Predictors of genomic differentiation within a hybrid taxon

Angelica Cuevas^{1*}, Fabrice Eroukhmanoff¹, Mark Ravinet¹², Glenn-Peter Sætre¹ and Anna Runemark³

Affiliations:

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

²School of Life Sciences, University of Nottingham, NG7 2RD, Nottingham, United Kingdom.

³Department of Biology, Lund University, Sölvegatan 37, SE-22362, Lund, Sweden.

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

1 ABSTRACT

2 Hybridization is increasingly recognized as an important evolutionary force. Novel genetic methods
3 now enable us to address how the genomes of parental species are combined to build hybrid
4 genomes. However, we still do not know the relative importance of contingencies, genome
5 architecture and local selection in shaping hybrid genomes. Here, we take advantage of the
6 genetically divergent island populations of Italian sparrow on Crete, Corsica and Sicily to
7 investigate the predictors of genomic variation within a hybrid taxon. We test if differentiation is
8 affected by recombination rate, selection, or variation in ancestry proportion from each parent
9 species. We find that the relationship between recombination rate and differentiation is less
10 pronounced within hybrid lineages than between the parent species, as expected if purging of minor
11 parent ancestry in low recombination regions reduces the variation available for differentiation. In
12 addition, we find that differentiation between islands is correlated with differences in selection in
13 two out of three comparisons. Patterns of within-island selection are correlated across all islands,
14 suggesting that shared selection may mould genomic differentiation. The best predictor of strong
15 differentiation within islands is the degree of differentiation from house sparrow, and hence loci
16 with Spanish sparrow ancestry may vary more freely. Jointly, this suggests that constraints and
17 selection interact in shaping the genomic landscape of differentiation in this hybrid species.

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 species 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 [2,10,11,12,13]. 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 increase in chromosome number [16,17,18,19]. Interestingly, the relative contributions of the parental species may vary within a hybrid taxon, as illustrated by the variable genome composition in sword-tail guppies [20], in *Lycaides* butterflies [21], and in isolated island populations of Italian sparrows [22].

Hybridization, at least in animals, was historically viewed as an evolutionary mistake [23], partly because hybrids are likely to suffer from incompatible combinations of derived alleles. While this view has changed over the last decades [17,24], hybrid lineages likely need to overcome a number of challenges to establish. Incompatibilities might mean low fertility, sterility or even inviability in some crosses [25]. This is shown by Haldane's rule [26], 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 [27,28,29].

43 For example, maladaptive metabolisms in hybrid species [30] suggest that mito-nuclear interactions
 44 could pose strong selection pressures on the genomic composition in hybrid taxa. Mito-nuclear
 45 interactions may also play a role in determining Italian sparrow genome composition [27]; hybrid
 46 Italian populations are largely fixed for house sparrow mitochondrial genomes, and there is
 47 evidence of an excess of house sparrow ancestry conserved in nuclear genome regions contributing
 48 to mitochondrion function [22]. Even in species that have successfully formed hybrid daughter
 49 lineages, early generation hybrids may still be inviable or infertile [31]. These findings suggest that
 50 only specific combinations may be viable in a hybrid lineage or only specific portions of the
 51 genome are free to vary, potentially resulting in convergent genomic compositions of hybrid
 52 lineages.

53

54 Although hybrid lineages in principle have a vast number of potential combinations of parental
 55 alleles and increased nucleotide diversity available as a source for adaptation, little is known about
 56 genome stabilization in hybrid taxa [19,32]. There could potentially be constraints on genomic
 57 mosaicism and bias on the overall composition of hybrid genomes. For instance, in nightingales
 58 [33], flycatchers [34] and field crickets [35] introgression on the sex chromosomes was reduced
 59 compared to genome-wide levels, consistent with selection against infertility. Experimental assays
 60 in sunflowers have shown that the same genetic combinations found in natural hybrid lineages re-
 61 emerge in experimental hybrid populations [36], possibly due to selection against alternative
 62 combinations. This raises the question of how easily hybrid lineages can achieve divergent genome
 63 compositions and phenotypes. Can different combinations of parental alleles easily be achieved due
 64 to selection for divergent local adaptation in homoploid hybrids? Or do patterns of differentiation at
 65 a local scale mirror those between strongly divergent populations, suggesting a role for constraints
 66 on which genome regions may differentiate? Exploiting the patterns of population differentiation
 67 within hybrid species may reveal novel insights into the selective forces shaping hybrid genomes.

