference of independent hybridization events, islands populations of the Italian sparrows diverge in the identity of the major- and minor-parental contributor. The share of house sparrow ancestry, estimated as admixture proportion, varies from 35% in the lineage on Malta, to 37% in Sicily, 62% in Corsica and 76% in the Cretan lineage (Runemark et al., 2018). Despite the large variation in parental ancestry in the nuclear genome of the hybrid, the Italian sparrow is nearly fixed for the house sparrow mitochondrial genome (Elgvin et al., 2011, 2017; Hermansen et al., 2011; Trier et al., 2014).

Reproductive barriers against both parental species are thought to have developed through the sorting of pre-existing parental incompatibilities, thus RI of the Italian sparrow consists of a subset of those isolating the parent species (Hermansen et al., 2014; Trier et al., 2014). Despite sympatry with the Spanish sparrow, a recent case of secondary contact in the Gargano peninsula (Italy), the lack of intermediate (backcrossed) individuals shows strong reproductive isolation against this parental species. In the Alps, the Italian sparrow is in parapatry with the house sparrow, presenting a narrow hybrid zone, with sharp clines of plumage coloration and nuclear and mitochondrial species-specific markers (Hermansen et al., 2011; Trier et al., 2014; Bailey et al., 2015). Overall, evidence on genomic clines at the hybrid-parent boundaries suggests that post-zygotic barriers are associated with mito-nuclear and sex-linked incompatibilities (Trier et al., 2014).

Despite our recently improved understanding on this hybrid lineage and the nature of its RI against parental species, little is known on the genomic variation of the Italian sparrow along its geographical distribution and the evolutionary processes that have taken and continue to take place in this lineage. We are yet scratching the surface as to how a hybrid lineage establishes and evolves at the genomic level after its formation; whether constraints linked to admixture have negatively impacted its evolvability, limiting the adaptive potential or alternatively, new genetic variation, via novel combinations of parental genomic blocks may have favoured adaptation and ultimately increased population divergence. This is what my thesis focuses on, using the Italian sparrow as a model organism for the study of an admixed lineage.

AIMS

AIMS

My goal was to shed light on how admixed genomes evolve after the formation of a hybrid species, focusing on the potential and challenges of a hybrid lineage. Within this scope, I am interested in studying the evolutionary path –and the factors that affect it– of the hybrid Italian sparrow after its formation; how genetic admixture may have increased genomic variability and thus its adaptive potential and to what extent sorting of incompatibilities has limited its evolvability and genomic differentiation.

In this thesis I asked the following questions:

a. What is the extent of genomic variation in the hybrid across its geographic distribution?

b. Which are the determining factors of the evolution, maintenance and genomic differentiation patterns in the hybrid Italian sparrow's genome?

c. What is the current impact of secondary contact between the hybrid species and one of its parent species, especially at the genome level?

Our knowledge on the intraspecific genomic variation of this hybrid species across its geographic distribution is scarce. Exploring patterns of population differentiation within hybrid species can inform us on the forces that shape the evolution of hybrid genomes, thus in **paper I** we aimed to study the population structure of the Italian sparrow in mainland Italy and evaluate potential factors that can explain population divergence at the genomic level. Previous studies have identified genomic regions under balancing selection (Elgvin et al., 2017; Runemark et al., 2018). Additionally, precipitation regimes have shown to be correlated to morphological traits, like beak shape variation (Eroukhmanoff et al., 2013). However our understanding of the genomic basis of adaptive variation in this hybrid is still limited. Thus, to shed light on the potential role of adaptation in the genomic evolution of this hybrid species, we evaluated the relationship between environmental variation and genomic differentiation; we also explored candidate regions, and adjacent genes, under selection. Considering that sorting of incompatibilities can have a limiting effect on the evolution and variation of a hybrid genome, we also analysed whether regions of divergence in the Italian sparrow were located on regions of high divergence (e.g. differentially fixed) between parental species.

In addition to drift and selection, recombination and rearrangements of parental genomic blocks and sorting of incompatibilities also have a determining effect on the composition of admixed

genomes, which in turn affects patterns of genomic differentiation among hybrid populations. Thus, studying if hybrid genomes can freely generate abundant novel combinations fuelling diversity, or whether local patterns of differentiation mirror those of more distant and divergent populations, can give us insights on the role of constraints in the evolution of this hybrid genome. To answer this, and to evaluate which other factors determine genomic differentiation, in **paper II,** we took advantage of the Italian sparrow populations in the Mediterranean islands. As they constitute independent, relatively isolated populations, likely to have originated from different hybridizations events, the islands system allowed us to compare hybrid lineages with varying ancestry proportions (between-islands populations) as well as populations with similar evolutionary history (similar ancestry contribution in within-islands populations). Apart from genomic patters consistent with incompatibilities we also examined the role of selection, recombination, geographic isolation and differential genomic parental contribution (purging of minor-parental ancestry in a major-parent genomic background), in shaping patterns of differentiation within and among islands. **Papers I and II** give us an overall understanding of the factors that determine the genome composition and the evolutionary path of this hybrid lineage since its formation, revealing the advantages and limitations that this hybrid genome has faced throughout its evolution.

Finally, in **paper III** we tackled genomic consequences of secondary contact. Even if a hybrid species is thriving ecologically, species interactions with the parental species could lead to challenging stages that test the integrity of the hybrid lineage. Contact zones, apart from testing the level of RI against the parent species, are an opportunity to evaluate the dynamics that secondary contact has on the hybrid genome. Even if, as in the case of the Italian sparrow, sufficiently strong RI mechanisms ensure the persistence of this hybrid species, the survival of the hybrid lineage is also determined by their fitness when facing competition from an evidently closely related species. Thus, a last objective of my research involves the effects that species interactions with the parental species may have on the phenotypic and genetic variation of the hybrid lineage. Previous studies have shown that the Italian sparrow is strongly isolated from the Spanish sparrow (Hermansen et al., 2011; Trier et al., 2014), however it is still unclear what are the genetic and phenotypic consequences of secondary contact on the hybrid when encountering one of its parental species. When closely related species co-occur in the same geographical space, ecological character displacement, resulting from competition, can cause evolutionary divergence in resource exploiting traits (Grant & Grant. 2006). This could further prevent interbreeding between the Italian sparrow and the Spanish sparrow in the Gargano peninsula. In this study we evaluated the genetic and

demographic changes occurring after an invasion of the Spanish sparrow in previously allopatric populations of the Italian sparrow.

PAPER SUMMARIES

PAPER SUMMARIES

Paper I

Intraspecific genomic variation and local adaptation in a young hybrid species

The establishment and further success of a hybrid lineage not only depends on the development of reproductive barriers against its parental species but the stabilization of their admixed genome, purging incompatibilities and maintaining arrangements that later might present an adaptive potential. Mathematical models and a few case studies have been use to evaluate the fate of a hybrid lineage after a handful of generations (Schumer et al., 2015; Comeault & Matute, 2018) looking to shed light on how an admixed genome evolves and stabilizes. However, studies addressing genomic differentiation and adaptive potential of hybrid lineages that have developed reproductive isolation towards both parent species are scarce.

In this study we aimed to assess the general population structure of the hybrid in mainland Italy, and whether differentiation patterns can be explained by environmental variation suggesting potential for adaptation, and to what extent the role of constraints in the evolution of this hybrid genome limits its variation. Variation in climatic and geographic factors can determine patterns of population genomic divergence, either through geographic isolation (isolation by distance, IBD), or through ecological isolation (isolation by environment IBE). In the latter adaptation to divergent habitats maintains populations ecologically separated, facilitating genomic differentiation. Evaluating IBD, IBE and Isolation by adaptation (IBA) we sampled a total of eight populations (N= 131 individuals) of the Italian sparrow across Italy, representing a large range of this hybrid geographic distribution.

Our results show that even though there is a weak population structure across mainland Italy and some evidence for genomic variation and adaptation in the Italian sparrow, constraints, possibly due to incompatibilities, still have an effect on the genomic variation in this hybrid lineage.

Variation in climatic factors partially explains the genomic differentiation found in the hybrid. We found a significant correlation between seasonality temperature and genomic variation. Moreover, genome scans and post-hoc analysis highlighted candidate loci under selection in the vicinity of some interesting genes, like the GD5F gene. This gene, part of the BMP gene family, is under selection in Darwin finches, showing a crucial role in craniofacial development and beak shape and size variation (Lamichhaney et al., 2015, 2016; Chaves et al., 2016). We also find evidence showing

that the majority of genomic variation in the hybrid is located in regions of low parental genomic divergence suggesting the effect of constraints in regions of high parental differentiation as those are regions that could harbour genomic incompatibilities. This suggests that either the effect of constraints in this hybrid has limited the potential for divergence or its young age has not allowed it to evolve further.

Conclusion: Results suggest that population divergence within this hybrid species, in mainland Italy, has evolved in response to climatic variation, suggesting ongoing local adaptation. We also report that most of the strongly divergent loci among hybrid populations are not strongly differentiated between the parental species, suggesting that novel rearrangements of parental alleles not necessarily enhance genomic differentiation in the hybrid. Rather, it is possible that constraints linked to incompatibilities have restricted the evolution of this admixed genome.

Paper II

Predictors of genomic differentiation within a hybrid taxon

Hybridization has the potential to provide high genomic variation that may enable rapid adaptation. However, novel rearrangement of parental blocks in admixed genomes can generate constraints due to genomic incompatibilities that would in turn limit genome evolution of admixed lineages. Thus, questions such as how easily can hybrid lineages achieve divergent genome compositions and whether different combinations of parental alleles can easily be achieved and fuel adaptive variation, become relevant. Populations of the Italian sparrow in the Mediterranean islands vary in their genomic composition differing in the proportion of genomic makeup contributed by each parental species. In islands like Corsica and Crete the hybrid genome present a higher proportion of house sparrow than Spanish sparrow alleles while Sicilian populations show the opposite pattern. It has been hypothesised that such differences are the result of independent hybridization events that gave raise to divergent hybrid populations. Nonetheless, certain genomic blocks have been found to be consistently inherited from the same parental species supporting the constraints hypothesis (Runemark et al., 2018). Thus, in this system it is still unknown how freely the hybrid genome can generate genomic variation available for adaptation and shaped by divergent selection; and whether constraints due to incompatibilities limit the evolution of genomic differentiation generating mirrored patterns between isolated islands' populations.

Hybrid genomes are shaped by a combination of processes, including selection for adaptive combinations, purging of incompatibilities and recombination rate. In this study we assessed the effect that these processes have in the genomic differentiation of the hybrid Italian sparrow. In this uniquely suited study system, where divergent populations vary in their ancestry proportion, we are also evaluating the effect that geographic isolation, genetic drift and differential genomic parental proportion have in the genomic differentiation in the hybrid lineage.

Our results show that, consistent with long periods of independent evolution due to geographic isolation, genomic divergence among islands is higher than within-islands differentiation. If constraints are a predominant factor in the evolution of the hybrid genome, limiting the independent variation of geographically isolated populations, mirrored patterns of genomic differentiation can be generated across islands. Our results suggest that even though purging of genomic incompatibilities has an effect on the genomic differentiation in the hybrid, parallelism in genome evolution in the hybrid Italian sparrow found across and within islands are also partially explained by selection. Thus, there is also room for adaptive variation in the genome evolution in this hybrid taxon.

We found that recombination rate and differences in minor-, major-parental ancestry have an interactive effect on the patterns of genomic differentiation. Our results show that in the hybrid lineages the relationship between recombination rate and differentiation, expected due to linked selection, is less pronounced than that in the parents. This could suggest lower differentiation in low recombination regions within the hybrid lineages, as expected if unbalanced purging of minor-parent alleles in low recombination regions reduces the variation available for divergence.

Supporting our previous findings in populations across mainland Italy (paper I), constraints on how freely genomic rearrangements contribute to the genomic differentiation seem to still limit the genomic variation of this hybrid taxa.

Conclusion: This study suggests that selection and constraints linked to admixture interact and play a determining role on the genomic composition and further differentiation among hybrid populations.

Paper III

Rapid polygenic response to secondary contact in a hybrid species

Interspecific competition of closely related species after secondary contact can lead to character displacement as a way of avoiding exploiting the same resources. In the case of hybrid taxa, secondary contact with a parental species can not only have an effect on the stage of phenotypic and/or genetic characters, but species fusion could also occur if reproductive barriers against the parental lineages are not well established. Early stages of interbreeding can also cause character convergence in traits that are under selection due to competition. Apart from a narrow belt in the Alps where the Italian sparrow coexists with the house sparrow in a hybrid zone, this hybrid taxon is in sympatry with the Spanish sparrow in the Gargano peninsula. The Italian sparrow came into contact with the Spanish sparrow approximately 20 years ago in a small area in the Italian east coast. More recently, in 2012, this parental species has expanded its distribution to localities where the Italian sparrow was previously in allopatry. In this study we took advantage of this recent secondary contact event and using a spatial-temporal design, we evaluated the effects that coexistence with a parental species has at the ecological, demographic, phenotypic and genetic level on the hybrid Italian sparrow. We sampled, a previously allopatric population of the Italian sparrow, before and after secondary contact with the Spanish sparrow and compared our findings with older sympatric and independent allopatric populations. We assessed 81 SNP markers on protein-coding genes that are species-informative between the parental species, the house and Spanish sparrows. Thus, changes in allele frequencies among Italian sparrow indicate to what extent there is genetic convergence or divergence from the parent species at each locus.

Our results confirmed the strong reproductive isolation between the Italian and the Spanish sparrow by means of genetic assignment analysis, demonstrating a lack of hybridization occurring between these two species. Our results also show that interspecific competition has strong consequences for the hybrid; secondary contact with the Spanish sparrow has a detrimental effect on ecology and demography of the hybrid population. We found clear evidence for habitat segregation, the Italian sparrows being displaced from a previously widely used feeding territory, resulting in poorer body condition. We also estimated that the effective population size in this site has dropped by 40% after only three years of sympatry. Drastic changes in feeding ecology and demography can also generate phenotypic and genetic changes. Although no significant morphological change could be detected, comparing to allopatric Italian sparrows, sympatric birds were shown to be more divergent from the parental species in the protein-coding genetic markers.

We found a consistent shift in allele frequency towards divergence from the Spanish sparrow. Reduction of population size can generate bottleneck effects and increase the rate of genetic drift. However drastic changes in allele frequency due to drift are expected to occur at random and not to present any specific directionality towards divergence from the parental species as found here. This result remained significant when only non-synonymous SNP loci were considered. Supporting the polygenic character of the genetic response to secondary contact genetic markers under displacement were evenly distributed across the genome, located in different chromosomes.

Changes in genetic divergence found in our focal temporal comparison are mirrored by a somewhat older sympatric population that has been in secondary contact for at least 20 years. Finally, significant outlier genes found in our study have previously been found to be associated with learning and neural development in other bird species. This may suggest a role for adaptive divergence in traits related to cognition, learning and/or behaviour following secondary contact. However, extensive behavioural studies are needed to investigate whether cognitive traits may actually be under selection, yet alone subject to character displacement after secondary contact.

Conclusion: We demonstrate strong demographic, ecological and genetic consequences of inter-specific competition between the Italian sparrow and one of its parental species, the Spanish sparrow. There is habitat segregation resulting in poorer body condition and a significant drop in population size for the hybrid species. Compared with allopatric birds, sympatric population of Italian sparrows are genetically more diverged from Spanish sparrows. The Italian sparrows had diverged significantly across a set of 81 protein-coding genes.

DISCUSSION

DISCUSSION

Hybrid lineages in the speciation continuum

Far from being an evolutionary dead end, we now know that hybridization has an important contribution to biodiversity, making it an important field of study in evolutionary research. The topic of hybrid speciation, although still controversial (Matute & Cooper, 2021), has given a new perspective to our understanding on the origin and evolution of species. Thanks to the advances in sequencing technology and novel opportunities that genomic studies offer, it is exciting to approach this evolutionary process from a genomic perspective.

Hybrid lineages have the potential to originate in a multitude of instances along the speciation continuum, as we have seen that various levels of RI allow hybridization between species (Fig. 1B). The origin of an independent hybrid lineage does not only depend on the strength of the reproductive barriers between the parents and how divergent they are, even if it provides the potential for novel rearrangements in the admixed genome that might facilitate the establishment and rapid adaptation of the hybrid. It also relies on the ecological opportunities available to an incipient hybrid population. Moreover, the development of a hybrid lineage could simply manifest itself in the form of introgression of small genomic blocks that confer a unique differentiation, such as in *Heliconious* butterflies (The Heliconius Genome Consortium et al., 2012), where introgression of a block containing coloration traits, potentially conferring RI, has determined the evolution of an independent population. Hybrid species have thus the possibility to occur as the result of hybridization events placed along this continuum (Fig. 1B), with their fate also depending on external factors that might or might not allow their establishment (e.g. ecological opportunity that isolate the hybrid from the parental species).

Evolution of a hybrid lineage: Potential for adaptation vs. genomic constraints

The evolutionary path that a hybrid lineage takes after its formation, has rarely been studied, mainly because the study of hybrid speciation has focused on identifying hybrid species and determining the strength and nature of reproductive barriers against its parental species (Salazar et al., 2005; Mavárez et al., 2006; Meyer et al., 2006; Jones et al., 2013; Schumer et al., 2014; Lamichhaney et al., 2018). Besides determining whether a species is of hybrid origin, understanding issues like genome stability after its formation, how easily an admixed genome can diverge and whether this confers adaptive potential, which genomic regions are free and contribute

to variation, would bring us closer to an understanding on how hybridization plays an important role on diversity beyond introgression. For instance, genomic variation in a hybrid species can be the result of novel combinations of parental genomic blocks, which could generate new epistatic interactions, or through the inheritance of standing genetic variation present in the parental lineages.

The Italian sparrow is currently one of the few well–documented homoploid hybrid species where RI against both of its parental species has been identified. However, knowledge about intrinsic genomic differentiation and the potential for adaptation in this lineage were still very limited when I started studying this system. In **papers I** and **II,** we assessed genomic variation in this hybrid lineage, identifying population structure across its geographic distribution. We found evidence for weak population structure across the Italian peninsula (**paper I**). This pattern of differentiation is consistent with ongoing gene flow between populations. However the young age of this hybrid lineage (~6000 years) (Hermansen et al., 2011; Ravinet et al., 2018a), can also explain the lack of genomic differentiation, as these populations may have not had enough time to diverge. Conversely, genomic constraints can have a limiting effect on differentiation as purging of e.g. BDMI can remove variation in regions of physical linkage with incompatibilities (Schumer et al., 2018). Previously, Elgvin et al., (2017) have also found evidence for balancing selection. However, such variation was mainly present in private alleles. In contrast, our results suggest that the potential for variation in the Italian sparrow is concentrated in regions where there is little parental differentiation, unlike in the parental taxa where population differentiation within each species occurred mainly at loci with high parental differentiation. This supports the expectation that regions of high parental differentiation can harbour potential incompatibilities in an admixed genome. Differentially divergent parental loci could be highly conserved, as negative epistatic interactions between them may lead to stabilizing selection on loci fixed for compatible alleles. This type of genetic constraint on hybrid species could reduce their evolutionary potential to diverge at the population level. Consequently, if genomic incompatibilities are pervasive, loci presenting low parental divergence are more likely to vary. However, our study do not present direct evidence on fitness effect that these varying vs. fixed loci have on the hybrid.

In **Paper II** we also evaluated genomic variation, but in this case across the remaining of the distribution of the Italian sparrow, the Mediterranean islands. By sampling three populations within each Mediterranean island that the hybrid species inhabits (Corsica, Sicily and Crete), we assessed in parallel genomic variation within and among independent hybrid lineages, thought to

be the result of independent hybridization events. Relatively high genomic differentiation among populations within islands, accounting for the dispersal ability of the species (Summers–Smith, 1988), supports the potential for genomic variation found in mainland Italy (**paper I**), despite the likely role of incompatibilities in genomic differentiation. It has been argued that the relatively low genome-wide divergence between the parental species could explain the possibility of having several independent hybrid lineages in the islands, as a low incident of potential incompatibilities in hybridization between less divergent species would be expected (Runemark et al., 2018; Trier, 2018). However, I found parallel patterns of population divergence between the islands, a result that could be consistent with a form of bias on which evolutionary paths a hybrid genome could evolve once hybrid speciation has occurred, as a result of incompatibilities. Combinations of parental blocks might not always favour the evolutionary potential of an admixed genome, rather, constraints linked to incompatibilities could restrict its evolution.

Besides characterizing the genomic differentiation in this hybrid lineage, which genomic regions are free to vary and the potential constraints limiting variation, our study in **paper I** also contributed to the understanding of the genomic basis of adaptation. Our results show that genomic variation and its distribution within mainland Italy is partially explained by climate, suggesting that, despite the apparent role that incompatibilities have had in the evolution of the Italian sparrow, this hybrid species has been able to respond to selective pressures. The potential role of climatic factors determining genomic differentiation is also concordant with the identification of a candidate gene involved in cranial and beak morphology, as well as previous results suggesting variation on regions involved in beak shape (Elgvin et al., 2017) and a role for climate in driving phenotypic divergence in beak morphology in these same populations (Eroukhmanoff et al., 2013). The beak is a trait known to be under strong selective pressure driving major evolutionary shifts in Darwin's finches (Lamichhaney et al., 2015, 2016; Chaves et al., 2016) as well as responding to environmental divergence affecting food availability (Grant & Grant, 2003, 2014). Thus, if climatic regimes are considered as a reliable proxy of food availability, it is possible that environmental pressures related to food resources might be reflected on the divergence of genes associated with beak morphology. These findings suggest that hybridization provides variation facilitating adaptation, consistent with previous studies across a wide range of taxa (Song et al., 2011; The Heliconius Genome Consortium et al., 2012; Sankararaman et al., 2014; Marques et al., 2019).

Factors determining genomic variation in a hybrid lineage

Apart from evaluating the capacity for differentiation and adaptation of the Italian sparrow, identifying the factors that have an effect on the potential for genomic variation of a hybrid lineage is crucial for our understanding on the evolution of admixed genomes.

Although genomic differentiation within the hybrid taxon can be triggered by the potential novel rearrangements of divergent parental genomic blocks, highly divergent loci are also prone to harbour incompatibilities when segregating in a single admixed genome. However, the flexibility of generating novel rearrangements and purging of incompatibilities also depend on factors like variation in recombination rate across the genome (Burri et al., 2015; Schumer et al., 2018) and ancestry proportions (Schumer et al., 2018). Moreover, extrinsic factors like selection also affect genomic variation. In **paper II** we analysed these factors, and their interactions, as they potentially have had a determining effect on the evolution of the Italian sparrow's admixed genome.

Genomic differentiation has shown to be affected by the recombination rate landscape in species of non-hybrid origin, generating correlated patterns of differentiation in closely related species (Burri et al., 2015; Ravinet et al., 2018b). For instance, linked selection (e.g. selective sweeps) could lead to higher differentiation in low recombination regions in comparison to regions of high recombination, giving rise to a negative correlation between recombination rate and genomic differentiation (Nachman & Payseur, 2012; Burri et al., 2015). Specific to hybrid taxa, purging of incompatibilities, more impactful in low recombination regions, could significantly reduce the variation available for subsequent differentiation. Moreover, ancestry sorting during genome stabilization in independent lineages of a hybrid species can lead to differences in admixture proportions of the major- and minor-parental species (Runemark et al., 2018; Schumer et al., 2018), which in turn could generate specific evolution within ancestry types. Using the populations from the Mediterranean islands of Corsica, Crete and Sicily, where divergent lineages vary in the genomic composition inherited from the parents, allowed us to assess the effect that differentiation in ancestry-proportions has in the genomic differentiation of the hybrid Italian sparrow, and its interaction with recombination rate.

Results in **paper II** suggest that purging of regions with minor-parent-ancestry, which potentially can harbour incompatibilities in a major-parent genomic background, may explain the less pronounced relationship found between recombination rate and genomic differentiation in the hybrid species, in contrast to a steeper pattern found in their parental species. This could support

the hypothesis that purging incompatible minor parent alleles in low recombination regions has an important impact on genomic variation, reducing the potential for differentiation in these regions. In fact, introgression has been found to be less permeable in low recombination regions in hybrid populations of swordtail fish, sticklebacks, *Heliconius* butterflies and even in humans (Sankararaman et al., 2014; Ravinet et al., 2018b; Schumer et al., 2018; Martin et al., 2019), such negative selection against minor parent ancestry could generate patterns of correlation between introgression and recombination rate (Schumer et al., 2018). However, in the Italian sparrow populations from the Islands, we do not find such a consistent pattern between recombination rate and minor–ancestry proportion across all three islands and very little of the differentiation within islands is explained by recombination rate.

Lastly, from previous work (Elgvin et al., 2017; Runemark et al., 2018) and our study in mainland Italy (**paper I**) there is evidence of genomic regions under balancing selection and candidate regions to adaptation in the Italian sparrow genome. Thus, in **paper II** we assess the effect that selection has on the genomic differentiation of populations with potentially similar evolutionary history (i.e. similar parental contributions in populations within islands) vs. populations among islands where ancestry proportions diverge. We find some evidence that same genomic regions repeatedly are involved in population divergence among and within–islands. Moreover, overall differentiation between islands is correlated with differences in signatures of selection and measures informative of selection are correlated among islands. These results jointly might be an indicative of parallelism in selection occurring among the islands' populations (Roesti et al., 2014) or a strong effect of constraints from incompatibilities that limit the regions flexible to diverge. In fact, similarities in genetic composition have been found in the wild ancestor of *Helianthus* sunflowers and arise repeatedly in lab–crosses (Rieseberg et al., 1996) and in younger and older lineages of Lycaides butterflies (Chaturvedi et al., 2020). Unfortunately our results evaluating parallel and background selection were inconclusive. Similarly to our findings in paper I, results in paper II also suggests that constraints and selection interact in shaping the genomic landscape of differentiation in this hybrid species.

Secondary contact – the integrity of a hybrid species

Secondary contact of closely related species can have different ecological and evolutionary outcomes at the phenotypic and genomic level (Chira et al., 2020). Sympatry can generate possibilities for introgression between species, or conversely reinforcement can also occur, although

identifying genomic traces of the latter process has been more controversial, as reinforcement often show signatures of character displacement (Ortiz-Barrientos et al., 2009; Garner et al., 2018). Competition for resources by the newly sympatric species –after secondary contact– becomes a strong selective pressure that can even cause extinction of one of the species (Connell, 1961) or limit the evolution of traits if there are restricted niche opportunities (Weir & Mursleen, 2013). Phenotypic variation can also be triggered by competition; one of the most studied consequences of competition is ecological character displacement (ECD), where rapid phenotypic divergence lessens competition and allows species to coexist (Schluter, 2000; Grant & Grant, 2006; Pfennig & Pfennig, 2010, 2012; Lamichhaney et al., 2016), a process that has been thought to play a causal role on trait diversification and even adaptive radiations. The study of ECD however, has rarely been studied within a single population before and after the secondary contact event.

In the case of secondary contact between a hybrid and its parental species, character convergence could also occur in the early stages of interbreeding or species fusion, if RI is weak. In **paper III** we evaluated the ecological and evolutionary consequences of secondary contact at the phenotypic and genetic level between the hybrid Italian sparrow and the Spanish sparrow, assessing changes in the same population before and after the encounter with the migrant Spanish sparrow and comparing with an allopatric and a somewhat older sympatric populations. First, we confirm strong RI between the hybrid and its parental species, the Spanish sparrow, through the absence of evidence for hybridization, as also previously suggested (e.g. Hermansen et al., 2011; Trier et al., 2014). Furthermore, our results suggest that the hybrid lineage has the capacity to diverge, coping with the novel selective pressures linked to competition with the Spanish sparrow, which results in the detrimental effect that species interaction has on its feeding ecology (i.e. habitat displacement), population dynamic and body condition. Even though our results do not support any character displacement in beak shape, a trait that often has been linked to resource competition (Grant & Grant, 2006; Olsen, 2017), it is important to consider that we only evaluated bi-dimensional traits, an approach that might have limitations. Perhaps, using tools like geometric morphometric analysis to evaluate beak shape might have given us more informative data on potential phenotypic changes.

At the genomic level we find a significant polygenetic change in the form of genetic character displacement, particularly in genes that have previously been linked to vision, cognitive and behavioural traits. Severe changes in population size could also have an impact on allele frequencies. However, drift would generate genetic changes at random and not the directional patterns in allele frequency changes we observe here, in parallel in both the temporal and the

spatial comparisons. Interestingly, the concordance of genetic character displacement found in the spatial as well as the temporal comparisons suggests that secondary contact has predictable effects at the genetic level in this hybrid species. We still however lack to evaluate the link between the candidate genes presenting character displacement and their role in phenotypic-behavioural traits to be able to draw any conclusion on the selective pressures acting in this system.

Overall, my work throughout this PhD, has provided a broader understanding of the potential and limiting factors that a hybrid species could encounter in its evolutionary path. While its evolution could be favourable, with the potential for variation added by the novel rearrangement of parental genomic blocks, constraints in the admixed genome are part of the nature of a hybrid species and as such can hinder the flexibility of variation. Nonetheless, as non-hybrid lineages, hybrid species will also encounter the stochastic effects of drift and selective pressures in e.g. divergent habitats, factors that may also have an important role on the evolution of a hybrid species. Thus, the evolutionary path of a hybrid lineage is far from being a deterministic one; it is contingent to the genomic composition of a particular admixed genome and the challenges and possibilities it encounters in nature through out its evolution.

CONCLUSION AND FUTHER DIRECTIONS

The studies forming this dissertation opened the door to many new research opportunities. By evaluating the potential for adaptation and genomic variation in the Italian sparrow, assessing the determining factors on the evolution of its admixed genome and its stability when co-existing with one of its parental species, new questions on the evolution of admixed genomes emerged. For instance, how to reconcile the constraints (e.g. via incompatibilities) and potential for adaptation (e.g. novel rearrangements) that are developing within a single admixed genome. Our studies showed that climate plays an important role in the genomic differentiation patterns found in Italian populations in mainland Italy, suggesting a potential for adaptive variation in the presence of gene flow, a phenomenon extensively reported in species of non-hybrid origin (de Leon et al., 2010; Marques et al., 2016; Martin et al., 2013; Raeymaekers et al., 2017). Furthermore, like for non-hybrid species, gene flow between these populations seems to be constrained by adaptation to climate (Edelaar et al., 2012). However, we also showed that incompatibilities might play an important role in the evolution of this admixed genome. Interestingly, it has been proposed previously that coupling mechanisms between incompatible loci and regions under selection could trigger a rapid genomic divergence and adaptation in hybrid populations (Seehausen, 2013). For

example, if populations of hybrid origin, early after the hybridization event and during genome stabilization, become partially isolated by e.g. environmentally divergent habitats, different sets of genomic incompatibilities might become trapped within such populations and coupling with loci involved in local adaptation may occur, which could potentially facilitate diversification within the hybrid lineage (Seehausen, 2004; Abbott et al., 2013; Bierne et al., 2013; Butlin & Smadja, 2018). This coupling mechanism, more prone to arise in hybrid lineages, could trigger rapid adaptive responses and processes of diversifying selection in comparison to non-hybrid species (Eroukhmanoff et al., 2013; Seehausen, 2013). Empirical studies and theoretical models addressing these hypotheses directly are needed to assess causality. Hybrid systems, like the Italian sparrow, where drastic differences in genomic composition (i.e. differences in parental-ancestry proportions), divergent selection and local adaptation are exposed, might present an ideal case to empirically test this hypothesis, looking for linkage between BDMIs and regions under natural selection.

Evidence of shared regions of divergence in genomes from independent lineages, with different genomic compositions, and correlation of measures of selection among islands (**paper II**), might be indicative of parallel selection. However, the nature of such selective pressures is still unknown, and such a pattern could also arise from shared incompatibility loci limiting genomic differentiation in specific regions. Thus, a formal evaluation on the variation of potential selective factors (e.g. environmental variation) among- and within- the Mediterranean islands inhabited by the Italian sparrows, and their correlation to shared regions of differentiation is needed. Other processes such as historical shared selection in the parental species (Runemark et al., 2018) and stabilizing selection linked to human commensalism, a trait shared with the parental, the house sparrow (Ravinet et al., 2018a), could also generate similar patters of differentiation and selection signatures. Crossing experiments, like those performed in *Helianthus* sunflowers (Rieseberg et al, 1996), and comparison of common genetic composition in wild populations (Chaturvedi et al., 2020), could give some clues about the restrictive role of genomic incompatibilities in the evolution of this hybrid species. Moreover, investigating whether structural variation is a determinant factor on the genomic divergence and constraints in this hybrid species could bring a wider scope on the study of genomic bases of differentiation in the Italian sparrow. As it has been shown that structural variation (e.g. inversions) can promote genomic divergence between ecotypes of non-hybrid origin (e.g. Berg et al., 2016).

Results from **paper III** pointed to candidate genes involved in cognitive and behavioural traits as loci under divergent selection in the Italian sparrow after secondary contact with the Spanish sparrow. However, it would be interesting to identify if cognitive traits are the phenotype under divergent selection, as a consequence of competition. Behavioural experiments comparing traits in allopatric populations against behavioural responses from the current Italian population living in sympatry with the Spanish sparrow could shed light on phenotypic traits that are potentially related to competition pressures. Moreover, new evaluation of beak shape using more informative tools like geometric morphometrics could be better able to detect potential character displacement following secondary contact, as it has exposed informative results on beak shape change in other birds species (Foster et al., 2008). Another interesting line of research is the evolution of reproductive barriers (most likely developed in allopatry) after secondary contact. We confirmed the strong RI of the Italian sparrow against the Spanish sparrow, also we found evidence of genetic character displacement in the Italian sparrow in populations in sympatry, these results could be indicative, among other mechanisms, of the reinforcement of reproductive barriers in this sympatric system. However, a rigorous study on whether loci exhibiting genetic character displacement are involved in any processes that facilitate assortative mating, using a whole-genome approach, would be crucial to elucidate whether reinforcement is occurring in this system. Using whole genome resequencing data to evaluate genome-wide temporal variation in this natural experiment would further shed light on the genomic consequences that secondary contact could have in a hybrid lineage.

Finally, in a more general context, it is interesting to ask whether the level of RI between parental species would determine the amount of constraints present in an admixed genome and the likelihood of hybrid speciation (Matute & Cooper, 2021). It has been shown that parental divergence has an effect on novelty in hybrids (Stelkens & Seehausen, 2009; Stelkens et al., 2009). However, the question remains of whether hybridization outcomes should be of a deterministic nature linked to the level of RI (i.e. where the parent species are situated along the speciation continuum). Alternatively, the potential for a hybrid lineage to persist and evolve may depend on whether particular admixed genomic regions can freely vary without constraining effects from the rest of the genome. A parallel could be drawn to the case of cichlids and *Heliconious* butterflies, where variation in key traits (and specific genomic regions) could generate novel adaptive variation (The Heliconius Genome Consortium et al., 2012; Meier et al., 2017).

In this thesis, by integrating ecological and genomic approaches to develop an understanding on the evolution of the Italian sparrow hybrid genome, after its formation, I aimed to shed light on the mechanisms and evolutionary processes that an admixed genome faces and its evolutionary implications. By analysing the genetic basis of adaptation I explored the evolutionary potential of a hybrid species at the genomic level. Correlating signatures of selection in genome–wide approaches with important ecological traits is a crucial step in evolutionary research.

