Intraspecific genomic variation and local adaptation in a young hybrid species

Angélica Cuevas[1*], Mark Ravinet[12], Glenn-Peter Sætre[1] and Fabrice Eroukhmanoff[1]

Affiliations:

[1]Department of Biosciences, Centre for Ecological and Evolutionary Synthesis, University of Oslo, PO Box 1066, N-0316, Oslo, Norway.

[2]School of Life Sciences, University of Nottingham, NG7 2RD, Nottingham, United Kindom.

*Corresponding author. Email: a.m.c.pulido@ibv.uio.no (A.C.).

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 non-hybrid 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: Local adaptation, hybrid species, *Passer* sparrows, genomic incompatibilities, hybrid constraints, genome evolution.

INTRODUCTION

Hybridization is an evolutionary process that has been increasingly studied in the last decade (Abbott et al., 2013; Marques, Meier & Seehausen, 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, Rosenthal & Andolfatto, 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, Morris, Asmussen, & Rieseberg, 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, Cui, Rosenthal, & Andolfatto, 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 re-arrangements 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 non-hybrid 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, Lucek, Young, & Seehausen, 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, et al., 2018a; Schumer, et al., 2014b; Schumer et al., 2018; Trier, Hermansen, Sætre, & Bailey, 2014). Sorting of incompatibilities, originally arising and driving reproductive isolation between parental species, can generate symmetrical incompatibilities isolating the hybrid from both parental species (Schumer, et al., 2015; Buerkle, et al., 2000). However, this process may also constrain hybrid lineages long after hybridization has occurred, affecting their evolutionary potential (Eroukhmanoff et al., 2017; Runemark et al., 2018a). 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, et al., 2018a). 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 et al., 2018a). 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., 2018b; Schumer, et al., 2014b). 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 non-hybrid 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, Colyn, Legendre, Nimon, & Flamand, 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, Alonso, Lagerveld, Senar & Björklund, 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, Waller & Svensson, 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, Welch, Loire, Bonhomme, & David, 2011), especially when genes under ecological selection are coupled with DMI loci (Seehausen, 2013), which in turn may facilitate local adaptation (Eroukhmanoff, Hermansen, Bailey, Sæther & Sætre, 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 likely occurred when the house sparrow spread into Europe alongside agriculture, approximately 6 kyr BP (Hermansen et al., 2011; Ravinet et al., 2018; Elgvin et

al., 2017). It is possible that this hybrid species originated through a period of multiple hybridization events (Runemark, et al., 2018a; Elgvin et al., 2017) with rapid evolution of reproductive barriers from both parental species (Trier et al., 2014; Hermansen 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 mito-nuclear 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 likely influenced by separate, independent hybridization events (Runemark, et al., 2018a). 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 isolation by environment (IBE), which appears to be driven primarily by temperature. We 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.

METHODS

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 8 populations across Italy (Fig.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).

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 ul 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 by Peterson, Weber, Kay, Fisher & Hoekstra (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 500-600bp 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 1x75bp sequencing format. On average, 2.4 $x10^6$ 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).

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 (Auwera et al., 2014; Mckenna et al., 2010) 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 5bp 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, non-variant 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 dataset 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 8 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 4 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.

Investigating population divergence within the Italian sparrow

To evaluate population structure and divergence in the hybrid species we used a SNP set containing 4387 loci identified across 8 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 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 8 populations of Italian sparrows. We calculated windowed F_{ST} and nucleotide diversity using a sliding window of 100kb in size with 25-kb steps. Nucleotide diversity was estimated retaining non-variant sites and avoiding minor allele frequency filtering. We also calculated Tajima's D on non-overlapping windows of 100kb, 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.

288

289 **Selection, local adaptation and environmental variation**

290

291 The Italian peninsula varies considerably in climate, thus we investigated whether genomic
292 divergence covaried with environmental variation. Pairwise differences in climatic variables
293 were regressed with the pairwise genetic distance between populations. We analysed five
294 climatic variables obtained from the global climate data server, WorldClim (v. 2.0,
295 http://www.worldclim.org) (Hijmans, Cameron, Parra, Jones & Jarvis, 2005), BIO1=Annual
296 Mean Temperature, BIO4=Temperature Seasonality (standard deviation *100),
297 BIO12=Annual Precipitation and BIO15=Precipitation Seasonality (Coefficient of Variation).
298 Values were retrieved using the R packages RGDAL (v 1.3-4, Bivand et al., 2017) and SP (v
299 1.2-4) (Pebesma & Bivand, 2005), with a resolution of 1km. Geographic distance was
300 obtained with the function spDistsN1 from the R package SP (v 1.2-4) and altitudinal data
301 was gathered from the R package RASTER (v 2.6-7) (Hijmans, 2014) and SP (v 1.2-4) using
302 the getData function. We also analysed phenotypic distance in two beak traits, mean beak
303 height (BH) and beak length (BL), in each population.

304

305 To test for associations between environmental factors, geographic, altitudinal and phenotypic
306 distances and genome-wide divergence we used univariate and multiple matrix regression
307 with randomization (UMRR and MMRR respectively) approaches (Wang, 2013) and a
308 modification implemented by Prunier et al., (2015), including commonality analysis (CA) to
309 account for multicollinearity (non-independence) among environmental factors. Data were Z-
310 transformed (i. standardization by subtracting the mean and dividing by the standard

311 deviation) to make regression coefficients of the predictor variables comparable (beta
312 weights, Prunier et al., 2015).
313
314 MMRR is a multiple regression analysis on distance matrices used to quantify the
315 contribution of environmental and geographic factors to patterns of genetic divergence
316 (Wang, 2013). It allows the quantification of isolation by distance (IBD), isolation by
317 environment (IBE) and even isolation by adaptation (IBA) when a phenotypic variable is
318 included as predictor. One advantage of the method is that it not only resolves whether the
319 dependent and independent variables are correlated but also quantifies the change and
320 directionality (regression coefficients, β_n) that the dependent variable (genomic distance) has
321 with respect to multiple independent variables, i.e. geographic and environmental distances
322 (Wang, 2013). The fit of the model is determined by the coefficient of determination (R^2).
323 Given the non-independent nature of the variables, the significance (*p-values*) of the
324 variable's effects (β_n) and fit of the model (R^2) are estimated by randomized permutations of
325 rows and columns of the dependent variable matrix (for more details see Wang, 2013).
326 However, strong multicollinearity among predictors is still a limitation of this approach.
327 Regression coefficients (β_n), fit of the model (R^2) and their significance can be affected by
328 multicollinearity among explanatory variables (Kraha, Turner, Nimon, Zientek & Henson,
329 2012; Nimon & Reio, 2011; Prunier et al., 2015). To overcome this caveat an incorporation of
330 variance-partitioning procedures via commonality analysis (CA) can be used, implemented by
331 Prunier et al., (2015). This method (CA) developed originally by Newton & Spurrell (1967)
332 decomposes the model coefficients into unique (U) and common (C) variance components
333 (Campbell & Tucker, 1992 in Prunier et al., 2015; K. F. Nimon & Oswald, 2013), allowing
334 identifying 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 approx. 100kb (Elgvin et al., 2017), thus we selected genes contained in regions at a maximum of 100kb 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 show 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 (2 populations per species) and 3 Italian sparrows populations; data retrieved from Elgvin et al., (2017) and Ravinet et al., (2018).