68 Interestingly, patterns of species differentiation are affected by the recombination rate landscape
 69 [37,38,39]. This can result in highly correlated patterns of divergence between closely related
 70 species pairs, such as that found in flycatchers [37]. Moreover, linked selection is more efficient in
 71 low recombination regions [37]. Selective sweeps in genomic regions of low recombination can
 72 give rise to a negative correlation between recombination rate and genomic differentiation [37,40].
 73 Specific to hybrid taxa, purging of minor parent ancestry in low recombination regions to reduce
 74 genomic incompatibilities could reduce the variation available for differentiation. Recent studies
 75 have indeed found reduced introgression in low recombination regions in hybrid swordtail fish,
 76 sticklebacks, *Heliconius* butterflies and humans, suggesting that the recombination landscape may
 77 indeed affect which regions are permeable to introgression [20,38,41]. Genomic blocks from the
 78 minor-parent may be retained in regions of high recombination rate due to their increased likelihood
 79 of breaking away from potential incompatibilities [20]. If purging of incompatibilities in low
 80 recombination regions is pervasive, the relationship between recombination rate and differentiation
 81 is expected to be less negative in hybrid lineages compared to the differentiation between parental
 82 taxa (Fig. 1A), but the relative importance of these two processes in shaping hybrid genomes
 83 remains unclear.

Figure 1. Expected predictors of within-island genomic differentiation. **A. Recombination rate:** *Parent-parent:* Since selective sweeps are more efficient 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 ancestry in low recombination

regions could reduce variation and the potential for differentiation. 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. Also, in higher recombination regions there is a potential for alternative blocks of ancestry to be fixed, which will inflate the differentiation among hybrids as recombination rate increases, relative to the parent-parent comparison, contributing to the flatter relationship in hybrids **B. Selection:** If selection is an important predictor of differentiation, we expect that differentiation among populations is positively correlated to differences in selection among populations. **C. Parental differentiation:** If differentiation is limited by potential incompatibilities between divergent parental loci we would expect that the potential for adaptive differentiation in the hybrid would be highest in genomic regions of low parental differentiation. In contrast, highly differentiated regions, more likely to be under constraint due to incompatibilities between parents, would have a greater likelihood of being fixed for the ancestry of one parent only.

84 There is evidence that hybrid taxa can use the variation that originated through hybridization for
85 local adaptation in the hybrid taxon. For instance, beak shape in Italian sparrows is explained by
86 local precipitation regime [42]. Beak size differences between island populations are best explained
87 by temperature seasonality [43] and the island populations are strongly differentiated for a gene
88 known to affect beak morphology in Darwin's Finches, FGF10 [22,44]. In addition, the gene GDF5,
89 part of the BMP gene family that has a fundamental role in beak shape and size variation in
90 Darwin's finches [45,46], is a candidate gene putatively under selection in the Italian sparrow
91 populations from mainland Italy [32]. However, beak shape is also affected by the proportion of the
92 genome inherited from each parent species suggesting a small albeit significant role for
93 contingencies in morphology [42,43]. Moreover, in a recent study investigating genetic
94 differentiation among Italian sparrow populations on mainland Italy, population differentiation was
95 associated with temperature and precipitation [32]. Interestingly, alleles segregating in both parental
96 species showed strong allele frequency differences within the Italian sparrow, suggesting that
97 adaptation is not completely contingent on the combination of alleles from different parent species
98 [32]. Selection, fixing alleles from either one or the other parent depending on which variant is
99 favoured locally, could contribute to the stabilization of hybrid genomes. If divergent selection
100 pressures between islands, favouring the fixation of alternate parental alleles, are important for
101 differentiation, we would expect stronger differentiation among islands in regions under divergent
102 selection (Fig 1B). However, genomic regions of parental divergence are more likely to harbour

incompatibilities, thus fixation for ancestry of one parental allele only is likely to occur in the hybrid due to constraints. Consequently, local differentiation in the hybrid could be expected to be highest in genomic regions of low parental differentiation (Fig 1C).