ACKNOWLEDGMENTS

REFERENCES

Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., et al. (2013). Hybridization and speciation. *J. Evol. Biol.* 26, 229–246. doi:10.1111/j.1420-9101.2012.02599.x.

Abbott, R. J., and Rieseberg, L. H. (2012). Hybrid Speciation. *eLS.* doi:110.1002/9780470015902.a0001753.pub2.

Abi-Rached, L., others, Jobin, M. J., Kulkarni, S., McWhinnie, A., Dalva, K., et al. (2011). The Shaping of Modern Human Immune Systems by Multiregional Admixture with Archaic Humans. *Science (80-.).* 334(89), 89–94.

Alexander, D. H., and Novembre, J. (2009). Fast Model-Based Estimation of Ancestry in Unrelated Individuals. 1655–1664. doi:10.1101/gr.094052.109.vidual.

Anderson, E., and Stebbins, G. L. (1954). Hybridization as an Evolutionary Stimulus. *Evolution (N. Y).* 8, 378. doi:10.2307/2405784.

Arnold, M. L., and Kunte, K. (2017). Adaptive Genetic Exchange: A Tangled History of Admixture and Evolutionary Innovation. *Trends Ecol. Evol.* 32, 601–611. doi:10.1016/j.tree.2017.05.007.

Baack, E. J., and Rieseberg, L. H. (2007). A genomic view of introgression and hybrid speciation. *Curr Opin Genet Dev* 17, 513–518.

Bailey, R. I., Tesaker, M. R., Trier, C. N., and Saetre, G. P. (2015). 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. doi:10.1111/jeb.12652.

Barrera-Guzmán, A. O., Aleixo, A., Shawkey, M. D., and Weir, J. T. (2018). Hybrid speciation leads to novel male secondary sexual ornamentation of an Amazonian bird. *Proc. Natl. Acad. Sci. U. S. A.* 115, E218–E225. doi:10.1073/pnas.1717319115.

Barton, N. H., and Hewitt, G. M. (1985). Analysis of hybrid zones. *Annu. Rev. Ecol. Syst.* 16, 113–148. doi:10.1146/annurev.es.16.110185.000553.

Berg, P. R., Star, B., Pampoulie, C., Sodeland, M., Barth, J. M. I., Knutsen, H., et al. (2016).

Three chromosomal rearrangements promote genomic divergence between migratory and stationary ecotypes of Atlantic cod. *Sci. Rep.* 6, 1–12. doi:10.1038/srep23246.

Bierne, N., Gagnaire, P., and David, P. (2013). The geography of introgression in a patchy environment and the thorn in the side of ecological speciation. 59, 72–86.

Bierne, N., Welch, J., Loire, E., Bonhomme, F., and David, P. (2011). The coupling hypothesis: Why genome scans may fail to map local adaptation genes. *Mol. Ecol.* 20, 2044–2072. doi:10.1111/j.1365-294X.2011.05080.x.

Bolnick, D., and Fitzpatrick, B. (2007). Sympatric Speciation: Models and Empirical Evidence. *Annu. Rev. Ecol. Evol. Syst.*, 459–87.

Buerkle, C. A., Morris, R. J., Asmussen, M. A., and Rieseberg, L. H. (2000). The likelihood of homoploid hybrid speciation. *Heredity (Edinb).* 84, 441–451.

Buerkle, C. A., and Rieseberg, L. H. (2008). The rate of genome stabilization in homoploid hybrid species. *Evolution* 62, 266–75. doi:10.1111/j.1558-5646.2007.00267.x.

Bull, V., Beltrán, M., Jiggins, C. D., McMillan, W. O., Bermingham, E., and Mallet, J. (2006). Polyphyly and gene flow between non-sibling Heliconius species. *BMC Biol.* 4, 1–17. doi:10.1186/1741-7007-4-11.

Burri, R., Nater, A., Kawakami, T., Mugal, C. F., Olason, P. I., Smeds, L., et al. (2015). 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. doi:10.1101/gr.196485.115.

Butlin, R. K., Galindo, J., and Grahame, J. W. (2008). Review. Sympatric, parapatric or allopatric: The most important way to classify speciation? *Philos. Trans. R. Soc. B Biol. Sci.* 363, 2997–3007. doi:10.1098/rstb.2008.0076.

Butlin, R. K., and Smadja, C. M. (2018). Coupling, Reinforcement, and Speciation. *Am. Nat.* 191, 000–000. doi:10.1086/695136.

Butlin, R. K., and Stankowski, S. (2020). Is it time to abandon the biological species concept? No. *Natl. Sci. Rev.* 7, 1400–1401.

Chaturvedi, S., Lucas, L. K., Buerkle, C. A., Fordyce, J. A., Forister, M. L., Nice, C. C., et al. (2020). Recent hybrids recapitulate ancient hybrid outcomes. *Nat. Commun.* 11. doi:10.1038/s41467-020-15641-x.

Chaves, J. A., Cooper, E. A., Hendry, A. P., Podos, J., De León, L. F., Raeymaekers, J. A. M., et al. (2016). Genomic variation at the tips of the adaptive radiation of Darwin's finches. *Mol. Ecol.* 25, 5282–5295. doi:10.1111/mec.13743.

Chira, A. M., Cooney, C. R., Bright, J. A., Capp, E. J. R., Hughes, E. C., Moody, C. J. A., et al. (2020). The signature of competition in ecomorphological traits across the avian radiation. *Proc. R. Soc. B Biol. Sci.* 287. doi:10.1098/rspb.2020.1585.

Comeault, A. A., and Matute, D. R. (2018). Genetic divergence and the number of hybridizing species affect the path to homoploid hybrid speciation. *Proc. Natl. Acad. Sci. U. S. A.* 115, 9761–9766. doi:10.1073/pnas.1809685115.

Connell, J. H. (1961). The Influence of Interspecific Competition and Other Factors on the Distribution of the Barnacle Chthamalus Stellatus. *Ecology* 42, 710–723. doi:doi:10.2307/1933500.

Consortium, T. H. G., Dasmahapatra, K. K., Walters, J. R., Briscoe, A. D., Davey, J. W., Whibley, A., et al. (2012). Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* 487, 94–98. doi:10.1038/nature11041.Butterfly.

Coyne, J. A. (1994). Ernst mayr and the origin of species. *Evolution (N. Y).* 48, 19–30.

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

Dagilis, A. J., and Matute, D. R. (2020). Incompatibilities between emerging species. *Science (80-.)*. 368, 710–711. doi:10.1126/science.abb8066.

Depew, D. J., and Weber, B. H. (1995). *Darwinism Evolving: Systems Dynamics and the Genealogy of Natural Selection.*

Drès, M., and Mallet, J. (2002). Host races in plant-feeding insects and their importance in sympatric speciation. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 357, 471–92. doi:10.1098/rstb.2002.1059.

Edelaar, P., Alonso, D., Lagerveld, S., Senar, J. C., and Björklund, M. (2012). Population differentiation and restricted gene flow in Spanish crossbills: Not isolation-by-distance but isolation-by-ecology. *J. Evol. Biol.* 25, 417–430. doi:10.1111/j.1420-9101.2011.02443.x.

Edelman, N. B., B., F. P., Miyagi, M., Clavijo, B., Davey, J., Dikow, R. B., et al. (2019). Genomic architecture and introgression shape a butterfly radiation. *Science (80-.)*. 366, 594–599.

Egan, S. P., Ragland, G. J., Assour, L., Powell, T. H. Q., Hood, G. R., Emrich, S., et al. (2015). Experimental evidence of genome-wide impact of ecological selection during early stages of speciation-with-gene-flow. *Ecol. Lett.* 18, 817–825. doi:10.1111/ele.12460.

Elgvin, T. O., Hermansen, J. S., Fijarczyk, A., Bonnet, T., Borge, T., Sæther, S. A., et al. (2011). Hybrid speciation in sparrows II: A role for sex chromosomes? *Mol. Ecol.* 20, 3823–3837. doi:10.1111/j.1365-294X.2011.05182.x.

Elgvin, T. O., Trier, C. N., Tørresen, O. K., Hagen, I. J., Lien, S., Nederbragt, A. J., et al. (2017). The genomic mosaicism of hybrid speciation. *Sci. Adv.* 3, 1–16. doi:10.1126/sciadv.1602996.

Eroukhmanoff, F., Hermansen, J. S., Bailey, R. I., Sæther, S. A., and Sætre, G. P. (2013). Local adaptation within a hybrid species. *Heredity (Edinb).* 111, 286–292. doi:10.1038/hdy.2013.47.

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., and Foll, M. (2013). Robust Demographic Inference from Genomic and SNP Data. *PLoS Genet.* 9. doi:10.1371/journal.pgen.1003905.

Excoffier, L., Marchi, N., Marques, D. A., Matthey-Doret, R., Gouy, A., and Sousa, V. C. (2021). Fastsimcoal2: Demographic inference under complex evolutionary scenarios. *Bioinformatics* 37, 4882–4885. doi:10.1093/bioinformatics/btab468.

Fairbanks, D. J., and Abbott, S. (2016). Darwin's influence on Mendel: Evidence from a new translation of Mendel's paper. *Genetics* 204, 401–405. doi:10.1534/genetics.116.194613.

Feliner, N. G., Álvarez, I., Fuertes-Aguilar, J., Heuertz, M., Marques, I., Moharrek, F., et al. (2017). Is homoploid hybrid speciation that rare? An empiricist's view. *Heredity (Edinb).* 118, 513–516. doi:10.1038/hdy.2017.7.

Fitzpatrick, B. M., Fordyce, J. A., and Gavrilets, S. (2009). Pattern, process and geographic modes of speciation. *J. Evol. Biol.* 22, 2342–2347. doi:10.1111/j.1420-9101.2009.01833.x.

Foster, D. J., Podos, J., and Hendry, A. P. (2008). A geometric morphometric appraisal of beak shape in Darwin's finches. *J. Evol. Biol.* 21, 263–275. doi:10.1111/j.1420-9101.2007.01449.x.

Funk, D. J., Nosil, P., and Etges, W. J. (2006). Ecological divergence exhibits consistently positive associations with reproductive isolation across disparate taxa. *Proc. Natl. Acad. Sci. U. S. A.* 103, 3209–3213. doi:10.1073/pnas.0508653103.

Garner, A. G., Goulet, B. E., Farnitano, M. C., Molina-Henao, Y. F., and Hopkins, R. (2018). Genomic signatures of reinforcement. *Genes (Basel).* 9. doi:10.3390/genes9040191.

Gompert, Z., and Buerkle, C. A. (2011). Bayesian estimation of genomic clines. *Mol. Ecol.* 20, 2111–2127. doi:10.1111/j.1365-294X.2011.05074.x.

Gompert, Z., Fordyce, J. A., Forister, M. L., Shapiro, A. M., and Nice, C. C. (2006). Homoploid hybrid speciation in an extreme habitat. *Science (80-.).* 314, 1923–1925. doi:10.1126/science.1135875.

Gompert, Z., Lucas, L. K., Nice, C. C., Fordyce, J. A., Forister, M. L., and Buerkle, C. A. (2012). Genomic regions with a history of divergent selection affect fitness of hybrids between two butterfly species. *Evolution (N. Y).* 66, 2167–2181. doi:10.1111/j.1558-5646.2012.01587.x.

Grant, B., and Grant, P. (2003). What Darwin's Finches can Teach Us about the Evolutionary Origin and Regualtion of Biodiversity. *Bioscience* 53, 965–975. doi:10.1641/0006-3568(2003)053[0965:WDFCTU]2.0.CO;2.

Grant, P. R., and Grant, B. R. (2006). Evolution of character displacement in Darwin's finches. *Science (80-.).* 313, 224–226. doi:10.1126/science.1128374.

Grant, P. R., and Grant, R. B. (2014). *40 Years of Evolution. Darwin's Finches on Daphne Major Island.* , ed. P. U. Press.

Gross, B. L., and Rieseberg, L. H. (2005). The ecological genetics of homoploid hybrid speciation. *J. Hered.* 96, 241–252. doi:10.1093/jhered/esi026.

Hendry, A. P. (2009). Ecological speciation! Or the lack thereof? *Can. J. Fish. Aquat. Sci.* 66, 1383–1398. doi:10.1139/F09-074.

Hendry, A. P., Bolnick, D. I., Berner, D., and Peichel, C. L. (2009a). Along the speciation continuum in sticklebacks. *J. Fish Biol.* 75, 2000–2036. doi:10.1111/j.1095-8649.2009.02419.x.

Hendry, A. P., Huber, S. K., De León, L. F., Herrel, A., and Podos, J. (2009b). Disruptive selection in a bimodal population of Darwin's finches. *Proc. R. Soc. B Biol. Sci.* 276, 753–759. doi:10.1098/rspb.2008.1321.

Hermansen, J. S., Haas, F., Trier, C. N., Bailey, R. I., Nederbragt, A. J., Marzal, A., et al. (2014). Hybrid speciation through sorting of parental incompatibilies in Italian sparrows. *Mol. Ecol.* 23, 5831–5842. doi:10.1111/mec.12910.

Hermansen, J. S., Sæther, S. A., Elgvin, T. O., Borge, T., Hjelle, E., and Sætre, G. P. (2011). Hybrid speciation in sparrows I: Phenotypic intermediacy, genetic admixture and barriers to gene flow. *Mol. Ecol.* 20, 3812–3822. doi:10.1111/j.1365-294X.2011.05183.x.

Huber, S. K., De León, L. F., Hendry, A. P., Bermingham, E., and Podos, J. (2007). Reproductive isolation of sympatric morphs in a population of Darwin's finches. *Proc. R. Soc. B Biol. Sci.* 274, 1709–1714. doi:10.1098/rspb.2007.0224.

Huerta-Sanchez, E., Jin, X., Asan, Bianba, Z., Peter, B., Vinckenbosch, N., et al. (2014). Altitude adaptation in Tibet caused by introgresson of Denosivan-like DNA. *Nature* 512, 194–197. doi:10.1038/nature13408.Altitude.

Jones, J. C., Fan, S., Franchini, P., Schartl, M., and Meyer, A. (2013). The evolutionary history of Xiphophorus fish and their sexually selected sword: A genome-wide approach using restriction site-associated DNA sequencing. *Mol. Ecol.* 22, 2986–3001. doi:10.1111/mec.12269.

Jones, M. R., Mills, L. S., Alves, P. C., Callahan, C. M., Alves, J. M., Lafferty, D. J. R., et al. (2018). Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe hares. *Science (80-.).* 1358, 1355–1358.

Juric, I., Aeschbacher, S., and Coop, G. (2016). The Strength of Selection against Neanderthal Introgression. *PLoS Genet.* 12, 1–25. doi:10.1371/journal.pgen.1006340.

Kottler, M. J. (1978). Charles darwin's biological species concept and theory of geographic speciation: The transmutation notebooks. *Ann. Sci.* 35, 275–297. doi:10.1080/00033797800200251.

Kronforst, M. R., Young, L. G., Blume, L. M., and Gilbert, L. E. (2006). Multilocus Analyses of Admixture and Introgression among Hybridizing Heliconius Butterflies. *Evolution (N. Y).* 60, 1254–1268.

Kulmuni, J., and Butlin, R. K. (2021). Isolating Mechanisms. *eLS* 2, 1–8. doi:10.1002/9780470015902.a0029304.

Kulmuni, J., Butlin, R. K., Lucek, K., Savolainen, V., and Westram, A. M. (2020). Towards the completion of speciation: The evolution of reproductive isolation beyond the first barriers. *Philos. Trans. R. Soc. B Biol. Sci.* 375. doi:10.1098/rstb.2019.0528.

Lamichhaney, S., Berglund, J., Almén, M. S., Maqbool, K., Grabherr, M., Martinez-Barrio, A., et al. (2015). Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* 518, 371–375. doi:10.1038/nature14181.

Lamichhaney, S., Han, F., Berglund, J., Wang, C., Sallman, A. M., Webster, M. ., et al. (2016). A beak size locus in Darwin's finches facilitated character displacement during a drought. *Science (80-.).* 6284, 470–474.

Lamichhaney, S., Han, F., Webster, M. T., Andersson, L., Grant, R. B., and Grant, P. R. (2018). Rapid hybrid speciation in Darwin's finches. *Science (80-.).* 359, 172. doi:10.1126/science.359.6372.172-d.

Lipson, M., Loh, P. R., Levin, A., Reich, D., Patterson, N., and Berger, B. (2013). Efficient moment-based inference of admixture parameters and sources of gene flow. *Mol. Biol. Evol.* 30, 1788–1802. doi:10.1093/molbev/mst099.

Malinsky, M., Challis, R. J., Tyers, A. M., Schiffels, S., Terai, Y., Ngatunga, B. P., et al. (2015). Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science (80-.).* 350, 1493–1498. doi:10.1126/science.aac9927.

Malinsky, M., Matschiner, M., and Svardal, H. (2021). Dsuite - Fast D-statistics and related admixture evidence from VCF files. *Mol. Ecol. Resour.* 21, 584–595. doi:10.1111/1755-0998.13265.

Mallet, J. (1995). A species definition for the modern synthesis. *Trends Ecol. Evol.* 10, 294–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21237047.

Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends Ecol. Evol.* 20, 229–237. doi:10.1016/j.tree.2005.02.010.

Mallet, J. (2007). Hybrid speciation. *Nature* 446, 279–283. doi:10.1038/nature05706.

Mallet, J. (2008a). Hybridization, ecological races and the nature of species: Empirical evidence for the ease of speciation. *Philos. Trans. R. Soc. B Biol. Sci.* 363, 2971–2986. doi:10.1098/rstb.2008.0081.

Mallet, J. (2008b). Mayr's view of Darwin: was Darwin wrong about speciation? *Biol. J. Linn. Soc.* 95, 3–16.

Mallet, J. (2010). Why was Darwin's view of species rejected by twentieth century biologists? *Biol. Philos.* 25, 497–527. doi:10.1007/s10539-010-9213-7.

Mallet, J. (2020). Alternative views of biological species: reproductively isolated units or genotypic clusters? *Natl. Sci. Rev.* 7, 1401–1407.

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

Marques, D. A., Meier, J. I., and Seehausen, O. (2019). A Combinatorial View on Speciation and Adaptive Radiation. *Trends Ecol. Evol.* 34, 531–544. doi:10.1016/j.tree.2019.02.008.

Martin, S. H., Davey, J. W., and Jiggins, C. D. (2015). Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Mol. Biol. Evol.* 32, 244–257. doi:10.1093/molbev/msu269.

Martin, S. H., Davey, J. W., Salazar, C., and Jiggins, C. D. (2019). Recombination rate

variation shapes barriers to introgression across butterfly genomes. *PLoS Biol.* 17, 1–28. doi:10.1101/297531.

Mathur, R., and Adlakha, N. (2016). A graph theoretic model for prediction of reticulation events and phylogenetic networks for DNA sequences. *Egypt. J. Basic Appl. Sci.* 3, 263–271. doi:10.1016/j.ejbas.2016.07.004.

Matute, D. R., and Cooper, B. S. (2021). Comparative studies on speciation: 30 years since Coyne and Orr. *Evolution (N. Y).* 75, 764–778. doi:10.1111/evo.14181.

Mavárez, J., Salazar, C. A., Bermingham, E., Salcedo, C., Jiggins, C. D., and Linares, M. (2006). Speciation by hybridization in Heliconius butterflies. *Nature* 441, 868–871. doi:10.1038/nature04738.

Mavárez, J., Salazar, C. A., Bermingham, E., Salcedo, C., Jiggins, C. D., and Linares, M. (2021). Author Correction: Speciation by hybridization in Heliconius butterflies (Nature, (2006), 441, 7095, (868-871), 10.1038/nature04738). *Nature* 592, E4–E5. doi:10.1038/s41586-021-03330-8.

McCarthy, E. M., Asmussen, M. A., and Anderson, W. W. (1995). A theoretical assessment of recombinational speciation. *Heredity (Edinb).* 74, 502–509. doi:10.1038/hdy.1995.71.

Meier, J. I., Marques, D. A., Mwaiko, S., Wagner, C. E., Excoffier, L., and Seehausen, O. (2017). Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nat. Commun.* 8, 1–11. doi:10.1038/ncomms14363.

Meier, J. I., Stelkens, R. B., Joyce, D. A., Mwaiko, S., Phiri, N., Schliewen, U. K., et al. (2019). The coincidence of ecological opportunity with hybridization explains rapid adaptive radiation in Lake Mweru cichlid fishes. *Nat. Commun.* 10. doi:10.1038/s41467-019-13278-z.

Meyer, A., Salzburger, W., and Schartl, M. (2006). Hybrid origin of a swordtail species (Teleostei: Xiphophorus clemenciae) driven by sexual selection. *Mol. Ecol.* 15, 721–730. doi:10.1111/j.1365-294X.2006.02810.x.

Moran, B. M., Payne, C., Langdon, Q., Powell, D. L., Brandvain, Y., and Schumer, M. (2021). The genomic consequences of hybridization. *Elife* 10, 1–33. doi:10.7554/elife.69016.

Nachman, M. W., and Payseur, B. A. (2012). Recombination rate variation and speciation: Theoretical predictions and empirical results from rabbits and mice. *Philos. Trans. R. Soc. B Biol. Sci.* 367, 409–421. doi:10.1098/rstb.2011.0249.

Norris, L. C., Main, B. J., Lee, Y., Collier, T. C., Fofana, A., Cornel, A. J., et al. (2015). Adaptive introgression in an African malaria mosquito coincident with the increased usage of insecticide-treated bed nets. *Proc. Natl. Acad. Sci. U. S. A.* 112, 815–820. doi:10.1073/pnas.1418892112.

Nosil, P. (2012). *Ecological speciation.*

Nosil, P., and Feder, J. L. (2012). Genomic divergence during speciation: Causes and consequences. *Philos. Trans. R. Soc. B Biol. Sci.* 367, 332–342. doi:10.1098/rstb.2011.0263.

Nosil, P., Harmon, L. J., and Seehausen, O. (2009). Ecological explanations for (incomplete) speciation. *Trends Ecol. Evol.* 24, 145–156. doi:10.1016/j.tree.2008.10.011.

Olsen, A. M. (2017). Feeding ecology is the primary driver of beak shape diversification in waterfowl. *Funct. Ecol.* 31, 1985–1995. doi:10.1111/1365-2435.12890.

Ortiz-Barrientos, D., Grealy, A., and Nosil, P. (2009). The Genetics and Ecology of Reinforcement. *Ann. N. Y. Acad. Sci.* 1168, 156–182. doi:10.1111/j.1749-6632.2009.04919.x.

Ottenburghs, J. (2018). Exploring the hybrid speciation continuum in birds. *Ecol. Evol.* 8, 13027–13034. doi:10.1002/ece3.4558.

Päckert, M., Ait Belkacem, A., Wolfgramm, H., Gast, O., Canal, D., Giacalone, G., et al.

(2019). Genetic admixture despite ecological segregation in a North African sparrow hybrid zone (Aves, Passeriformes, Passer domesticus × Passer hispaniolensis). *Ecol. Evol.* 9, 12710–12726. doi:10.1002/ece3.5744.

Papadopulos, A. S. T., Baker, W. J., Crayn, D., Butlin, R. K., Kynast, R. G., Hutton, I., et al. (2011). Speciation with gene flow on Lord Howe Island. *Proc. Natl. Acad. Sci. U. S. A.* 108. doi:10.1073/pnas.1106085108.

Papakostas, S., Vøllestad, L. A., Bruneaux, M., Aykanat, T., Vanoverbeke, J., Ning, M., et al. (2014). Gene pleiotropy constrains gene expression changes in fish adapted to different thermal conditions. *Nat. Commun.* 5, 1–9. doi:10.1038/ncomms5071.

Patrick Gainey, D., Kim, J. Y., and Maroja, L. S. (2018). Mapping reduced introgression loci to the X chromosome of the hybridizing field crickets, Gryllus firmus and G. Pennsylvanicus. *PLoS One* 13, 1–7. doi:10.1371/journal.pone.0208498.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., et al. (2012). Ancient admixture in human history. *Genetics* 192, 1065–1093. doi:10.1534/genetics.112.145037.

Payseur, B. A. (2010). Using differential introgression in hybrid zones to identify genomic regions involved in speciation. *Mol. Ecol. Resour.* 10, 806–820. doi:10.1111/j.1755-0998.2010.02883.x.

Peter, B. M. (2016). Admixture, population structure, and f-statistics. *Genetics* 202, 1485–1501. doi:10.1534/genetics.115.183913.

Pfennig, D. W., and Pfennig, K. S. (2010). Character Displacement and the Origins of Diversity. 176, 26–44. doi:10.1086/657056.

Pfennig, D. W., and Pfennig, K. S. (2012). Development and evolution of character displacement. 1256, 89–107. doi:10.1111/j.1749-6632.2011.06381.x.

Pinho, C., and Hey, J. (2010). Divergence with gene flow: Models and data. *Annu. Rev. Ecol. Evol. Syst.* 41, 215–230. doi:10.1146/annurev-ecolsys-102209-144644.

Powell, D. L., García-Olazábal, M., Keegan, M., Reilly, P., Du, K., Díaz-Loyo, A. P., et al. (2020). Natural hybridization reveals incompatible alleles that cause melanoma in swordtail fish. *Science (80-.).* 368, 731–736. doi:10.1126/science.aba5216.

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics* 155, 945–959. doi:10.1007/s10681-008-9788-0.

Ravinet, M., Elgvin, T. O., Trier, C., Aliabadian, M., Gavrilov, A., and Sætre, G. P. (2018a). Signatures of human-commensalism in the house sparrow genome. *Proc. R. Soc. B Biol. Sci.* 285. doi:10.1098/rspb.2018.1246.

Ravinet, M., Yoshida, K., Shigenobu, S., Toyoda, A., Fujiyama, A., and Kitano, J. (2018b). *The genomic landscape at a late stage of stickleback speciation: High genomic divergence interspersed by small localized regions of introgression.* doi:10.1371/journal.pgen.1007358.

Rieseberg, L. H. (1991). Homoploid Reticulate Evolution in Helianthus (Asteraceae): Evidence from Ribosomal Genes. *Am. J. Bot.* 78, 1218. doi:10.2307/2444926.

Rieseberg, L. H. (1997). Hybrid origins of plant species. *Annu. Rev. Ecol. Syst.* 28, 359–389. doi:10.1146/annurev.ecolsys.28.1.359.

Rieseberg, L. H., Raymond, O., Rosenthal, D. M., Lai, Z., Livingstone, K., Nakazato, T., et al. (2003). Major ecological transitions in wild sunflowers facilitated by hybridization. *Science (80-.).* 301, 1211–1217.

Roesti, M., Gavrilets, S., Hendry, A. P., Salzburger, W., and Berner, D. (2014). The genomic signature of parallel adaptation from shared genetic variation. *Mol. Ecol.* 23, 3944–3956. doi:10.1111/mec.12720.

Roux, C., Fraïsse, C., Romiguier, J., Anciaux, Y., Galtier, N., and Bierne, N. (2016). Shedding

Light on the Grey Zone of Speciation along a Continuum of Genomic Divergence. *PLoS Biol.* 14, 1–22. doi:10.1371/journal.pbio.2000234.

Runemark, A., Trier, C. N., Eroukhmanoff, F., Hermansen, J. S., Matschiner, M., Ravinet, M., et al. (2018). Variation and constraints in hybrid genome formation. *Nat. Ecol. Evol.* 2, 549–556. doi:10.1038/s41559-017-0437-7.

Sætre, G. P., Borge, T., Lindroos, K., Haavie, J., Sheldon, B. C., Primmer, C., et al. (2003). Sex chromosome evolution and speciation in Ficedula flycatchers. *Proc. R. Soc. B Biol. Sci.* 270, 53–59. doi:10.1098/rspb.2002.2204.

Sætre, G. P., Cuevas, A., Hermansen, J. S., Elgvin, T. O., Fernández, L. P., Sæther, S. A., et al. (2017). Rapid polygenic response to secondary contact in a hybrid species. *Proc. R. Soc. B Biol. Sci.* 284. doi:10.1098/rspb.2017.0365.

Salazar, C. A., Jiggins, C. D., Arias, C. F., Tobler, A., Bermingham, E., and Linares, M. (2005). Hybrid incompatibility is consistent with a hybrid origin of Heliconius heurippa Hewitson from its close relatives, Heliconius cydno Doubleday and Heliconius melpomene Linnaeus. *J. Evol. Biol.* 18, 247–256. doi:10.1111/j.1420-9101.2004.00839.x.

Sambatti, J. B. M., Strasburg, J. L., Ortiz-Barrientos, D., Baack, E. J., and Rieseberg, L. H. (2012). Reconciling extremely strong barriers with high levels of gene exchange in annual sunflowers. *Evolution (N. Y).* 66, 1459–1473. doi:10.1111/j.1558-5646.2011.01537.x.

Sankararaman, S., Mallick, S., Dannemann, M., Prüfer, K., Kelso, J., Pääbo, S., et al. (2014). The genomic landscape of Neanderthal ancestry in present-day humans. *Nature* 507, 354–357. doi:10.1038/nature12961.

Schluter, D. (2000). Ecological character displacement in adaptive radiation. *Am. Nat.* 156. doi:10.2307/3079223.

Schumer, M., Cui, R., Powell, D. L., Rosenthal, G. G., and Andolfatto, P. (2016). Ancient hybridization and genomic stabilization in a swordtail fish. *Mol. Ecol.* 25, 2661–2679. doi:10.1111/mec.13602.

Schumer, M., Cui, R., Rosenthal, G. G., and Andolfatto, P. (2015). Reproductive Isolation of Hybrid Populations Driven by Genetic Incompatibilities. *PLoS Genet.* 11. doi:10.1371/journal.pgen.1005041.

Schumer, M., Rosenthal, G. G., and Andolfatto, P. (2014). How common is homoploid hybrid speciation? *Evolution (N. Y).* 68, 1553–1560. doi:10.1111/evo.12399.

Schumer, M., Xu, C., Powell, D. L., Durvasula, A., Skov, L., Holland, C., et al. (2018). Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science (80-.).* 360, 656–660.

Schwarz, D., Matta, B. M., Shakir-Botteri, N. L., and McPheron, B. A. (2005). Host shift to an invasive plant triggers rapid animal hybrid speciation. *Nature* 436, 546–549. doi:10.1038/nature03800.

Schwarzbach, A. E., Donovan, L. A., and Riesberg, L. H. (2001). Transgressive character expression in a hybrid sunflower species. *Am. J. Bot.* 88, 270–277. doi:10.2307/2657018.

Seehausen, O. (2004). Hybridization and adaptive radiation. *Trends Ecol. Evol.* 19, 198–207. doi:10.1016/j.tree.2004.01.003.

Seehausen, O. (2013). Conditions when hybridization might predispose populations for adaptive radiation. *J. Evol. Biol.* 26, 279–281. doi:10.1111/jeb.12026.

Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., et al. (2014). Genomics and the origin of species. *Nat. Rev. Genet.* 15, 176–192. doi:10.1038/nrg3644.

Shaw, K. L., and Mullen, S. P. (2014). Speciation continuum. *J. Hered.* 105, 741–742. doi:10.1093/jhered/esu060.

Song, Y., Endepols, S., Klemann, N., Richter, D., Matuschka, F. R., Shih, C. H., et al. (2011). Adaptive introgression of anticoagulant rodent poison resistance by hybridization between old world mice. *Curr. Biol.* 21, 1296–1301. doi:10.1016/j.cub.2011.06.043.

Stankowski, S., and Ravinet, M. (2021). *Defining the speciation continuum.* doi:10.1111/evo.14215.

Stelkens, R. B., Schmid, C., Selz, O., and Seehausen, O. (2009). Phenotypic novelty in experimental hybrids is predicted by the genetic distance between species of cichlid fish. *BMC Evol. Biol.* 9, 1–13. doi:10.1186/1471-2148-9-283.

Stelkens, R., and Seehausen, O. (2009). GENETIC DISTANCE BETWEEN SPECIES PREDICTS NOVEL TRAIT EXPRESSION IN THEIR HYBRIDS. *Evolution (N. Y).* doi:10.1111/j.1558-5646.2008.00599.x.

Storchová, R., Reif, J., and Nachman, M. W. (2010). Female heterogamety and speciation: Reduced introgression of the z chromosome between two species of nightingales. *Evolution (N. Y).* 64, 456–471. doi:10.1111/j.1558-5646.2009.00841.x.

Svardal, H., Quah, F. X., Malinsky, M., Ngatunga, B. P., Miska, E. A., Salzburger, W., et al. (2020). Ancestral hybridization facilitated species diversification in the lake malawi cichlid fish adaptive radiation. *Mol. Biol. Evol.* 37, 1100–1113. doi:10.1093/molbev/msz294.

Svardal, H., Salzburger, W., and Malinsky, M. (2021). Genetic Variation and Hybridization in Evolutionary Radiations of Cichlid Fishes. *Annu. Rev. Anim. Biosci.* 9, 1–25. doi:10.1146/annurev-animal-061220-023129.

Taylor, E. B., Boughman, J. W., Groenenboom, M., Sniatynski, M., Schluter, D., and Gow, J. L. (2006). Speciation in reverse: Morphological and genetic evidence of the collapse of a three-spined stickleback (Gasterosteus aculeatus) species pair. *Mol. Ecol.* 15, 343–355. doi:10.1111/j.1365-294X.2005.02794.x.

Taylor, S. A., Curry, R. L., White, T. A., Ferretti, V., and Lovette, I. (2014). Spatiotemporally consistent genomic signatures of reproductive isolation in a moving hybrid zone. *Evolution (N. Y).* 68, 3066–3081. doi:10.1111/evo.12510.

Taylor, S. A., and Larson, E. L. (2019). Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nat. Ecol. Evol.* 3, 170–177. doi:10.1038/s41559-018-0777-y.

Trier, C. N., Hermansen, J. S., Sætre, G. P., and Bailey, R. I. (2014). Evidence for Mito-Nuclear and Sex-Linked Reproductive Barriers between the Hybrid Italian Sparrow and Its Parent Species. *PLoS Genet.* 10. doi:10.1371/journal.pgen.1004075.

Turner, T. L., Hahn, M. W., and Nuzhdin, S. V. (2005). Genomic islands of speciation in Anopheles gambiae. *PLoS Biol.* 3, 1572–1578. doi:10.1371/journal.pbio.0030285.

Veller, C., Edelman, N. B., Muralidhar, P., and Nowak, M. A. (2019). Recombination and selection against introgressed DNA. *bioRxiv*, 1–35. doi:10.1101/846147.

Vorzimmer, P. (1969). Darwin, Malthus, and the Theory of Natural Selection. *J. Hist. Ideas* 30, 527–542. doi:10.2307/2708609.

Wang, X., He, Z., Shi, S., and Wu, C. I. (2020). Genes and speciation: Is it time to abandon the biological species concept? *Natl. Sci. Rev.* 7, 1387–1397. doi:10.1093/nsr/nwz220.

Weir, J. T., and Mursleen, S. (2013). Diversity-dependent cladogenesis and trait evolution in the adaptive radiation of the auks (Aves: Alcidae). *Evolution (N. Y).* 67, 403–416. doi:10.1111/j.1558-5646.2012.01786.x.