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 non-hybrid species, we compared population genomic parameters of the parental species to the Italian sparrow. We also estimated ancestry patterns in the Italian

406 sparrow looking to shed light on the source of the genomic variation found in this hybrid
407 lineage.
408
409 To identify how highly divergent loci in the hybrid are distributed, for instance whether they
410 are located in genomic regions of high parent species divergence or not, we selected the top
411 1% loci with the highest F_{ST} among all 8 Italian sparrow populations across the different
412 localities and estimated ancestry as well as hybrid-parent F_{ST} and between-parents (SH) F_{ST}
413 values for these same loci. Similarly, we extracted the top 1% loci with the highest F_{ST} among
414 house sparrow populations and among Spanish sparrow populations and as for the hybrid
415 species, hybrid-parent F_{ST} and between-parents (SH) F_{ST} values were estimated for these
416 highly variable loci. We also compared the observed patterns of Tajima's D between species.
417 As for the hybrid, Tajima's D for the parental species was estimated using VCF files that
418 were not filtered for minor allele frequency.
419
420 To evaluate whether loci involved in population divergence within the Italian sparrow
421 correspond to loci of high or low genetic differentiation between the Italian and Spanish (IS
422 F_{ST}) sparrows, Italian and house (IH F_{ST}) sparrows or Spanish and house (SH F_{ST}) sparrows,
423 we performed logistic regressions on the probability of being an Italian F_{ST} outlier. In these
424 models, the outlier status (outlier/non-outlier) of each locus (SNP) is the response variable,
425 while additive and interaction effects of pairwise F_{ST} between the three species were tested as
426 predictors.
427
428 We also used whole genome resequencing (WGS) data from Elgvin et al., (2017) and Ravinet
429 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 3 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, Mairal, & Blum, 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 8 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 100kb window. We calculated house sparrow ancestry proportion across all 8 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 & James, 2017; Ortiz-Barrientos, Engelstädter & Rieseberg, 2016). 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.

RESULTS

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 (Fig.1C, 1D). To evaluate the genomic variation among populations of the Italian sparrow we performed a PCA and admixture analyses from 8 locations across the Italian peninsula (N=131 individuals, 4387 SNPs. Fig.1B, S1), covering a wide range of its mainland geographic distribution (Fig.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 8 localities = 0.013, π=2.595x10^{-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 (Fig.S2A, S2C).

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

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 dataset (Table.1, Table.2, Table.S3). In UMRR (Table.1) geographical distance (GEO) showed a non-significant relationship (R^2=0.053, β=0.004) to genetic differentiation among populations. Its contribution in the multivariate model (MMRR) was non-significant (β=0.003, *P*=0.34) 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, Fig.S3), with a R^2=0.163 and β weight of β=0.007. The multivariate model including all the climatic factors, altitude and geographic distances as predictors (MMRR – model 1, Table.2),

502 explained 25% of the inter-population variation in F_{ST} within the Italian sparrow (R^2=0.25).
503 Consistent with the results from UMRR, temperature seasonality yielded the highest β weight,
504 with a considerable explanatory power (β=0.007) (Table.2), accounting for 8% of the
505 variation explained by the model. However, variance partitioning (CA) showed its unique
506 contribution was almost negligible, meaning the interaction with other variables (collinearity)
507 had a larger effect (U=0.003, C=0.2, Table.2, Fig.S4).
508
509 While mean annual temperature explained a considerable amount of the variance (A.TEMP,
510 Table.2) most of it fell into the unique factor (U=0.14) and its beta weight was non-significant
511 (β=0.001, *P*=0.89). Mean annual precipitation showed similar results (A.PREC, T=24%,
512 Table.2). This suggests that there is collinearity between climatic factors. Unique (U) and
513 common (C) contributions to the variation, estimated by CA (Table.2, Fig.S4), showed mean
514 annual temperature (T=0.16) and mean annual precipitation (T=0.06), as the major
515 contributors, accounting for 64% and 24% of the variation explained by the model,
516 respectively (Table.2). However, beta weights for these predictors were not significant.
517 Moreover, when removing mean annual temperature from the model (MMRR – model 2,
518 Table.S3) temperature seasonality was no longer significant (*P*=0.1), supporting the
519 collinearity effect among climatic variables.
520
521 Finally, evaluating IBA, incorporating beak morphology as predictors, the univariate (UMRR,
522 Table.1) and multivariate (MMRR, Table.S3) models showed that these phenotypic traits do
523 not explain a significant amount of the genomic divergence among Italian sparrow
524 populations. The univariate models for each of the beak traits showed a non-significant

R^2=0.036 (P >0.34), 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 non-significant.

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 (Fig.2A) while presenting high, although non-significant, q-values of $\boldsymbol{g}$ for mean annual temperature and altitude (Fig.S5A and S5C). 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 (Fig.2B). We also found three candidate loci under selection related to beak morphology, associated with population divergence in beak height (Fig.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 or phenotypic traits. 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

549 the Italian sparrow; one locus on chromosome 6, a second locus on chromosome 20 and
550 another in chromosome 15. The latter was previously identified as associated with mean
551 annual precipitation by BayeScEnv (Fig.3A).
552
553 Within the putative regions under selection (i.e. 100kb around the outlier loci) we identified
554 potential genes of interest that may be associated to climatic variation (Table.S4). To further
555 assess signatures of selection, specifically at the gene level, we used WGS data from Elgvin et
556 al., (2017) and Ravinet et al., (2018) to calculate per-gene population statistics (F_{ST}, Tajima's
557 D, Pi, *d*xy, (Table.S4)). The gene GDF5 was identified as a 5% gene-F_{ST} outlier (one-tailed
558 test) presenting a gene-F_{ST}=0.047 (Table.S4). The GDF5 gene, also known as BMP-14,
559 involved in bone and cartilage development, encodes a growth differentiation factor protein
560 related to the BMP (bone morphogenetic protein) gene family (Reddi & Reddi, 2009), a gene
561 family involved in skeletal and jaw development (Bleuming et al., 2007; Cerny et al., 2010;
562 Kaucka & Adameyko, 2019).
563
564 In the house sparrow (75 individuals, 6503 SNPs, 4 localities) 8 candidate loci on
565 chromosomes 1, 5 and 8 were inferred to be significantly under selection (Fig.3B, Table.S4).
566 Similarly, in the Spanish sparrow (1320 SNPs across 82 individuals from 4 localities) 8
567 candidate loci (on chromosomes 1, 2, 3 and 5) were also identified using Bayescan (Fig.3C,
568 Table.S4).
569
570 Only one of the outlier loci was simultaneously identified by both genome scan approaches
571 (Bayescan and BayeScEnv) for the Italian sparrow. The lack of overlapping outlier loci under
572 selection among the three species may be due to differential selective pressures 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.