One way to further our understanding of the selective forces acting on hybrid genomes is to investigate patterns of differentiation within hybrid lineages and the factors that best explain them. Here, we focus on the Italian sparrow, a clear example of a homoploid hybrid species, with reproductive barriers to the parent species consisting of a subset of those isolating the parent species [27,47]. This is a uniquely suited study system as there are genetically divergent island populations, potentially originating from independent hybridization events, with different proportions of their genomes inherited from their parental species, house and Spanish sparrows (*P. domesticus* and *P. hispaniolensis*) [22] (Fig. 2). We use the island Italian sparrows to investigate how differentiation *among populations within each island* compares to *among island* differentiation. Our aim is to provide insight into how readily within-island local populations of hybrid species differentiate in relation to differentiation between islands. We also address which factors best predict differentiation, specifically addressing whether recombination rate, differences in selection or ancestry contribution best predict differentiation within- and among islands.

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 has led to a correlation between recombination rate and differentiation that is lower than that between the parent species (Fig. 1A) III) that regions under stronger differential selection between islands also are more differentiated *among* islands (Fig. 1B) IV) that constraints on how freely genomic regions are able to diverge have led to a correlated landscape of *within-island* and *among-island* differentiation, and V) that differences in

128 minor – major parental ancestry in the hybrid have a direct effect on the genomic differentiation
 129 between populations.
 130

Figure 2. Sampling design and population structure. **A.** Map of sampling locations (map redrawn from Runemark et al., 2018a). Sparrows were sampled in three populations in each island, 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 with highest support. Three geographically separated populations of the parental species, the house sparrow and Spanish sparrow, were included. **C.** Principal component analysis illustrating the relationship between the Italian sparrow populations and the parental species. ♦ indicate the reference parental populations with lower levels of introgression.

RESULTS

I) Genomic differentiation within- and between islands

Consistent with [22] we find strong differentiation between the studied islands. From our principal component analysis, 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). F_{ST} estimates are consistent with this, with Crete being overall more differentiated from Sicily (mean $F_{ST} = 0.043$) and Corsica (0.042), while mean F_{ST} between Corsica and Sicily is lower (0.025; Fig. 3A).

Figure 3 (next page). Genomic differentiation in islands. **A.** Levels of genomic differentiation *among*-islands compared to levels of differentiation between populations *within*-islands. Means and 95% CI depicted. **B.** Discriminant analysis of principal components (DAPC) for populations in Corsica, Crete and Sicily. **C.** Map of studied islands with the three populations sampled within each island.

140 F_{ST} estimates between islands are higher than those within islands (Monte-Carlo permutation paired
 141 t-test ($t=33.21$, $df=7927$, $P=1.98\text{e-}15$; Table S3; Fig. 3A). A discriminant function analysis recovers

some differentiation among local populations in each island (Fig. 3B, Fig. 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 *within*-island populations of origin. Within-island F_{ST} differs significantly among islands (All $P_s < 0.5e-3$; Table S3), with Corsican populations exhibiting the highest mean F_{ST} of 0.018, as well as the highest nucleotide diversity (π : $3.021e-06$; Table S3). Differentiation within Sicily is intermediate at 0.013, while Crete has the lowest within-island F_{ST} of 0.011 (Fig. 3A, Table S3). 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, Fig. S1).

Table 1. Analysis of Molecular Variance (AMOVA) across islands and populations within islands. Several cut-offs for missing-ness per loci were use (5%, 10%, 20%), but the results from the AMOVA did not change substantially. Significance of variation partitioning elements was maintained.