Wright, S. (1931). Evolution in mendelian populations. *Genetics*, 241–295.

Wright, S. (1932). The Roles of Mutation, Interbreeding, Crossbreeding and Selection in Evolution. *Proc. Sixth Annu. Congr. Genet.* 1, 356–366.

Wu, C.-I. (2001). The genic view of the process of speciation. *Evol. Biol.*, 851–865.

Wu, C.-I., Wang, X., He, Z., and Shi, S. (2020). Replies to the commentaries on the question of 'Is it time to abandon the biological species concept? *Natl. Sci. Rev.* 7, 1407–1409. doi:10.1093/nsr/nwaa116.

Wu, C. I., and Ting, C. T. (2004). Genes and speciation. *Nat. Rev. Genet.* 5, 114–122. doi:10.1038/nrg1269.

Zheng, Y., and Janke, A. (2018). Gene flow analysis method, the D-statistic, is robust in a wide parameter space. *BMC Bioinformatics* 19, 1–19. doi:10.1186/s12859-017-2002-4.

PAPER I

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Intraspecific genomic variation and local adaptation in a young hybrid species

Angélica Cuevas[1] | Mark Ravinet[1,2] | Glenn-Peter Sætre[1] | Fabrice Eroukhmanoff[1]

[1]Department of Biosciences, Centre for Ecological and Evolutionary Synthesis, University of Oslo, Oslo, Norway

[2]School of Life Sciences, University of Nottingham, Nottingham, UK

Correspondence
Angélica Cuevas, Department of Biosciences, Centre for Ecological and Evolutionary Synthesis, University of Oslo, PO Box 1066, N-0316 Oslo, Norway.
Email: a.m.c.pulido@ibv.uio.no

Abstract

Hybridization increases genetic variation, hence hybrid species may have greater evolutionary potential once their admixed genomes have stabilized and incompatibilities have been purged. Yet, little is known about how such hybrid lineages evolve at the genomic level following their formation, in particular their adaptive potential. Here we investigate how the Italian sparrow (*Passer italiae*), a homoploid hybrid species, has evolved and locally adapted to its variable environment. Using restriction site-associated DNA sequencing (RAD-seq) on several populations across the Italian peninsula, we evaluate how genomic constraints and novel genetic variation have influenced population divergence and adaptation. We show that population divergence within this hybrid species has evolved in response to climatic variation, suggesting ongoing local adaptation. As found previously in other nonhybrid species, climatic differences appear to increase population differentiation. We also report strong population divergence in a gene known to affect beak morphology. Most of the strongly divergent loci among Italian sparrow populations do not seem to be differentiated between its parent species, the house and Spanish sparrows. Unlike in the hybrid, population divergence within each of the parental taxa has occurred mostly at loci with high allele frequency difference between the parental species, suggesting that novel combinations of parental alleles in the hybrid have not necessarily enhanced its evolutionary potential. Rather, our study suggests that constraints linked to incompatibilities may have restricted the evolution of this admixed genome, both during and after hybrid species formation.

KEYWORDS
genome evolution, genomic incompatibilities, hybrid constraints, hybrid species, local adaptation, *Passer* sparrows

1 | INTRODUCTION

Hybridization is an evolutionary process that has been increasingly studied in the last decade (Abbott et al., 2013; Marques et al., 2019; Taylor & Larson, 2019). It can have a wide array of consequences, ranging from speciation reversal, reinforcement of prezygotic barriers to gene exchange, adaptive introgression and hybrid speciation.

In particular, hybrid speciation – the formation of new species as a result of hybridization (Mallet, 2007) – can be seen as one of the most creative outcomes of hybridization. Especially the case of homoploid hybrid speciation (HHS) is thought to be rare given that reproductive isolation from the parental species does not automatically derive from differences in ploidy levels. Nevertheless, in the last decade, several compelling cases of HHS have been described

in animals (Abbott et al., 2013; Mallet, 2007; Schumer et al., 2014). Mathematical models have addressed the mechanisms by which hybrid populations develop reproductive isolation from the parental lineages leading to HHS. Some studies suggest that geographic isolation of the hybrid from the parental taxa (Buerkle et al., 2000), as well as the genetic architecture and selection pressures on adaptive loci linked to incompatibility loci, is needed for the development of reproductive isolation from the parental taxa (Comeault, 2018). Other studies argue that HHS can occur solely by the rapid development of reproductive barriers via sorting of genetic incompatibilities (Schumer et al., 2015).

However, most of these theoretical and empirical studies focused on making a case for demonstrating HHS while little focus has been placed on analysing the evolutionary fate and adaptive potential of hybrid species. In the long term, the establishment and success of a homoploid hybrid species only partially depends on the fast evolution of reproductive barriers that isolate it from its parental species and the purging of incompatibilities. Selection should also favour locally adapted allelic combinations to ensure the hybrid's ecological persistence and further adaptation to a potentially variable environment.

Genetic variability in hybrid lineages can be enhanced by the admixture process itself through the generation of heterozygosity at loci that are differentially fixed in the parental species, novel rearrangements of parental ancestry blocks, or the inheritance of parental standing genetic variation (Abbott et al., 2013). These processes can produce genetic variation in the hybrid that later may display a higher evolutionary potential than that found in nonhybrid species. Studies have shown that novel genetic combinations in hybrid lineages can substantially increase phenotypic variation and even lead to adaptive radiations (Keller et al., 2013; Meier et al., 2017; Rieseberg et al., 2003; Selz et al., 2014). However, the evolutionary potential of a hybrid species can be hampered by genetic incompatibilities (i.e., Dobzhansky-Muller incompatibilities - DMIs) inherent to the formation of admixed genomes (Runemark, Trier et al., 2018; Schumer, Cui et al., 2014; Schumer et al., 2018; Trier et al., 2014). Sorting of incompatibilities, originally arising and driving reproductive isolation between parental species, can generate symmetrical incompatibilities isolating the hybrid from both parental species (Buerkle et al., 2000; Schumer et al., 2015). However, this process may also constrain hybrid lineages long after hybridization has occurred, affecting their evolutionary potential (Eroukhmanoff et al., 2017; Runemark, Trier et al., 2018). For instance, selection against DMIs can reduce the availability of variation responsive to adaptive evolution and hence, reduce population divergence and the potential for local adaptation (Runemark, Trier et al., 2018). DMIs and incompatibilities in general often involve alleles at different loci that have never coexisted within the same genome. Accordingly, genomic variation in a hybrid species could be reduced at loci where alleles are differentially fixed between the parents, through purging of incompatible alleles. This may in turn constrain or fix certain genomic blocks through linkage with incompatibility loci and reduce the evolutionary potential in these genomic regions (Runemark, Trier et al.,

2018). Thus, the process of HHS includes both the sorting of incompatibilities and fixation of favourable genetic combinations to generate viable and functional genomes (Rieseberg et al., 2003; Runemark et al., 2018; Schumer, Cui et al., 2014). In this study we aim to provide insights to how admixture may ultimately constrain or facilitate adaptive divergence in a hybrid lineage and how genetic variation is generated and made accessible to selection.

In addition to constraints inherent to admixed genomes, hybrid lineages experience the same challenges as nonhybrid species do. The examination of factors that may mediate population differentiation (i.e., environmental variation or geography) in conjunction with inference regarding the role of drift and selection is therefore crucial to understand population divergence (Prunier et al., 2015; Seeholzer & Brumfield, 2018; Wang, 2013). Heterogeneity in abiotic factors such as climate and geography can determine patterns of population genomic divergence, either through geographic isolation (isolation by distance, IBD) where gene flow is limited due to physical distance and geographic barriers (Meirmans, 2012; Slatkin, 1993; Wang, 2013; Wang & Bradburd, 2014), or through ecological isolation (isolation by environment IBE) (Shafer & Wolf, 2013; Wang & Bradburd, 2014), where individuals locally adapting to divergent habitats remain separated, facilitating genomic differentiation. Specific selective pressures, like those in IBE, could result in differential changes in phenotypic traits that can also contribute to population genomic divergence; a process that in time could lead to isolation by adaptation (IBA) (Edelaar et al., 2012; Nosil et al., 2008).

In the absence of geographic isolation, genetic and phenotypic population divergence can be hampered by gene flow (Hendry & Taylor, 2004; Räsänen & Hendry, 2008; Stuart et al., 2017), limiting local adaptation, although the directionality of causation of these processes is debatable. The opposite process can also occur; local adaptation may constrain gene flow, favouring divergence between populations and even lead to ecological speciation (Gosden et al., 2015; Räsänen & Hendry, 2008; Nosil, 2012). In the specific case of hybrid lineages, it has also been argued that incompatibilities could reduce gene flow between hybrid populations (Bierne et al., 2011), especially when genes under ecological selection are coupled with DMI loci (Seehausen, 2013), which in turn may facilitate local adaptation (Eroukhmanoff et al., 2013; Trier et al., 2014).

In this study we investigate how the homoploid hybrid Italian sparrow (*Passer italiae*) has evolved since its formation. We focus on how constraints and novel genetic variation, linked to admixture, have impacted its genomic evolvability, limiting or favouring its adaptive potential and ultimately its population divergence. The Italian sparrow is a homoploid hybrid species resulting from past hybridization between the house sparrow (*Passer domesticus*) and the Spanish sparrow (*Passer hispaniolensis*) (Hermansen et al., 2014; Trier et al., 2014). This hybridization event probably occurred when the house sparrow spread into Europe alongside agriculture, approximately 6 kyr BP (Elgvin et al., 2017; Hermansen et al., 2011; Ravinet et al., 2018). It is possible that this hybrid species originated through a period of multiple hybridization events (Elgvin et al., 2017; Runemark, Trier et al., 2018) with rapid
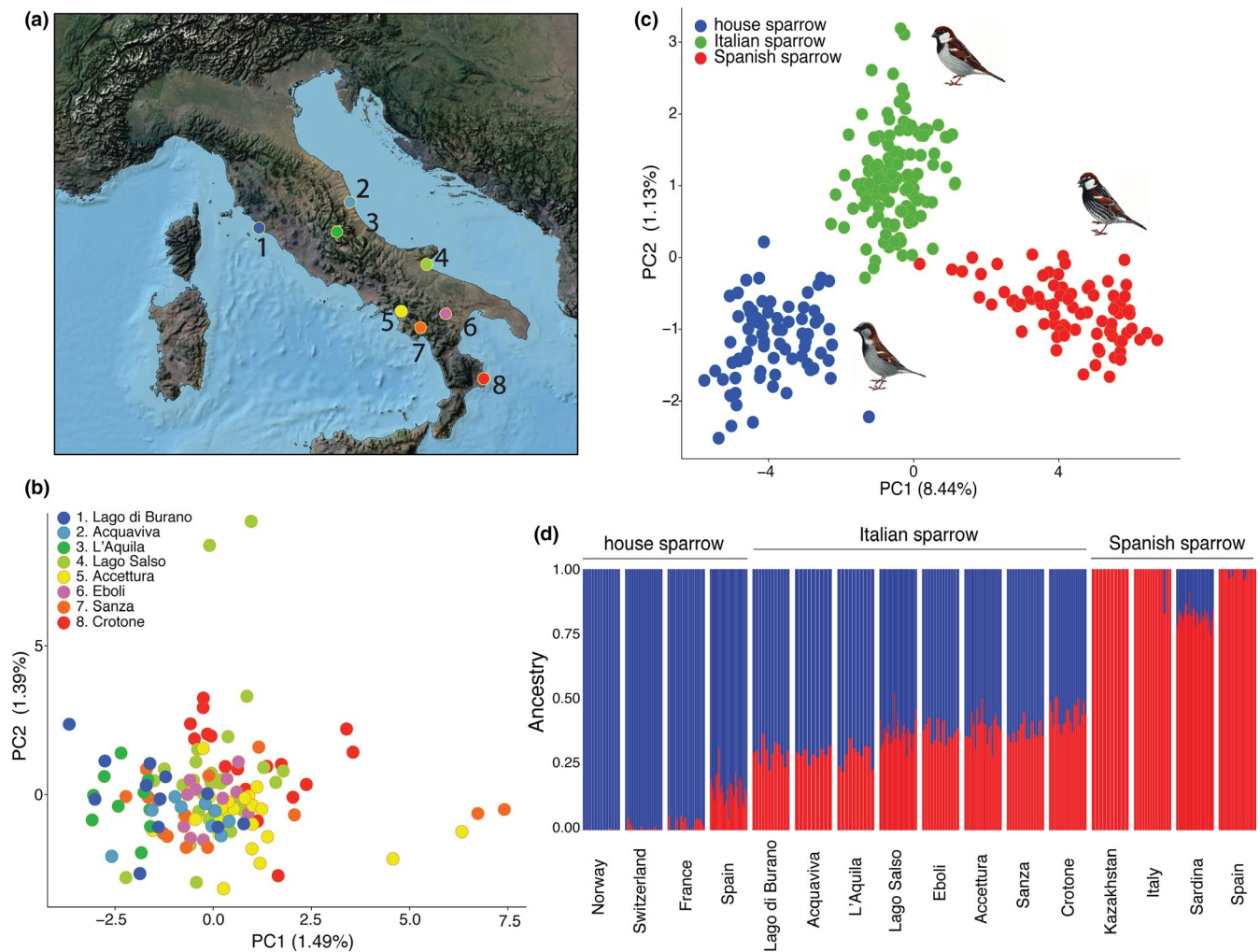
FIGURE 1 (a) Geographic distribution of sampled Italian sparrow populations. (b) Principal component analysis (PCA) to explore genetic variation within the Italian sparrow (eight Italian populations, 131 individuals and 4387 SNPs). (c) PCA assessing the three focal species. Spanish sparrow (red), house sparrow (blue) and Italian sparrow (green). (d) Admixture analysis based on a VCF file containing 288 individuals (131 Italian, 82 Spanish and 75 House sparrows) and 2737 high-quality SNPs. Localities are ordered following latitudinal distribution [Colour figure can be viewed at wileyonlinelibrary.com]

evolution of reproductive barriers from both parental species (Hermansen et al., 2014; Trier et al., 2014), despite some localized ongoing gene flow in parts of Italy and Europe. In mainland Italy the genome is admixed with a slightly higher contribution from the house sparrow (Elgvin et al., 2017). It is reproductively isolated from its parental species, with strong post-zygotic barriers associated with mitonuclear and sex-linked incompatibilities (Elgvin et al., 2017; Trier et al., 2014).

Patterns of population divergence and local adaptation at the genomic level have not yet been investigated in the Italian sparrow, nor the extent to which genomic constraints might have affected population divergence in this species. We limited our study to mainland populations across the Italian peninsula, excluding populations from Mediterranean islands as they are probably influenced by separate, independent hybridization events (Runemark, Trier et al., 2018). We assessed population divergence and the role of climatic variation on genomic divergence. Our results suggest that genetic divergence

within the Italian sparrow is driven by climatic variation. We report patterns of IBE, which appear to be driven primarily by temperature, and identify some outlier loci of adaptive divergence associated with precipitation and beak height variation. To determine the nature of the genomic divergence patterns found in the hybrid species, we examined the ancestry of the hybrid genome and genomic divergence in its parental species. Our results demonstrate that most loci involved in local adaptation in the hybrid species are little differentiated between the parental species, suggesting that incompatibilities may play a role in constraining population divergence. Conversely, loci involved in local adaptation within each parent species seem to have previously been under divergent selection between the parental taxa, which is consistent with the natural history of both species (Ravinet et al., 2018). Overall, genomic divergence and local adaptation seem to be highly polygenic both in the hybrid and the parent species, albeit different loci are involved in adaptive intraspecific divergence.

2 | MATERIALS AND METHODS

2.1 | Study species and sampling

The Italian sparrow is distributed across the Italian peninsula and a few Mediterranean islands. Of its parental species the house sparrow has a wider native distribution, extending throughout large parts of Eurasia, whereas the Spanish sparrow is located around the Mediterranean Sea and eastwards to Central Asia (Summers-Smith, 1988). We concentrated on the mainland distribution of the Italian sparrow sampling several populations across the Italian peninsula.

Birds were caught using mist nets. Blood samples were obtained by puncturing the left brachial vein and stored in standard Queen's lysis buffer. Individuals were released immediately after sampling. All relevant sampling permits were obtained from the regional authorities.

We sampled a total of 131 (68 males and 63 females) Italian sparrows from eight populations across Italy (Figure 1a, Table S1). These populations are geographically well spread representing most of the mainland distribution of the Italian sparrow. In addition, we sampled 82 Spanish sparrows (51 males and 31 females) from Spain, Italy, Kazakhstan and Sardinia and 75 house sparrows (49 males, 26 females) from Norway, Switzerland, Spain and France. Per location between 13 and 27 individuals were sampled (Table S1).

2.2 | DNA extraction and sequencing

Genomic DNA was purified from blood samples using Qiagen DNeasy 96 Blood and Tissue Kits (Qiagen N.V., Venlo, The Netherlands) according to the manufacturer's instructions. The protocol was slightly modified by adding 125 µl of blood stored in Queen's lysis buffer and warming the Qiagen elution buffer (EB) to 40°C to increase yield of DNA. DNA isolates were stored in EB. Double digestion of the genomic DNA for ddRAD sequencing was performed using EcoRI and MseI restriction enzymes following the protocol of Peterson et al. (2012). Genomic DNA was digested and ligated to respective adapters comprising EcoRI and MseI restriction overhangs. Molecular identifier tags were added with PCR amplification. Resulting individual sample libraries were pooled and library pools were size selected for fragments between 500–600 base pairs (bp) with gel electrophoresis and extraction of the respective size range. The size selected library pools were then sequenced using an Illumina Nextseq500 machine and the 1 × 75 bp sequencing format. On average, 2.4×10^6 single reads were produced per sample. Library preparation, sequencing, demultiplex and trimming of the adapters were performed by Ecogenics GmbH (Balgach, Switzerland) (www.ecogenics.ch).

2.3 | Mapping to reference genome and variant calling

RAD sequences were quality checked by FASTQC (Andrews, 2010) and mapped to the house sparrow reference genome, assembled by Elgvin et al. (2017), with BWA-MEM (v 0.7.8) (Li & Durbin, 2009) using the default parameters with the exception of using the -M flag allowing Picard compatibility for further analysis. Bam files were sorted by coordinates using Picardtools (v 1.72) SortSam (https://broadinstitute.github.io/picard/). Identification of indels and local realignment was run using genome analysis tool kit (GATK)'s RealignerTargetCreator and IndelRealigner (McKenna et al., 2010; Auwera et al., 2014) with default parameters. We validated bam files with the Picardtools (v 1.72) ValidateSamFile tool.

From the realigned bam file a set of variants were called by GATK (v 3.7) HaplotypeCaller using the following cut off for filtering: a Phred based mapping quality score of 10, soft clipping of the last 5 bp without the need to soft clip both ends (-rf OverclippedRead --filter_is_too_short_value and --do_not_require_softclips_both_ends). The resulting individual genomic variant files (gVCF) were then combined by CombineGVCFs and merged using the GenotypeGVCFs tools. As our analyses were based on single nucleotide polymorphisms (SNPs), all indels were excluded using the GATK's SelectVariants tool. Variants in unplaced scaffolds were removed using SelectVariants. Individuals with a proportion of missing data greater than 0.75 were excluded at this early stage before further filtering.

SNPs were subsequently filtered by quality using vcftools v. 0.1.14 (Danecek et al., 2011) as follows: proportion of missing data <0.8, genotype quality >20, Depth of coverage >10 and minor allele frequency of 0.02. Finally, nonvariant sites present after filtering and excluding missing-data-individuals, were removed using GATK's SelectVariants with the -env parameter.

After filtering, we obtained a final VCF file including the Italian sparrow and its parental species (288 individuals, 131 Italian, 82 Spanish and 75 house sparrows) containing 2737 high-quality SNPs and with mean proportion of per individual missing data of 0.13. This data set was used to identify genomic divergence among species.

Within-species analyses were conducted using species-specific VCF files by selecting the correspondent samples, merging individual genomic variant files (gVCF) and genotyping using the GenotypeGVCFs and finally recalling variants within species. Filtering was conducted as described above. The Italian sparrow-only VCF file contains 131 individuals and 4387 SNPs from eight localities. VCF files for each parental species were additionally filtered by minor allele frequency of 0.01. The house sparrow-only VCF includes 75 individuals across four localities and 6503 high-quality-SNPs and a Spanish sparrow VCF file with 1320 SNPs across 82 individuals from Spain, Kazakhstan and two localities in Italy; Fontanarosa in the Gargano peninsula and Sardinia. The average proportion of individual missing data for these species-specific VCF files are 0.12, 0.12 and 0.13 for the Italian-only, house-only and Spanish-only files, respectively.

2.4 | Investigating population divergence within the Italian sparrow

To evaluate population structure and divergence in the hybrid species we used an SNP set containing 4387 loci identified across eight

Italian localities (N = 131). We ran admixture analysis and principal component analysis (PCA) using glPca in the R package ADEGENET 2.0 (Jombart, 2008). We used vcftools (Danecek et al., 2011) and PLINK v. 1.9 (Chang et al., 2015) to transform the VCF file into format files (MAP, RAW, PED and BED) required by ADEGENET.

To assess the potential for isolation by distance among these Italian sparrow populations at different locations we used a multiple (and univariate) matrix regression with randomization (MMRR and UMRR, respectively) approach (Prunier et al., 2015; Wang, 2013), correlating geographic distance and genomic divergence (mean pairwise F_{ST}) across all pairwise comparison of Italian sparrow populations. This method is described in the next section.

We used Tajima's D statistics to investigate signals of selection and potentially recent demographic change, which may have occurred post-hybridization. We also calculated mean values of Tajima's D, nucleotide diversity (π) and F_{ST} for the Italian sparrow, using vcftools v. 0.1.14 (Danecek et al., 2011).

To identify regions of divergence in the hybrid species, genome scan analyses were performed across the genome for the eight populations of Italian sparrows. We calculated windowed F_{ST} and nucleotide diversity (π) using a sliding window of 100 kb in size with 25 kb steps. Nucleotide diversity was estimated retaining nonvariant sites and avoiding minor allele frequency filtering. We also calculated Tajima's D on nonoverlapping windows of 100 kb, given that linkage disequilibrium (LD) tends to decay within this distance in sparrows (Elgvin et al., 2017) using a VCF file without minor allele frequency filtering to avoid bias by removing rare variants.

2.5 | Selection, local adaptation and environmental variation

The Italian peninsula varies considerably in climate, thus we investigated whether genomic divergence covaried with environmental variation. Pairwise differences in climatic variables were regressed with the pairwise genetic distance between populations. We analysed five climatic variables obtained from the global climate data server, WorldClim (v. 2.0, http://www.worldclim.org) (Hijmans et al., 2005), BIO1 = annual mean temperature, BIO4 = temperature seasonality (standard deviation × 100), BIO12 = annual precipitation and BIO15 = precipitation seasonality (coefficient of variation). Values were retrieved using the R packages RGDAL (v 1.3-4, Bivand et al., 2017) and SP (v 1.2-4) (Pebesma & Bivand, 2005), with a resolution of 1 km. Geographic distance was obtained with the function spDistsN1 from the R package SP (v 1.2-4) and altitudinal data was gathered from the R package RASTER (v 2.6-7) (Hijmans, 2014) and SP (v 1.2-4) using the getData function. We also analysed phenotypic distance in two beak traits, mean beak height (BH) and beak length (BL), in each population.

To test for associations between environmental factors, geographic, altitudinal and phenotypic distances and genome-wide divergence we used univariate and multiple matrix regression with randomization (UMRR and MMRR respectively) approaches (Wang,

2013) and a modification implemented by Prunier et al. (2015), including commonality analysis (CA) to account for multicollinearity (nonindependence) among environmental factors. Data were Z-transformed (i.e., standardization by subtracting the mean and dividing by the standard deviation) to make regression coefficients of the predictor variables comparable (beta weights, Prunier et al., 2015).

MMRR is a multiple regression analysis on distance matrices used to quantify the contribution of environmental and geographic factors to patterns of genetic divergence (Wang, 2013). It allows the quantification of IBD, IBE and even IBA when a phenotypic variable is included as predictor. One advantage of the method is that it not only resolves whether the dependent and independent variables are correlated but also quantifies the change and directionality (regression coefficients, β_n) that the dependent variable (genomic distance) has with respect to multiple independent variables, i.e., geographic and environmental distances (Wang, 2013). The fit of the model is determined by the coefficient of determination (R^2). Given the nonindependent nature of the variables, the significance (p-values) of the variable's effects (β_n) and fit of the model (R^2) are estimated by randomized permutations of rows and columns of the dependent variable matrix (for more details see Wang, 2013). However, strong multicollinearity among predictors is still a limitation of this approach. Regression coefficients (β_n), fit of the model (R^2) and their significance can be affected by multicollinearity among explanatory variables (Kraha et al., 2012; Nimon & Reio, 2011; Prunier et al., 2015). To overcome this caveat an incorporation of variance-partitioning procedures via commonality analysis (CA) can be used, implemented by Prunier et al. (2015). This method (CA) developed originally by Newton and Spurrell (1967) decomposes the model coefficients into unique (U) and common (C) variance components (Campbell & Tucker, 1992 in Prunier et al., 2015; Nimon & Oswald, 2013), allowing identification of the magnitude of collinearity and the unique (U) effect that a predictor variable has on the dependent variable. The common (C) effect represents the proportion of variance, in the dependent variable, explained by the collinearity of the predictor evaluated and another explanatory variable, while the unique component (U) quantifies the variance explained by the unique effect of the predictor (Prunier et al., 2015).

CA allows determining unique (U) and common (C) contributions of each predictor to the response variable (pairwise F_{ST}) while accounting for collinearity among predictors. The total effect ($T = U + C$) of each predictor corresponds to the total effect that a predictor has to the variance explained by the model, independently of collinearity with other predictors, and the total variation a specific predictor accounts for is determined by T/R^2, which would be a portion of the variation explained by the model.

These methods have been shown to provide a better resolution of the effects of environment, geographic distance and phenotype, allowing us to identify patterns of IBD, IBE and IBA (Seeholzer & Brumfield, 2018). This approach is ideal for our analysis given the nature of our data. We are interested in understanding whether genomic divergence and gene flow within the Italian sparrow is linked to climatic, geographic and phenotypic variation. We ran UMRR

and MMRR with 1000 permutations to estimate significance. We also performed variance-partitioning analysis by CA, 95% coefficient intervals of the commonality coefficient were calculated by bootstrapping 1000 replicates, as implemented by Seeholzer and Brumfield (2018).

We used pairwise geographical distance, altitudinal difference, climate disparity per environmental factor and pairwise mean phenotypic distance as predictor matrices and a genomic distance matrix (pairwise F_{ST}) as the dependent variable. As the number of predictor variables cannot be greater than the number of populations analysed in the MMRR analysis, two models were run. In model 1 only geographic and climate variables were used as predictors, while in model 2, altitude and one of the temperature variables were replaced by the phenotypic variables.

To identify SNP candidate loci under selection we ran an outlier analysis using Bayescan (v. 2.1 – Foll & Gaggiotti, 2008), for the Italian sparrow and its parental species independently. Bayescan is a Bayesian approach based on the multinomial-Dirichlet model that uses differences in allele frequency to identify candidate loci under selection by decomposing F_{ST} coefficients into population (β) and loci (α) components; a reversible-jump MCMC evaluates models with and without selection and calculates posterior probabilities of the parameters under the different models (Foll & Gaggiotti, 2008).

Associations of genomic divergence and environmental (and phenotypic) variation can differ across the genome. Therefore we also evaluated such associations at the locus level (SNP), in the hybrid taxon, performing outlier analyses with BayeScEnv, version 1.1 (de Villemereuil & Gaggiotti, 2015). We used the same environmental variables ran on MMRR as predictors, including beak height and length. BayeScEnv, as Bayescan, is a genome-scan software based on Bayesian inference. To account for population structure it uses the F-model and to control for multiple testing, it returns false discovery rate statistics (posterior error probability [PEP], q-value). This method allows the incorporation of environmental information so that the associations between allele frequencies and environmental variables can be evaluated.

We ran BayeScEnv using the default parameters. As in Bayescan, the parameters β used in the neutral model as well as the locus-specific effect using α are estimated. However, a third model of local adaptation, estimating the parameter g, uses the environmental differentiation information. Significantly associated loci were determined by setting a FDR significance threshold of 5% for the correlation q-value of g (de Villemereuil & Gaggiotti, 2015).

To identify candidate genes associated to local adaptation we used the house sparrow annotation file developed by Elgvin et al. (2017). In the house sparrow linkage decays at approximately 100 kb (Elgvin et al., 2017), thus we selected genes contained in regions at a maximum of 100 kb distance from the Bayescan/BayeScEnv outlier loci. To further assess signatures of selection at the gene level, we identified all genes from the house sparrow annotation file and calculated values of F_{ST}, Tajima's D, Pi and Dxy per-gene across the whole genome. Later we assessed whether our candidate genes showed extreme values of the population statistics in comparison

to the other genes genome-wide. For this per-gene analysis we used WGS data from the house and Spanish sparrow (two populations per species) and three Italian sparrow populations; the data was retrieved from Elgvin et al. (2017) and Ravinet et al. (2018).

2.6 | Investigating genomic constraints to population divergence linked to hybridization

To determine the nature of the genomic divergence patterns in the hybrid species, and how they differ from nonhybrid species, we compared population genomic parameters of the parental species to the Italian sparrow. We also estimated ancestry patterns in the Italian sparrow looking to shed light on the source of the genomic variation found in this hybrid lineage.

To identify how highly divergent loci in the hybrid are distributed, for instance whether they are located in genomic regions of high parent species divergence or not, we selected the top 1% loci with the highest F_{ST} among all eight Italian sparrow populations across the different localities and estimated ancestry as well as hybrid-parent F_{ST} and between-parents (SH) F_{ST} values for these same loci. Similarly, we extracted the top 1% loci with the highest F_{ST} among house sparrow populations and among Spanish sparrow populations and as for the hybrid species, hybrid-parent F_{ST} and between-parents (SH) F_{ST} values were estimated for these highly variable loci. We also compared the observed patterns of Tajima's D between species. As for the hybrid, Tajima's D for the parental species was estimated using VCF files that were not filtered for minor allele frequency.

To evaluate whether loci involved in population divergence within the Italian sparrow correspond to loci of high or low genetic differentiation between the Italian and Spanish (IS F_{ST}) sparrows, Italian and house (IH F_{ST}) sparrows or Spanish and house (SH F_{ST}) sparrows, we performed logistic regressions on the probability of being an Italian F_{ST} outlier. In these models, the outlier status (outlier/nonoutlier) of each locus (SNP) is the response variable, while additive and interaction effects of pairwise F_{ST} between the three species were tested as predictors.

We also used whole genome resequencing (WGS) data from Elgvin et al. (2017) and Ravinet et al. (2018) to estimate ancestry for the Italian sparrow genome. A total of 54 genomes were used, a single population per parental lineage (10 Spanish sparrows from Kazakhstan and 14 house individuals from Norway) and three Italian sparrow populations (Crotone, Guglionesi and Rimini) with 10 genomes per population. Data was phased prior to analysis (see Ravinet et al., 2018) and ancestry estimates were performed using the software LOTER (Dias-Alves et al., 2018), a software package for local ancestry inference (LAI) that uses a copying model based on an optimization problem where switches of parental haplotypes are penalized by the regularization parameter λ. A final ancestry estimate is found by averaging results from different values of λ and several runs of the algorithm. Moreover, this package does not require statistical or biological parameters (i.e., recombination rate) to be specified, making it more accessible to non-model species.

Following this, we identified ancestry estimates for the 4387 RAD loci found across the eight Italian sparrow populations. When it was not possible to identify the ancestry estimate of a specific RAD locus its value was instead taken from the closest identified locus within a 100 kb window. We calculated house sparrow ancestry proportion across all eight populations of the Italian sparrow. Also, given that LOTER assigns a specific ancestry estimate (house or Spanish ancestry) for each haplotype we weighted those estimates using the parental allele frequency difference (AFD), calculated from the WGS data, as a measure of certainty. Thus, the sign of the estimate symbolizes parental ancestry (negative values for house ancestry and positive for Spanish ancestry) and the value represents the degree of AFD between parental species. Values of zero show loci where alleles are segregating equally in the parental lineages, while values of 1 (or –1) occur on loci that are differentially fixed between the parents.

Evolution of recombination rate variation across the genome may have an effect on patterns of differentiation within and among species (Burri et al., 2015; Ortiz-Barrientos et al., 2016; Ortiz-Barrientos & James, 2017). Therefore, we evaluated whether there was a correlation between recombination rate (estimates taken from a linkage map from Elgvin et al., 2017) and genomic differentiation (F_{ST}) among populations for each of the species (house, Spanish and Italian sparrows) respectively.

3 | RESULTS

3.1 | Genomic landscape of population divergence in the Italian sparrow

As found in previous studies (Elgvin et al., 2017; Hermansen et al., 2011) our results support the mosaic nature of the hybrid Italian sparrow genome (Figure 1c, d). To evaluate the genomic variation among populations of the Italian sparrow, we performed a PCA and admixture analyses from eight locations across the Italian peninsula (N = 131 individuals, 4387 SNPs. Figure 1b, Figure S1), covering a wide range of its mainland geographic distribution (Figure 1a). We found no evidence for genome-wide population structure, only moderate among-population clustering.

Estimated parameters of population divergence among Italian sparrows also showed a moderate genome-wide population divergence (mean F_{ST} across all eight localities = 0.013, π = 2.595 × 10^{-6}, Table S2). Nonetheless, it was possible to identify regions of higher divergence in autosomes, with maximum F_{ST} values of ~0.17 across populations and high nucleotide diversity (Figure S2A, C).

Genome-wide average of Tajima's D for the Italian sparrow was negative, as well as for the parental species; however, there is a significant difference between species. Genome-wide average of Tajima's D for the Italian sparrow was negative, as well as for the parental species; however, there is a significant difference between species. 1% F_{ST} outliers between Italian populations had higher nucleotide diversity than the genome-wide average and, interestingly,

TABLE 1 Univariate matrix regression with randomization (UMRR) across eight populations of the Italian sparrow. Pairwise F_{ST} between populations as the response variable

	R^2	β	t	p-value
TEMP.S	0.163	0.007	2.251	.048[*]
A.PREC	0.061	−0.004	−1.302	.260
GEO	0.053	0.004	1.201	.264
BEAK.L	0.036	0.003	0.991	.358
BEAK.H	0.036	0.003	0.991	.341
PREC.S	0.031	0.003	0.910	.422
ALT	0.020	−0.002	−0.732	.513
A.TEMP	0.001	−0.001	−0.172	.864

Note: Predictor variables are as following. ALT: altitude; A.PREC: annual mean precipitation; A.TEMP: annual mean temperature; BEAK.H: beak height; BEAK.L: beak length; GEO: geographic distance; PREC.S: precipitation seasonality; TEMP.S: temperature seasonality.

these loci also showed elevated nucleotide diversity in the parental species, especially in the house sparrow (Table S2).