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 x10^{-6} (Fig.S2D, S2F, Table.S2) was similar to that in the Spanish sparrow (mean F_{ST}=0.021, π=1.642x10^{-6}), with a maximum F_{ST} of 0.34 (Fig.S2G, S2I, 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=2.595x10^{-6}; Fig.S2A, S2C, 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=0.56, Fig.5A). However, Italian outlier loci (1% F_{ST} outliers) segregated for alleles from both parents, as most of the genome presents a mosaic pattern (Fig.5B, S6, 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: Chi2=806789.041, *P*=0.00, Fig.5B), also supporting the low SH F_{ST} values in highly divergent loci in the Italian sparrow, in comparison with the parental taxa (Fig.4A, 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 (Fig.4A, 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*=0.0127). A negative, yet non-significant, correlation was also found between the highly divergent regions within the hybrid species and between parental species genetic divergence (SH F_{ST}, *P*=0.0926) (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 (Fig.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, Fig.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 (Fig.5C). Furthermore, highly divergent loci within each of the parental species did not correspond to those found within the hybrid Italian sparrow (Fig.4C, 4D).

We find some evidence suggesting that recombination rate could explain part of the genomic divergence pattern found within the Italian sparrow (R^2= 0.00085, *P*= 0.033) and within the Spanish sparrow (R^2= 0.003211, *P*= 0.026). 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 not significant correlation (R^2= -9.67e-07, *P*= 0.319) (Fig.S8).

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

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, et al., 2018a), 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.

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 approx. 6.000 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 likely 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 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 the 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,

692 therefore, it is difficult to conclusively identify the processes that could have generated this
693 pattern.
694
695 **Selection, local adaptation and environmental variation**
696
697 Assessing genomic patterns across a spatially heterogeneous distribution, in correlation with
698 factors that can play a role in genomic divergence, can help us elucidate the processes that
699 have determined population differentiation in hybrid lineages. It can also give insights to the
700 adaptive potential of the species (local adaptation and gene flow reduction) or whether
701 genomic differentiation is essentially a result of genetic drift, where patterns of genetic
702 variation are shaped by low gene flow (Prunier et al., 2015; Seeholzer & Brumfield, 2018;
703 Wang, 2013).
704
705 To assess adaptive divergence and gene flow, we evaluated IBE, IBA through beak
706 divergence and IBD. We did not find evidence for IBD or IBA, but the significant correlation
707 between genetic distance and climatic variation is consistent with IBE. Our results suggest
708 that climatic differences, with temperature as the main factor, likely contribute to reduced
709 gene flow between populations in the Italian sparrow, possibly as a result of local adaptation.
710 Previously, precipitation has been found to correlate with beak morphology variation in this
711 species (Runemark, Fernández, Eroukhmanoff, & Sætre, 2018b), and could indirectly be
712 mediating gene flow between phenotypically divergent populations (Eroukhmanoff et al.,
713 2013). Differential changes in phenotypic traits responding to selective pressures can have an
714 effect on local adaptation that may sometimes lead to IBA (Edelaar, Alonso, Lagerveld, Senar
715 & Björklund, 2012). However, when directly evaluating beak trait variation as a predictor of

716 overall genomic differentiation among populations of the Italian sparrow we did not find
717 evidence for IBA.
718
719 Patterns of adaptive divergence with ongoing gene flow have also been extensively reported
720 in species of non-hybrid origin (de Leon, Bermingham, Podos, & Andrew, 2010; Marques et
721 al., 2016; Martin et al., 2013; Raeymaekers et al., 2017), which suggests that despite the
722 possibility of constraints reducing the evolvability of this hybrid species (Runemark, et al.,
723 2018a), there is also potential for adaptive divergence leading to local adaptation, as in non-
724 hybrid lineages. In fact, theory suggests that incompatibilities could facilitate local adaptation
725 by the coupling of genes under ecological selection and DMI loci (Seehausen, 2013). For
726 example, if genomic incompatibilities become trapped in environmentally divergent habitats,
727 coupling with loci involved in local adaptation may occur, which could potentially facilitate
728 diversification within the hybrid lineage (Abbott et al., 2013; Bierne, Gagnaire & David,
729 2013; Butlin & Smadja, 2018; Seehausen, 2004). This coupling mechanism, more prone to
730 arise in hybrid lineages around regions of interspecific incompatibilities, could facilitate rapid
731 local adaptation in comparison to other processes of diversifying selection in non-hybrid
732 species (Eroukhmanoff et al., 2013; Seehausen, 2013). To the best of our knowledge, there
733 are no empirical studies that report such linkage between DMIs and regions under natural
734 selection. However, our results and previous studies (e.g. Runemark, et al., 2018a) show that
735 genomic constraints play an important role in the formation of the admixed Italian sparrow
736 genome.
737
738 Here, we present for the first time direct evidence for the role that environmental variation has
739 in mediating genomic variation in a hybrid species, a phenomenon well described in non-

740 hybrid species (Wang & Bradburd, 2014). We also report loci where high levels of adaptive
741 genetic differentiation has occurred, some of which are covarying directly with climate
742 variation, suggesting that they are situated in genomic regions linked to local adaptation. For
743 example on chromosome 20 an outlier locus for adaptive divergence between Italian sparrow
744 populations (via Bayescan) was found to be in the vicinity of the GDF5 gene (growth
745 differentiation factor 5, also known as BMP14 (NCBI), a gene also identified as a 5% F_{ST}
746 outlier in the per-gene analysis based on whole genome-resequencing data. This gene is
747 known to be involved in jaw development in vertebrates (Bleuming et al., 2007; Cerny et al.,
748 2010; Kaucka & Adameyko, 2019) and related to the BMP (bone morphogenic protein) gene
749 family (Buxton, Edwards, Archer & Francis-West, 2001; Francis-West et al., 1999a; Francis-
750 West, Philippa, Parish, Lee & Archer, 1999b). The BMP gene family has a fundamental role
751 in craniofacial development and beak shape and size variation in Darwin's finches
752 (Abzhanov, Protas, Grant, Grant & Tabin, 2004; Lamichhaney et al., 2016).
753
754 The beak is a trait known to be under strong selective pressure (Lamichhaney et al., 2016;
755 Lamichhaney et al., 2015). Beak size has been shown to be a crucial trait underlying the
756 survival of Darwin's finches after a drought (Lamichhaney et al., 2016) and beak traits in
757 general act as drivers of major evolutionary shifts in Darwin's finches (Almén et al., 2016;
758 Chaves et al., 2016; Lamichhaney et al., 2016; Lamichhaney et al., 2015). Beak shape
759 variation has been found to respond to environmental divergence affecting food availability in
760 the medium ground finch (*Geospiza fortis*) (Grant & Grant, 2003; Grant & Grant, 2014).
761 Thus, climatic factors could be considered a reasonable proxy for food availability in
762 sparrows (Runemark, et al., 2018b). It is possible that divergence of genes associated with
763 beak morphology may reflect an adaptive response to variation in food resources found in