Analysis of Molecular Variance					Randomization by	
AMOVA					Permutation	
					Monte Carlo test	
Variance partitioning	Df	Sum Sq	Sigma	% of covariance	Std. Observed	<i>P-value</i>
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

II) The relationship between genomic differentiation and recombination rate

We tested the difference in the slopes of the relation between recombination rate and genomic

155 differentiation for each island and the parental divergence. We also evaluated whether there was a
 156 significant interaction between comparison (e.g. parent-parent vs. within island) and recombination
 157 rate on differentiation for each island. This allows us to test the hypothesis from Fig 1A; i.e. we
 158 predicted that the slopes would be less steep in the hybrids if variation and hence differentiation is
 159 decreased in low recombination regions in hybrids due to selection against minor parent ancestry,
 160 and find that this is the case (Table S4; Table S5; Fig. 4A). The slope generated by the relation
 161 between the parental differentiation and recombination rate differs from those found in each within-
 162 island comparison (Table S4; Fig. 4A). We find a significant interaction of recombination rate and
 163 comparison (e.g. parent-parent vs. within island) in all independent linear models per island (Table
 164 S5). We find no significant correlation between recombination rate and within-island genomic
 165 differentiation for Corsica (correlation=-0.012, $R^2=1.4e-4$, $P=1$; Fig. 4B) or Crete (correlation= -
 166 0.003, $R^2=0.9e-5$, $P=1$; Fig. 4C). However, differentiation within Sicily is weakly but significantly
 167 negatively correlated with recombination rate, (correlation=-0.048, $R^2=0.002$, $P=0.037$; Fig. 4D).

Figure 4 (next page). 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 *within*-island genomic differentiation for **B.** Corsica, **C.** Crete and **D.** Sicily.

168 A GLM examining the influence of recombination rate, parental differentiation and differentiation
 169 to the two parent species, respectively, did not reveal any significant relationship between
 170 recombination rate and differentiation within islands (Estimate=-9.66e-04, Std.Error=7.84e-04,
 171 $P=0.22$; Table S7). A corresponding binomial model revealed that recombination rate did not affect
 172 the probability of loci being F_{ST} outliers within islands either (Table 2, Table S6). However, a
 173 corresponding GLM evaluating individual islands separately shows that recombination rate
 174 significantly explains differentiation between Sicilian populations (Estimate=-3.48e-
 175 03, Std.Error=1.34e-03, $P=9.3e-3$; Table S8; Fig. 4D).

Table 2. Logistic regressions on the probability of being a local F_{ST} outlier within island.

Model:				
Pr(outlier)= Hybrid Index + 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	0.327
	Hybrid Index	1.411e-02	6.336e-02	0.824
	Island v.s House F_{ST}	-1.401e+00	6.674e-01	0.036*
	Island v.s Spanish F_{ST}	5.728e-01	4.698e-01	0.223
	House v.s Spanish F_{ST}	-6.752e-06	8.770e-06	0.441

Post-hoc Estimated marginal (Least-squares) means for the predictor variable “island”

Comparison	Estimate	Std. Error	Z ratio	P-value
Corsica - Crete	0.0207	0.133	0.159	0.9866
Corsica - Sicily	-0.1135	0.133	-0.877	0.6684
Crete - Sicily	-0.1343	0.134	-1.025	0.5756

p-value adjustment: Tukey

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 test if differentiation is significantly correlated to differences in selection. We find that genomic differentiation between island pairs was significantly correlated to divergent selection. F_{ST} and x_p -EHH were negatively correlated for the Corsica – Sicily (correlation = -0.061, $R^2=0.0037$, $P = 0.014$), and Crete and Sicily (correlation = -0.059, $R^2=0.004$, $P = 0.019$; Fig. 5A; Table S9A) comparisons. However, we do not find any significant correlation between differentiation and selection differences for the Corsica – Crete combination (correlation = -0.042, $R^2=0.002$, $P =$

184 0.148).

Figure 5. Effects of selection on genomic differentiation, similarity in selection pressures across islands and shared patterns of genomic differentiation. A. Relationship between divergent selection (Extended Haplotype Homozygosity statistic - XP-EHH) and genomic differentiation *between-island*. **B. Similarity in patterns of selection.** Correlations of *within-island* selection, measured as integrated haplotype homozygosity score (iHS) between islands. **C.** Shared outlier loci among populations *within-islands* (right)

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

We also addressed how consistent selection is within and across islands. Patterns of *within*-island selection (iHS) showed a significant positive correlation in all pairwise comparisons between islands, with R^2 ranging from 0.095 to 0.17 for iHS (Fig. 5B; Table S9C). This suggests shared patterns of selection across the genome among islands, potentially driven by similar selection pressures or genomic constraints reducing the availability of genomic variation. However, differentiation within island populations is not significantly correlated to selection (iHS) for Crete or Corsica (Fig. S2; Table S9B), although there is a weak significant correlation for Sicily ($R^2=0.005$, $P=0.01$; Fig. S2; Table S9B). Interestingly, mean Tajima's D differs considerably among islands, with Sicilian and Corsican populations exhibiting negative estimates (range -0.25 to -0.048; Fig. S3; Table S3). In contrast, populations on Crete exhibit higher values of Tajima's D (-0.11 to 0.015; Fig. S3; Table S3).