3.2 | Selection, local adaptation and environmental variation

To further understand the genetic differentiation found among populations of the hybrid we tested patterns of IBD, IBE and IBA using the distances of several climatic factors and phenotypic traits, as well as altitudinal and geographic distances as predictor variables. We ran UMRR and MMRR models (Wang, 2013) and variance partitioning through commonality analyses (CA) (Prunier et al., 2015; Seeholzer & Brumfield, 2018). We found no evidence for IBD in our data set (Tables 1 and 2; Table S3). In UMRR (Table 1) geographical distance (GEO) showed a nonsignificant relationship (R^2 = 0.053, β = 0.004) to genetic differentiation among populations. Its contribution in the multivariate model (MMRR) was nonsignificant (β = 0.003, p = .34) and under the commonality analysis the unique (U = 0.03) and common (C = 0.02) effects were considerably small (Table 2). Isolation by environment (IBE) appeared to be a more determining factor. Results from UMRR and MMRR yielded evidence that climate is driving genetic differentiation within the Italian sparrow, suggesting adaptation to climate (or some unmeasured factor correlate of climate). In particular, temperature seasonality explained a significant proportion of the genetic variation, (Table 1, Figure S3), with a R^2 = 0.163 and β weight of 0.007. The multivariate model including all the climatic factors, altitude and geographic distances as predictors (MMRR – model 1, Table 2), explained 25% of the interpopulation variation in F_{ST} within the Italian sparrow (R^2 = 0.25). Consistent with the results from UMRR, temperature seasonality yielded the highest β weight, with a considerable explanatory power (β = 0.007) (Table 2), accounting for 8% of the variation explained by the model. However, variance partitioning (CA) showed its unique contribution was almost negligible, meaning the interaction with other variables

MODEL 1: F_{ST} ~ GEO + A.TEMP + A.PREC +TEMP.S + PREC.S + ALT						R^2 = 0.25
Predictor	β	t	p-value	Unique (U)	Common (C)	Total (T)
GEO	0.003	0.93	.34	0.03 (12%)	0.02 (8%)	0.05 (20%)
A.TEMP	0.001	0.15	.89	0.14 (56%)	0.03 (12%)	0.16 (64%)
A.PREC	−0.004	−1.16	.32	0.05 (20%)	0.01 (4%)	0.06 (24%)
TEMP.S	0.007	1.96	.05*	0.003 (0%)	0.02 (8%)	0.02 (8%)
PREC.S	−0.003	−0.60	.57	0.001 (0%)	0.00 (0%)	0.00 (0%)
ALT	0.002	−0.29	.78	0.01 (4%)	0.02 (8%)	0.03 (12%)

TABLE 2 Multiple matrix regression with randomization (MMRR) and coefficients from commonality analysis (CA) – MODEL 1. Unique (U), common (C) and total (T) variance partitioning coefficients of each predictor variable to genomic divergence (pairwise F_{ST}), in parentheses the per cent contribution of the predictor to the total variance explained by the model (100 × partition coefficient (U, C or T)/R^2). Pairwise F_{ST} between eight populations of the Italian sparrow as the response variable.

Note: Predictor variables are following. ALT: altitude; A.PREC: annual mean precipitation; A.TEMP: annual mean temperature; GEO: geographic distance; PREC.S: precipitation seasonality; TEMP.S: temperature seasonality.

(collinearity) had a larger effect (U = 0.003, C = 0.2, Table 2, Figure S4).

While mean annual temperature explained a considerable amount of the variance (Table 2) most of it fell into the unique factor (U = 0.14) and its beta weight was nonsignificant (β = 0.001, p = .89). Mean annual precipitation showed similar results (T = 24%, Table 2). This suggests that there is collinearity between climatic factors. Unique (U) and common (C) contributions to the variation, estimated by CA (Table 2, Figure S4), showed mean annual temperature (T = 0.16) and mean annual precipitation (T = 0.06), as the major contributors, accounting for 64% and 24% of the variation explained by the model, respectively (Table 2). However, beta weights for these predictors were not significant. Moreover, when removing mean annual temperature from the model (MMRR – model 2, Table S3) temperature seasonality was no longer significant (p = .1), supporting the collinearity effect among climatic variables.

Finally, evaluating IBA, incorporating beak morphology as predictors, the univariate (UMRR, Table 1) and multivariate (MMRR, Table S3) models showed that these phenotypic traits do not explain a significant amount of the genomic divergence among Italian sparrow populations. The univariate models for each of the beak traits showed a non significant R^2 = 0.036 (p > .34), and in the multivariate model (MMRR - model 2, Table S3) beta weights were low (β = 0.001 for BEAK.H and β = 0.002 BEAK.L) and nonsignificant.

To determine whether highly divergent genomic regions are associated with environmental factors and identify potential genes associated to local adaptation to climate we used a genome scan approach implemented by the software BayeScEnv (de Villemereuil & Gaggiotti, 2015). Five loci were found to be under selection through correlation with environmental variables. On chromosome 5 two outlier loci were associated with mean annual precipitation. One of these displayed values of Tajima's D = −0.833 and F_{ST} = 0.136 among Italian sparrow populations. A locus on chromosome 15 (with values of F_{ST} = 0.172 among Italian populations) was also found to associate significantly with mean annual precipitation (Figure 2a) while presenting high, although nonsignificant, q-values of **g** for mean annual temperature and altitude (Figure S5A and C). Consistently, divergence between species pairs for these loci was low (Table S4). Similarly, chromosome 3 and 2 contained one outlier locus each

(with across Italian localities F_{ST} = 0.050 and F_{ST} = 0.084, respectively) associated to precipitation seasonality (Figure 2b). We also found three candidate loci under selection related to beak morphology, associated with population divergence in beak height (Figure 2c, Table S4).

Further, we used the software Bayescan (Foll & Gaggiotti, 2008) to identify loci under selection across the Italian sparrow populations, independently on whether they are associated to specific environmental factors, phenotypic traits or other unmeasured parameter. We also performed the same analysis in each of the parental species to evaluate whether the hybrid lineage presents similar loci under selection as those in the parental taxa. Three outlier loci were identified as under selection in the Italian sparrow; one locus on chromosome 6, a second locus on chromosome 20 and another in chromosome 15. The latter was previously identified as associated with mean annual precipitation by BayeScEnv (Figure 3a).

Within the putative regions under selection (i.e., 100 kb around the outlier loci) we identified potential genes of interest that may be associated to climatic variation (Table S4). To further assess signatures of selection, specifically at the gene level, we used WGS data from Elgvin et al. (2017) and Ravinet et al. (2018) to calculate per-gene population statistics (F_{ST}, Tajima's D, Pi, dxy; Table S4). The gene GDF5 was identified as a 5% gene-F_{ST} outlier (one-tailed test) presenting a gene-F_{ST} of 0.047 (Table S4). The GDF5 gene, also known as BMP-14, involved in bone and cartilage development, encodes a growth differentiation factor protein related to the BMP (bone morphogenetic protein) gene family (Reddi & Reddi, 2009), a gene family involved in skeletal and jaw development (Bleuming et al., 2007; Cerny et al., 2010; Kaucka & Adameyko, 2019).

In the house sparrow (75 individuals, 6503 SNPs, four localities), eight candidate loci on chromosomes 1, 5 and 8 were inferred to be significantly under selection (Figure 3b, Table S4). Similarly, in the Spanish sparrow (1320 SNPs across 82 individuals from four localities), eight candidate loci (on chromosomes 1, 2, 3 and 5) were also identified using Bayescan (Figure 3c, Table S4).

Only one of the outlier loci was simultaneously identified by both genome scan approaches (Bayescan and BayeScEnv) for the Italian sparrow. The lack of overlapping outlier loci under selection among the three species may be due to differential selective pressures
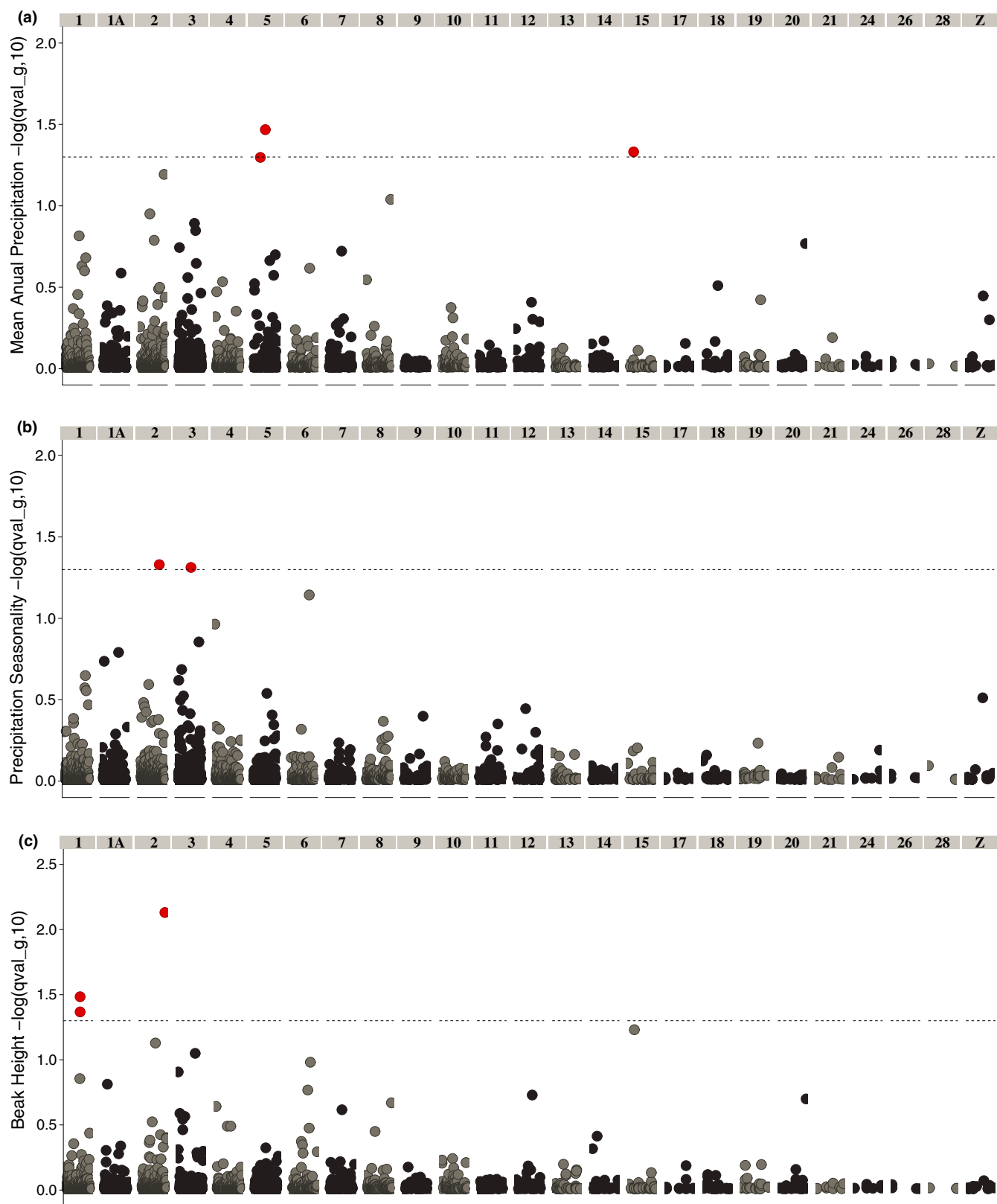
FIGURE 2 Outlier analysis of local adaptation to climate (BayeScEnv). Manhattan plots of correlation *q*-values for genetic divergence (SNPs) within the Italian sparrow showing association to climatic factors and one phenotypic trait. Significance level (FDR-corrected) is set at a *q*-value of <0.05 (−log10 = 1.3). (a) Mean annual precipitation; (b) precipitation seasonality; and (c) beak height [Colour figure can be viewed at wileyonlinelibrary.com]
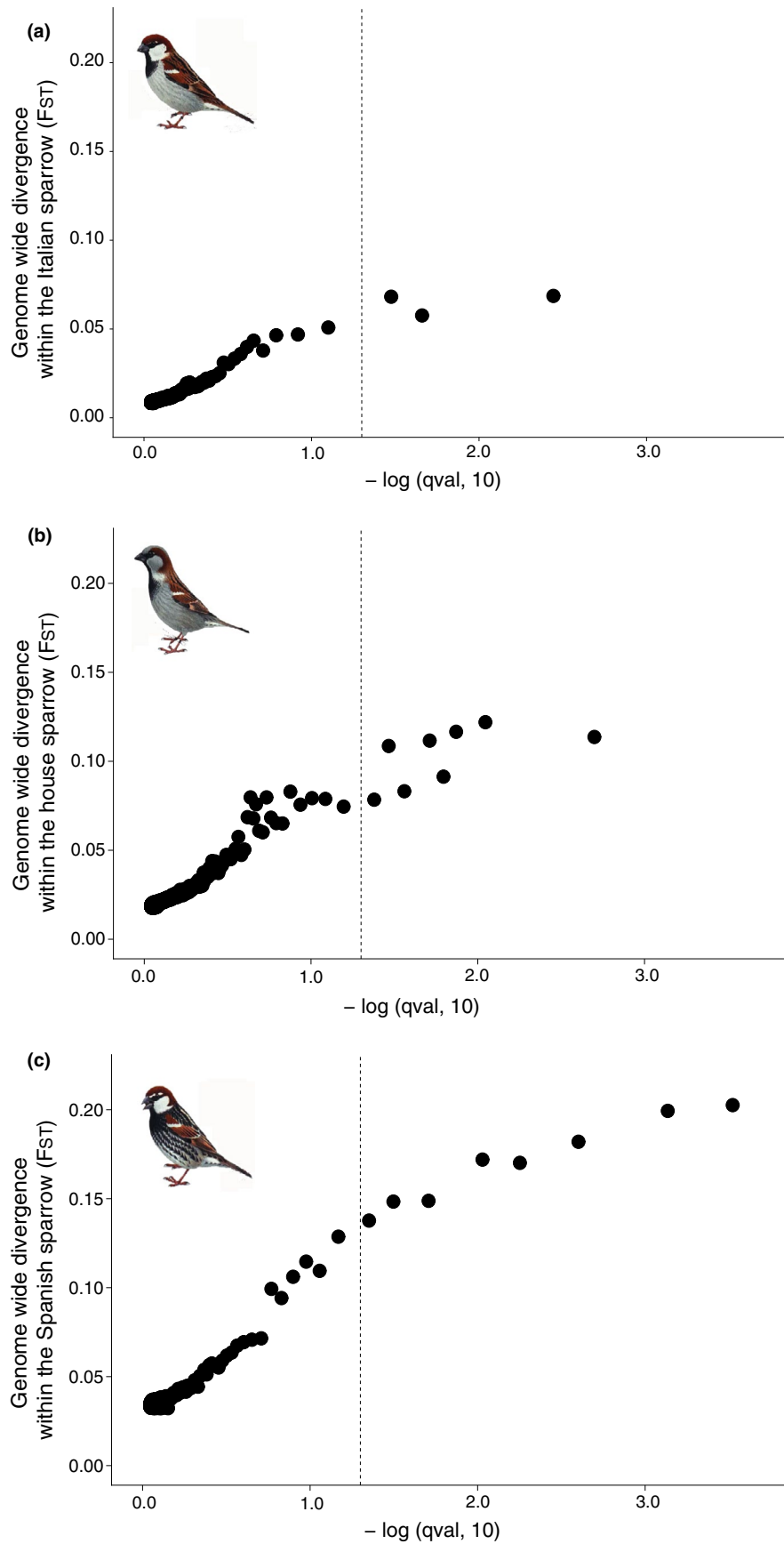
FIGURE 3 Outlier analysis (BayeScan). Correlation *q*-values for genetic divergence (SNPs). Significance level (FDR-corrected) is set at a *q*-value of <0.05 (−log10 = 1.3). (a) Within the Italian sparrow; (b) the house sparrow; and (c) the Spanish sparrow. [Colour figure can be viewed at wileyonlinelibrary.com]
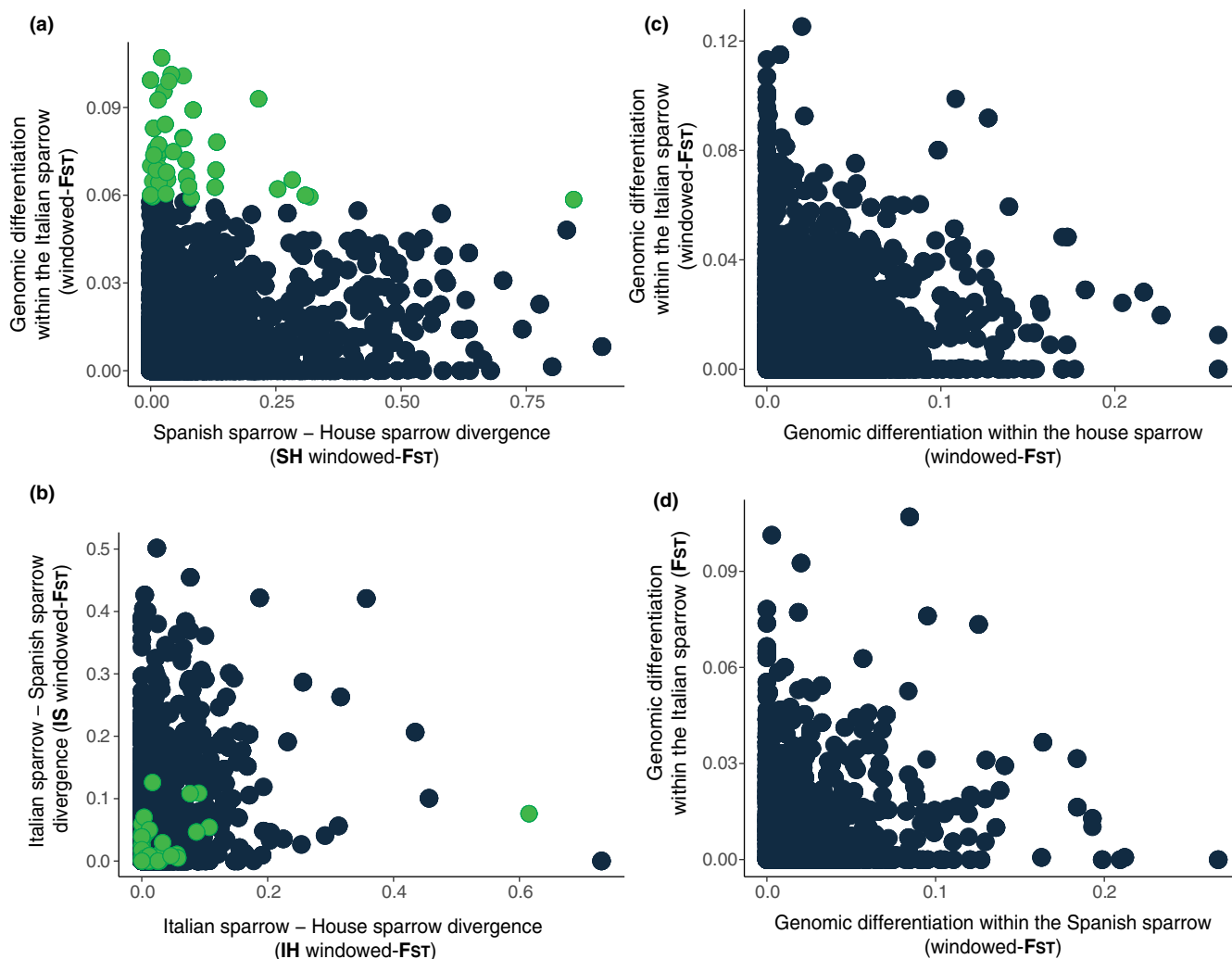
FIGURE 4 Hybrid constraints to population divergence. (a) Genomic differentiation within the Italian sparrow (windowed-F_{ST}) and divergence of its parental species (SH windowed-F_{ST}). (b) Genomic divergence of the Italian sparrow and each of its parental species (Italian–house sparrow divergence [IH windowed-F_{ST}] and Italian–Spanish sparrow divergence [IS windowed-F_{ST}]), with highlighted within-Italian-sparrow-F_{ST} outliers in green. Genomic differentiation within the Italian sparrow vs genomic differentiation within each of the parental species. (c) The house; and (d) the Spanish sparrows [Colour figure can be viewed at wileyonlinelibrary.com]

acting in the hybrid and its parental species. However, further work specifically investigating these loci is necessary to properly assess the role of selection in generating this pattern.

3.3 | Hybrid constraints to population divergence

We compared population genomic parameters between the house, Spanish and Italian sparrows and estimated ancestry of the hybrid loci to determine whether genomic constrains are playing an important role in the genomic divergence of the hybrid species or whether genomic variation, boosted by the hybridization event, facilitates population structuring. We also looked to identify differences in genetic variation patterns between the hybrid and its parent species.

Population divergence in the house sparrow, with a maximum value of F_{ST} = 0.33 across all chromosomes (mean F_{ST} = 0.019) and mean nucleotide diversity of π = 2.996 × 10^{-6} (Figure S2D, F, Table

S2) was similar to that in the Spanish sparrow (mean F_{ST} = 0.021, π = 1.642 × 10^{-6}), with a maximum F_{ST} of 0.34 (Figure S2G, I, Table S2). In contrast, divergence in the Italian sparrow was lower, with a maximum F_{ST} value of ~0.17 (mean F_{ST} = 0.013, π = 2.595 × 10^{-6}; Figure S2A, C, Table S2).

Ancestry estimates showed that the probability of being an Italian F_{ST} outlier was not related to the ancestry proportion across Italian sparrow populations (logistic regression estimate = 0.229, p = .56, Figure 5a). However, Italian outlier loci (1% F_{ST} outliers) segregated for alleles from both parents, as most of the genome presents a mosaic pattern (Figure 5b, Figures S6 and S7). Yet, outlier positions showed low allele frequency differentiation (AFD) between the parents, as the majority of weighted ancestry values for outlier loci were around zero (Bartlett's test of homogeneity of variances using absolute values of weighted ancestry: χ^2 = 806789.041, p = .00, Figure 5b), also supporting the low SH F_{ST} values in highly divergent loci in the Italian sparrow, in comparison with the parental
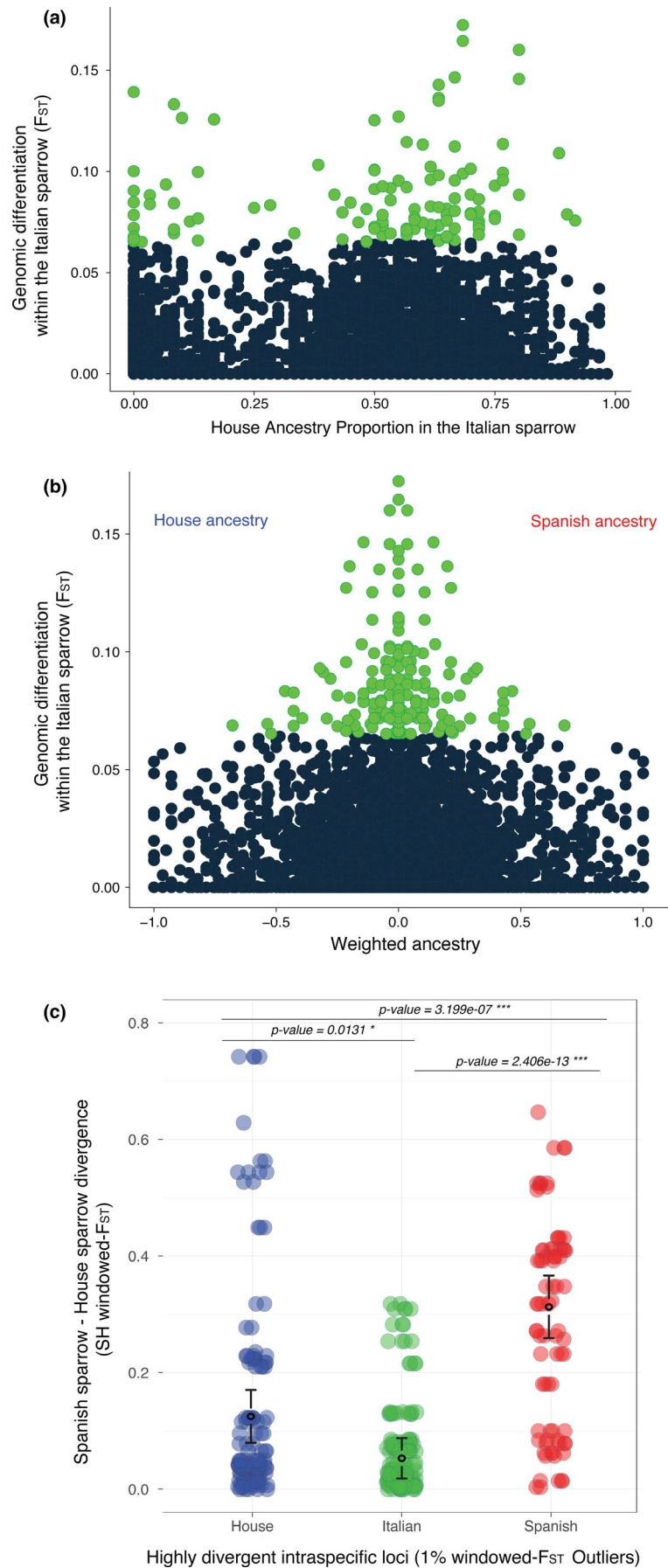
taxa (Figures 4a and 5c). In contrast, inherited parental blocks that are differentially fixed (regions with weighted ancestry values of 1 or –1 and with high values of between-parent-species-differentiation [SH F_{ST}]) showed lower levels of genetic differentiation within the Italian sparrow (Figures 4a and 5b).

Moreover, in the additive model, where both comparisons of the hybrid and each of the parent species are evaluated (IH F_{ST} + IS F_{ST}), the probability of being an F_{ST} outlier within the Italian sparrow decreased with Italian-Spanish (IS F_{ST}) genetic divergence (Table 3, $p = .0127$). A negative, yet nonsignificant, correlation was also found between the highly divergent regions within the hybrid species and between parental species genetic divergence (SH F_{ST}, $p = .0926$) (Table 3).

Additionally, none of the highly divergent regions within the hybrid lineage differed substantially from both of the parental species simultaneously, indicating that private alleles do not account for most of the population differentiation in the hybrid species (Figure 4b). Furthermore, the majority of private alleles have extremely low frequencies and were removed from the analysis when applying MAF filtering.

In contrast to the patterns found for the highly divergent regions in the Italian sparrow, 1% F_{ST} outliers within each of the parental species present high parental genomic divergence (high SH F_{ST} values, Figure 5c). The 1% outlier loci of within house sparrow F_{ST} showed higher divergence between the parental species Spanish-House (SH F_{ST}) than those within the hybrid species, and the same pattern was found for the Spanish sparrow (Figure 5c). Furthermore, highly divergent loci within each of the parental species did not correspond to those found within the hybrid Italian sparrow (Figure 4c, d).

We found some evidence suggesting that recombination rate could explain part of the genomic divergence pattern found within the Italian sparrow ($R^2 = 0.00085$, $p = .033$) and within the Spanish sparrow ($R^2 = 0.003211$, $p = .026$). However, despite significance, extremely low level of variation in divergence between loci is explained by recombination rate (exemplified by the low R^2, 0.3% at most). For the house sparrow, there was no significant correlation ($R^2 = -9.67e{-}07$, $p = .319$) (Figure S8).

We also found an overall higher proportion of negative genome-wide Tajima's D in the Italian sparrow (Figure S2B) as well as in the house (Figure S2E) and Spanish sparrows (Figure S2H). However, the hybrid species differed significantly from the parental species (Table S2).

4 | DISCUSSION

Little is known about how a newly formed hybrid species evolves beyond just a handful of generations. The majority of genomic variation

in a hybrid lineage will be derived from admixture, standing genetic variation inherited from the parental species and novel mutations after hybridization. This variation may ultimately facilitate rapid divergence, whereas genetic incompatibilities may constrain hybrid genome evolution (Runemark, Trier et al., 2018), including their potential for local adaptation. Purging of incompatibilities can remove adaptive variation in regions in physical linkage to DMIs (Schumer et al., 2018). In this study we investigated the extent to which populations of a relatively young hybrid lineage have diverged in response to climatic variation. We further investigated to what extent divergence in the hybrid occurs at loci where variation is generated by admixture itself, in turn fuelling local adaptation.

4.1 | Population divergence in the Italian sparrow

We found moderate, but significant genome-wide population divergence, in line with what has been previously found using neutral markers (Eroukhmanoff et al., 2013), and consistent with ongoing gene flow between populations of Italian sparrows across the Italian peninsula, although other scenarios could also explain this pattern. The young age of this hybrid lineage, thought to be of approximately 6000 years (Hermansen et al., 2011; Ravinet et al., 2018), may explain this pattern as there may not have been sufficient time for populations to strongly diverge. Given the hybrid nature of the Italian sparrow, genomic constraints may also be an important factor in its evolution, hampering population divergence. Consistently, we found negative values of Tajima's D suggesting that regions in the genome are experiencing purifying selection, potentially linked to purging of incompatibilities. Nonetheless, genetic variation may also have been maintained by balancing selection, as we found regions harbouring high nucleotide diversity and loci exhibiting high divergence among populations, suggesting that there is room for variation in the hybrid genome. Also, variation in recombination rate could in part explain some of genomic differentiation identified.

Interestingly, this general pattern of differentiation was comparable but somewhat lower than the pattern of population divergence (F_{ST}) we report for within each of the parent species. Yet, it is difficult to draw further conclusions on the within-species divergence in the parental lineages since the populations sampled are separated by greater geographic distances than those of the hybrid species, which probably affects relative divergence.

Tajima's D differed between the hybrid and the parental lineages; however, all three species exhibited a negative genome-wide average. In the house sparrow, this result supports recent work demonstrating a population expansion about 6 kya (Ravinet et al., 2018). A negative Tajima's D in the Italian sparrow

TABLE 3 Logistic regression on the probability to be an Italian sparrow F_{ST} outlier. Top 1% intraspecific F_{ST} outlier loci selected from a vcf file including the three focal species (131 Italian, 82 Spanish and 75 house sparrows). Outlier loci were identified in a data set of 2,737 shared SNPs between the three species. Outlier status (Italian F_{ST} outlier) used as response variable. F_{ST} outlier threshold = 0.06275, Genomic divergence between parental species (Spanish–House [SH F_{ST}]) and between the hybrid lineage and each of its parents (Italian – House (IH F_{ST}), Italian–Spanish [IS F_{ST}]), additive and interaction effects, are used as predictors

Model	Predictor	Parameter estimate	SE	p-Value
Italian F_{ST} outlier ~SH F_{ST}	SH F_{ST}	–2.1391	1.2719	.0926
~ IH F_{ST} +IS F_{ST}	IH F_{ST}	1.9886	2.4321	.4135
	IS F_{ST}	–7.6170	3.0571	.0127[*]
~ IH F_{ST} × IS F_{ST}	IH F_{ST}	3.2764	2.6091	.2092
	IS F_{ST}	–5.9250	3.5668	.0967
	IH F_{ST}: IS F_{ST}	–49.4795	66.2468	.4551

could also suggest recent population expansion that could mask the high nucleotide diversity expected from the hybridization event itself. Tajima's D in the Italian sparrow has been found to be negative overall and positive values were mostly located in regions of novel divergence, putatively under balancing selection (Elgvin et al., 2017).

Loci of high differentiation among Italian sparrow populations had higher nucleotide diversity in the parental species than their corresponding genome-wide average. One explanation could be that hybrid genetic variation has its origin in standing genetic variation inherited from the parental species, maintained by balancing selection and divergent natural selection following hybridization, leading to population differentiation possibly through the selection of variants playing a role in local adaption to climate (Guerrero & Hahn, 2017). Tajima's D in these outlier loci was negative, yet higher than the genome-wide average. However, there are a variety of processes, including demography, purifying selection and the break-up of parental blocks, that can have cofounding effects on the Tajima's D patterns observed in these regions, therefore, it is difficult to conclusively identify the processes that could have generated this pattern.

4.2 | Selection, local adaptation and environmental variation

Assessing genomic patterns across a spatially heterogeneous distribution, in correlation with factors that can play a role in genomic divergence, can help us elucidate the processes that have determined population differentiation in hybrid lineages. It can also give insights to the adaptive potential of the species (local adaptation and gene flow reduction) or whether genomic differentiation is essentially a result of genetic drift, where patterns of genetic variation are shaped by low gene flow (Prunier et al., 2015; Seeholzer & Brumfield, 2018; Wang, 2013).

To assess adaptive divergence and gene flow, we evaluated IBE, IBA through beak divergence and IBD. We did not find evidence for IBD or IBA, but the significant correlation between genetic distance and climatic variation is consistent with IBE. Our results suggest that

climatic differences, with temperature as the main factor, probably contribute to reduced gene flow between populations in the Italian sparrow, possibly as a result of local adaptation. Previously, precipitation has been found to correlate with beak morphology variation in this species (Runemark, Fernández et al., 2018), and could indirectly be mediating gene flow between phenotypically divergent populations (Eroukhmanoff et al., 2013). Differential changes in phenotypic traits responding to selective pressures can have an effect on local adaptation that may sometimes lead to IBA (Edelaar et al., 2012). However, when directly evaluating beak trait variation as a predictor of overall genomic differentiation among populations of the Italian sparrow we did not find evidence for IBA.

Patterns of adaptive divergence with ongoing gene flow have also been extensively reported in species of nonhybrid origin (de Leon et al., 2010; Marques et al., 2016; Martin et al., 2013; Raeymaekers et al., 2017), which suggests that despite the possibility of constraints reducing the evolvability of this hybrid species (Runemark, Trier et al., 2018), there is also potential for adaptive divergence leading to local adaptation, as in nonhybrid lineages. In fact, theory suggests that incompatibilities could facilitate local adaptation by the coupling of genes under ecological selection and DMI loci (Seehausen, 2013). For example, if genomic incompatibilities become trapped in environmentally divergent habitats, coupling with loci involved in local adaptation may occur, which could potentially facilitate diversification within the hybrid lineage (Abbott et al., 2013; Bierne et al., 2013; Butlin & Smadja, 2018; Seehausen, 2004). This coupling mechanism, more prone to arise in hybrid lineages around regions of interspecific incompatibilities, could facilitate rapid local adaptation in comparison to other processes of diversifying selection in nonhybrid species (Eroukhmanoff et al., 2013; Seehausen, 2013). To the best of our knowledge, there are no empirical studies that report such linkage between DMIs and regions under natural selection. However, our results and previous studies (e.g., Runemark, Trier et al., 2018) show that genomic constraints play an important role in the formation of the admixed Italian sparrow genome.

Here, we present for the first time direct evidence for the role that environmental variation has in mediating genomic variation in a hybrid species, a phenomenon well described in nonhybrid species

(Wang & Bradburd, 2014). We also report loci where high levels of adaptive genetic differentiation has occurred, some of which are co-varying directly with climate variation, suggesting that they are situated in genomic regions linked to local adaptation. For example on chromosome 20 an outlier locus for adaptive divergence between Italian sparrow populations (via Bayescan) was found to be in the vicinity of the GDF5 gene (growth differentiation factor 5, also known as BMP14 (NCBI), a gene also identified as a 5% F_{ST} outlier in the per-gene analysis based on whole genome-resequencing data. This gene is known to be involved in jaw development in vertebrates (Bleuming et al., 2007; Cerny et al., 2010; Kaucka & Adameyko, 2019) and related to the BMP (bone morphogenic protein) gene family (Buxton et al., 2001; Francis-West, Abdelfattah et al., 1999; Francis-West et al., 1999). The BMP gene family has a fundamental role in cra-niofacial development and beak shape and size variation in Darwin's finches (Abzhanov et al., 2004; Lamichhaney et al., 2016).