764 environmentally different habitats. However, further analyses need to be conducted in order to
765 determine the true underlying mechanisms of divergence between population both at the
766 genetic and phenotypic level.
767
768 **Hybrid constraints to population divergence**
769
770 Evaluating patterns of ancestry and divergence in the hybrid genome can provide important
771 insights on whether population differentiation is facilitated by novel genetic variation or
772 hampered by genomic constraints linked to hybrid incompatibilities. Genomic variation
773 within a hybrid lineage can be generated by novel genetic combinations through
774 rearrangements of parental blocks, potentially generating novel epistatic interactions, or
775 through heterozygosity at parental divergent loci. In this case, highly differentiated loci within
776 the hybrid taxon can expected to be located in regions where the parental species have
777 diverged strongly. On the other hand, negative epistatic interactions between inherited
778 parental blocks (in particular if these interactions involve genetic incompatibilities) may lead
779 to strong stabilizing selection on loci fixed for compatible alleles and, through linkage
780 disequilibrium, on other loci situated in their vicinity. Thus, inherited parental genomic blocks
781 would be expected to be highly conserved, as these are more likely to harbour candidate loci
782 for genetic incompatibilities. This type of genetic constraint on hybrids could reduce the
783 evolutionary potential of the hybrid species to diverge at the population level. However, this
784 may depend on variation in recombination rate across the genome, which is also known to
785 affect the extent of purging and population divergence.
786

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 parental 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 divergent allele frequencies seem to be preferentially fixed for one parental allele across Italian sparrow populations (also evidenced by Runemark et al., 2018a). This supports the hypothesis of constraints 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 estimates 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 hybrid taxon and each of its parental species seem to differ, suggesting that the admixed nature of the hybrid species may be somewhat restricted compared to its non-hybrid parental species. In contrast to the hybrid

811 species, intraspecific genomic variation in the parental lineages is located mainly in regions of
812 parental divergence. Additionally, there is no overlap of outlier loci under selection among
813 the three species. This could suggest that differential selective pressures may be operating in
814 addition to specific genomic constraints in the hybrid species. However, an important factor
815 to be considered in admixed genomes is the inheritance of traces of different evolutionary
816 histories as well as the individual evolutionary path that the hybrid species has taken since its
817 formation (and eventual further introgression with parent species). Thus, processes other than
818 differential selective pressures could generate this pattern.
819
820 These results provide a new perspective on how hybridization may impact adaptive evolution,
821 more specifically on how novel genomic variation evolves and is utilized in a hybrid lineage
822 post hybrid speciation, not only through genomic rearrangements linked to admixture and
823 incompatibilities.
824
825 **CONCLUSION**
826
827 Genetic variation within the Italian sparrow appears to be driven by climatic variation,
828 temperature being the main factor; we find evidence for isolation by environment (IBE),
829 which could facilitate ongoing local adaptation. Our study supports previous findings
830 suggesting that local adaptation nonetheless can occur, albeit in a biased and constrained
831 manner. Indeed, genetic differentiation in the hybrid species is mainly found in loci that are
832 not divergent between the parental species and hence possibly less prone to be incompatible
833 in the hybrid. This suggests that purging of incompatibilities could be an important element in
834 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.

ACKNOWLEDGEMENTS

We thank S. Vitulano, M. Caldarella, and M. Griggio for assistance in the field. Jo S. Hermansen, Tore O. Elgvin, Stein A. Sæther, Anna Runemark and Cassandra N. Trier for help obtaining the data. Camilla Lo Cascio Sætre and Caroline Øien Guldvog for help with lab work. This work was done in collaboration with the Centre for Naturalistic Studies ONLUS (CSN, Italy) and ISPRA (Italy). This work was funded by The Research Council of Norway grant 297137, the Swedish Research Council and the European Union Marie Sklodowska Curie Action 2011-302504. The Faculty of Mathematics and Natural Sciences (University of Oslo) and Molecular Life Science (MLS) at the University of Oslo.

REFERENCES

Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., … 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

Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Auwera, G. A. Van Der, 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

Bierne, N., Gagnaire, P., & David, P. (2013). The geography of introgression in a patchy environment and the thorn in the side of ecological speciation, *59*(1), 72–86.

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*.

Bleuming, S. A., He, X. C., Kodach, L. L., Hardwick, J. C., Koopman, F. A., Ten Kate, F. J., … 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), 000–000. 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. *Proc Natl Acad Sci*, *107*, 17262–7. https://doi.org/10.1073/pnas.1009304107/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1009304107

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., … 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., … Vcf, T. (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. *Mol. Biol. Evol.*, *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., … Sætre, G. P. (2017). The genomic mosaicism of hybrid speciation. *Science Advances*, *3*(6). https://doi.org/10.1126/sciadv.1602996
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
Eroukhmanoff, Fabrice, 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*, 1–48. https://doi.org/10.1101/102970
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, Philippa 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. https://doi.org/10.1641/0006-3568(2003)053[0965:WDFCTU]2.0.CO;2
Grant, P. R., & Grant, R. B. (2014). *40 Years of Evolution. Darwin's Finches on Daphne Major Island*. (P. U. Press, Ed.).
Guerrero, R. F., & Hahn, M. W. (2017). Speciation as a sieve for ancestral polymorphism. *Molecular Ecology*, (May), 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,