IV) Distribution of differentiation across the genome

We found that patterns of differentiation are significantly correlated between Sicily and Corsica (R^2 : 0.0066, $P = 1.16e-10$), but not between Corsica and Crete (R^2 : $5.8e-4$, $P = 0.16$), or Sicily and Crete (R^2 : $1.96e-4$, $P = 0.81$; Fig. S4). Hence, levels of differentiation are not correlated between all islands. Interestingly, we find that the outlier loci that are most differentiated within one island occur more frequently among other island outliers than expected by chance in two out of three comparisons (Fig. 5C; Table S10). 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 is found in within-Sicily F_{ST} outliers (9.3%; Chi-squared: 6.80, P : 0.009; Table S10). However, Crete and Sicily outliers do not significantly overlap (Chi-squared: 0.09, P : 0.767). We also tested whether individual island outliers also have a higher mean distribution of F_{ST} values within other islands. We find that

208 outliers from Sicily have a higher F_{ST} than the mean among Corsican populations ($t = -3.082$, $df =$
 209 444.7 , $P = 0.002$; Fig. 5D) and among populations from Crete ($t = -1.997$, $df = 457.9$, $P = 0.046$).
 210 Similarly, Corsica F_{ST} outlier loci have higher F_{ST} within Sicily ($t = -3.393$, $df = 392.2$, $P = 7.6e-4$)
 211 and within Crete ($t = -3.586$, $df = 385.1$, $P = 3.8e-4$). However, outliers from Crete do not have
 212 higher F_{ST} values than expected by chance within Corsica ($t = -1.403$, $df = 400.3$, $P = 0.161$), or
 213 Sicily ($t = 0.078$, $df = 425.1$, $P = 0.938$; Fig. 5D).

214 Pair-wise correlations between differentiation within- and among islands suggest that while
 215 differentiation within Corsica is correlated to differentiation between Corsica and Sicily (R^2 :
 216 0.0025 , P : 0.013) and Crete and Sicily (R^2 : 0.0025 , P : 0.026 ; Fig. S5), none of the other
 217 comparisons are significant. Consistent with this pattern, the proportion of the most differentiated
 218 loci within Corsica that overlap with the most differentiated loci between Corsica-Sicily (9.8%) and
 219 between Crete-Sicily (7.4%) are also higher than expected by chance (Chi-square tests: P : $8.1e-05$
 220 and P : 0.04 , respectively; Table S11). Moreover, we also find that a higher proportion (10.2%) of
 221 loci that are most differentiated within Crete than expected by chance are among those most
 222 differentiated between Crete and Sicily (P : $4.3e-06$, Table S11).

223

224 **V) Patterns of local genomic differentiation in relation to parental contributions to the** 225 **genome**

226 The factor that best predicts the probability that a SNP is an outlier or not within islands is the
 227 extent of differentiation to the house sparrow (Logistic regression estimate: $-1.401e+00$, P : 0.036 ;
 228 Table 2). However, when evaluating factors that may affect the within-island differentiation (F_{ST}),
 229 using a GLM, the differentiation to the house sparrow was found to be non-significant (Table S7).
 230 Neither the extent of differentiation from the Spanish sparrow, parental differentiation,
 231 recombination rate, nor hybrid index (HI) contributed significantly to differentiation, in both

models, the logistic regression or the GLM (Table 2; Table S7). In separate logistic regressions for each island, none of the studied factors significantly affected the probability of being an outlier (Table S6), potentially because of reduced statistical power. While in separate GLMs per island (Table S8), general distribution of differentiation (F_{ST}) between Corsican populations is explained by the differentiation to the Spanish sparrow (Estimate=1.166e-02, $P=0.036$) and within-Sicily differentiation is significantly explained by recombination rate (Estimate=-3.475e-03, $P=9.3e-3$; Table S8).