The beak is a trait known to be under strong selective pressure (Lamichhaney et al., 2015, 2016). Beak size has been shown to be a crucial trait underlying the survival of Darwin's finches after a drought (Lamichhaney et al., 2016) and beak traits in general act as drivers of major evolutionary shifts in Darwin's finches (Almén et al., 2016; Chaves et al., 2016; Lamichhaney et al., 2015, 2016). Beak shape variation has been found to respond to environmental divergence affecting food availability in the medium ground finch (*Geospiza fortis*) (Grant & Grant, 2003, 2014). Thus, climatic factors could be considered a reasonable proxy for food availability in spar-rows (Runemark, Fernández et al., 2018). It is possible that diver-gence of genes associated with beak morphology may reflect an adaptive response to variation in food resources found in environ-mentally different habitats. However, further analyses need to be conducted in order to determine the true underlying mechanisms of divergence between population both at the genetic and phenotypic level.

4.3 | Hybrid constraints to population divergence

Evaluating patterns of ancestry and divergence in the hybrid genome can provide important insights on whether population differentia-tion is facilitated by novel genetic variation or hampered by genomic constraints linked to hybrid incompatibilities. Genomic variation within a hybrid lineage can be generated by novel genetic combi-nations through rearrangements of parental blocks, potentially gen-erating novel epistatic interactions, or through heterozygosity at parental divergent loci. In this case, highly differentiated loci within the hybrid taxon can expected to be located in regions where the parental species have diverged strongly. On the other hand, negative epistatic interactions between inherited parental blocks (in particu-lar if these interactions involve genetic incompatibilities) may lead to strong stabilizing selection on loci fixed for compatible alleles and, through linkage disequilibrium, on other loci situated in their vicinity. Thus, inherited parental genomic blocks would be expected to be

highly conserved, as these are more likely to harbour candidate loci for genetic incompatibilities. This type of genetic constraint on hy-brids could reduce the evolutionary potential of the hybrid species to diverge at the population level. However, this may depend on vari-ation in recombination rate across the genome, which is also known to affect the extent of purging (Schumer et al., 2018) and population divergence (Burri et al., 2015).

We found that a large proportion of the hybrid genome presents a mosaic pattern where polymorphic sites seem to be generated either by the inheritance of differential parental alleles or standing genetic variation already present in the parents. We also report that genetic variation present in loci that are not divergent between pa-rental species accounts for most of the high genomic differentiation found within the hybrid at the population level and that some of this variation may play a role in local adaptation. Furthermore, loci where the parent species are fixed for different alleles or have highly di-vergent allele frequencies seem to be preferentially fixed for one parental allele across Italian sparrow populations (also evidenced by Runemark, Trier et al., 2018). This supports the hypothesis of con-straints biasing evolution to loci that are not differentiated between the parental species and hence are less likely to be incompatible, although we note in this case we have no direct evidence that such loci have any fitness effects on hybrids.

Despite the potentially constrained nature of the hybrid genome, the Italian sparrow has been able to diverge and locally adapt as a response to environmental variation. Our results on ancestry esti-mates suggest that inheritance of parental standing genetic variation is a plausible source of the genetic divergence found in the hybrid species. This variation could be neutral in the parental species, as it seems to not be involved in population divergence in either parent species. Additionally, genomic variation generated in the hybrid (i.e., private alleles) does not seem to contribute to population structure.

Interestingly, patterns of population divergence within the hy-brid taxon and each of its parental species seem to differ, suggest-ing that the admixed nature of the hybrid species may be somewhat restricted compared to its nonhybrid parental species. In contrast to the hybrid species, intraspecific genomic variation in the pa-rental lineages is located mainly in regions of parental divergence. Additionally, there is no overlap of outlier loci under selection among the three species. This could suggest that differential selective pres-sures may be operating in addition to specific genomic constraints in the hybrid species. However, an important factor to be considered in admixed genomes is the inheritance of traces of different evolu-tionary histories as well as the individual evolutionary path that the hybrid species has taken since its formation (and eventual further introgression with parent species). Thus, processes other than dif-ferential selective pressures could generate this pattern.

These results provide a new perspective on how hybridization may impact adaptive evolution, more specifically on how novel ge-nomic variation evolves and is utilized in a hybrid lineage post hybrid speciation, not only through genomic rearrangements linked to ad-mixture and incompatibilities.

4.4 | Conclusions

Genetic variation within the Italian sparrow appears to be driven by climatic variation, temperature being the main factor; we find evidence for isolation by environment (IBE), which could facilitate ongoing local adaptation. Our study supports previous findings suggesting that local adaptation nonetheless can occur, albeit in a biased and constrained manner. Indeed, genetic differentiation in the hybrid species is mainly found in loci that are not divergent between the parental species and hence possibly less prone to be incompatible in the hybrid. This suggests that purging of incompatibilities could be an important element in the evolution of this species. Standing genetic variation inherited from the parental species is a likely explanation for much of the genomic variation in the hybrid species, and some of the variation may be involved in subsequent local adaptation. In contrast, we find little or no evidence that novel variation (private alleles - new mutations occurring after HHS) has been important in local adaptation. Coupling of incompatibilities and loci under natural selection may also have facilitated the rapid genomic divergence observed in the Italian sparrow and its effect on gene flow. However, studies addressing these hypotheses directly are necessary to assess causality.

AUTHOR CONTRIBUTIONS
A.C., and F.E. designed the study; A.C., F.E., and M.R. analysed the data; A.C. conducted laboratory work; A.C., F.E., and G-P.S. collected field data; A.C. wrote the manuscript. F.E., M.R., and G-P.S. contributed and commented on earlier drafts of the manuscript.

DATA AVAILABILITY STATEMENT
Genomic data produced in this study has been deposited at the NCBI Sequence Read Archive under BioProject PRJNA680598 BioSample accessions numbers SAMN16886216- SAMN16886520 (raw RADseq reads in fastq format). VCF files, scripts to process the genomic data as well as scripts used for the statistical analysis and other final data set generated have been deposited in the Dryad Digital Repository at https://doi.org/10.5061/dryad.q573n5th7.

ORCID
Angélica Cuevas https://orcid.org/0000-0003-0862-4582
Mark Ravinet https://orcid.org/0000-0002-2841-1798

REFERENCES

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

Abzhanov, A., Protas, M., Grant, B. R., Grant, P. R., & Tabin, C. J. (2004). Bmp4 and morphological variation of beaks in Darwin's finches. *Science*, *305*, 1462–1466. https://doi.org/10.1126/science.1098095.

Almén, M. S., Lamichhaney, S., Berglund, J., Grant, B. R., Grant, P. R., Webster, M. T., & Andersson, L. (2016). Adaptive radiation of Darwin's finches revisited using whole genome sequencing. *BioEssays*, *38*(1), 14–20. https://doi.org/10.1002/bies.201500079.

Andrews, S. (2010). *FastQC: A quality control tool for high throughput sequence data*. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Bierne, N., Gagnaire, P., & David, P. (2013). The geography of introgression in a patchy environment and the thorn in the side of ecological speciation. *Current Zoology*, *59*(1), 72–86. https://doi.org/10.1093/czoolo/59.1.72

Bierne, N., Welch, J., Loire, E., Bonhomme, F., & David, P. (2011). The coupling hypothesis: Why genome scans may fail to map local adaptation genes. *Molecular Ecology*, *20*(10), 2044–2072. https://doi.org/10.1111/j.1365-294X.2011.05080.x.

Bivand, R., Keitt, T., Rowlingson, B., Pebesma, E., Sumner, M., Hijmans, R., & Rouault, E. (2017). *rgdal: Bindings for the "Geospatial" data abstraction library*. R Package Version 1.3-4. https://www.rdocumentation.org/packages/rgdal

Bleuming, S. A., He, X. C., Kodach, L. L., Hardwick, J. C., Koopman, F. A., ten Kate, F. J., van Deventer, S. J. H., Hommes, D. W., Peppelenbosch, M. P., Offerhaus, G. J., Li, L., & van den Brink, G. R. (2007). Bone morphogenetic protein signaling suppresses tumorigenesis at gastric epithelial transition zones in mice. *Cancer Research*, *67*(17), 8149–8155. https://doi.org/10.1158/0008-5472.CAN-06-4659.

Buerkle, C. A., Morris, R. J., Asmussen, M. A., & Rieseberg, L. H. (2000). The likelihood of homoploid hybrid speciation. *Heredity*, *84*, 441–451.

Burri, R., Nater, A., Kawakami, T., Mugal, C. F., Olason, P. I., Smeds, L., & Ellegren, H. (2015). Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of Ficedula flycatchers. *Genome Research*, *25*(11), 1656–1665. https://doi.org/10.1101/gr.196485.115.

Butlin, R. K., & Smadja, C. M. (2018). Coupling, reinforcement, and speciation. *The American Naturalist*, *191*(2), 155–172. https://doi.org/10.1086/695136.

Buxton, P., Edwards, C., Archer, C., & Francis-West, P. (2001). Growth/Differentiation Factor-5 (GDF-5) and Skeletal Development. *The Journal of Bone & Joint Surgery*, *83-A*, S1–23.

Campbell, K., & Tucker, M. (1992). The use of commonality analysis in multivariate canonical correlation analysis. Annual Meeting of the Southwest Educational Research Association, Houston, Texas.

Cerny, R., Cattell, M., Sauka-spengler, T., Bronner-fraser, M., Yu, F., & Medeiros, M. (2010). Evidence for the prepattern / cooption model of vertebrate jaw evolution. *Proceedings of the National Academy of Sciences*, *107*, 17262–17267.

Chang, C. C., Chow, C. C., Tellier, L. C. A. M., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, *4*(7), https://doi.org/10.1186/s13742-015-0047-8.

Chaves, J. A., Cooper, E. A., Hendry, A. P., Podos, J., De León, L. F., Raeymaekers, J. A. M., MacMillan, W. O., & Uy, J. A. C. (2016).

Genomic variation at the tips of the adaptive radiation of Darwin's finches. *Molecular Ecology*, 25(21), 5282–5295. https://doi.org/10.1111/mec.13743.

Comeault, A. A. (2018). The genomic and ecological context of hybridization affects the probability that symmetrical incompatibilities drive hybrid speciation. *Ecology and Evolution*, 8(5), 2926–2937.

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

de Leon, L. F., Bermingham, E., Podos, J., & Andrew, P. H. (2010). Divergence with gene flow as facilitated by ecological differences : within-island variation in Darwin ' s finches. *Philosophical Transactions of the Royal Society B*, 365, 1041–1052. https://doi.org/10.1098/rstb.2009.0314.

de Villemereuil, P., & Gaggiotti, O. E. (2015). A new FST-based method to uncover local adaptation using environmental variables. *Methods in Ecology and Evolution*, 6(11), 1248–1258. https://doi.org/10.1111/2041-210X.12418.

Dias-Alves, T., Mairal, J., & Blum, M. G. B. (2018). Loter: A software package to infer local ancestry for a wide range of species. *Molecular Biology and Evolution*, 35(9), 2318–2326. https://doi.org/10.1093/molbev/msy126.

Edelaar, P., Alonso, D., Lagerveld, S., Senar, J. C., & Björklund, M. (2012). Population differentiation and restricted gene flow in Spanish crossbills: Not isolation-by-distance but isolation-by-ecology. *Journal of Evolutionary Biology*, 25(3), 417–430. https://doi.org/10.1111/j.1420-9101.2011.02443.x.

Elgvin, T. O., Trier, C. N., Tørresen, O. K., Hagen, I. J., Lien, S., Nederbragt, A. J., Ravinet, M., Jensen, H., & Sætre, G.-P. (2017). The genomic mosaicism of hybrid speciation. *Science Advances*, 3(6), e1602996. https://doi.org/10.1126/sciadv.1602996

Eroukhmanoff, F., Bailey, R. I., Elgvin, T. O., Hermansen, J. S., Runemark, A. R., Trier, C. N., & Sætre, G.-P. (2017). Resolution of conflict between parental genomes in a hybrid species. *BioRxiv*, 102970, 1–48. https://doi.org/10.1101/102970

Eroukhmanoff, F., Hermansen, J. S., Bailey, R. I., Sæther, S. A., & Sætre, G. P. (2013). Local adaptation within a hybrid species. *Heredity*, 111(4), 286–292. https://doi.org/10.1038/hdy.2013.47.

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(2), 977–993. https://doi.org/10.1534/genetics.108.092221.

Francis-West, P. H., Abdelfattah, A., Chen, P., Allen, C., Parish, J., Ladher, R., & Archer, C. W. (1999). Mechanisms of GDF-5 action during skeletal development. *Development*, 126(6), 1305–1315. https://doi.org/10.1001/jamaoto.2015.3595.

Francis-West, P. H., Parish, J., Lee, K., & Archer, C. W. (1999). BMP/GDF-signalling interactions during synovial joint development. *Cell and Tissue Research*, 296(1), 111–119. https://doi.org/10.1007/s004410051272.

Gosden, T. P., Waller, J. T., & Svensson, E. I. (2015). Asymmetric isolating barriers between different microclimatic environments caused by low immigrant survival. *Proceedings of the Royal Society B: Biological Sciences*, 282(1802), 1–7. https://doi.org/10.1098/rspb.2014.2459.

Grant, B., & Grant, P. (2003). What Darwin's Finches can Teach Us about the Evolutionary Origin and Regualtion of Biodiversity. *BioScience*, 53(10), 965–975.

Grant, P. R., & Grant, R. B. (2014). *40 Years of Evolution. Darwin's Finches on Daphne Major Island*. Princetown University Press.

Guerrero, R. F., & Hahn, M. W. (2017). Speciation as a sieve for ancestral polymorphism. *Molecular Ecology*, 26, 5362–5368. https://doi.org/10.1111/mec.14290.

Hendry, A. P., & Taylor, E. B. (2004). How Much of the Variation in Adaptive Divergence Can Be Explained by Gene Flow ? An Evaluation Using Lake-Stream Stickleback Pairs. *Evolution*, 58(10), 2319–2331. https://doi.org/10.1554/04-376.

Hermansen, J. S., Haas, F., Trier, C. N., Bailey, R. I., Nederbragt, A. J., Marzal, A., & Sætre, G. P. (2014). Hybrid speciation through sorting of parental incompatibilities in Italian sparrows. *Molecular Ecology*, 23(23), 5831–5842. https://doi.org/10.1111/mec.12910.

Hermansen, J. S., Sæther, S. A., Elgvin, T. O., Borge, T., Hjelle, E., & Sætre, G. P. (2011). Hybrid speciation in sparrows I: Phenotypic intermediacy, genetic admixture and barriers to gene flow. *Molecular Ecology*, 20(18), 3812–3822. https://doi.org/10.1111/j.1365-294X.2011.05183.x.

Hijmans, R. J. (2014). *raster: Geographic Data Analysis and Modeling*. R Package Version 2.6-7. Retrieved from https://cran.r-project.org/web/packages/raster/vignettes/Raster.pdf

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25, 1965–1978. https://doi.org/10.1002/joc.1276.

Jombart, T. (2008). Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129.

Kaucka, M., & Adameyko, I. (2019). Evolution and development of the cartilaginous skull : From a lancelet towards a human face. *Seminars in Cell and Developmental Biology*, 91, 2–12. https://doi.org/10.1016/j.semcdb.2017.12.007.

Keller, I., Wagner, C. E., Greuter, L., Mwaiko, S., Selz, O. M., Sivasundar, A., Wittwer, S., & Seehausen, O. (2013). Population genomic signatures of divergent adaptation, gene flow and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes. *Molecular Ecology*, 22(11), 2848–2863. https://doi.org/10.1111/mec.12083.

Kraha, A., Turner, H., Nimon, K., Zientek, L. R., & Henson, R. K. (2012). Tools to support interpreting multiple regression in the face of multicollinearity. *Frontiers in Psychology*, 3, 1–16. https://doi.org/10.3389/fpsyg.2012.00044.

Lamichhaney, S., Berglund, J., Almén, M. S., Maqbool, K., Grabherr, M., Martinez-Barrio, A., Promerová, M., Rubin, C.-J., Wang, C., Zamani, N., Grant, B. R., Grant, P. R., Webster, M. T., & Andersson, L. (2015). Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*, 518(7539), 371–375. https://doi.org/10.1038/nature14181.

Lamichhaney, S., Han, F., Berglund, J., Wang, C., Almen, M. S., Webster, M. T., Grant, B. R., Grant, P. R., & Andersson, L. (2016). A beak size locus in Darwin's finches facilitated character displacement during a drought. *Science*, 6284, 470–474.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows – Wheeler transform. *Bioinformatics*, 25(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

Mallet, J. (2007). Hybrid speciation. *Nature*, 446(7133), 279–283. https://doi.org/10.1038/nature05706.

Marques, D. A., Lucek, K., Meier, J. I., Mwaiko, S., Wagner, C. E., Excoffier, L., & Seehausen, O. (2016). Genomics of Rapid Incipient Speciation in Sympatric Threespine Stickleback. *PLOS Genetics*, 12(2), 1–34. https://doi.org/10.1371/journal.pgen.1005887.

Marques, D. A., Meier, J. I., & Seehausen, O. (2019). A Combinatorial View on Speciation and Adaptive Radiation. *Trends in Ecology & Evolution*, 34(6), 531–544. https://doi.org/10.1016/j.tree.2019.02.008.

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

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., & Depristo, M. A. (2010). The Genome Analysis Toolkit : A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297–1303. https://doi.org/10.1101/gr.107524.110.20.

Meier, J. I., Marques, D. A., Mwaiko, S., Wagner, C. E., Excoffier, L., & Seehausen, O. (2017). Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature Communications*, 8, 1–11. https://doi.org/10.1038/ncomms14363.

Meirmans, P. G. (2012). The trouble with isolation by distance. *Molecular Ecology*, 21(12), 2839–2846. https://doi.org/10.1111/j.1365-294X.2012.05578.x.

Newton, R. G., & Spurrell, D. J. (1967). A Development of Multiple Regression for the Analysis of Routine Data. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 16(1), 51–64.

Nimon, K. F., & Oswald, F. L. (2013). Understanding the results of multiple linear regression : beyond standardized regression coefficients. *Organizational Research Methods*, 16(4), 650–674. https://doi.org/10.1177/1094428113493929.

Nimon, K., & Reio, T. G. (2011). Regression Commonality Analysis : A Technique for Quantitative Theory Building. *Human Resource Development Review*, 10, 329–340. https://doi.org/10.1177/1534484311411077.

Nosil, P., Egan, S. P., & Funk, D. J. (2008). Heterogeneous genomic differentiation between walking-stick ecotypes : " isolation by adaptation " and multiple roles for divergent selection. *Evolution*, 62, 316–336. https://doi.org/10.1111/j.1558-5646.2007.00299.x.

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

Ortiz-Barrientos, D., & James, M. E. (2017). Evolution of recombination rates and the genomic landscape of speciation. *Journal of Evolutionary Biology*, 30(8), 1519–1521. https://doi.org/10.1111/jeb.13116.

Pebesma, E., & Bivand, R. (2005). Package "sp". Classes and Methods for Spatial Data in R. *R News*, 5(2), 1–21.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One*, 7(5), e37135. https://doi.org/10.1371/journal.pone.0037135.

Prunier, J. G., Colyn, M., Legendre, X., Nimon, K. F., & Flamand, M. C. (2015). Multicollinearity in spatial genetics: separating the wheat from the chaff using commonality analyses. *Molecular Ecology*, 24, 263–283. https://doi.org/10.1111/mec.14283.

Raeymaekers, J. A. M., Chaturvedi, A., Hablützel, P. I., Verdonck, I. O., Hellemans, B., Maes, G. E., De Meester, L., & Volckaert, F. A. M. (2017). Adaptive and non-adaptive divergence in a common landscape. *Nature Communications*, 8(267), https://doi.org/10.1038/s41467-017-00256-6.

Räsänen, K., & Hendry, A. P. (2008). Disentangling interactions between adaptive divergence and gene flow when ecology drives diversification. *Ecology Letters*, 11(6), 624–636. https://doi.org/10.1111/j.1461-0248.2008.01176.x.

Ravinet, M., Elgvin, T. O., Trier, C., Aliabadian, M., Gavrilov, A., & Sætre, G. P. (2018). Signatures of human-commensalism in the house sparrow genome. *Proceedings of the Royal Society B: Biological Sciences*, 285(1884), https://doi.org/10.1098/rspb.2018.1246.

Reddi, A. H., & Reddi, A. (2009). Bone morphogenetic proteins (BMPs): From morphogens to metabologens. *Cytokine and Growth Factor Reviews*, 20(5–6), 341–342. https://doi.org/10.1016/j.cytogfr.2009.10.015.

Rieseberg, L. H., Raymond, O., Rosenthal, D. M., Lai, Z., Livingstone, K., Nakazato, T., & Lexer, C. (2003). Major ecological transitions in wild sunflowers facilitated by hybridization. *Science*, 301(5637), 1211–1217.

Runemark, A., Fernández, L. P., Eroukhmanoff, F., & Sætre, G.-P. (2018). Genomic Contingencies and the Potential for Local Adaptation in a Hybrid Species. *The American Naturalist*, 192(1), 10–22. https://doi.org/10.1086/697563.

Runemark, A., Trier, C. N., Eroukhmanoff, F., Hermansen, J. S., Matschiner, M., Ravinet, M., Elgvin, T. O., & Sætre, G.-P. (2018). Variation and constraints in hybrid genome formation. *Nature Ecology and Evolution*, 2(3), 549–556. https://doi.org/10.1038/s41559-017-0437-7.

Schumer, M., Cui, R., Powell, D. L., Dresner, R., Rosenthal, G. G., & Andolfatto, P. (2014). High-resolution mapping reveals hundreds of genetic incompatibilities in hybridizing fish species. *elife*, 3, 1–21. https://doi.org/10.7554/eLife.02535.

Schumer, M., Cui, R., Rosenthal, G. G., & Andolfatto, P. (2015). Reproductive Isolation of Hybrid Populations Driven by Genetic Incompatibilities. *PLOS Genetics*, 11(3), e1005041. https://doi.org/10.1371/journal.pgen.1005041.

Schumer, M., Rosenthal, G. G., & Andolfatto, P. (2014). How common is homoploid hybrid speciation? *Evolution*, 68(6), 1553–1560. https://doi.org/10.1111/evo.12399.

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

Seehausen, O. (2004). Hybridization and adaptive radiation. *Trends in Ecology & Evolution (Personal Edition)*, 19(4), 198–207. https://doi.org/10.1016/j.tree.2004.01.003.

Seehausen, O. (2013). Conditions when hybridization might predispose populations for adaptive radiation. *Journal of Evolutionary Biology*, 26(2), 279–281. https://doi.org/10.1111/jeb.12026.

Seeholzer, G. F., & Brumfield, R. T. (2018). Isolation by distance, not incipient ecological speciation, explains genetic differentiation in an Andean songbird (Aves: Furnariidae: Cranioleuca antisiensis, Line-cheeked Spinetail) despite near threefold body size change across an environmental gradien. *Molecular Ecology*, 27(1), 279–296. https://doi.org/10.1111/mec.14429.

Selz, O. M., Lucek, K., Young, K. A., & Seehausen, O. (2014). Relaxed trait covariance in interspecific cichlid hybrids predicts morphological diversity in adaptive radiations. *Journal of Evolutionary Biology*, 27(1), 11–24. https://doi.org/10.1111/jeb.12283.

Shafer, A. B. A., & Wolf, J. B. W. (2013). Widespread evidence for incipient ecological speciation: a meta-analysis of isolation-by-ecology. *Ecology Letters*, 16, 940–950. https://doi.org/10.1111/ele.12120.

Slatkin, M. (1993). Isolation by distance in equilibrium and non-equilibrium populations. *Evolution*, 47(1), 264–279.

Stuart, Y. E., Veen, T., Weber, J. N., Hanson, D., Ravinet, M., Lohman, B. K., Thompson, C. J., Tasneem, T., Doggett, A., Izen, R., Ahmed, N., Barrett, R. D. H., Hendry, A. P., Peichel, C. L., & Bolnick, D. I. (2017). Contrasting effects of environment and genetics generate a continuum of parallel evolution. *Nature Ecology and Evolution*, 1(6), 1–7. https://doi.org/10.1038/s41559-017-0158.

Summers-Smith, J. D. (1988). *The Sparrows: a study of the genus Passer*. T & AD Poyser.

Taylor, S. A., & Larson, E. L. (2019). Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nature Ecology and Evolution*, 3(2), 170–177. https://doi.org/10.1038/s41559-018-0777-y.

Trier, C. N., Hermansen, J. S., Sætre, G. P., & Bailey, R. I. (2014). Evidence for Mito-Nuclear and Sex-Linked Reproductive Barriers between the Hybrid Italian Sparrow and Its Parent Species. *PLOS Genetics*, 10(1), e1004075. https://doi.org/10.1371/journal.pgen.1004075.

Van Der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Levy-moonshine, A., Jordan, T., & Depristo, M. A. (2014). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics*, 11(1110), 11.10.1–11.10.33. https://doi.org/10.1002/0471250953.bi1110s43.From

Wang, I. J. (2013). Examining the full effects of landscape heterogeneity on spatial genetic variation: A multiple matrix regression approach

for quantifying geographic and ecological isolation. *Evolution*, *67*(12), 3403–3411. https://doi.org/10.1111/evo.12134.

Wang, I. J., & Bradburd, G. S. (2014). Isolation by environment. *Molecular Ecology*, *23*(23), 5649–5662. https://doi.org/10.1111/mec.12938.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

PAPER II

RESEARCH ARTICLE

Predictors of genomic differentiation within a hybrid taxon

Angélica Cuevas[1]*, **Fabrice Eroukhmanoff**[1], **Mark Ravinet**[1,2], **Glenn-Peter Sætre**[1],
Anna Runemark[3]

1 Department of Biosciences, Centre for Ecological and Evolutionary Synthesis, University of Oslo, Oslo,
Norway, **2** School of Life Sciences, University of Nottingham, Nottingham, United Kingdom, **3** Department of
Biology, Lund University, Lund, Sweden

* a.m.c.pulido@ibv.uio.no, angelica.cuevas@evobio.eu
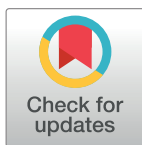
Abstract

Hybridization is increasingly recognized as an important evolutionary force. Novel genetic
methods now enable us to address how the genomes of parental species are combined in
hybrid lineages. However, we still do not know the relative importance of admixed propor-
tions, genome architecture and local selection in shaping hybrid genomes. Here, we take
advantage of the genetically divergent island populations of Italian sparrow on Crete, Cor-
sica and Sicily to investigate the predictors of genomic variation within a hybrid taxon. We
test if differentiation is affected by recombination rate, selection, or variation in ancestry pro-
portions. We find that the relationship between recombination rate and differentiation is less
pronounced within hybrid lineages than between the parent species, as expected if purging
of minor parent ancestry in low recombination regions reduces the variation available for dif-
ferentiation. In addition, we find that differentiation between islands is correlated with differ-
ences in signatures of selection in two out of three comparisons. Signatures of selection
within islands are correlated across all islands, suggesting that shared selection may mould
genomic differentiation. The best predictor of strong differentiation within islands is the
degree of differentiation from house sparrow, and hence loci with Spanish sparrow ancestry
may vary more freely. Jointly, this suggests that constraints and selection interact in shaping
the genomic landscape of differentiation in this hybrid species.

Author summary

Genomes of hybrid lineages are mosaics of those of their parent species and harbour varia-
tion that has the potential to facilitate adaptation when hybrids encounter diverse envi-
ronments. However, genetic incompatibilities between parental species can also act to
limit possible combinations of parental alleles, constraining hybrid genome formation.
What is the relative importance of selection and constraints in form of admixture propor-
tions and genomic architecture in this process? We investigated this in the Italian sparrow,
a hybrid species resulting from past hybridization between the house and Spanish spar-
row. Using three independent hybrid lineages, we addressed how their genomes, harbour-
ing different parental combinations, have evolved. We examined the roles of selection due

to divergent local adaptation, recombination and purging of genetic incompatibilities in predicting differentiation. We found that selection against incompatibilities may constrain hybrid genome composition. In addition, signals of local selection as well as estimates of differentiation were correlated across populations, and outliers were shared among the hybrid lineages more often than expected by chance. Overall, our results suggest that in the Italian sparrow selection interacts with constraints linked to genetic incompatibilities affecting which sections of the genome can readily diverge among hybrid lineages.

Introduction

Heritable variation is the substrate on which natural selection acts, and hybridization is increasingly recognized as an important process providing such variation in fish [1,2], insects [3], birds [4,5] and even humans [6]. Hybridization can enable lineages to combine parental genomes in adaptive ways, for instance contributing alleles linked to insecticide resistance in mosquitoes malaria vectors [7], adaptive fur colour in hares [8] and MHC immune defence diversity in modern humans [9]. Similarly, the variation created by hybridization has provided the raw materials for the extensive adaptive radiations of African lake cichlids [2,10–13]. Lineages resulting from hybridization may even outcompete the parental species in certain environments and colonize new niches as documented in *Helianthus* sun flowers where hybrid taxa colonize extreme environments [14,15]. One outcome of hybridization is hybrid speciation, resulting in the formation of a taxon that is reproductively isolated from its parent species [16]. Hybrid speciation can arise both through allopolyploidization and homoploid hybrid speciation, without an increase in chromosome number in the latter case [16–19]. Interestingly, the relative contributions of the parental species may vary within a hybrid, as illustrated by the variable genome composition in sword-tail guppies [20], in Lycaides butterflies [21], and, as is the focus of this study, in isolated island populations of Italian sparrows [22]. Here, we focus on the Italian sparrow, a well known example of a homoploid hybrid species, with reproductive barriers to the parent species consisting of a subset of those isolating the parent species [23,24]. Genetically divergent island populations of the Italian sparrow, potentially originating from independent hybridization events, differ in proportions of their genomes inherited from their parental species, house and Spanish sparrows (*P. domesticus* and *P. hispaniolensis*) [22]. The share of house sparrow ancestry, estimated as admixture proportion, ranges from 37% in the lineage on Sicily, to 62% in Corsica and 76% in the Cretan lineage [22].

Hybridization, was, at least in animals, historically viewed as an evolutionary mistake [25], partly because hybrids are likely to suffer from incompatible allelic combinations. While this view has changed over the last decades [17,26], hybrid lineages likely need to overcome a number of challenges to successfully establish. Incompatibilities might mean low fertility, sterility or even inviability in some crosses [27]. This is shown by Haldane's rule [28], when species have heterogametic sex chromosomes, the heterogametic sex is more likely to be sterile or inviable. In addition, evidence for a role of mito-nuclear interactions causing fitness reduction in hybrids is mounting [23,29,30]. For example, maladaptive metabolisms in hybrids [31] suggest that mito-nuclear interactions could pose strong selection pressures on the genomic composition in hybrid taxa. Mito-nuclear interactions may also play a role in determining the Italian sparrow genome composition [23]; hybrid Italian populations are largely fixed for the house sparrow mitochondrial genome, and there is evidence of an excess of house sparrow ancestry conserved in nuclear genome regions contributing to mitochondrion function [22]. Even in

species that have successfully formed hybrid daughter lineages, early generation hybrids may
still be inviable or infertile [32]. These findings suggest that fitness losses due to incompatible
parental combinations, i.e. Bateson-Dobzhansky-Muller incompatibilities (BDMI) [33–37],
may be restored through fixation of compatible pairs of alleles from either of the parent spe-
cies. Alternatively, if the portion of the genome that is free to vary (i.e. where constrains may
be reduced) is reduced, it could potentially result in convergent allelic compositions at specific
genomic regions among independent hybrid lineages.

Although hybrid lineages in principle have a vast number of potential combinations of
parental alleles and increased nucleotide diversity available as a source for adaptation, little is
known about genome stabilization in hybrid taxa [19,38]. After reproductive isolation from
the parental species develops, stabilization of the hybrid genome will occur, removing ancestry
blocks by purging of incompatibilities and fixing genomic combinations [39,40]. The speed of
genomic stabilization varies between hybrid taxa and will also occur at different rates in differ-
ent parts of the genome. It could take several hundred of generations for neutral loci [39] or
occur very quickly in functionally important regions [40,41]. In addition to drift and selection,
ancestry sorting during genome stabilization also has a determining effect on the composition
of admixed genomes [42], which in turn affects patterns of genomic differentiation among
hybrid populations. There could also potentially be constraints- here defined as effects of geno-
mic architecture, including recombination rate, or incompatibilities due to ancestry admix-
ture- on which genomic regions are free to vary and bias on the overall composition of hybrid
genomes. For instance, introgression on the sex chromosomes is commonly reduced com-
pared to genome-wide levels in species where one sex is heterogametic, consistent with selec-
tion against infertility [43–45]. Experimental assays in sunflowers and recent studies of
Lycaeides butterflies have shown that the same genetic combinations found in natural hybrid
lineages re-emerge in experimental hybrid populations [46] and younger lineages [47], possi-
bly due to selection against alternative combinations and recombination effects. This raises the
question of how easily hybrid lineages can achieve divergent genome compositions and pheno-
types. Can different combinations of parental alleles easily be achieved due to selection for
divergent local adaptation in homoploid hybrids? Or do patterns of differentiation at a local
scale mirror those between strongly divergent populations, suggesting a role for constraints
from recombination rate on which genome regions may differentiate? Exploring the patterns
of population differentiation within hybrid species may reveal novel insights into the forces
shaping hybrid genomes.