960 G. P. (2014). Hybrid speciation through sorting of parental incompatibilities in Italian
961 sparrows. *Molecular Ecology*, *23*(23), 5831–5842. https://doi.org/10.1111/mec.12910
962 Hermansen, J. S., Sæther, S. A., Elgvin, T. O., Borge, T., Hjelle, E., & Sætre, G. P. (2011).
963 Hybrid speciation in sparrows I: Phenotypic intermediacy, genetic admixture and
964 barriers to gene flow. *Molecular Ecology*, *20*(18), 3812–3822.
965 https://doi.org/10.1111/j.1365-294X.2011.05183.x
966 Hijmans, R.J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high
967 resolution interpolated climate surfaces for global land areas. *International Journal of*
968 *Climatology*, (25), 1965–1978. https://doi.org/10.1002/joc.1276
969 Hijmans, Robert J. (2014). raster: Geographic Data Analysis and Modeling. *R Package*
970 *Version 2.6-7*. Retrieved from https://cran.r-
971 project.org/web/packages/raster/vignettes/Raster.pdf
972 Jombart, T. (2008). Adegenet: A R package for the multivariate analysis of genetic markers.
973 *Bioinformatics*, *24*(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129
974 Kaucka, M., & Adameyko, I. (2019). Evolution and development of the cartilaginous skull :
975 From a lancelet towards a human face. *Seminars in Cell and Developmental Biology*, *91*,
976 2–12. https://doi.org/10.1016/j.semcdb.2017.12.007
977 Keller, I., Wagner, C. E., Greuter, L., Mwaiko, S., Selz, O. M., Sivasundar, A., … Seehausen,
978 O. (2013). Population genomic signatures of divergent adaptation, gene flow and hybrid
979 speciation in the rapid radiation of Lake Victoria cichlid fishes. *Molecular Ecology*,
980 *22*(11), 2848–2863. https://doi.org/10.1111/mec.12083
981 Kraha, A., Turner, H., Nimon, K., Zientek, L. R., & Henson, R. K. (2012). Tools to support
982 interpreting multiple regression in the face of multicollinearity. *Frontiers in Psychology*,
983 *3*, 1–16. https://doi.org/10.3389/fpsyg.2012.00044
984 Lamichhaney, S., Han, F., Berglund, J., Wang, C., Sallman, A. M., Webster, M. ., …
985 Anderson, L. (2016). A beak size locus in Darwin's finches facilitated character
986 displacement during a drought. *Science*, *6284*, 470–474.
987 Lamichhaney, Sangeet, Berglund, J., Almén, M. S., Maqbool, K., Grabherr, M., Martinez-
988 Barrio, A., … Andersson, L. (2015). Evolution of Darwin's finches and their beaks
989 revealed by genome sequencing. *Nature*, *518*(7539), 371–375.
990 https://doi.org/10.1038/nature14181
991 Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows – Wheeler
992 transform. *Bioinformatics*, *25*(14), 1754–1760.
993 https://doi.org/10.1093/bioinformatics/btp324
994 Luquet, E., Vorburger, C., Hervant, F., Joly, P., Kaufmann, B., Schmeller, D. S., … Plénet, S.
995 (2011). Invasiveness of an introduced species: the role of hybridization and ecological
996 constraints. *Biological Invasions*, *13*(8), 1901–1915. https://doi.org/10.1007/s10530-011-
997 0010-2
998 Mallet, J. (2007). Hybrid speciation. *Nature*, *446*(7133), 279–283.
999 https://doi.org/10.1038/nature05706
1000 Marques, D. A., Lucek, K., Meier, J. I., Mwaiko, S., Wagner, C. E., Excoffier, L., &
1001 Seehausen, O. (2016). Genomics of Rapid Incipient Speciation in Sympatric Threespine
1002 Stickleback. *PLoS Genetics*, *12*(2), 1–34. https://doi.org/10.1371/journal.pgen.1005887
1003 Marques, D. A., Meier, J. I., & Seehausen, O. (2019). A Combinatorial View on Speciation
1004 and Adaptive Radiation. *Trends in Ecology & Evolution*, *34*(6), 531–544.
1005 https://doi.org/10.1016/j.tree.2019.02.008
1006 Martin, S. H., Dasmahapatra, K. K., Nadeau, N. J., Salazar, C., Walters, J. R., Simpson, F., …
1007 Jiggins, C. D. (2013). Genome-wide evidence for speciation with gene flow in

1008 Heliconius butterflies. *Genome Research*, *23*(11), 1817–1828.
1009 https://doi.org/10.1101/gr.159426.113
1010 Mckenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., …
1011 Depristo, M. A. (2010). The Genome Analysis Toolkit : A MapReduce framework for
1012 analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303.
1013 https://doi.org/10.1101/gr.107524.110.20
1014 Meier, J. I., Marques, D. A., Mwaiko, S., Wagner, C. E., Excoffier, L., & Seehausen, O.
1015 (2017). Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature*
1016 *Communications*, *8*, 1–11. https://doi.org/10.1038/ncomms14363
1017 Meirmans, P. G. (2012). The trouble with isolation by distance. *Molecular Ecology*, *21*(12),
1018 2839–2846. https://doi.org/10.1111/j.1365-294X.2012.05578.x
1019 Newton, R. G., & Spurrell, D. J. (1967). A Development of Multiple Regression for the
1020 Analysis of Routine Data. *Journal of the Royal Statistical Society. Series C (Applied*
1021 *Statistics)*, *16*(1), 51–64.
1022 Nimon, K. F., & Oswald, F. L. (2013). Understanding the Results of Multiple Linear
1023 Regression : Beyond Standardized Regression Coefficients. *Organizational Research*
1024 *Methods*, *16*(4), 650–674. https://doi.org/10.1177/1094428113493929
1025 Nimon, K., & Reio, T. G. (2011). Regression Commonality Analysis : A Technique for
1026 Quantitative Theory Building. *Human Resource Development Review*, *10*, 329–340.
1027 https://doi.org/10.1177/1534484311411077
1028 Nosil, P., Egan, S. P., & Funk, D. J. (2008). Heterogeneous genomic differentiation between
1029 walking-stick ecotypes : " isolation by adaptation " and multiple roles for divergent
1030 selection. *Evolution*, *62*, 316–336. https://doi.org/10.1111/j.1558-5646.2007.00299.x
1031 Ortiz-Barrientos, D., Engelstädter, J., & Rieseberg, L. H. (2016). Recombination Rate
1032 Evolution and the Origin of Species. *Trends in Ecology and Evolution*, *31*(3), 226–236.
1033 https://doi.org/10.1016/j.tree.2015.12.016
1034 Ortiz-Barrientos, D., & James, M. E. (2017). Evolution of recombination rates and the
1035 genomic landscape of speciation. *Journal of Evolutionary Biology*, *30*(8), 1519–1521.
1036 https://doi.org/10.1111/jeb.13116
1037 Pebesma, E., & Bivand, R. (2005). Package "sp". Classes and Methods for Spatial Data in R.
1038 *R News*, *5*(2), 1–21. https://doi.org/10.1.1.160.9361
1039 Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double
1040 digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in
1041 model and non-model species. *PLoS ONE*, *7*(5).
1042 https://doi.org/10.1371/journal.pone.0037135
1043 Prunier, J. G., Colyn, M., Legendre, X., Nimon, K. F., & Flamand, M. C. (2015).
1044 Multicollinearity in spatial genetics: separating the wheat from the chaff using
1045 commonality analyses. *Molecular Ecology*, *24*, 263–283.
1046 https://doi.org/10.1111/mec.14283
1047 Raeymaekers, J. A. M., Chaturvedi, A., Hablützel, P. I., Verdonck, I., Hellemans, B., Maes,
1048 G. E., … Volckaert, F. A. M. (2017). Adaptive and non-adaptive divergence in a
1049 common landscape. *Nature Communications*, *8*(267). https://doi.org/10.1038/s41467-
1050 017-00256-6
1051 Räsänen, K., & Hendry, A. P. (2008). Disentangling interactions between adaptive divergence
1052 and gene flow when ecology drives diversification. *Ecology Letters*, *11*(6), 624–636.
1053 https://doi.org/10.1111/j.1461-0248.2008.01176.x
1054 Ravinet, M., Elgvin, T. O., Trier, C., Aliabadian, M., Gavrilov, A., & Sætre, G. P. (2018).
1055 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*(August), 1211–1217.