Taking advantage of the differences in parental contribution among islands, we addressed whether differentiation within islands was correlated to differentiation between the focal island population and the minor-ancestry parent species. For Corsica and Sicily, there is a significant correlation between within-island differentiation and that against 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. S6B). Population differentiation within Crete was not correlated to differentiation to any of the parental species (Fig. 6A; Fig. S6B). Similarly, loci involved in high population differentiation *within*-islands (1% F_{ST} outliers) were more differentiated to the minor-ancestry parent than the genome wide neutral expectation (Fig. 6B). Outliers from Corsica and Crete present higher genomic differentiation to the Spanish sparrow, parent contributing the least ancestry, than the overall genome-wide average. However, the same pattern is not found in Sicily, where the house sparrow is the minor-ancestry parent, *within*-island outliers do not show higher differentiation than the genome-wide average (Fig. 6B). Also, outliers from Corsica were significantly more differentiated from Spanish- than house sparrow ($t = -6.22$, $df = 519.4$, $P = 1.01e-09$), as were outliers from Crete ($t = -2.96$, $df = 668.2$, $P = 3.17e-3$). And, outliers within Sicily, where the Spanish sparrow is the major-ancestry parent, are more differentiated from house sparrow than from Spanish sparrow ($t = 3.76$, $df = 679.8$, $P = 1.81e-4$; Fig. 6B). While divergence

257 between the parental species (house-Spanish F_{ST}) is weakly correlated to population differentiation
 258 within Corsica ($R^2 = 3.4e-3$, $P = 5.34e-5$) and it is not significantly correlated to differentiation within
 259 the other islands (Fig. S6A). Similarly, within Corsica, outlier loci also have higher parental
 260 differentiation than expected by chance ($t = 2.15$, $P = 0.033$; Fig. 6B), but this does not hold true for
 261 Crete ($t = 1.852$, $P = 0.065$) or Sicily ($t = -0.811$, $P = 0.42$). Differentiation *within* Corsica and
 262 *within* Sicily was not correlated to differentiation *within* either of the parental species (Fig. S7) and
 263 differentiation *within* the parental species was not higher than expected by chance in outlier loci for
 264 these two islands (Fig. S8). Differentiation *within* Crete, and highly differentiated loci, was
 265 however, negatively correlated to differentiation *within* the house sparrow (Fig. S7; Fig. S8). These
 266 results suggest that overall, outlier loci for *within*-island differentiation are not involved in the
 267 genomic differentiation among populations within each parent species.

Figure 6 (next page). Effects of divergence from parental species and minor –major parent ancestry on differentiation. **A.** Correlations of between populations *within*-island F_{ST} and genomic differentiation from the minor-ancestry parent (island v.s minor-parent F_{ST}). **B.** Between parents divergence (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 hybrid index (HI) for within-island F_{ST} outliers compared to the genome wide distribution. Means and 95% confident intervals of outlier and non-outlier loci depicted.

268 Finally, we assessed the effect that ancestry divergence can have on the genomic differentiation
 269 within islands. The distribution of the hybrid index (HI) does not differ between the *within*-island
 270 F_{ST} outliers and the non-outlier loci for any of the islands (t-test of $t=0.27$, 1.04 and -0.53 , for
 271 Corsica, Crete and Sicily, respectively. $P>0.05$ for all islands; Fig. 6C). A post hoc correlation
 272 analysis shows no evidence to support that the overall genomic differentiation within islands is
 273 largely explained by ancestry (R^2 varying between $5.1e-4$ and $1.8e-5$, with $P>0.05$ for all islands;
 274 Fig. S9A). We further tested if the HI affected the probability of a locus to be highly differentiated
 275 within islands (1% F_{ST} outlier loci; Fig. S9B). Examining outliers with extreme values of HI only,
 276 we found that outliers within Sicily more frequently have excess Spanish ancestry (mean frequency
 277 proportion based on 10000 random draws = 0.31), whereas Corsica and Crete have outliers with

278 excess house ancestry (mean frequency proportion based 10000 random draws = 0.804 and 0.799,
279 respectively; Fig. S8B). Even though comparison with proportions from a similar resampling
280 analysis for non-outlier loci shows that distributions differ (Corsica: $t = 27.175$, Crete: $t = -30.846$,
281 Sicily: $t = -4.369$, all $P < 1.25e-05$), the ancestry pattern of outliers is generally very similar to that
282 of the general genomic background (Fig. S9B).