Interestingly, patterns of species differentiation are affected by the recombination rate land-
scape [48–50]. This can result in highly correlated patterns of divergence between closely
related species pairs, such as that found in flycatchers [48]. Moreover, linked selection has a
greater effect in regions of low recombination [48]. Selective sweeps in genomic regions of low
recombination can give rise to a negative correlation between recombination rate and genomic
differentiation [48,51]. Specific to hybrid taxa, evolutionary processes occurring during
genome stabilization could have an impact on the distribution of genome diversity and later
potential differentiation between independent hybrid populations [19,40]. For example, initial
ancestry sorting could lead to differences in admixture proportions. This could give rise to
subsequent lineage specific evolution within ancestry types. Purging of minor parent ancestry
in low recombination regions to reduce genomic incompatibilities could reduce the variation
available for subsequent differentiation. Recent studies have indeed found reduced introgres-
sion in low recombination regions in hybrid swordtail fish, sticklebacks, *Heliconius* butterflies
and humans, suggesting that the recombination landscape may indeed affect which regions
are permeable to introgression [20,49,52–54]. Genomic blocks with ancestry from the minor-
parent may be retained in regions of high recombination rate, due to their increased likelihood
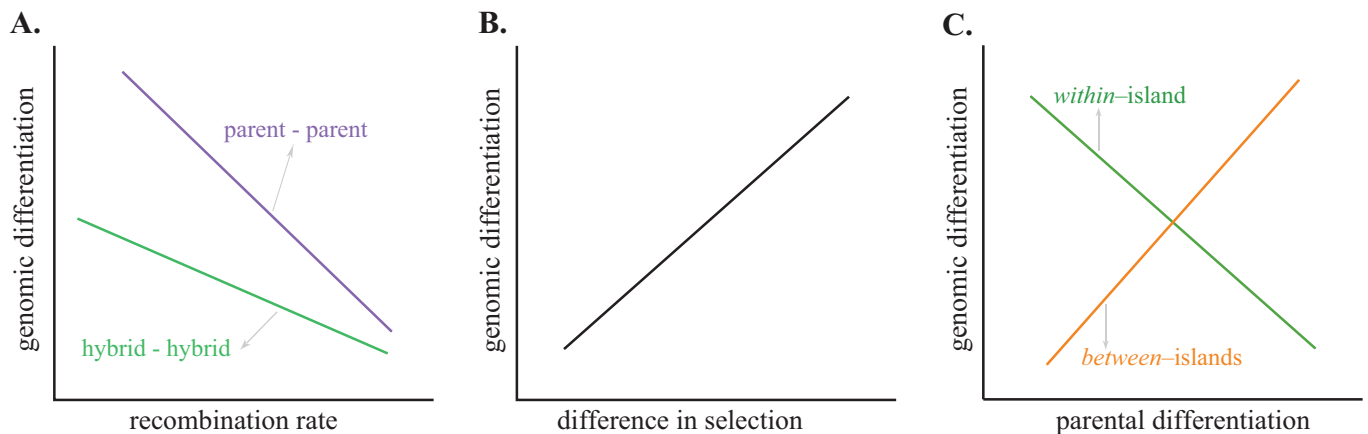
Fig 1. Expected patterns of genomic differentiation for different levels of recombination rates, divergent selection and parental differentiation. A. Recombination rate. *Parent-parent*: Since selective sweeps have a greater effect in low recombination regions we expect a negative relationship between recombination rate and differentiation between the non-hybrid parental lineages. *Hybrid—hybrid*: Specific to hybrid taxa, additional selection against minor parent ancestry in low recombination regions could reduce the variation available for differentiation in these regions. As this process does not affect parent taxa, we would expect a flatter relationship between recombination rate and differentiation in hybrids, compared to their parent taxa, if purging of minor parent ancestry is pervasive. In addition, in higher recombination regions there is a potential for alternative blocks of ancestry to be fixed, which will inflate the differentiation among hybrids relative to the parent-parent comparison, contributing to the flatter relationship in hybrids **B. Selection.** If divergent selection is an important predictor of differentiation, we expect differentiation between two islands to be positively correlated to the strength of divergent selection between them. **C. Parental differentiation.** If differentiation is limited by potential incompatibilities between divergent parental loci, we would expect the potential for adaptive differentiation in hybrid populations of similar origin to be highest in genomic regions of low parental divergence. Regions that are strongly differentiated between parents, more likely to be incompatible in combination with other parental alleles, would have a greater likelihood of being fixed for the ancestry of one parent only during genome stabilization. This process is expected to act *within islands*, where the overall genomic composition and the resolution of incompatibilities through fixation of alleles from one parent species, are most likely similar. In contrast, in independent hybrid populations with different proportions of ancestry (like the ones found in different islands), fixation of alternative parental alleles could increase the potential for differentiation and local adaptation *between islands*. These processes, with influence from processes not addressed in this study, affect the genomic composition in hybrid lineages. However, their relative importance may be inferred from the signatures of differentiation in relation to the predictors.

of breaking away from potential incompatibilities in these regions [20]. If, during genome stabilization, purging of incompatibilities in low recombination regions is pervasive, the resulting reduced diversity in these regions could also act as a constraint on genomic differentiation in the hybrid, decreasing differentiation at a greater rate in low- than high recombination regions. Thus, the relationship between recombination rate and differentiation is expected to be less negative in hybrid lineages compared to the differentiation between the non-hybrid parental taxa should such purging be important (Fig 1A). In contrast, local selective sweeps, having a greater impact on regions of low recombination due to linked selection, could lead to higher differentiation in low recombination regions in comparison to regions of high recombination. However, the relative importance of these two processes in shaping hybrid genomes remains unclear.

There is evidence that hybrid taxa can use the variation that originated through hybridization for local adaptation. For instance, beak shape in Italian sparrows is explained by local precipitation regime [55]. Beak size differences among island populations are best explained by temperature seasonality [56], and some island populations are strongly differentiated for a gene known to affect beak morphology in Darwin's Finches, FGF10 [22,57]. In addition, the gene GDF5, part of the BMP gene family that has a fundamental role in beak shape and size variation in Darwin's finches [58,59], is a candidate gene putatively under selection in the Italian sparrow populations from mainland Italy [38]. However, beak shape is also affected by the proportion of the genome inherited from each parent species suggesting a small albeit significant role for admixture proportions in morphology [55,56]. Moreover, in a recent study

investigating genetic differentiation among Italian sparrow populations on mainland Italy, population differentiation was associated with temperature and precipitation [38]. Interestingly, alleles segregating in both parental species showed strong allele frequency differences within the Italian sparrow, suggesting that adaptation is not completely dependent on the combination of alleles from different parent species [38]. If selection has a significant role in genomic differentiation of the Italian sparrow and selection pressures diverge among islands, favouring the fixation of alternate alleles, we would expect stronger differentiation in regions under divergent selection among islands (Fig 1B). Furthermore, depending on the variation of parental ancestry proportions in hybrid populations, parental divergence may affect genomic differentiation in the hybrid in different ways. In populations with similar parental ancestry, like those within islands, the resolution of incompatibilities during genome stabilization is likely to be similar. For these, genomic differentiation may be most easily achieved from standing genetic variation inherited from the parents, thus from variants that are segregating in the parent species (Fig 1C). Alternatively, in independent hybrid populations with divergent proportions of parental ancestry, differentiation is more likely to be found in regions of strong parental differentiation (Fig 1C).

One way to further our understanding of the evolutionary forces acting on hybrid genomes is to investigate patterns of differentiation within hybrid lineages and the factors that best explain them. The Italian sparrow is a uniquely suited study system, as it provides independent populations with varying ancestry proportions (Fig 2). These island lineages are likely to have originated as a result of different hybridizations events [22]. This is supported by the low pairwise correlations of ancestry tracts among islands and significant albeit small differences in ancestry tract sizes [22], suggesting at least long periods of independent evolution. This unique system enables comparison of hybrid lineages with divergent ancestry proportions, as well as comparison of populations with potentially similar parental contributions i.e., populations within islands with similar evolutionary history. Here, we use the island Italian sparrows to investigate how differentiation within island compares to that among islands. Our overachieving aim is to address which factors best predict differentiation within and among islands to disentangle the evolutionary forces shaping hybrid genomes.

We test the hypotheses that I) long periods of independent evolution have resulted in significantly higher divergence *among*-islands than *within*-islands; II) selective sweeps and purging of minor parent ancestry in low recombination regions in the hybrid has led to a less steep relationship between recombination rate and differentiation than that between the parent species (Fig 1A) III) that genomic regions experiencing stronger divergent selection *among* islands also show elevated differentiation (Fig 1B) IV) that constraints on how freely genomic regions are able to diverge have led to correlated landscapes of differentiation *within*- and *among*-islands, and V) that differences in minor–major parental ancestry in the hybrid have a direct effect on the genomic differentiation between populations.

Results

I) Genomic differentiation within- and between islands

Consistent with [22] we find strong differentiation between the focal island populations based on RAD data. From our principal component analysis, based on 2224 SNPs (S2 Table), the first main axis of differentiation largely reflects the proportion of the genome inherited from each parent species, and Crete diverges along the second axis (Fig 2C). Interestingly, the ADMIXTURE analysis supports the presence of three clusters rather than two, with Crete forming a separate cluster (Fig 2B). Average of windows-based F_{ST} estimates, based on 2856 SNPs (S2 Table), are consistent with this, Crete is more strongly differentiated from Sicily
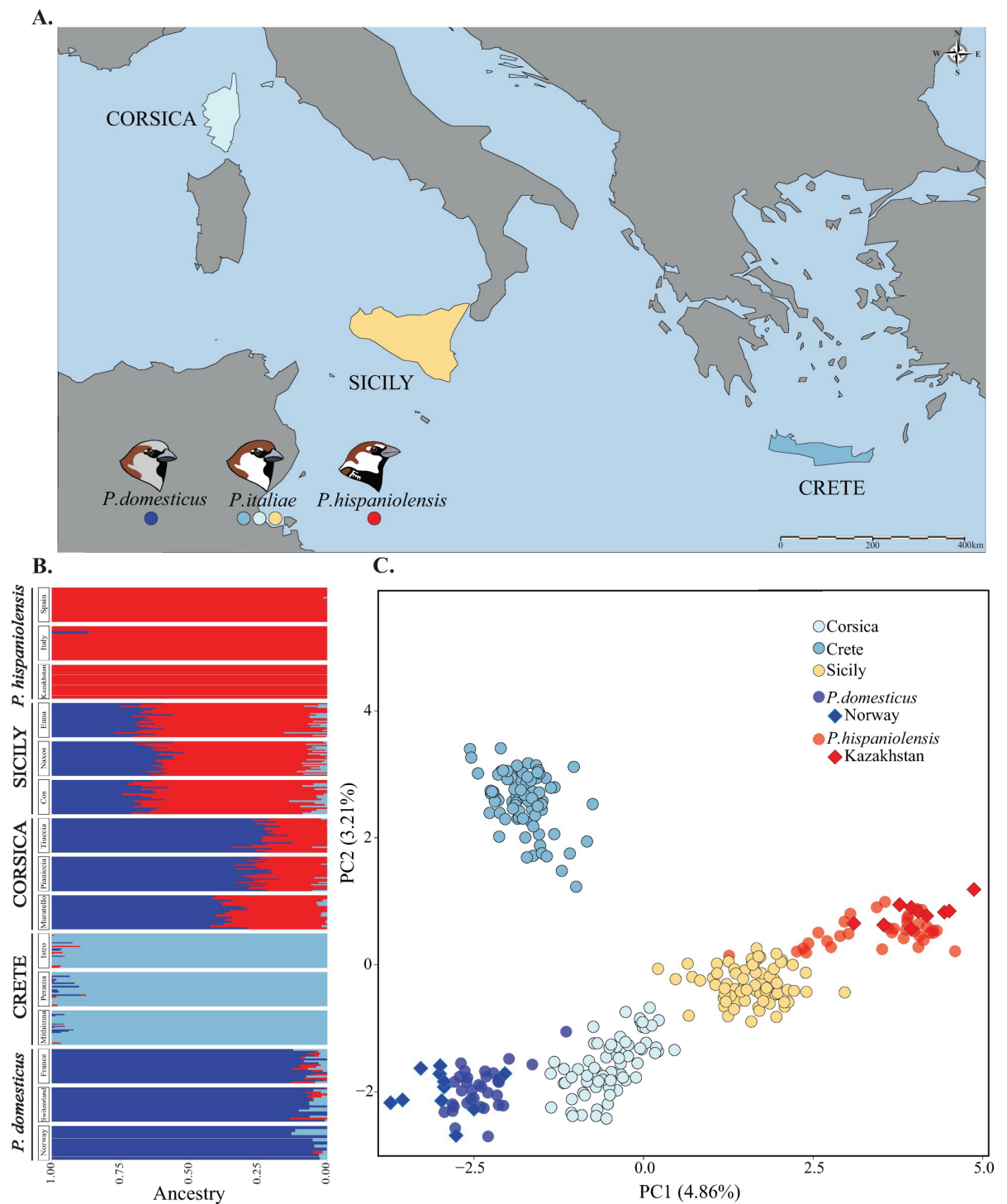
Fig 2. Sampling design and population structure. A. Map of sampling locations. Sparrows were sampled from three populations on each of the islands of Corsica (N = 70), Crete (N = 77) and Sicily (N = 76) **B.** Admixture analysis illustrating the clustering of the island populations of the Italian sparrow and their parental species for three clusters (K = 3), the value that received the highest support. Three geographically separated populations of each of the parental species, the house and Spanish sparrow, were included. **C.** Principal component analysis (PCA) illustrating the relationship between the Italian sparrow populations and the parental species. ◆ indicate the reference parental populations with lower levels of introgression. Analyses were based on a VCF containing 2224 SNPs. Map base layer was retrieved using the R-package "rworldmap" and the function getMap() - https://cran.r-project.org/web/packages/rworldmap/.

https://doi.org/10.1371/journal.pgen.1010027.g002

(mean F_{ST} = 0.043) and Corsica (0.042), whereas mean F_{ST} between Corsica and Sicily is comparatively lower (0.025; Fig 3A).

Differentiation is more pronounced between islands than within islands (Monte-Carlo permutation paired t-test: t = 33.21, df = 7927, P = 1.98e-15; Table A in S1 Text; Fig 3A), with a medium effect size estimate (Cohen's d = 0.523; 95% CI 0.49–0.56). A discriminant function analysis, using the dataset of 2856 SNPs (S2 Table), recovers some differentiation among local populations in each island (Fig 3B and 3C), and correctly assigns 95.3% of Corsican individuals, 78.4% of the individuals from Crete and 75.5% of the Sicilian individuals to their populations of origin within each island. *Within*-island F_{ST} differs significantly among islands (All Ps < 0.5e-3; Table A in S1 Text), with Corsican populations exhibiting the highest mean F_{ST} of 0.018, as well as the highest nucleotide diversity (π: 3.021e-06; Table A in S1 Text). Differentiation within Sicily is intermediate at 0.013, while F_{ST} among the Cretan populations is the lowest at 0.011 (Fig 3A and Table A in S1 Text). While most variation segregates within individuals and populations, an AMOVA reveals that 4.84% of the variation is found among islands whereas 0.91% of the variation is found among populations within islands (both fractions are statistically significant P: 0.001, as estimated from a randomization Monte Carlo test with 1000 permutations; Table 1 and Fig A in S1 Text).

II) The relationship between genomic differentiation and recombination rate

To evaluate the hypothesis that hybrid-specific purging of minor ancestry blocks in low recombination regions can reduce genomic diversity (and in turn genomic differentiation) in these regions in contrast to high recombination regions, where the effect of purging is expected to be weaker; we tested if genomic differentiation decreases less rapidly with recombination rate between populations within islands than between the parent species (Fig 1A). We evaluated this by comparing the slopes of the relationship between recombination rate and differentiation and through testing for a significant interaction effect between the type of comparison (parent-parent vs. *within*–island) and recombination rate on differentiation. Using differentiation among populations within islands implies that relatively similar resolution of incompatibilities across populations can be assumed. We predicted that the slopes would be less steep among populations from the same island (*within*-island F_{ST}) than between the parent species, if selection against minor parent ancestry is an important selection pressure in the hybrid lineages, reducing variation in low recombination regions (Fig 1A). Indeed, we find that this is the case (Tables B and C in S1 Text; Fig 4A). The slope generated by the relationship between the parental differentiation and recombination rate differs from those found in each *within*-island comparison (Table B in S1 Text; Fig 4A). We find a significant interaction of recombination rate and comparison (parent-parent vs. within-island) in all independent linear models per island (Table C in S1 Text). We find no significant correlation between recombination rate and *within*-island genomic differentiation for Corsica (correlation = -0.012, R^2 = 1.4e-4, P = 1; Fig 4B) or Crete (correlation = -0.003, R^2 = 0.9e-5, P = 1; Fig 4C). However, differentiation within Sicily is weakly but significantly negatively correlated with recombination rate, albeit the effect size is very small (correlation = -0.048, R^2 = 0.002, P = 0.037; Fig 4D).

When evaluating the influence of recombination rate, parental differentiation and differentiation to the two parent species, a GLM did not reveal any significant relationship between recombination rate and differentiation within islands (GLM, Estimate = -9.66e-04, Std. Error = 7.84e-04, P = 0.22; Table D in S1 Text). Corresponding binomial models revealed that recombination rate did not affect the probability of loci being F_{ST} outliers (1% outliers of the F_{ST} distribution) within islands either (Table 2, Table E in S1 Text). However, evaluating
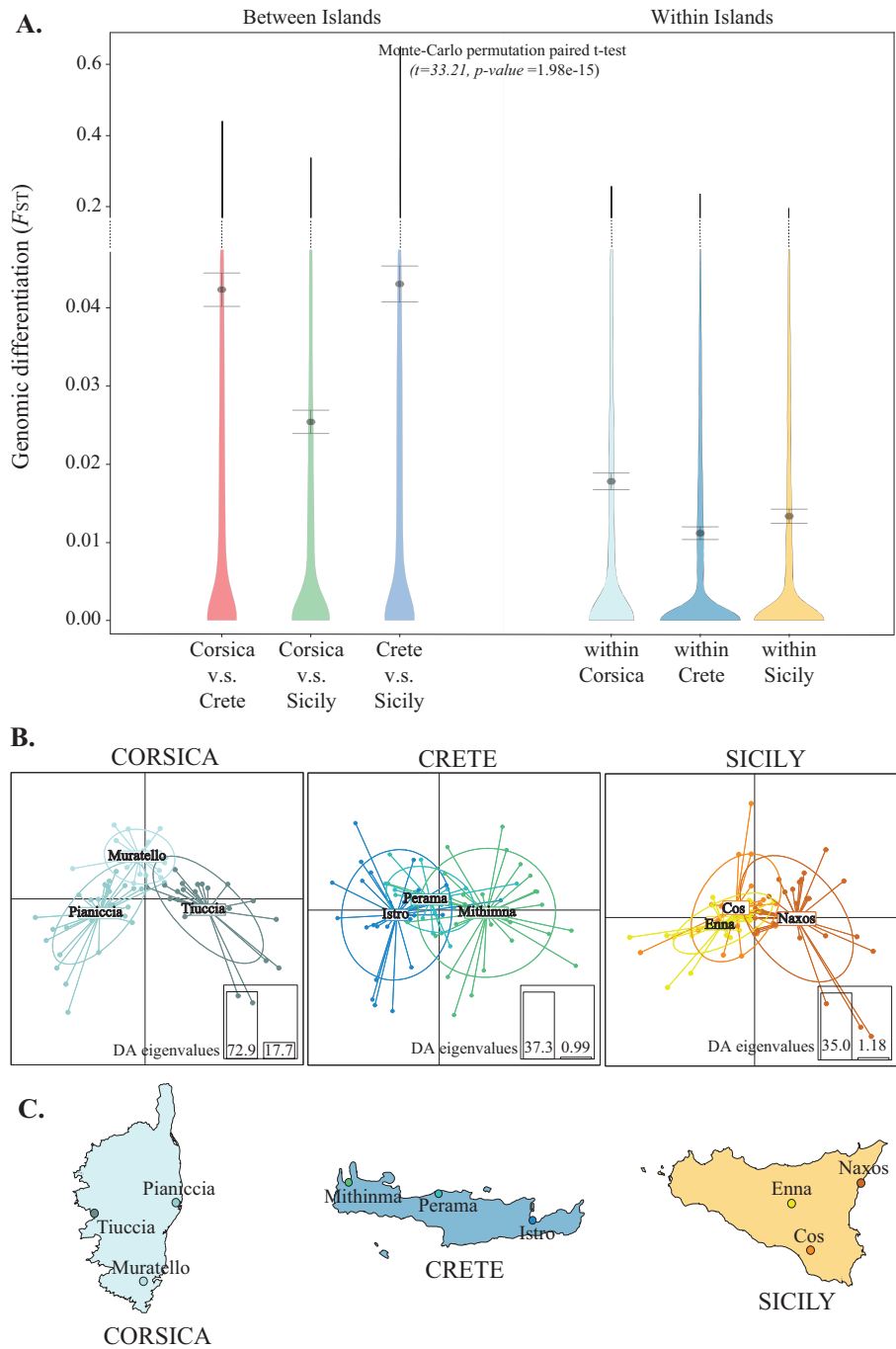
Fig 3. Genomic differentiation *within-* and *among*-islands. A. Levels of genomic differentiation among the three islands (window based F_{ST} between islands pairs) compared to levels of differentiation within islands (window based F_{ST}). Points and bars depict the means and 95% CI, respectively. Continuity of the y-axis is broken (dashed line) to minimize the size of the figure in order to include the extreme values of the F_{ST} distribution. **B.** Discriminant analysis of principal components (DAPC) for populations within Corsica, Crete and Sicily, respectively. **C.** Map of the focal islands with the three populations sampled within each island denoted by filled circles. All analyses were based on a VCF containing 2856 SNPs. Maps base layer was retrieved using the R-package "rworldmap" and the function getMap (). - https://cran.r-project.org/web/packages/rworldmap/.

https://doi.org/10.1371/journal.pgen.1010027.g003

Table 1. Analysis of Molecular Variance (AMOVA) across islands and populations within islands. Several cut-offs for missing-ness per loci were used (5%, 10%, 20%), but the results from the AMOVA did not change substantially. Here we present the results from a cut-off of 5% (Table L in S1 Text). The significance of variation partitioning in each element was maintained. Analyses are based on a VCF containing 2856 SNPs.

Analysis of Molecular Variance					Randomization by Permutation	
AMOVA					Monte Carlo test	
Variance partitioning	Df	Sum Sq	Sigma	% of covariance	Std. Observed	*P-value*
Among islands	2	778.79	2.22	4.84	4.33	0.001**
Between populations within island	6	365.37	0.42	0.91	9.42	0.001**
Between individuals within populations	213	8654.88	-2.54	-5.54	-4.17	1.000
Within individuals	222	10148.04	45.71	99.80	1.26	0.897

https://doi.org/10.1371/journal.pgen.1010027.t001
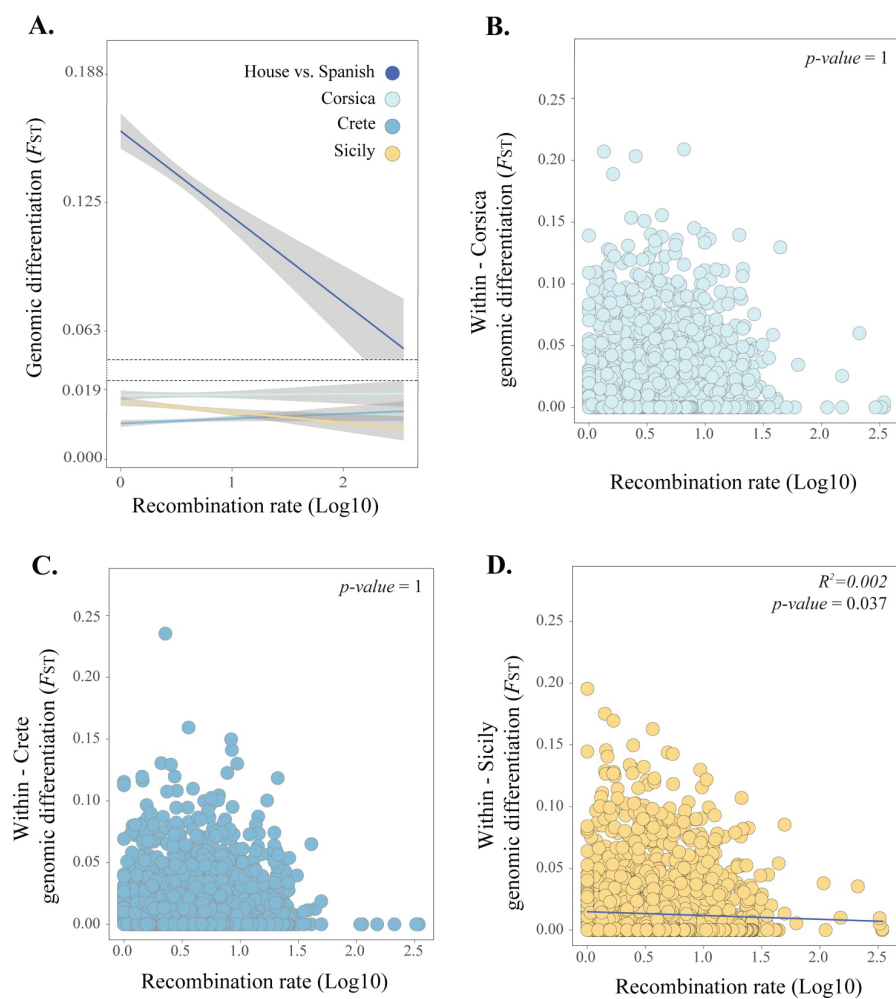
Fig 4. The influence of recombination rate on genomic differentiation. A. Comparison of the relationship between recombination rate and genomic differentiation between the parental species (dark blue) and between Italian sparrow populations within each island. Correlation between recombination rate and genomic differentiation within island for **B**. Corsica, **C.** Crete and **D.** Sicily. Analyses based on a VCF containing 2804 SNPs for **A** and one containing 2856 SNPs for **B, C** and **D**.

https://doi.org/10.1371/journal.pgen.1010027.g004

Table 2. Logistic regression assessing the predictors on the probability of being a *within*–island F_{ST} outlier and post-hoc estimated marginal (least-square) means.

Model:				
Pr(*within*-island F_{ST} outlier) = per-loci local ancestry proportion (LLAP) + Recombination Rate + island + island.house F_{ST} + island.Spanish F_{ST} + house.Spanish F_{ST}				
Response variable	**Predictor**	**Estimate**	**Std. Error**	***P-value***
Pr (*within*-island F_{ST} outlier)	Recombination Rate	-1.402e-01	1.431e-01	0.33
	LLAP	1.411e-02	6.336e-02	0.82
	Island v.s House F_{ST}	-1.401e+00	6.674e-01	0.04*
	Island v.s Spanish F_{ST}	5.728e-01	4.698e-01	0.22
	House v.s Spanish F_{ST}	-6.752e-06	8.770e-06	0.44
Post-hoc Estimated marginal (Least-squares) means for the predictor variable "island"				
Comparison	**Estimate**	**Std. Error**	**Z ratio**	***P-value***
Corsica—Crete	0.021	0.133	0.16	0.99
Corsica—Sicily	-0.11	0.133	-0.88	0.67
Crete—Sicily	-0.13	0.134	-1.03	0.58
P-value adjustment: Tukey's HSD				

https://doi.org/10.1371/journal.pgen.1010027.t002

individual islands separately shows that recombination rate significantly explains differentiation between Sicilian populations, with higher divergence in low recombination regions, as revealed by a GLM (Parameter estimate = -3.48e-03, Std. Error = 1.34e-03, P = 9.3e-03; Fig 4D; Table F in S1 Text).

III) The concordance of patterns of selection and genomic differentiation

To assess the role of selection in shaping genomic differentiation in the Italian sparrow, we tested if differences in selection were correlated to genetic differentiation. If divergent selective pressures among the islands have a large influence on the formation of their hybrid genomes, we expect a positive correlation between measures of divergent selection and genomic differentiation between islands (Fig 1B). Here, we tested this prediction, using cross population haplotype homozygosity (xp-EHH) a statistic that measures putative patterns of divergent selection by comparing haplotype lengths between populations to detect potential selective sweeps [60]. Genomic differentiation between two island pairs was significantly correlated to xp-EHH. This measure was negatively correlated with F_{ST} for the Corsica–Sicily (correlation = -0.061, R^2 = 0.0037, P = 0.014), and Crete–Sicily (correlation = -0.059, R^2 = 0.004, P = 0.019; Fig 5A; Table G1 in S1 Text) comparisons. However, the effect of xp-EHH on differentiation between these islands was very small with an almost marginal effect size, and there was no relationship between xp-EHH estimates and differentiation in the Corsica–Crete comparison (correlation = -0.042, R^2 = 0.002, P = 0.148; Fig 5A).

We also addressed how consistent selection is within- and across islands. Patterns of selection within islands (estimated by the integrated haplotype homozygosity score, iHS) were positively correlated in all pairwise comparisons between islands, with R^2 ranging from 0.095 to 0.174 (Fig 5B; Table G3 in S1 Text). This suggests shared patterns of selection across the genomes, potentially driven by similar selection pressures or genomic constraints arising from the distribution of variation and incompatibilities in the parent species, reducing the availability of genomic variation. However, differentiation within island populations was not significantly correlated to iHS for neither Crete nor Corsica (Fig B and Table G2 in S1 Text), although there was a weak correlation with a small effect size for Sicily (R^2 = 0.005, P = 0.01; Fig B and Table G2 in S1 Text). Interestingly, mean Tajima's D differed considerably among islands, with Sicilian and Corsican populations exhibiting negative estimates (range from -0.25
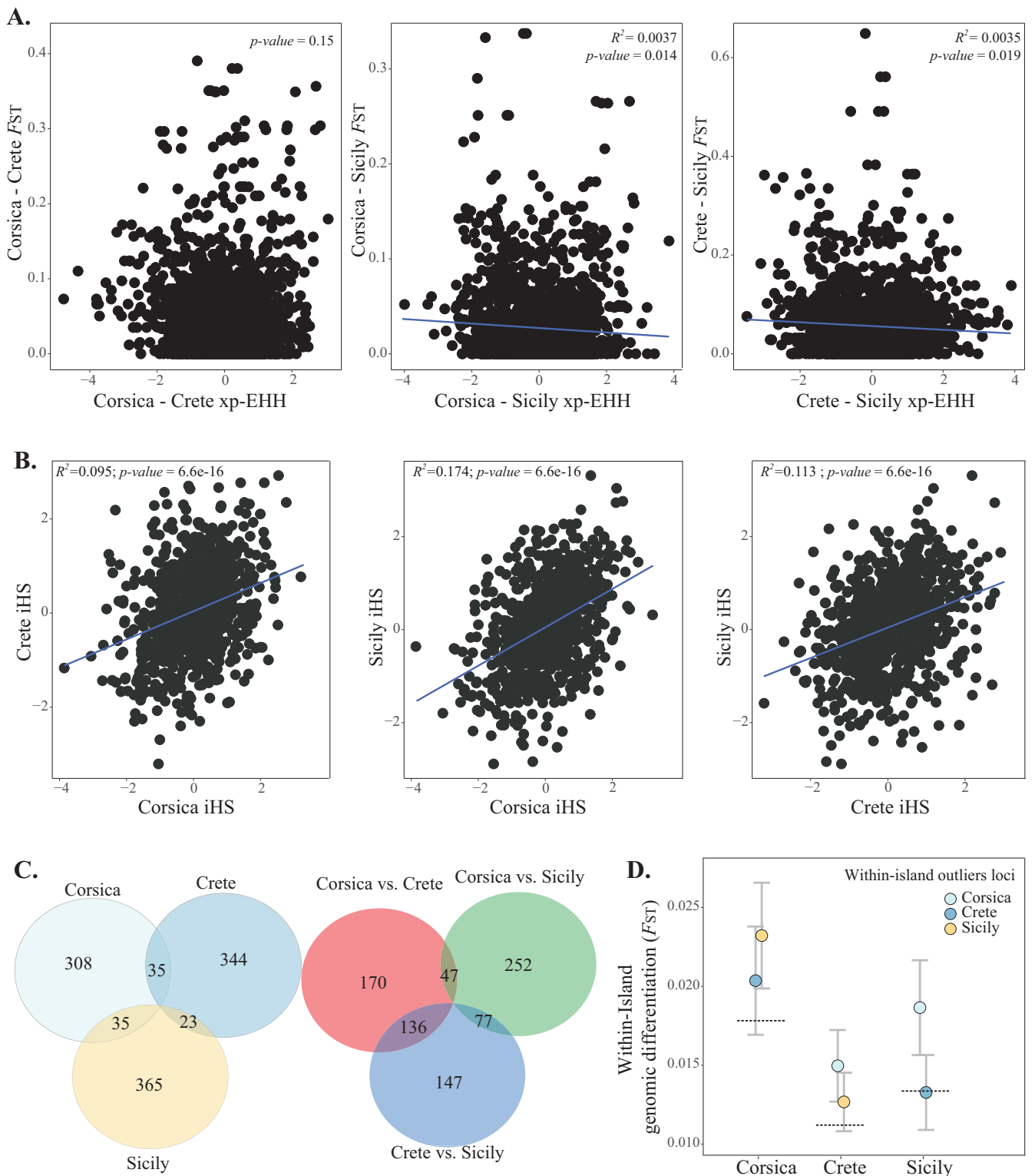
Fig 5. Effects of selection on genomic differentiation, similarity in selection pressures across islands and shared patterns of genomic differentiation. A. Relationship between divergent selection (estimated as Extended Haplotype Homozygosity statistic—XP-EHH) and genomic differentiation between islands. **B. Similarity in patterns of selection.** Between islands correlations of the *within*-island selection measure, estimated as integrated haplotype homozygosity score

(iHS). **C.** Shared outlier loci among populations within islands (left) and among islands (right). **D.** F_{ST} for outlier loci from each of the other islands within each island. Dashed lines represent the global within-island F_{ST} mean. Error bars denote 95% CI.

to -0.048; Fig C and Table A in S1 Text). In contrast, populations on Crete exhibited higher values of Tajima's D (range from -0.11 to 0.015; Fig C and Table A in S1 Text).

IV) Distribution and repeatability of differentiation across the genome

To evaluate concordance in the differentiation landscape, as expected if genomic differentiation is affected by recombination rate, constraints or similar selective pressures, background or parallel selection, we ran correlations of *within*-islands and *among*-islands F_{ST}. We found that patterns of *within*-island differentiation are significantly correlated between Sicily and Corsica (correlation = 0.081, R^2: 0.0066, P = 1.16e-10), but not between Corsica and Crete (correlation = 0.02, R^2: 5.8e-4, P = 0.16), or Sicily and Crete (correlation = 0.014, R^2: 1.96e-4, P = 0.81; Fig D in S1 Text). Hence, levels of differentiation are not correlated between all islands. Interestingly, we found that the outlier loci within one island were more frequently outliers within other islands than expected by chance in two out of three comparisons (Fig 5C; Table H in S1 Text). A total of 9.3% of the outliers within Corsica overlap with those from Crete (Chi-squared: 7.18, P: 0.007) and a similar percentage (9.3%) in within–Sicily F_{ST} outliers (Chi-squared: 6.80, P: 0.009; Table H in S1 Text). However, outliers from Crete and Sicily are not shared to a higher extent than expected by chance (Chi-squared: 0.09, P: 0.767). We also tested whether individual island outliers have a higher mean F_{ST} within other islands. We found outliers from Sicily to have a higher F_{ST} values within Crete and within Corsica than expected by chance (Crete: t = -1.997, df = 457.9, P = 0.046; Corsica: t = -3.082, df = 444.7, P = 0.002; Fig 5D). Similarly, Corsica F_{ST} outlier loci have elevated F_{ST} within Sicily (t = -3.393, df = 392.2, P = 7.6e-4) and within Crete (t = -3.586, df = 385.1, P = 3.8e-4). However, outliers from Crete do not have higher F_{ST} values than expected by chance in any of the other island populations (Fig 5D).