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

Runemark, A., Fernández, L. P., Eroukhmanoff, F., & Sætre, G.-P. (2018b). 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

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

Schumer, M., Cui, R., Powell, D. L., Dresner, R., Rosenthal, G. G., & Andolfatto, P. (2014b). High-resolution mapping reveals hundreds of genetic incompatibilities in hybridizing fish species. *ELife*, 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). https://doi.org/10.1371/journal.pgen.1005041

Schumer, M., Xu, C., Powell, D. L., Durvasula, A., Skov, L., Holland, C., … Sankararaman, S. (2018). Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science*, *360*(May), 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., … 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). https://doi.org/10.1371/journal.pgen.1004075

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

DATA ACCESSIBILITY: 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 dataset generated have been deposited in the Dryad Digital Repository at DOI (doi:10.5061/dryad.q573n5th7).

AUTHORS' CONTRIBUTIONS: A.C. and F.E. designed the study; A.C., F.E., 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.

COMPETING INTERESTS

The authors declare that they have no competing interests.

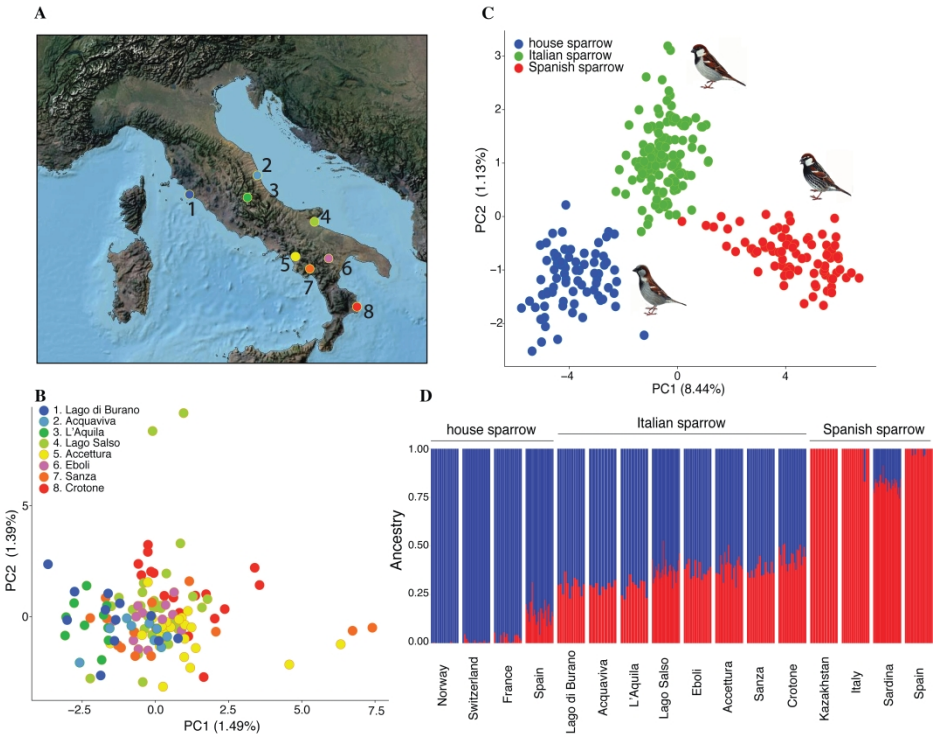

Figure 1. A. Geographic distribution of sampled Italian sparrow populations B. Principal component analysis (PCA) to explore genetic divergence within the Italian sparrow (8 Italian populations, 131 individuals and 4387 SNPs). **C. PCA assessing the three focal species.** Spanish sparrow (red), house sparrow (blue) and Italian sparrow (green), and **D. Admixture analysis** based in 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.