283

284 **DISCUSSION**

285 While evidence for a creative role of hybridization in evolution is piling up, little is known about
286 how the genomes of hybrid taxa can differentiate in response to local selection pressures.
287 Investigating the factors that explain how hybrid taxa can differentiate within lineages of the Italian
288 sparrow, we find high within-island differentiation at relatively short distances in the light of the
289 dispersal ability of the species [48]. A discriminant function analysis classifies 75-95% of the
290 individuals to the correct population within islands. This local differentiation suggests that there is
291 potential for population structure and adaptive divergence within this hybrid species. However,
292 there is also pronounced between-island differentiation, approximately five times higher than
293 within-island differentiation. Interestingly, genomic differentiation and a measure of between-
294 islands selective sweeps are significantly correlated in two out of three between-island-
295 comparisons, supporting a role for natural selection in shaping the differentiation of the hybrid
296 genomes of Italian sparrows in geographically isolated islands. However, contrary to our
297 expectation of positive correlations, the negative nature of the relationship do not support the
298 scenario where highly differentiated parental alleles are involved in divergent natural selection
299 facilitating genomic differentiation between islands. Genomic differentiation within island is poorly
300 explained by signatures of selective sweeps, with the exception of Sicily, where patterns of
301 selection are weakly correlated with local genomic differentiation. In this species, we know from
302 previous findings that local differences in beak shape among island populations is driven by climate

and diet [43] and adaptive differentiation in mainland Italy is also driven by climatic variables [32,42]. This is also consistent with the large body of work suggesting that hybridization enables adaptive variation across a range of taxa [2,49,50,51].

Interestingly, there is some evidence 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 with 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 (Ravinet *et al.*, in prep). This introgression may also have contributed to Crete forming a third cluster in the Admixture analysis. In addition, Corsica also shares a higher proportion of the most strongly differentiated outliers 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 linked to local genomic differentiation may be limited to specific genomic regions and are reused across hybrid lineages.

Interestingly, patterns of selection are consistent across populations and correlated between islands. Hence, the findings of some degree of shared differentiation could partially be explained by a similar selection landscape 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 [36]. 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 (Ravinet *et al.* in prep), selection for a functional admixed genome [22,29], stabilizing selection, linked to human commensalism [52] or

parallel selection linked to adaptation to insularity. For instance, such shared ancestral selection landscape could lead to bias in which parental alleles are retained or more prone to be lost or selected against (i.e. the Spanish sparrow being considered a non-human commensal in most of its range, compared to the house sparrow).

A major finding is the limited evidence of genome structure shaping local differentiation. Variation in the underlying recombination rate landscape may mould the landscape of differentiation [53], and has been shown to affect e.g. the genomic differentiation generating correlated patterns of differentiation in divergent populations of mice, rabbits [40] and flycatchers [37]. In admixed species, selection against minor parent ancestry has been hypothesised to generate patterns of strong correlation between measures of introgression and recombination rate [20,38,41]. 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 during stabilization of hybrid genomes (Fig. 1A). We find that within island divergence is significantly less affected by recombination rate than divergence between the parent species. This could suggest that selection against incompatible minor parent alleles in low recombination regions reduces the potential for differentiation within the hybrid species in these regions. Overall, very little of the divergence between populations within islands is explained by recombination rate, despite its weak correlation with differentiation 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 (Fig. 1C). Low parental divergence is expected to reduce the number of segregating

alleles for selection to act on, consistent with findings that hybrids from more divergent parent species are morphologically more novel [54,55]. However, genomic regions of high divergence between parental species can also 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). Indeed, we find that differentiation between the parent species explains neither the degree of differentiation within islands, nor improves the fit of the models evaluating *within-island* differentiation (logistic and GLM-models). Interestingly, a similar pattern has previously been found in the populations of the Italian sparrow in mainland Italy [32], suggesting that constraints might have played an important role during the stabilization of the hybrid genome, limiting genomic variation. The repeatability of this result across the range of the Italian sparrow also suggests that hybridization may indeed lead to predictable patterns of genomic variation long after genome stabilization and hybrid species formation. We find that differentiation from the house sparrow is the factor that best explains the probability of being a highly differentiated locus within islands. An additional source of variation in hybrid species could stem from variants that segregate within the individual parental species. However, we did not find any compelling evidence that within-parent differentiation predicts differentiation within the hybrid species.