We further tested whether loci differentiated within islands also are more differentiated among islands. Pair-wise correlations between *within*– and *among*–islands F_{ST} suggest the same regions are differentiated, but the effect is weak and varied. While differentiation within Corsica is correlated to Corsica–Sicily F_{ST} (correlation = 0.05, R^2: 0.0025, P: 0.013) and Crete–Sicily F_{ST} (correlation = 0.05, R^2: 0.0025, P: 0.026; Fig E in S1 Text), none of the other seven comparisons are significant. Consistent with this pattern, the proportion of *within*–Corsica outlier loci that overlap with the most differentiated loci in Corsica–Sicily F_{ST} (9.8%) and in Crete–Sicily F_{ST} (7.4%) are also higher than expected by chance (Chi-square tests: X^2: 15.53, P: 8.1e-05 and X^2: 4.09, P: 0.04, respectively; Table I in S1 Text). Moreover, we found a higher proportion (10.2%) of Crete's outlier loci than expected among the Crete–Sicily F_{ST} outlier loci (X^2: 21.13, P: 4.3e-06, Table I in S1 Text). Among the 56 putative genes located in the shared regions of differentiation there is one presenting mitochondrial functions (S3 Table).

To evaluate whether background selection or adaptive parallel selection shape the patterns of shared differentiation, we tested correlations across all possible pair-wise comparisons of subpopulations within each island and compared these to all pair-wise correlations between populations on different islands. The rationale for this is that background selection should result in significant correlations in all analyses, as the correlations would reflect a conserved recombination rate landscape resulting in elevated differentiation in low recombination regions. We find variation in the strength of the relationships depending on comparison, with stronger relationships between differentiation in some comparisons, including some borderline significant ones (Figs F and G and Table J in S1 Text).

V) Patterns of local genomic differentiation in relation to parental contributions to the genome

Multiple factors may affect which loci are free to vary within the Italian sparrow. For example, variation in parental contributions to the genomes of the different island populations, the level of differentiation between the parent species across the genome, and the recombination rate. We tested to what extent these factors explain the patterns of *within*-island differentiation by performing a generalized linear model (GLM) and a logistic model using *within*-island F_{ST} as the response variable. The factor that best predicts the probability of a SNP to be an F_{ST} outlier within islands is the extent of differentiation to the house sparrow (Logistic regression estimate: -1.401e+00, *P*: 0.036; Table 2). However, when evaluating factors that may affect the *within*-island differentiation (F_{ST}), using a GLM, differentiation to the house sparrow was found to be non-significant (Table D in S1 Text). Neither the extent of differentiation from the Spanish sparrow, parental differentiation, recombination rate, nor the per-locus local ancestry proportion (LLAP) contributed significantly to differentiation in either, the logistic regression or the GLM (Table 2; Table D in S1 Text). Furthermore, in separate logistic regressions for each island, none of the studied factors significantly affected the probability of being an outlier (Table E in S1 Text), potentially because of reduced statistical power. However, in separate GLMs run for each island (Table F in S1 Text) including all the predictors mentioned above, the general distribution of differentiation (F_{ST}) within Corsica is explained by the differentiation to the Spanish sparrow (Estimate for Spanish F_{ST} = 1.166e-02, *P* = 0.036, Multiple R^2 of the model = 0.003) and differentiation within Sicily is weakly but significantly negative correlated to recombination rate (Estimate for recombination rate = -3.475e-03, *P* = 9.3e-3, Multiple R^2 of the model = 0.004; Table F in S1 Text).

Parental contributions to ancestry differ among islands. For example, the Spanish sparrow is the minor-ancestry parent to Corsican and Cretan populations, while the house sparrow is the minor-ancestry parent for the Sicilian populations. Taking advantage of this variation, we addressed whether *within*-islands differentiation was correlated to the differentiation between the focal island and their minor-ancestry parent species. For Sicily and Corsica, there is a significant correlation between *within*-island differentiation and differentiation to their minor-ancestry parent; the house and Spanish sparrow, respectively (Corsica: R^2 = 0.002, *P*: 3.58e-4; Sicily R^2 = 0.001, *P*: 0.022; Fig 6A) but not to the alternative parent species in either case (Fig H1 in S1 Text). Differentiation within Crete was not correlated to differentiation to any of the parental species (Fig 6A; Fig H1 in S1 Text). Similarly, there is a pattern where *within*-island outliers are more differentiated to the minor-parent than the genome-wide neutral expectation in two out of three islands (Fig 6B) as outliers within Crete and within Corsica are more differentiated from the Spanish sparrow than expected based on the overall genome-wide average. Conversely, in Sicily *within*-island outlier loci are not more differentiated to the minor-parent, the house sparrow (Fig 6B). We also find higher divergence of *within*-islands outlier loci (1% F_{ST} outliers) to the minor- than to the major-ancestry parent. Outlier loci within Corsica were significantly more differentiated from the Spanish sparrow (the minor-parent) than from the house sparrow (t = -6.22, df = 519.4, *P* = 1.01e-09), as were outlier loci within Crete (t = -2.96, df = 668.2, *P* = 3.17e-3). Outliers within Sicily, where the house sparrow is the minor-ancestry parent, are significantly more differentiated from house sparrow than from Spanish sparrow than expected (t = 3.76, df = 679.8, *P* = 1.81e-4; Fig 6B).

The degree of genomic stabilization can affect the potential for genomic differentiation in hybrid lineages, as both parental alleles are expected to segregate in populations with genomes that are not stabilized. To evaluate whether island populations differ in the degree of genomic stabilization we estimated the rate of fixation of differentially fixed parental alleles. Crete
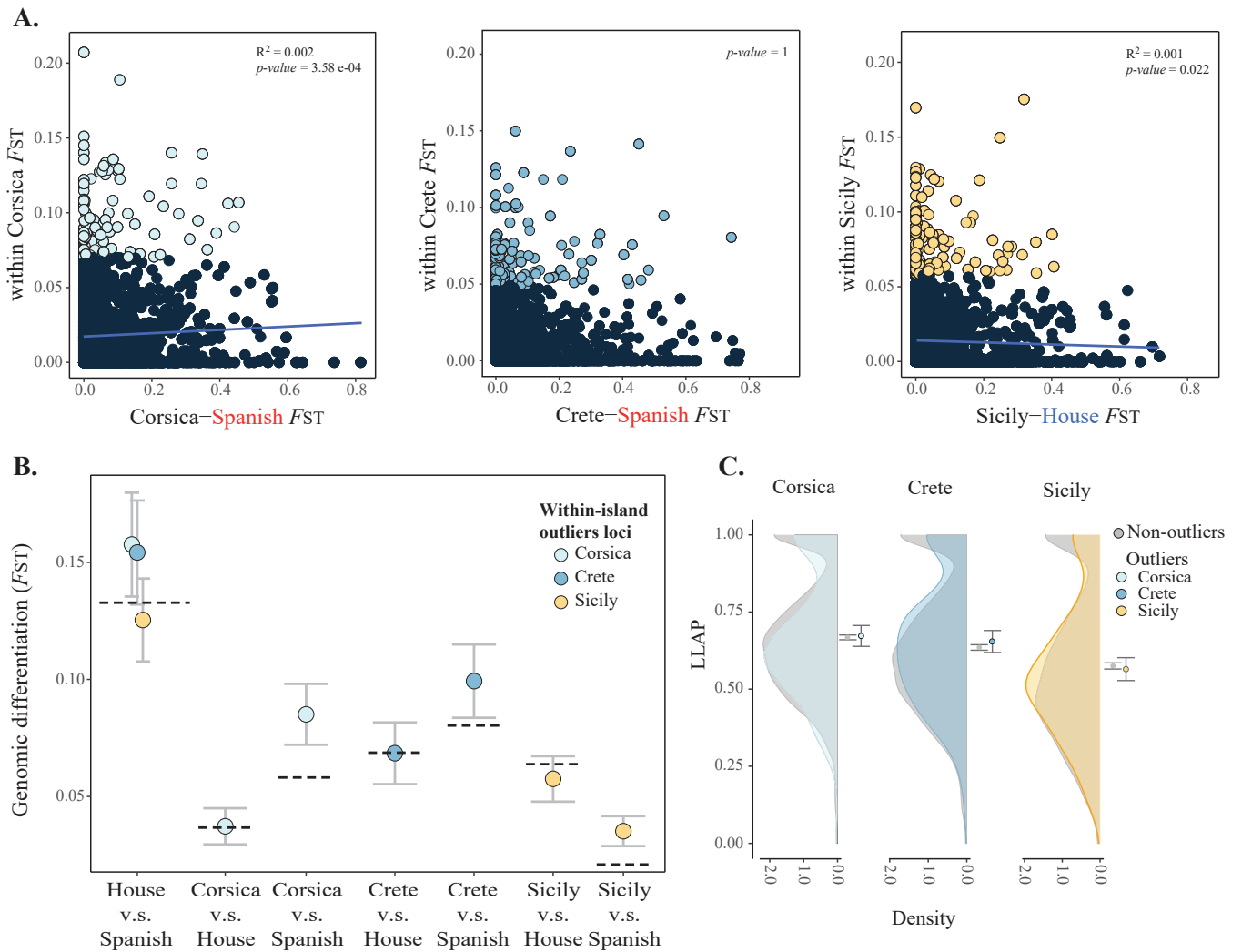
Fig 6. Effects of divergence from parental species and minor–, major parent ancestry on differentiation. A. Correlations of *within*-island F_{ST} and genomic differentiation from the minor-ancestry parent (island vs. minor-parent F_{ST}). 1% F_{ST} outliers are indicated in coloured dots in contrast to the non-outlier loci, in black. **B.** Parental differentiation (house-Spanish F_{ST}) and island-parent F_{ST} for *within*-island outlier loci compared to the genome wide average (dashed line). **C.** Density of per locus local ancestry proportion (LLAP) for within-island F_{ST} outliers, compared to the genome wide distribution. Means and 95% confident intervals of outlier and non-outlier loci are depicted.

https://doi.org/10.1371/journal.pgen.1010027.g006

shows the highest degree of fixation with a rate of 0.17, while Sicily has a fixation rate of 0.003 and Corsica with a negligible number of fixed loci parentally diverging. We further evaluated variation on the fixation rate of major and minor parental alleles independently, to address if there is evidence of preferential fixation of alleles from one of the parent species to reduce incompatible interactions. As expected, the ratio of fixed loci in the islands is higher for the ancestry from the major parent (Fig I in S1 Text). This pattern is found for two out of the three islands. While Crete has inherited more fixed loci from the house sparrow, Sicily present a higher ratio of fixed loci inherited from the Spanish sparrow. Hence loci from the major parent are more frequently fixed in these populations. Among the islands, Crete shows the highest rate of fixation of the major-parent ancestry (ratio = 0.168 for house sparrow ancestry), follow

by Sicily (ratio = 0.0014 for Spanish ancestry), while Corsica has an approximately equal number of fixed sites inherited from the two parent species (Fig I in S1 Text).

We also evaluated if divergence between the parental species (house-Spanish F_{ST}) affected *within*-island differentiation. Parental differentiation was weakly correlated to differentiation within Corsica (correlation = 0.058, R^2 = 3.4e-3, P = 5.34e-5) but not to any other *within*-island differentiation (Fig H2 in S1 Text). Consistent with this, Corsica outlier loci also had higher parental differentiation than expected by chance (t = 2.15, P = 0.033; Fig 6B), but this does not hold true for Crete (t = 1.852, P = 0.065) or for Sicily (t = -0.811, P = 0.42). Finally, differentiation *within*-Corsica and *within*-Sicily was not correlated to differentiation among populations within either of the parental species (Fig J in S1 Text) and differentiation within the parental species was not higher than expected by chance for outlier loci from these islands (Fig K in S1 Text). Interestingly, genome-wide differentiation within Crete was negatively correlated to differentiation within the house sparrow, and differentiation to the house sparrow was lower than expected by chance for outliers within Crete (Figs J and K in S1 Text). Jointly, these results suggest that outlier loci among populations within islands are not dependent on differentiation among populations within each of the parent species.

We assessed the effect of ancestry divergence on genomic differentiation within islands. We use a per locus local ancestry proportion (LLAP, estimated using whole genome sequencing data from [5], [22] and [61]). Whereas a LLAP of 0 corresponds to only Spanish ancestry, 1 corresponds to pure house sparrow ancestry. The distribution of the LLAP does not differ between the *within*-island F_{ST} outlier and the non-outlier loci for any of the islands (t-tests with t = 0.27, 1.04 and -0.53, for Corsica, Crete and Sicily, respectively; P>0.05 for all islands; Fig 6C). A post hoc correlation analysis shows that *within*-island differentiation is not affected by ancestry (R^2 ranging from 5.1e-4 to 1.8e-5, with P>0.05 for all islands; Fig L1 in S1 Text). We further tested if the local ancestry (LLAP), estimated from whole genome data, affected the probability of a locus to be highly differentiated within islands (1% F_{ST} outlier loci; Fig L2 in S1 Text). Examining outliers with extreme values of LLAP only, we found that outliers within Sicily more frequently have excess of Spanish ancestry compared to the genome wide expectation (mean frequency proportion based on 10000 random draws = 0.31), whereas Corsica and Crete outliers display an excess of house ancestry (mean frequency proportion based 10000 random draws = 0.804 and 0.799, respectively; Fig L2 in S1 Text). Even though comparison with proportions from a similar resampling analysis for non-outlier loci shows that the excess ancestry is higher than expected given the genome wide levels of ancestry (Corsica: t = 27.175, Crete: t = -30.846, Sicily: t = -4.369, all P<1.25e-05), the ancestry pattern of outliers is generally very similar to that of the general genomic background (Fig L2 in S1 Text).

Finally we evaluated whether genomic blocks with minor-parent-ancestry were more frequently found in regions with high recombination rate, and if such relation has an effect on genomic differentiation, by performing genome-wide correlations between recombination rate and per locus local ancestry proportion (LLAP). Overall recombination rate only explains a small proportion of the variation on minor-ancestry frequency (adjusted R2 < 0.008, Table K in S1 Text). There is a significant positive relationship between minor-ancestry and recombination rate in the Corsica population (whole genome Pearson's correlation: adjusted R2 < 0.0006, P < 2.2e-16), implying that there are more minor-ancestry blocks in higher recombination regions.

For Crete and Sicily the results are inconclusive, while the variation explained is low (R2: 8.0e-05 and 0.008, respectively) the relationships appears to be negative (Fig M and Table K in S1 Text), contrary to what would be expected if there is a higher rate of purging minor parent ancestry in low recombination regions than in regions of high recombination.

Discussion

While evidence for a creative role of hybridization in evolution is piling up, little is known about how the genomes of hybrid taxa can freely differentiate in response to local selection pressures. Investigating the factors that explain how hybrid taxa can differentiate within lineages of the Italian sparrow, we find surprisingly high genomic differentiation among populations within islands, separated by relatively short distances in light of the dispersal ability of the species [62]. A discriminant function analysis classifies 75–95% of the individuals to the correct population within islands. This local differentiation suggests that there is potential for adaptive divergence within this hybrid species. However, there is more pronounced differentiation among islands, approximately five times higher than *within*-island differentiation, as expected from populations isolated by strong physical barriers.

Interestingly, we find a weak albeit significant correlation between genomic differentiation and a measure of divergent selective sweeps between islands in two out of three comparisons of the islands pairs. However, contrary to our expectation of positive correlations between signatures of divergent selection and genomic differentiation (Fig 1B), we found a weak negative relationship. This does not support a scenario where divergent parental alleles are fixed in response to divergent natural selection. Initial genome stabilization processes, determining admixture proportions, may have been more important than divergent ecological selection. For instance, purging of parental incompatibilities during genome stabilization may have limited the variation for selection to act upon. However, the unexpected direction of the relationship might also reflect that the effects of selection are weak, as selective sweep statistics only explain a very small proportion of the variation in genomic differentiation among islands. Differentiation within islands, likely to have arisen after initial genome stabilization resulting in the island specific admixture proportions, is poorly explained by signatures of selective sweeps. However, within Sicily, haplotype homozygosity (iHS) is weakly correlated with local genomic differentiation. While the weak patterns found in this study offer little support for an important role for divergent selection in population diversification, previous findings are consistent with a role for selection in population differentiation in the Italian sparrow. For instance, local differences in beak shape are best explained by climate and diet for island populations [56]. On the Italian peninsula, population variation in beak shape is best explained by precipitation and genomic differentiation is best explained by temperature [38,55]. These findings are consistent with the large body of work suggesting that hybridization provides the variation facilitating adaptive variation across a range of taxa [2,63–65]. The extent to which signals of selection may be confounded by historical selection acting in the parent lineages, or more recent selection occurring on the hybrid and whether the time frame of hybridization is too short for haplotype based signals to develop, is however not known. However, genomic regions identified as being under selection in the hybrid lineages, using haplotype-based tests, are similar as those previously detected in the house sparrows using whole genome data [61]. This could suggest that observed signals of selection reflect historical selection pressures, but does not exclude the possibility that additional contemporary selection is also reflected in these signatures.

Differences in the degree of genome stabilization can also influence patterns of differentiation among the islands, as purging of incompatibilities and stochastic fixations of parental alleles affect the composition of hybrid genomes. A vast majority of Italian sparrows have house sparrow mitochondrial genome, and among regions that are fixed for house sparrow ancestry across all island populations of Italian sparrow, an excess of nuclear regions with mitochondrial function have been identified [22,23]. This suggests that there has been stabilization of at least parts of the genomes of these hybrid lineages. We find additional evidence suggesting differences in the degree of genome stabilization among the islands. Overall fixation

rates as well as fixation of loci from the major-ancestry parent varied across the islands. Crete has the highest fixation rate, with an elevated fixation of house sparrow alleles follow by Sicily that has a higher fixation rate of Spanish sparrow alleles. Corsica presents the lowest overall fixation rates and did not have differentially fixed alleles from either of the parent species.

We find some evidence suggesting that the same genomic regions repeatedly are involved in population divergence. Differentiation within Corsica is significantly correlated to that within Sicily, but differentiation within Crete is not correlated to that of the other islands. Although our analyses may lack statistical power to detect such correlations, this could also be due to the contribution of *P. domesticus biblicus*, a house sparrow subspecies distributed across the Middle-East, to the population on Crete. This introgression may also have contributed to Crete forming a third cluster in the Admixture analysis. In addition, Corsica shares a higher proportion of the outlier loci than expected by chance with both Crete and Sicily, while the proportion of outliers shared between Crete and Sicily is not higher than expected by chance. These results may, to some degree, support the hypothesis that loci involved in differentiation may be limited to specific genomic regions and are reused across hybrid lineages. Differentiation within island populations could occur in genomic regions that are not under strong selection to fix alleles that are divergent between the parents, after initial genome stabilization where major incompatibilities are sorted, as these regions are likely to be under less strong negative selection.

Measures indicative of selection are consistent across populations and correlated between islands. Hence, the findings of some degree of shared differentiation could partially be explained by similar selection landscapes for all populations of this hybrid taxon or by specific allelic combinations available to selection. Earlier work has shown that the same genetic composition as in the wild ancestor repeatedly arise in lab-crosses of *Helianthus* sunflowers [46] and in younger and older lineages of Lycaides butterflies [47]. It remains to be investigated to what extent the similarity in selection landscapes is caused by historically shared selection in the ancestral populations of the parental species, selection for a functional admixed genome [22,30], stabilizing selection linked to human commensalism [61] or parallel selection for adaptation to insularity. A shared ancestral selection landscape could lead to bias in which parental alleles are retained or more prone to be lost or selected against. For instance, the Spanish sparrow is not considered commensal across most of its range, whereas the Italian and the house sparrow share a commensal ecology. Potentially resulting in consistent selection for specific house sparrow alleles in the independent island lineages of Italian sparrow. We also find variation in the strength of correlation of differentiation among subpopulations comparisons. This pattern is consistent with some degree of parallelism in selection rather than background selection, as we would expect differentiation to be correlated across all comparisons in case background selection strongly limits which areas of the genome are free to vary. If parallel selection is pronounced, we would instead expect the relationship between genomic differentiations to be stronger in some pairs than in the other pair-wise comparisons, and hence variation in the strength of correlations as observed. However, neither of these forces strongly affected the distribution of differentiation, as none of the comparisons were significant when correcting for multiple testing.

A major finding is the limited evidence for genome structure in shaping local differentiation. Variation in the underlying recombination rate landscape may mould the landscape of differentiation [66]. It has been shown to affect the genomic differentiation, generating correlated patterns of differentiation in divergent populations of mice, rabbits [51], flycatchers [48], stonechats [67] and warblers [68], among others. In admixed lineages, selection against minor parent ancestry has been hypothesised to generate patterns of strong correlation between measures of introgression and recombination rate [20,49,52]. This type of selection might be

expected to reduce the genetic variation available for differentiation among hybrid populations. As only the latter process is hybrid specific, a decoupling of the correlation between recombination rate and differentiation present among the parent species is expected if purging of minor parent alleles is important in hybrid taxa (Fig 1A). Recombination rate only explains a small fraction of the variation of minor-parent ancestry proportion. While we find a weak positive relationship between recombination rate and minor-parent-ancestry proportion in Corsica, the pattern is reverse in the other two islands. However, an interesting finding is the stepper correlation between differentiation and recombination rate for the parent species than that among Italian sparrows within islands (Fig 4A). This could suggest that selection against incompatible minor parent alleles in low recombination regions reduces the potential for differentiation in these regions within the hybrid species. However, overall very little of the differentiation within islands is explained by recombination rate, despite the observation of a weak correlation in Sicily. Furthermore, recombination rate overall did not significantly improve models explaining within island differentiation.

Differentiation between the parent species could potentially affect the diversity available for adaptation in the hybrid (Fig 1C), as sorting of ancestry blocks during the genome stabilization process could lead to either fixation of a single ancestry across the hybrid lineage, or of alternative parental blocks in independent hybrid populations. In regions of low parental divergence a lower number of segregating alleles for selection to act on is expected, especially if within-parent diversity is low. A higher evolutionary potential for more divergent loci would be consistent with findings that hybrids from more divergent parent species are morphologically more novel [69,70]. Nevertheless, dominance patterns could also affect the resulting phenotypes in early generation of hybrids [71]. On the other hand, genomic regions of high divergence between parental species can harbour potential genomic incompatibilities in the hybrid taxon. This could generate a negative relationship between the genomic differentiation in the hybrid populations and highly divergent parental loci as the hybrid can only fix ancestry from one of the parent species (Fig 1C). Our data does not lend support to any of these predictions, as we find that overall differentiation between the parent species explains neither the degree of differentiation within islands, nor improves the fit of the models evaluating differentiation within islands (logistic and GLM-models). This could partly reflect the high levels of polymorphism segregating in both parent species and low levels of fixed differences between parent species in this data set. However, highly differentiated loci across populations of the Italian sparrow in mainland Italy have previously been found to present low parental differentiation [38], suggesting that constraints might have played an important role during the stabilization of the hybrid genome, limiting the variation available to selection. Finally, an additional source of variation in hybrid species could stem from variants that segregate within the individual parental species, but we found no evidence for within-parent differentiation affecting differentiation within the hybrid species.

Whether ancestry is a determining factor for how genomic differentiation is distributed in the hybrid genome is not easily disentangled. The divergence in ancestry proportion from the minor–major parent among island populations of the Italian sparrow [22] enables us to test whether differences in ancestry has affected population differentiation after establishing the admixture proportions during early stages of genome stabilization. Purging of genomic incompatibilities, in form of minor parent ancestry blocks, plays an important role in determining the genetic variation in the Italian sparrow [23,24]. A range of studies has suggested that the probability of retaining neutral ancestry is higher in genomic regions with a high recombination rate [20,41,49,52]. To address if minor parent ancestry, in spite of selection against incompatibilities, could be involved in adaptation within the Italian sparrow, we investigated whether minor parent ancestry was important for differentiation. We did not find any

clear effect of ancestry in population differentiation within islands, as highly differentiated outlier loci were not found in minor-parental ancestry blocks more frequently than expected by chance. However, we found significant correlations between local differentiation and the differentiation to the minor-ancestry-parent for two out of three islands, but with opposing signs. Overall F_{ST} outliers are also more differentiated from the minor-parent blocks than expected based on genome-wide levels of differentiation. This suggests that alleles from the minor ancestry parent segregate at loci that are strongly differentiated within islands. As our findings are mixed this would be interesting to investigate further with e.g. whole genome data.

Recombination rate can determine how ancestry is distributed across the hybrid genome [20,49] and may affect the effect that ancestry has on genomic differentiation. The probability for minor-ancestry blocks to rapidly decouple from potential incompatibilities with the major-parent genetic background increases with recombination rate [20]. This affects how easily regions with minor parent ancestry are retained in low recombination areas, and hence affects the variation available for selection that can fuel divergence between hybrid lineages. Interestingly, the probability of being among the 1% most differentiated loci is best explained by how differentiated a given island population is to the house sparrow. Neither recombination rate, the ancestry of the region, nor the differentiation to the Spanish sparrow significantly affected the degree of differentiation or the probability that the locus was an outlier. This is an interesting finding as Runemark *et al.*, [22] previously also found a bias towards house sparrow ancestry in loci consistently inherited from one parent species across island populations. Specifically, they identified an enrichment of mito-nuclear loci and loci involved in DNA-repair. Potentially, these findings could be indicative of some constraints on differentiation from the house sparrow, as most Italian sparrows are fixed for house sparrow mitochondrial haplotypes [5,22]. Another factor that could contribute to this pattern is the overall lower nucleotide diversity and population size [5,38,61] of the Spanish sparrow that could be consistent with a higher incidence of fixation of mildly deleterious alleles.

Conclusion

Taken together, our findings of correlated differentiation patterns among islands and sharing of outlier loci as well as similar selection pressures signatures within islands may suggest that similarity in selection pressures and/or constraints can contribute to parallelism in genome evolution in the hybrid Italian sparrow. Interestingly, we find that the negative relationship between recombination rate and differentiation expected due to linked selection, being stronger in low recombination regions, was significantly stronger in the parent-parent comparison than within the three hybrid lineages. This could be consistent with a lower differentiation in low recombination regions within the hybrid lineages, as expected if purging of minor parent alleles reduces the variation available for divergence. However, a logistic model revealed that differentiation to the house sparrow is the overall best predictor of the probability of outlier status. Jointly, this suggests that selection interacts with constraints linked to admixture during the stabilization of hybrid genomes.

Materials and methods

Ethics statement

All relevant sampling permits were obtained from the regional authorities and handling of birds was conducted according to their guidelines. (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)–Prot 11177, 23557, Consejería de Industria, Energía y Medio Ambiente (Spain), Norwegian Food Safety Authority (Norway),

Bundesamt für Umwelt BAFU, Abteilung (Switzerland)) and Ministry of Education and Science (Republic of Kazakhstan). Permits approval was granted by the above named boards in the corresponding country of sampling.

Background

The Italian sparrow originated from hybridization between the house and Spanish sparrow, likely during the spread of the commensal house sparrow to Europe in the wake of the introduction of agriculture [61,72]. The parental species diverged approximately 0.68 million years ago [61]. In addition to the distribution on the Italian peninsula, Italian sparrow populations are also found on some Mediterranean islands. These insular populations have strongly differentiated genomes, with different contributions from each parent species [22], and exhibit phenotypic divergence with island specific beak shape matching local temperature and diet [56]. Furthermore, the island populations are evolutionarily independent and are hypothesized to have arisen from individual hybridization events [22]. Runemark *et al.*, [22] show low concordance (pairwise correlations between islands) of f_d statistic [73] across the same windows along the genome, as well as significant differences in ancestry tract lengths between islands, suggesting that the islands populations have evolved independently. These approaches have previously been used to suggest that a single ancient hybridization event resulted in differential lineages of cichlid fishes [2].

Sampling and sequencing

Three populations of Italians sparrows were sampled from each of the islands, Sicily (n = 76), Crete (n = 77) and Corsica (n = 70) in March-June 2013 (Fig 2A and S1 Table). On each island we sampled individuals from three geographically separated populations (Figs 2B and 3C). Population sample size varied between 16 and 30 (S1 Table). We sampled reference house sparrow parent populations from Norway (n = 11), and Spanish sparrows from Kazakhstan (n = 10). To increase the number of sampled individuals from the parent species, for analyses that work better with approximately equal sample sizes of all taxa, we added house sparrow samples from Switzerland (n = 17) and France (n = 18) and Spanish sparrow samples from the Gargano peninsula (n = 14) and Spain (n = 23); (S1 Table). All birds were caught using mist nests, and blood was sampled from the brachial vein and stored in Queen's lysis buffer. All necessary permits were obtained from relevant local authorities prior to sampling. DNA was extracted using the Qiagen DNeasy Blood and Tissue Kit, (Qiagen N.V., Venlo, The Netherlands) and the product was stored in Qiagen's buffer EB prior to sequencing. We used a RAD-tag approach; library preparation, sequencing, de-multiplexing and removal of adapters were done by Ecogenics GmbH (Balgach, Switzerland; www.ecogenics.ch). Specifically, the restriction enzymes EcoRi and MseI were used for double digest restriction-site associated DNA sequencing (ddRAD). Fragments between 500-600bp were selected with gel electrophoresis and then sequenced using an Illumina Nextseq500 machine with a 1x75bp read sequencing format.

Data processing and variant calling

First, the quality of all RAD sequences was checked using FASTQC [74]. Raw reads were filtered using the module process_radtag from the software Stacks [75]. Reads shorter than 73 base pairs were discarded as well as those with an uncalled base. To ensure high confidence-based calls, a Phred quality score of 20 (99% accuracy) was used as threshold across a sliding window fraction of 0.1 of the read length. We used BWA-MEM (v 0.7.8) [76] to map the reads to the house sparrow reference genome [5] using default parameters. We re-aligned indels with GATKs (v 3.7) RealignerTargetCreator and IndelRealigner [77,78] and called the variants using HaplotypeCaller [78]. For a detailed description of the variant calling pipeline, see

Cuevas *et al.*, [38]. We filtered SNPs using Vcftools v. 0.1.14 [79], setting the filter parameters to —max-missing 0.8 (20% missing data allowed), —minDP 10.00, —minGQ 20.00 and —maf 0.02. Using PLINK v. 1.9 [80] we pruned linked sites with an $R^2 > 0.1$, calculated from 100 kb sliding windows and a step of 25 bp. VCF-files containing different set of individuals were generated to suit the different analyses (S2 Table). After filtering VCF files contain between 2224 and 2856 high-quality SNPs and with mean proportion of per individual missing data not larger than 0.13.

I) Genomic differentiation within- and between islands

We tested the hypothesis that the degree of divergence is significantly higher between islands than within islands, reflecting long periods of independent evolution. To this end, we first illustrated the overall divergence between the islands and populations using a Principal Component Analysis as implemented in glPca() in the R package ADEGENET 2.0 [81]. We also evaluated the level of clustering in the data through estimating the cross-validation error for K = 1 to K = 9, and estimated the probability of each individual belonging to these clusters using ADMIXTURE v.1.3.0 [82]. To illustrate the extent to which the divergence was aligned with the axis of parental differentiation, three parental populations of each species were included in these analyses, resulting in 316 individuals approximately equally distributed across the three species (S2 Table) in a VCF file containing 2224 SNPs (S2 Table). To further illustrate the degree of differentiation within islands, we also performed a Discriminant Analysis of Principal Components (DAPC) within each island to address to which extent the local populations can be correctly classified based on the available variation, we used the dapc() function from ADEGENET 2.0 [81]). To characterize potential variation in genomic diversity we also estimated nucleotide diversity for each population. The analyses were performed in 100kb sliding windows with 25-kb steps using vcftools v. 0.1.14 [79]. For estimates of nucleotide diversity non-variant sites were retained, and we did not filter on minor allele frequency.

We investigated whether the differentiation was stronger between islands than within islands, using two approaches. First, we estimated global F_{ST} among populations within islands, as well as pair-wise F_{ST} among islands in 100kb windows using vcftools v. 0.1.14 [79]. The window size was selected as linkage disequilibrium in sparrows is known to decay within this distance [5], and the windows contained on average 1,5 (SD: ±0.89) SNPs. We used a Monte Carlo permutation paired t-test to investigate if pairwise F_{ST}-values were higher among- than within islands.

Second, we used an Analysis of Molecular Variance (AMOVA) to formally address what proportion of genetic variance is explained by differentiation among islands, among local populations within islands, within local populations and variation within individuals. We transformed the VCF to a genlight object, where levels of divergence were defined, using the ADEGENET R-package and ran an AMOVA with the poppr.amova() function from the POPPR R-package [83,84]. We assessed significance by randomization of population assignments using a Monte Carlo test with 1000 permutations implemented in the randtest() function from the ADE4 R-package [85]. Several cut-off of missing-ness per loci were also use (5%, 10% and 20%) with the missingno() function to evaluate the sensitivity of the test (Table L in S1 Text). Loci with high percentages of missing data can disturb the Euclidian distance matrix performed by AMOVA.

II) The relationship between genomic differentiation and recombination rate

We examined the hypothesis that hybrid genome formation influences the association between differentiation and recombination rate. Our rationale was that if purging of minor ancestry is

stronger in low recombination regions, this reduces diversity in these regions and therefore acts as a constraint on differentiation such that we expect greater differentiation in higher recombination regions where the effect of purging is weaker. Alternatively, if local selective sweeps play a more important role in shaping hybrid genomes, we would expect greater differentiation in low recombination regions [48,51]. As the relative strength of these processes is unknown, we used the relationship between differentiation and recombination rate between the parent species as a null expectation, and tested if there was a deviation from this relationship in the direction expected from purging of minor parent ancestry in the hybrid populations compared to the parent species (see Fig 1A). To this end, we tested for differences in the slopes of individual linear regressions of F_{ST} and recombination rate. We also evaluated a significant interaction between lineage combination (parent-parent vs. within island) and recombination rate on F_{ST} using independent linear model per island. We used recombination rate estimates from Elgvin *et al.*, [5]. We also evaluated the significance of the relation between genomic differentiation and recombination rate within each island using Pearson's correlation tests. We used Bonferroni corrected *P*-values to account for multiple comparisons.

III) The concordance of patterns of selection and genomic differentiation

To address if elevated genomic differentiation is driven by strong divergent selection, we performed Bonferroni corrected Pearson's correlations of F_{ST} between island pairs to their cross-population Extended Haplotype Homozygosity statistic (xp-EHH) [60], which is designed to compare haplotype lengths between populations (between islands in this case) in order to detect selective sweeps. We further investigated whether regions putatively under selection within-island are independent across islands and whether they coincide with areas of elevated differentiation. We performed pairwise Bonferroni corrected Pearson's correlations between each island pair of the integrated haplotype homozygosity score (iHS) [86] developed for detecting positive selection within a population, in this case calculated within each island. Then, we tested if putative concordance in selection may result in correlated patterns of differentiation in islands, through investigating the correlation between iHS-scores and genomic differentiation for each island. We estimated long range haplotype statistics through phasing data with SHAPEIT/v2.r837 [87,88] and converted the resulting VCF-file using the vcfR R-package [89]. We then used the functions data2haplohh(), ihh2ihs() and ies2xpehh() from the rehh R- package [90,91] to prepare the data, estimate the integrated haplotype homozygosity score (iHS) and estimate Extended Haplotype Homozygosity (XP-EHH), respectively.

IV) Distribution and repeatability of differentiation across the genome

To test if the differentiation landscape between populations within islands is correlated to that within other islands and between islands, as would be expected if differentiation is affected by the underlying recombination rate landscape and constraints or similar selection pressures acting on the populations, we performed pairwise Pearson's correlation tests on F_{ST} estimates. We tested if global F_{ST} estimates within one island were significantly correlated to these within another island, as well as if between-island differentiation was significantly correlated to global F_{ST} within any of the islands using a resampling approach and Bonferroni corrections for multiple testing.