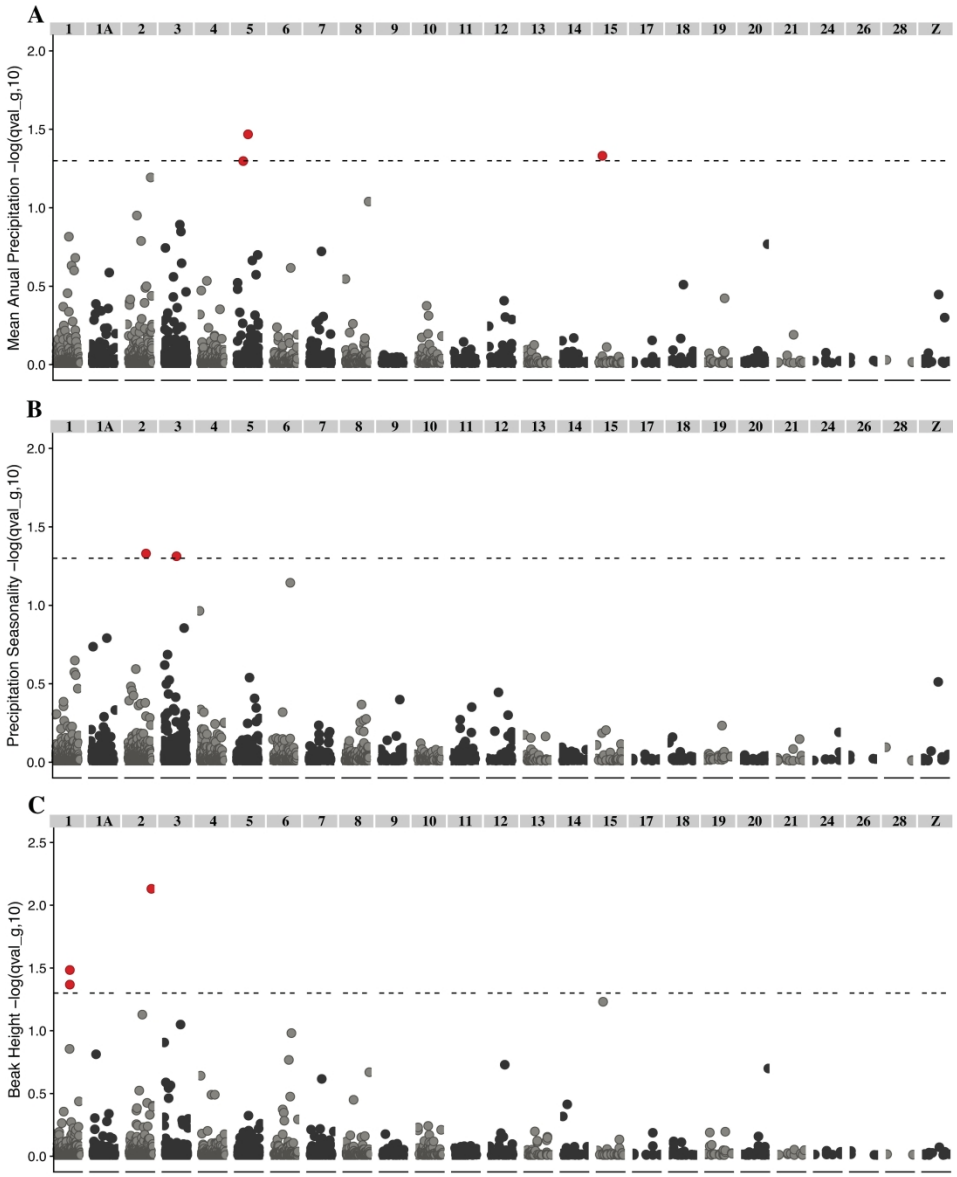

Figure 2. Outlier analysis of local adaptation to climate (BayeScEnv). Manhattan plots of correlation q-values for genetic divergence (SNPs) within the Italian sparrow showing association to climatic factors and one phenotypic trait. Significance level (FDR-corrected) is set at a q-value of < 0.05 (-log10 = 1.3). **A.** Mean Annual Precipitation **B.** Precipitation seasonality and **C.** Beak height.

300x379mm (300 x 300 DPI)

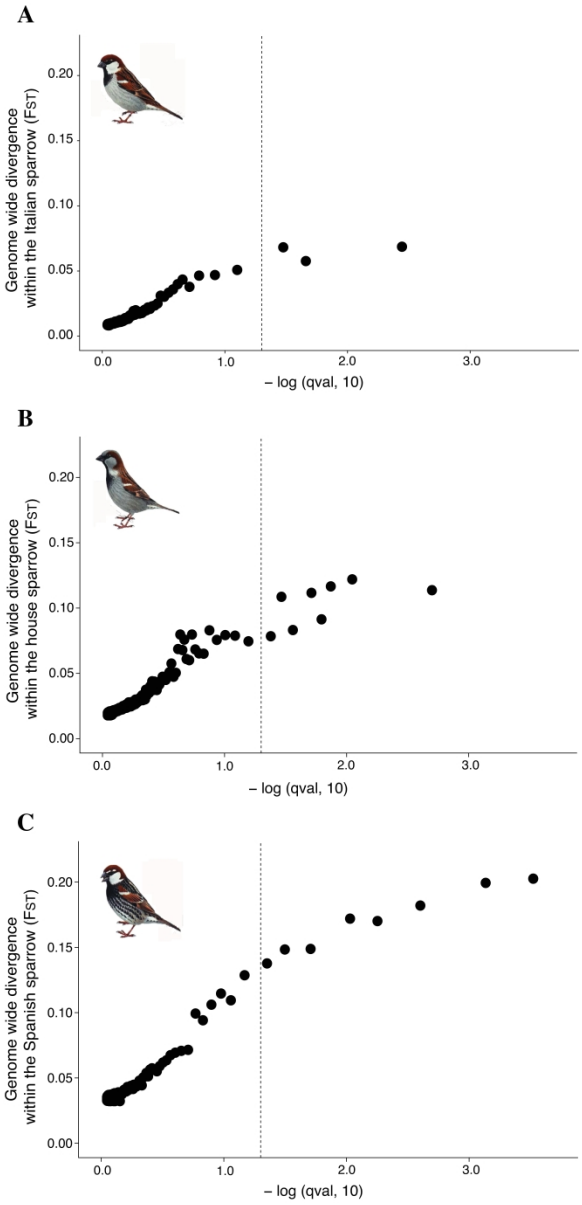

Figure 3. Outlier analysis (BayeScan). Correlation q-values for genetic divergence (SNPs). Significance level (FDR-corrected) is set at a q-value of < 0.05 (-log10 = 1.3). **A.** within the Italian sparrow **B.** within the house sparrow and **C.** within the Spanish sparrow.

351x731mm (600 x 600 DPI)