In addition, we investigated to what extent the same loci were among the most strongly differentiated on different islands. We estimated the proportions of the 1% most differentiated loci that were shared between each island pair. We then investigated if this proportion of shared F_{ST} outliers was higher than expected by chance using a series of χ^2-test for each pairwise comparison, applying Bonferroni corrections for multiple testing. We also provide a list

of candidate genes that are in the vicinity of outliers shared between comparisons. We extracted coding regions within 100kb distance from the shared loci, as linkage decays at approx. 100kb in the house sparrow [5], using the house sparrow annotation file developed by Elgvin et al. (2017).

To further differentiate whether background selection or adaptive parallel selection determine shared patterns of differentiation we run correlations of all possible pair-wise comparison of subpopulations *within-* and *between*-islands. The rationale is that background selection is expected to give rise to correlations in all comparisons as the recombination rate landscape is projected to be constant, whereas parallel selection pressures would generate correlations only in the comparisons where these selection pressures are shared. To correct for multiple testing we performed a resampling approach by running 100 iterations of the correlations.

V) Patterns of local genomic differentiation in relation to parental contributions to the genome

To evaluate how multiple factors, like genomic parental contribution, parental differentiation and recombination rate among others, may affect which loci are free to vary within the Italian sparrow we performed a generalized linear model (GLM) using within-island F_{ST} as the response variable: F_{ST} = per locus local ancestry proportion (LLAP) + recombination rate + island + island to house sparrow differentiation (F_{ST}) + island to Spanish sparrow differentiation (F_{ST}) + parental differentiation (house-Spanish F_{ST}). We also evaluated how these factors affected the probability of a locus belonging to the 1% most differentiated loci within an island using a similar model with a logistic regression where the response variable was the Pr(outlier). In addition, we performed logistic regressions and GLM individually for each island, excluding the island term. As post hoc tests, we examined Bonferroni corrected Pearson correlations of within-island differentiation against differentiation of the island to each of the parental taxa as well as between the parent species. We also assessed whether highly differentiated loci found in the Italian sparrow are also involved in the genomic differentiation among populations within each parent species (within-house F_{ST} and within-Spanish F_{ST}).

We evaluated the degree of genomic stabilization in the different island populations by comparing fixation rates of parentally divergent loci in the Italian sparrow. We also investigated fixation of major- and minor-ancestry parent individually. Loci fixed for different alleles for the two parent species (F_{ST} = 1) were identified from whole genome sequencing (WGS) data for the parental species retrieved from [61] and [5]. For these loci fixation rates were evaluated on WGS data from [22] for Crete, Corsica and Sicily. A total of 17887 SNPs were found to be differentially fixed between parental species and these loci were used to calculated fixation levels in one subpopulation of the Italian sparrow per island.

To address if variation in minor parent ancestry affects within-island differentiation, we tested the correlation between genomic differentiation and the proportion of per locus local ancestry (LLAP) reflecting the relative contribution of each parent species. We estimate a per locus local ancestry proportion (LLAP) using whole genome data from [22], [61] and [5]. To this end we phased data using SHAPEIT/v2.r837 [87,88] and inferred ancestry estimates using LOTER [92]. These were then translated into a per locus local ancestry proportion (LLAP), where values of 0 correspond to loci where only Spanish ancestry is present across all individuals in the population evaluated and 1 corresponds to pure house sparrow ancestry. We estimated the LLAP separately for each island. We also tested if highly differentiated loci were found in blocks with high allele frequencies from major- (greater than 65% major parent alleles) or minor parent ancestry (greater than 65% minor parent alleles) more frequently than expected by chance. This was achieved by comparing the confidence intervals from 10000

resamplings of 8 outlier loci to the value for the entire F_{ST}-distribution to assess significance. The same analysis was run for the distribution of non-outlier loci to assess whether the outliers diverge from the neutral expectations.

Finally, we evaluated whether genomic blocks of minor-parent ancestry are more common in regions with high recombination rates, as high recombination rate allows target loci to escape linkage with loci incompatible with the major-parent genomic background. We evaluated to which extent recombination rate explained the proportion of minor parent ancestry through Pearson's correlations between recombination rate estimates retrieved from [5] and the proportion on minor ancestry (LLAP). All data generated in this study can be found in [93].

Dryad DOI

https://doi.org/10.5061/dryad.wpzgmsbns

Supporting information

S1 Table. Sampled individuals from the parent species (the house and Spanish sparrows) and the Italian sparrow.
(XLSX)

S2 Table. VCF files.
(XLSX)

S3 Table. Genes in the vicinity of shared regions of differentiation within island.
(XLSX)

S1 Text. Supporting Figures and Statistics. It includes: **Table A. Per-island population genomic statistics.** Left panel: Mean values of π and within-island genomic differentiation (F_{ST}). Middle panel: t-test for pairwise comparison between genome wide within island F_{ST}, evaluating a significance difference between genome wide *within*-island genomic differentiation (F_{ST}) across islands. Right panel: Mean values of Tajima's D per population within each island. **Table B. Intercept, slope and confidence intervals of the slope of individual linear regression of *within*-island genomic differentiation and recombination rate as well as parent-parent differentiation and recombination rate. Table C. Evaluating the effect that the interaction between recombination rate and the type of comparison (parental differentiation (house-Spanish), which is the null model, and within-island differentiation) has on genomic differentiation (F_{ST}).** Individual linear models per island were run to test if there is a significant interaction between **recombination rate** and **comparison**, as expected if the relationship between recombination rate and differentiation differs between parent species and the hybrid Italian sparrow ([Fig 1A](#)). **Table D. Generalized linear model on *within*-island F_{ST}. Table E. Logistic regressions per island, on the probability of being a local F_{ST} outlier within island. Table F. Generalized linear models, separated by island on within-island F_{ST}. Table G. Concordance of 1. *between*-island divergent selection (xp-EHH) and 2. *within*-island selection (iHS) with genomic differentiation (F_{ST}). 3.** Correlation between islands of their correspondent *within*-island selection (iHS) estimates. **Table H. Number and percentage of *within*-island F_{ST} outlier loci shared between islands.** Chi-squared denote tests for overrepresentation compared to the genome wide average. **Table I. Number and percentage of *within*-island F_{ST} outlier loci identical to *between*-island outliers.** Chi-squared denote tests for overrepresentation compared to the genome wide average. **Table J. Parallel vs. background selection.** F_{ST} comparisons between within-island subpopulations across all islands. *P*-value, correlations estimates and t-estimates are corrected for multiple testing by

resampling and taking mean estimates after 100 iterations of correlations. **Table K. Linear model of recombination rate and minor-parent ancestry across islands.** The models are performed using values of Log10 of recombination rate as a predictor of local ancestry (LLAP) and dividing these in quartile bins to group the recombination rate values and facilitate interpretation. **Table L. Different cut-offs for the Analysis of Molecular Variance (AMOVA) across islands and populations within islands.** Several cut-offs for missing-ness per loci were used: 5% (see Table 1), A. 10% and B. 20%, but the results from the AMOVA did not change substantially. **Fig A. AMOVA significance—Randomization via permutation.** Monte Carlo test with 1000 permutations implemented in the randtest() function from the ADE4 R-package to evaluate significance. Black line denotes the observed values of Sigma (Variance in each hierarchical level). **Fig B. Concordance of patterns of selection and genomic differentiation.** Correlations of the integrated haplotype homozygosity score (iHS) and genomic differentiation (within-island FST). 1% FST outliers are indicated in coloured dots in contrast to the non-outlier loci, in black. **Fig C. Distribution of Tajima's D per population in each island. Fig D. Correlation of within-islands differentiation across the three Mediterranean islands.** Bonferroni corrections of the p-values are reported. **Fig E. Correlation of within-islands differentiation vs. between-islands divergence.** Adjusted p-values after resampling and Bonferroni corrections. **Fig F. Parallelism of *within*-island pairwise FST.** Pairwise F_{ST} correlations between populations within island "A" to pairwise F_{ST} estimates of populations within island "B", highlighted in green. Significant correlations before correction for multiple testing highlighted in red. Abbreviations of the comparisons are as follow: CORSICA populations: Muratello (Mur), Pianiccia (Pi), Tiuccia (Pi). CRETE populations: Istro (Is), Mithimna (Mi), Perama (Pe). SICILY populations: Cos (Co), Enna (En), Naxos (Na). Thus pair-wise F_{ST} between Muratello vs. Pianiccia is abbreviated as "*Cor_Mur.Ti*". Similarly, pair-wise F_{ST} between Enna vs. Naxos is abbreviated as "*Sic_En.Na*". Estimate values are corrected for multiple testing using a resampling approach (Table J in S1 Text). **Fig G. Parallelism of *between*-island pairwise F_{ST} across all subpopulations.** Correlations of pairwise-F_{ST} between subpopulation a (from island "A") and b (from island "B") and its contrast pairwise-F_{ST} between subpopulation a' (from island "A") and b' (from island "B"). **1.** Correlations between Corsican vs. Sicilian subpopulations. **2.** Corsican vs. Cretan subpopulations and **3.** Sicilian vs. Cretan subpopulations. Populations' name of each island are presented in Fig F. Abbreviations of the comparisons are as follow: As an example, pair-wise F_{ST} between Muratello (from Corsica) vs. Enna (from Sicily) is abbreviated as "*Cor.Mur_Sic.En*". Similarly, pair-wise F_{ST} between Perama (from Crete) vs. Naxos (from Sicily) is abbreviated as "*Cre.Pe_Sic.Na*". **Fig H. Correlation of within-islands differentiation and the parental species.** Adjusted p-values after Bonferroni corrections. **Fig I. Fixation rate of parentally differentiated fixed sites across the islands Italian sparrow populations.** Fixation rate is presented individually by ancestry. Continuity of the y-axis is broken (dashed line) to minimize the size of the figure in order to include the extreme values of the distribution. **Fig J. Correlations of within-island differentiation and within-parent differentiation (within-house or and within-Spanish sparrow).** 1% F_{ST} outliers are indicated in coloured dots in contrast to the non-outlier loci, in black. **Fig K. 1.** Intraspecific genomic differentiation in the parental species for the within-island F_{ST} outlier loci. Dash lines represent the within-parent F_{ST} global mean. **2.** t-tests evaluating whether within-island outlier loci present higher/lower values than expected by chance in the within-parent differentiation. **Fig L. 1.** Relation between within-island F_{ST} and per locus local ancestry proportion (LLAP). Results of linear regression reported. Dashed lines depict the 1% outliers threshold. **2.** Frequency proportion of outlier loci found in regions of mainly house ancestry (0.65<LLAP) and mainly Spanish ancestry (LLAP<0.35) (minor-major parental ancestry). Distribution of 10.000 random resampling draws of 8 outlier loci. **Fig M. Recombination rate**

v.s proportion of minor-ancestry (using LLAP, where values of 1 = 100% house ancestry and 0 = 100% Spanish ancestry). Recombination rate is presented in quartiles using whole genome resequencing data retrieved from Ravinet et al (2018), Elgvin et al (2017) and Runemark et al (2018a). Mean and confident intervals of LLAP are shown per recombination rate quantile. Minor ancestors are as follow: Corsica: minor-ancestry from the Spanish sparrow (LAAP = 0 to 0.5), n = 237.523 SNPs. Crete: minor-ancestry Spanish sparrow (LAAP = 0 to 0.5), n = 294.749SNPs and Sicily: minor-ancestry the house sparrow (LAAP = 0.5 to 1), n = 424.739SNPs.
(DOCX)

Acknowledgments

We thank Jo S. Hermansen and Maria Tesaker for help with fieldwork.

Author Contributions

Conceptualization: Fabrice Eroukhmanoff, Glenn-Peter Sætre, Anna Runemark.

Data curation: Angélica Cuevas, Fabrice Eroukhmanoff, Anna Runemark.

Formal analysis: Angélica Cuevas, Fabrice Eroukhmanoff, Mark Ravinet, Anna Runemark.

Funding acquisition: Fabrice Eroukhmanoff, Glenn-Peter Sætre, Anna Runemark.

Methodology: Angélica Cuevas, Fabrice Eroukhmanoff, Mark Ravinet, Anna Runemark.

Resources: Angélica Cuevas, Anna Runemark.

Software: Angélica Cuevas, Mark Ravinet.

Supervision: Fabrice Eroukhmanoff, Anna Runemark.

Visualization: Angélica Cuevas.

Writing – original draft: Angélica Cuevas, Anna Runemark.

Writing – review & editing: Angélica Cuevas, Fabrice Eroukhmanoff, Mark Ravinet, Glenn-Peter Sætre, Anna Runemark.

References

1. McGee M.D., Borstein S.R., Neches R.Y., Buescher H.H., Seehausen O. & Wainwright P.C. (2015). A pharyngeal jaw evolutionary innovation facilitated extinction in Lake Victoria cichlids. *Science* 350: 1077–1079. https://doi.org/10.1126/science.aab0800 PMID: 26612951

2. Meier J.I., Marques D.A., Mwaiko S., Wagner C.E., Excoffier L. & Seehausen O. (2017). Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature Communications* 8: 1–11. https://doi.org/10.1038/s41467-016-0009-6 PMID: 28232747

3. Salazar C., Baxter S.W., Pardo-Diaz C., Wu G., Surridge A., Linares M., et al. (2010). Genetic evidence for hybrid trait speciation in heliconius butterflies. *PLoS Genet* 6: e1000930. https://doi.org/10.1371/journal.pgen.1000930 PMID: 20442862

4. Hermansen J. S., Sæther S. A., Elgvin T. O., Borge T., Hjelle E., & Sætre G. P. (2011). Hybrid speciation in sparrows I: Phenotypic intermediacy, genetic admixture and barriers to gene flow. *Molecular Ecology*, 20(18), 3812–3822. https://doi.org/10.1111/j.1365-294X.2011.05183.x PMID: 21771138

5. Elgvin T. O., Trier C. N., Tørresen O. K., Hagen I. J., Lien S., Nederbragt A. J., et al. (2017). The genomic mosaicism of hybrid speciation. *Science Advances*, 3(6). https://doi.org/10.1126/sciadv.1602996 PMID: 28630911

6. Sankararaman S., Mallick S., Dannemann M., Prüfer K., Kelso J., Pääbo S., et al. (2014). The genomic landscape of Neanderthal ancestry in present-day humans. 507: 354–357. https://doi.org/10.1038/nature12961 PMID: 24476815

7. Norris L. C., Main B. J., Lee Y., Collier T. C., Fofana A., Cornel A. J., et al. (2015). Adaptive introgression in an African malaria mosquito coincident with the increased usage of insecticide-treated bed nets. *Proceedings of the National Academy of Sciences of the United States of America*, 112(3), 815–820. https://doi.org/10.1073/pnas.1418892112 PMID: 25561525

8. Jones M.R., Mills L.S., Alves P.C., Callahan C.M., Alves J.M., Lafferty D.J.R., et al. (2018). Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe hares. *Science* 360: 1355–1358. https://doi.org/10.1126/science.aar5273 PMID: 29930138

9. Abi-Rached L., Jobin M.J., Kulkarni S., McWhinnie A., Dalva K., Gragert L., et al. (2011). The shaping of modern human immune systems by multiregional admixture with archaic humans. *Science* 334: 89–94. https://doi.org/10.1126/science.1209202 PMID: 21868630

10. Malinsky M., Svardal H., Tyers A. M., Miska E. A., Genner M. J., Turner G. F., et al. (2018). Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nature Ecology and Evolution*, 2(12), 1940–1955. https://doi.org/10.1038/s41559-018-0717-x PMID: 30455444

11. Ronco F., Matschiner M., Böhne A., Boila A., Büscher H. H., El Taher A., et al. (2020). Drivers and dynamics of a massive adaptive radiation in cichlid fishes. *Nature*, 589(January). https://doi.org/10.1038/s41586-020-2930-4 PMID: 33208944

12. Svardal H., Quah F. X., Malinsky M., Ngatunga B. P., Miska E. A., Salzburger W., et al. (2020). Ancestral hybridization facilitated species diversification in the lake malawi cichlid fish adaptive radiation. *Molecular Biology and Evolution*, 37(4), 1100–1113. https://doi.org/10.1093/molbev/msz294 PMID: 31821500

13. Svardal H., Salzburger W., & Malinsky M. (2021). Genetic Variation and Hybridization in Evolutionary Radiations of Cichlid Fishes. *Annual Review of Animal Biosciences*, 9(1), 1–25. https://doi.org/10.1146/annurev-animal-061220-023129 PMID: 33197206

14. Schwarzbach A.E., Donovan L.A. & Rieseberg L.H. (2001). Transgressive character expression in a hybrid sunflower species. *Am. J. Bot.* 88: 270–277. PMID: 11222249

15. Rieseberg L.H. (2003). Major Ecological Transitions in Wild Sunflowers Facilitated by Hybridization. *Science* 301: 1211–1216. https://doi.org/10.1126/science.1086949 PMID: 12907807

16. Schumer M., Rosenthal G. G., & Andolfatto P. (2014). How common is homoploid hybrid speciation? *Evolution*, 68(6), 1553–1560. https://doi.org/10.1111/evo.12399 PMID: 24620775

17. Abbott R., Albach D., Ansell S., Arntzen J.W., Baird S.J.E., Bierne N., et al. (2013). Hybridization and speciation. *J. Evol. Biol.* 26: 229–246. https://doi.org/10.1111/j.1420-9101.2012.02599.x PMID: 23323997

18. Mallet J. (2007). Hybrid speciation. *Nature*, 446(7133), 279–283. https://doi.org/10.1038/nature05706 PMID: 17361174

19. Runemark A., Vallejo-Marín M. & Meier J.I. (2019). Eukaryote hybrid genomes. *PLoS Genet* 15 (11), 1–22. https://doi.org/10.1371/journal.pgen.1008404 PMID: 31774811

20. Schumer M., Xu C., Powell D.L., Durvasula A., Skov L., Holland C., et al. (2018). Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science* 360: 656–660. https://doi.org/10.1126/science.aar3684 PMID: 29674434

21. Nice C.C., Gompert Z., Fordyce J.A., Forister M.L., Lucas L.K. & Buerkle C.A. (2013). Hybrid speciation and independent evolution in lineages of alpine butterflies. *Evolution* 67: 1055–1068. https://doi.org/10.1111/evo.12019 PMID: 23550755

22. Runemark A., Trier C.N., Eroukhmanoff F., Hermansen J.S., Matschiner M., Ravinet M., et al. (2018a). Variation and constraints in hybrid genome formation. *Nature Ecology & Evolution* 2: 549–556.

23. Trier C. N., Hermansen J. S., Sætre G. P., & Bailey R. I. (2014). Evidence for Mito-Nuclear and Sex-Linked Reproductive Barriers between the Hybrid Italian Sparrow and Its Parent Species. *PLoS Genetics*, 10(1). https://doi.org/10.1371/journal.pgen.1004075 PMID: 24415954

24. Hermansen J. S., Haas F., Trier C. N., Bailey R. I., Nederbragt A. J., Marzal A., et al. (2014). Hybrid speciation through sorting of parental incompatibilities in Italian sparrows. *Molecular Ecology*, 23(23), 5831–5842. https://doi.org/10.1111/mec.12910 PMID: 25208037

25. Mayr E. (1963) Animal Species and Evolution, *Harvard University Press*.

26. Taylor S.A. & Larson E.L. (2019). Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. Nature Ecology & *Evolution* 3: 170–177. https://doi.org/10.1038/s41559-018-0777-y PMID: 30697003

27. Coyne J.A. & Orr H.A. (1989). *Patterns of Speciation in Drosophila*. Evolution 43: 362.

28. Haldane J.B.S. (1922). Sex ratio and unisexual sterility in hybrid animals. *Journal of Genetics* 12: 101–109. Hill, G.E. (2016). Mitonuclear coevolution as the genesis of speciation and the mitochondrial DNA barcode gap. Ecol Evol 6: 5831–5842.

29. Hill G.E. (2016). Mitonuclear coevolution as the genesis of speciation and the mitochondrial DNA barcode gap. *Ecol Evol* 6: 5831–5842. https://doi.org/10.1002/ece3.2338 PMID: 27547358

30. Runemark A., Eroukhmanoff F., Nava-Bolaños A., Hermansen J.S. & Meier J.I. (2018b). Hybridization, sex-specific genomic architecture and local adaptation. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 373: 20170419.

31. Wagner D.N., Curry R.L., Chen N., Lovette I.J. & Taylor S.A. (2020). Genomic regions underlying metabolic and neuronal signaling pathways are temporally consistent in a moving avian hybrid zone. *Evolution* 74: 1498–1513. https://doi.org/10.1111/evo.13970 PMID: 32243568

32. Eroukhmanoff F., Rowe M., Cramer E.R.A., Haas F., Hermansen J.S., Runemark A., et al. (2016). Experimental evidence for ovarian hypofunction in sparrow hybrids. *Avian Research* 7: 229.

33. Bateson W. (1909). Heredity and variation in modern lights. In *Darwin and Modern Science* (Seward A. C., ed.), pp. 85–101, Cambridge University Press. PMID: 19973580

34. Dobzhansky T. (1936). Studies on hybrid sterility. II. Localization of sterility factors in Drosophila pseudoobscura hybrids. *Genetics* 21, 113–135. https://doi.org/10.1093/genetics/21.2.113 PMID: 17246786

35. Muller H.J.(1942). Isolating mechanisms, evolution and temperature. *Biol. Symp.* 6, 71–125.

36. Orr H. A. (1995). Population genetics of speciation in two closely related wild tomatoes (Solanum section Lycopersicon). *Genetics*, 139, 1805–1813. https://doi.org/10.1093/genetics/139.4.1805 PMID: 7789779

37. Orr H. A. (1996). Dobzhansky, Bateson, and the Genetics of Speciation. *Genetics*, 144, 1331–1335. https://doi.org/10.1093/genetics/144.4.1331 PMID: 8978022

38. Cuevas A., Ravinet M., Sætre G-P., Eroukhmanoff F (2020) Intraspecific genomic variation and local adaptation in a young hybrid species. *Molecular Ecology*. https://doi.org/10.1111/mec.15760

39. Buerkle C. A., & Rieseberg L. H. (2008). The rate of genome stabilization in homoploid hybrid species. Evolution; *International Journal of Organic Evolution*, 62(2), 266–275. https://doi.org/10.1111/j.1558-5646.2007.00267.x

40. Schumer M., Cui R., Powell D. L., Rosenthal G. G., & Andolfatto P. (2016). Ancient hybridization and genomic stabilization in a swordtail fish. *Molecular Ecology* (2016) 25, 2661–2679. https://doi.org/10.1111/mec.13602 PMID: 26937625

41. Moran B. M., Payne C., Langdon Q., Powell D. L., Brandvain Y., & Schumer M. (2021). The genomic consequences of hybridization. *ELife*, 10, 1–33. https://doi.org/10.7554/elife.69016

42. Veller C., Edelman N. B., Muralidhar P., & Nowak M. A. (2019). Recombination, variance in genetic relatedness, and selection against introgressed DNA. *BioRxiv*, 1–35. https://doi.org/10.1101/846147

43. Storchová R., Reif J. & Nachman M.W. (2010). Female heterogamety and speciation: reduced introgression of the z chromosome between two species of nightingales. *Evolution* 64: 456–471. https://doi.org/10.1111/j.1558-5646.2009.00841.x PMID: 19796142

44. Sætre G. P., Borge T., Lindroos K., Haavie J., Sheldon B. C., Primmer C., et al. (2003). Sex chromosome evolution and speciation in Ficedula flycatchers. Proceedings of the Royal Society B: *Biological Sciences*, 270(1510), 53–59. https://doi.org/10.1098/rspb.2002.2204 PMID: 12590771

45. Gainey D.P., Kim J.Y. & Maroja L.S. (2018). Mapping reduced introgression loci to the X chromosome of the hybridizing field crickets, Gryllus firmus and G. pennsylvanicus. *PLoS ONE* 13: e0208498. https://doi.org/10.1371/journal.pone.0208498 PMID: 30566487

46. Rieseberg L., Sinervo B., Linder C., Ungerer M. & Arias D. (1996). Role of Gene Interactions in Hybrid Speciation: Evidence from Ancient and Experimental Hybrids. *Science* 272: 741–745. https://doi.org/10.1126/science.272.5262.741 PMID: 8662570

47. Chaturvedi S., Lucas L. K., Buerkle C. A., Fordyce J. A., Forister M. L., Nice C. C., & Gompert Z. (2020). Recent hybrids recapitulate ancient hybrid outcomes. *Nature Communications*, 11(1). https://doi.org/10.1038/s41467-020-15641-x PMID: 32358487

48. Burri R., Nater A., Kawakami T., Mugal C. F., Olason P. I., Smeds L., et al. (2015). Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of Ficedula flycatchers. *Genome Research*, 25(11), 1656–1665. https://doi.org/10.1101/gr.196485.115 PMID: 26355005

49. Ravinet M., Yoshida K., Shigenobu S., Toyoda A., Fujiyama A., & Kitano J. (2018a). The genomic landscape at a late stage of stickleback speciation: High genomic divergence interspersed by small localized regions of introgression. *PLoS Genetics* (Vol. 14). https://doi.org/10.1371/journal.pgen.1007358

50. Wolf J. B. W., & Ellegren H. (2017). Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics*, 18(2), 87–100. https://doi.org/10.1038/nrg.2016.133 PMID: 27840429

51. Nachman M. W., & Payseur B. A. (2012). Recombination rate variation and speciation: Theoretical predictions and empirical results from rabbits and mice. Philosophical Transactions of the Royal Society B: *Biological Sciences*, 367(1587), 409–421. https://doi.org/10.1098/rstb.2011.0249 PMID: 22201170

52. Martin S. H., Davey J. W., Salazar C., & Jiggins C. D. (2019). Recombination rate variation shapes barriers to introgression across butterfly genomes. *PLoS Biology*, 17(2), 1–28. https://doi.org/10.1371/journal.pbio.2006288 PMID: 30730876

53. Brandvain Y., Kenney A. M., Flagel L., Coop G., & Sweigart A. L. (2014). Speciation and Introgression between Mimulus nasutus and Mimulus guttatus. *PLoS Genetics*, 10(6). https://doi.org/10.1371/journal.pgen.1004410 PMID: 24967630

54. Nelson T. C., Stathos A. M., Vanderpool D. D., Finseth F. R., Yuan Y. W., & Fishman L. (2021). Ancient and recent introgression shape the evolutionary history of pollinator adaptation and speciation in a model monkeyflower radiation (Mimulus section Erythranthe). *PLoS Genetics*, 17(2), 1–26. https://doi.org/10.1371/journal.pgen.1009095

55. Eroukhmanoff F., Hermansen J. S., Bailey R. I., Sæther S. A., & Sætre G. P. (2013). Local adaptation within a hybrid species. *Heredity*, 111(4), 286–292. https://doi.org/10.1038/hdy.2013.47 PMID: 23695379

56. Runemark A., Fernández L.P., Eroukhmanoff F. & Sætre G.-P. (2018c). Genomic Contingencies and the Potential for Local Adaptation in a Hybrid Species. *The American Naturalist* 192: 10–22.

57. Lamichhaney S., Berglund J., Almén M.S., Maqbool K., Grabherr M., Martinez-Barrio A., et al. (2015). Evolution of Darwin's finches and their beaks revealed by genome sequencing. 518: 371–375. https://doi.org/10.1038/nature14181 PMID: 25686609

58. Lamichhaney S., Han F., Berglund J., Wang C., Sallman A. M., Webster M., et al. (2016). A beak size locus in Darwin's finches facilitated character displacement during a drought. *Science*, 6284, 470–474. https://doi.org/10.1126/science.aad8786 PMID: 27102486

59. Abzhanov A., Protas M., Grant B. R., Grant P. R., & Tabin C. J. (2004). Bmp4 and morphological variation of beaks in Darwin's finches. *Science*, 305, 1462–1466. https://doi.org/10.1126/science.1098095 PMID: 15353802

60. Sabeti P. C., Varilly P., Fry B., Lohmueller J., Hostetter E., Cotsapas C., et al. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature*, 449(7164), 913–918. https://doi.org/10.1038/nature06250 PMID: 17943131

61. Ravinet M., Elgvin T.O., Trier C., Aliabadian M., Gavrilov A. & Sætre G.-P. (2018b). Signatures of human-commensalism in the house sparrow genome. *Proc. Biol. Sci.* 285: 20181246

62. Summers-Smith J. D. (1988). *The Sparrows: a study of the genus Passer*. T & AD Poyser.

63. Song Y., Endepols S., Klemann N., Richter D., Matuschka F.-R., Shih C.-H., et al. (2011). Adaptive Introgression of Anticoagulant Rodent Poison Resistance by Hybridization between Old World Mice. *Current Biology* 21: 1296–1301. https://doi.org/10.1016/j.cub.2011.06.043 PMID: 21782438

64. Consortium T.H.G. (2012). Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. 487: 94–98. https://doi.org/10.1038/nature11041 PMID: 22722851

65. Marques D.A., Meier J.I. & Seehausen O. (2019). A Combinatorial View on Speciation and Adaptive Radiation. *Trends in Ecology & Evolution* 34: 531–544. https://doi.org/10.1016/j.tree.2019.02.008 PMID: 30885412

66. Cutter A. D., & Payseur B. A. (2013). Genomic signatures of selection at linked sites: Unifying the disparity among species. *Nature Reviews Genetics*, 14(4), 262–274. https://doi.org/10.1038/nrg3425 PMID: 23478346

67. Van Doren B. M., Campagna L., Helm B., Illera J. C., Lovette I. J., & Liedvogel M. (2017). Correlated patterns of genetic diversity and differentiation across an avian family. *Molecular Ecology*, 26, 3982–3997 https://doi.org/10.1111/mec.14083 PMID: 28256062

68. Irwin D. E., Alcaide M., Delmore K. E., Irwin J. H., & Owens G. L. (2016). Recurrent selection explains parallel evolution of genomic regions of high relative but low absolute differentiation in a ring species. *Molecular Ecology*, 25(18), 4488–4507. https://doi.org/10.1111/mec.13792 PMID: 27484941

69. Stelkens R. & Seehausen O. (2009). Genetic distance between species predicts novel trait expression in their hybrids. *Evolution* 63: 884–897. https://doi.org/10.1111/j.1558-5646.2008.00599.x PMID: 19220450

70. Stelkens R.B., Schmid C., Selz O. & Seehausen O. (2009). Phenotypic novelty in experimental hybrids is predicted by the genetic distance between species of cichlid fish. *BMC Evol. Biol.* 9: 283. https://doi.org/10.1186/1471-2148-9-283 PMID: 19961584

71. Thompson K. A., Urquhart-Cronish M., Whitney K. D., Rieseberg L. H., & Schluter D. (2021). Patterns, predictors, and consequences of dominance in hybrids. *American Naturalist*, 197(3), E72–E88.

72. Sætre G. P., Riyahi S., Aliabadian M., Hermansen J. S., Hogner S., Olsson U., et al. (2012). Single origin of human commensalism in the house sparrow. *Journal of Evolutionary Biology*, 25(4), 788–796. https://doi.org/10.1111/j.1420-9101.2012.02470.x PMID: 22320215

73. Martin S. H., Davey J. W., & Jiggins C. D. (2015). Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Molecular Biology and Evolution*, 32(1), 244–257. https://doi.org/10.1093/molbev/msu269 PMID: 25246699

74. Andrews S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

75. Catchen J., Hohenlohe P. A., Bassham S., Amores A., & Cresko W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, 22(11), 3124–3140. https://doi.org/10.1111/mec.12354 PMID: 23701397

76. Li H., & Durbin R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, 25(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324 PMID: 19451168

77. Auwera G. A. Van Der, Carneiro M. O., Hartl C., Poplin R., Levy-moonshine A., Jordan T., et al. (2014). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics*, 11(1110), 11.10.1–11.10.33. https://doi.org/10.1002/0471250953.bi1110s43.

78. Mckenna A., Hanna M., Banks E., Sivachenko A., Cibulskis K., Kernytsky A., et al. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297–1303. https://doi.org/10.1101/gr.107524.110 PMID: 20644199

79. Danecek P., Auton A., Abecasis G., Albers C. A., Banks E., Depristo M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330 PMID: 21653522

80. Chang C. C., Chow C. C., Tellier L. C. A. M., Vattikuti S., Purcell S. M., & Lee J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, 4(7). https://doi.org/10.1186/s13742-015-0047-8

81. Jombart T. (2008). Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129 PMID: 18397895

82. Alexander D. H., Novembre J., & Lange K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19(9), 1655–1664. https://doi.org/10.1101/gr.094052.109 PMID: 19648217

83. Kamvar Z. N., Tabima J. F., & Grünwald N. J. (2014). Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ, 2014(1), 1–14. https://doi.org/10.7717/peerj.281 PMID: 24688859

84. Kamvar Z. N., Brooks J. C., & Grünwald N. J. (2015). Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. Frontiers in Genetics, 6(JUN), 1–10. https://doi.org/10.3389/fgene.2015.00208

85. Dray S., & Dufour A. B. (2007). The ade4 package: Implementing the duality diagram for ecologists. Journal of Statistical Software, 22(4), 1–20. https://doi.org/10.18637/jss.v022.i04

86. Voight B. F., Kudaravalli S., Wen X., & Pritchard J. K. (2006). A map of recent positive selection in the human genome. *PLoS Biology*, 4(3), 0446–0458. https://doi.org/10.1371/journal.pbio.0040072 PMID: 16494531

87. Delaneau O., Marchini J., & Zagury J. F. (2012). A linear complexity phasing method for thousands of genomes. *Nature Methods*, 9(2), 179–181. https://doi.org/10.1038/nmeth.1785

88. O'Connell J., Gurdasani D., Delaneau O., Pirastu N., Ulivi S., Cocca M., et al. (2014). A General Approach for Haplotype Phasing across the Full Spectrum of Relatedness. *PLoS Genetics*, 10(4). https://doi.org/10.1371/journal.pgen.1004234 PMID: 24743097

89. Knaus B. J., & Grünwald N. J. (2017). vcfr: a package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources*, 17(1), 44–53. https://doi.org/10.1111/1755-0998.12549 PMID: 27401132

90. Gautier M., Klassmann A., & Vitalis R. (2017). rehh 2.0: a reimplementation of the R package rehh to detect positive selection from haplotype structure. *Molecular Ecology Resources*, 17(1), 78–90. https://doi.org/10.1111/1755-0998.12634 PMID: 27863062

91. Gautier M., & Vitalis R. (2012). Rehh An R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics*, 28(8), 1176–1177. https://doi.org/10.1093/bioinformatics/bts115 PMID: 22402612

92. Dias-alves T., Mairal J., & Blum M. G. B. (2018). Loter: A Software Package to Infer Local Ancestry for a Wide Range of Species. *Mol. Biol. Evol.*, 35(9), 2318–2326. https://doi.org/10.1093/molbev/msy126 PMID: 29931083

93. Cuevas, et al. (2021), Predictors of genomic differentiation within a hybrid taxon, Dryad, Dataset, https://doi.org/10.5061/dryad.wpzgmsbns