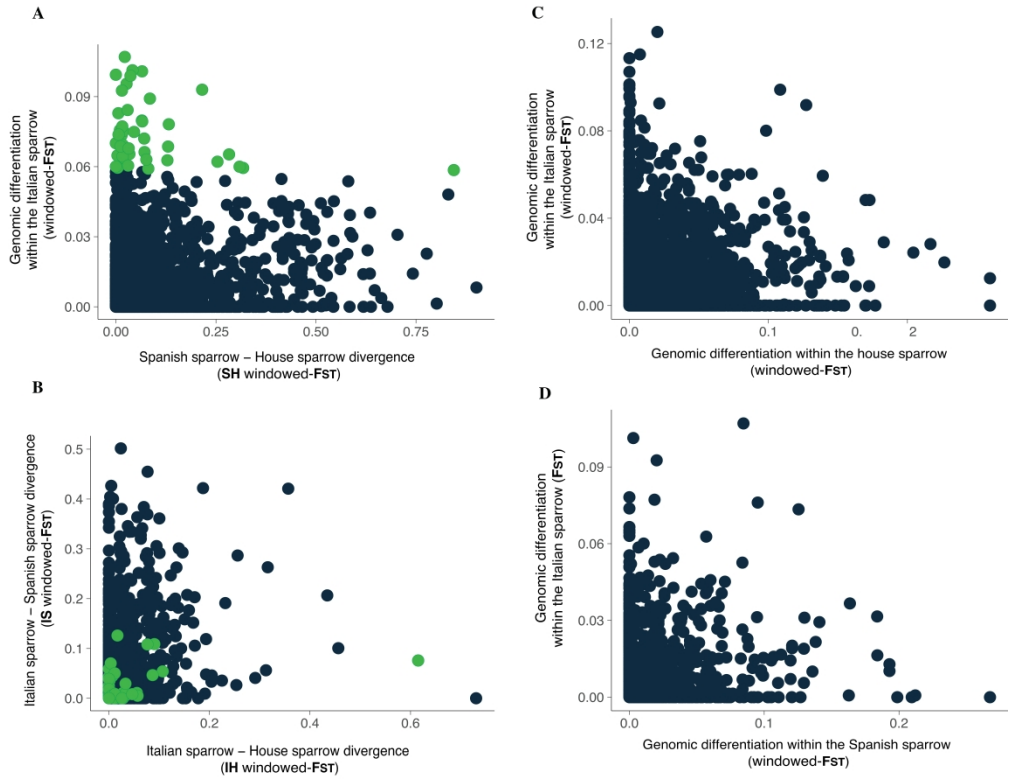

Figure 4. Hybrid constraints to population divergence. A. Genomic differentiation within the Italian sparrow (windowed-F_{ST}) and divergence of its parental species (SH windowed-F_{ST}) **B.** Genomic divergence of the Italian sparrow and each of its parental species (Italian – House sparrow divergence (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 v.s. genomic differentiation within each of the parental species, **C.** the house and **D.** the Spanish. sparrows.

831x644mm (600 x 600 DPI)

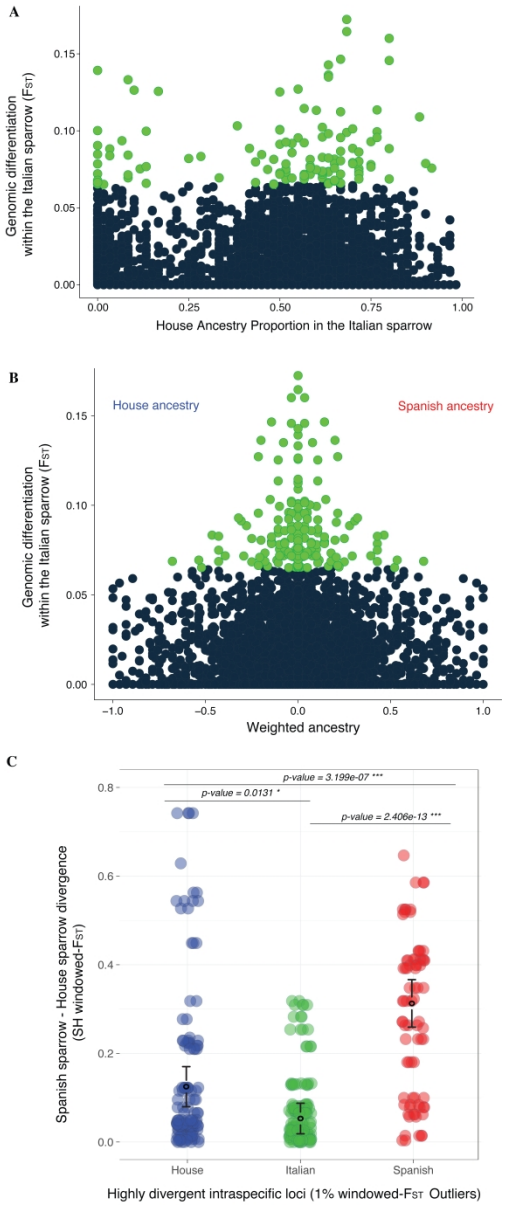

Figure 5. A. Ancestry proportion v.s. Italian sparrow F_{ST}. House ancestry proportion calculated across all Italian sparrow populations. **B. Weighted ancestry v.s. intraspecific F_{ST} in the Italian sparrow.** Ancestry weighted by parental allele frequency difference. Green points represent the 1% Italian F_{ST} outliers. Negative values correspond to loci with house ancestry, while positive ones show Spanish ancestry. **C. Parental genomic divergence (SH windowed-F_{ST})** presented on the intraspecific 1% windowed-F_{ST} outlier loci from the three focal species (house sparrow F_{ST} outliers in blue, Italian sparrow in green and Spanish sparrow in red).

Table1. Univariate matrix regression with randomization (UMRR). 8 populations of the Italian sparrows. Pairwise F_{ST} between populations as the response variable. Predictor variables are as following: Annual mean temperature (A.TEMP), Temperature seasonality (TEMP.S), Altitude (ALT), Geographic distance (GEO), Annual mean precipitation (A.PREC), Precipitation seasonality (PREC.S), Beak height (BEAK.H) and Beak length (BEAK.L)

	R^2	β	*t*	*p-value*
TEMP.S	0.163	0.007	2.251	0.048 *
A.PREC	0.061	-0.004	-1.302	0.260
GEO	0.053	0.004	1.201	0.264
BEAK.L	0.036	0.003	0.991	0.358
BEAK.H	0.036	0.003	0.991	0.341
PREC.S	0.031	0.003	0.910	0.422
ALT	0.020	-0.002	-0.732	0.513
A.TEMP	0.001	-0.001	-0.172	0.864

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 FST), in parentheses the per cent contribution of the predictor to the total variance explained by the model (100 * partition coefficient (U, C or T) / R2). Pairwise FST between 8 populations of the Italian sparrow as the response variable. Predictor variables are the following: Annual mean temperature (A.TEMP), Temperature seasonality (TEMP.S), Altitude (ALT), Geographic distance (GEO), Annual mean precipitation (A.PREC), Precipitation seasonality (PREC.S).

MODEL 1: Fst ~ 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	0.34	0.03 (12%)	0.02 (8%)	0.05 (20%)
A.TEMP	0.001	0.15	0.89	0.14 (56%)	0.03 (12%)	0.16 (64%)
A.PREC	-0.004	-1.16	0.32	0.05 (20%)	0.01 (4%)	0.06 (24%)
TEMP.S	0.007	1.96	0.05 *	0.003 (0%)	0.02 (8%)	0.02 (8%)
PREC.S	-0.003	-0.60	0.57	0.001 (0%)	0.00 (0%)	0.00 (0%)
ALT	-0.002	-0.29	0.78	0.01 (4%)	0.02 (8%)	0.03 (12%)

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 dataset of 2737 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	Std. Error	p-value
Italian F_{ST} Outlier	~ SH F_{ST}	SH F_{ST}	-2.1391	1.2719	0.0926
	~ IH F_{ST} + IS F_{ST}	IH F_{ST}	1.9886	2.4321	0.4135
		IS F_{ST}	-7.6170	3.0571	0.0127 *
	~ IH F_{ST} * IS F_{ST}	IH F_{ST}	3.2764	2.6091	0.2092
		IS F_{ST}	-5.9250	3.5668	0.0967
		IH F_{ST} : IS F_{ST}	-49.4795	66.2468	0.4551