370

Whether ancestry is a determining factor on how genomic differentiation is distributed in the hybrid genome is not easily disentangled. The divergence in ancestry proportion from the minor – major parent in the Italian sparrow [22] enables us to test whether the differences in ancestry has affected population differentiation. Purging of genomic incompatibilities of the minor parental ancestry blocks plays an important role in determining the genetic variation in the Italian sparrow [27,47]. A range of studies has suggested that the probability of retaining neutral ancestry is higher in genome regions with a high recombination rate [20,38,41,56]. To address if minor parent ancestry, in spite

378 of selection against incompatibilities, could be involved in adaptation within the Italian sparrow, we
 379 investigated whether minor parent ancestry was important for differentiation within the Italian
 380 sparrow. We did not find a clear effect of ancestry in population differentiation within islands;
 381 nevertheless, overall highly differentiated loci are not found more frequently in minor-parental
 382 ancestry blocks. However, we found significant correlations between local differentiation and the
 383 differentiation to the minor-ancestry-parent for two out of three islands. Overall F_{ST} -outliers are also
 384 more differentiated from the minor-parental blocks than expected based on genome-wide levels of
 385 differentiation. This suggests that loci linked to local differentiation *within*-islands are also involved
 386 in divergence to the minor-ancestry parent. Recombination rate can determine how ancestry is
 387 distributed across the hybrid genome [20,38] and this may affect the effect ancestry has on genomic
 388 differentiation. The probability for minor-ancestry blocks to rapidly decouple from potential
 389 incompatibilities with the major-parent genetic background, increases with recombination rate [20],
 390 which in turn can have an effect on how easily regions with minor-ancestry in high/low
 391 recombination areas are able to harbour genomic differentiation among populations.

392
 393 Interestingly, the probability of being among the 1% most strongly differentiated loci is best
 394 explained by how differentiated a given island population is to the house sparrow. Neither
 395 recombination rate, the ancestry of the region, nor the differentiation to the Spanish sparrow
 396 significantly affected the degree of differentiation or the probability that the locus was an outlier.
 397 This is an interesting finding as [22] previously also found a skew towards house sparrow ancestry
 398 in loci consistently inherited from one parent species, particularly identifying an enrichment of
 399 mito-nuclear loci and loci involved in DNA-repair. Potentially, this could be indicative of some
 400 constraints on differentiation from the house sparrow, as most Italian sparrows are fixed for house
 401 sparrow mitochondrial haplotypes [5,22].

402

Conclusion

Taken together, our findings suggest that hybrid lineages may use their elevated levels of variation for local adaptation, with differentiation among islands being significantly correlated to differences in signatures of selective sweeps in two out of three comparisons. Moreover, the findings of correlated differentiation among islands as well as similar selection pressures 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 would 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 and past hybridization during the stabilization of hybrid genomes. To further understand how constraints and selection shape hybrid genomes, one interesting future line of inquiry would be investigating whether alleles inherited from the parent contributing least to the genome in genomic regions involved in population differentiation are more likely to occur in high recombination rate regions.

MATERIALS AND METHODS

The Italian sparrow originated from hybridization between house and Spanish sparrow, likely during the spread of the commensal house sparrow to Europe in the wake of the introduction of agriculture [52,57]. The parental species diverged approximately 0.68 million years ago [52]. 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 [43]. Furthermore, the island populations are evolutionarily independent and are hypothesized to have arisen from individual hybridization events.

Sampling and sequencing

Three populations of Italian sparrows were sampled from each of the islands Sicily (n=76), Crete (n=77) and Corsica (n=70) in March-June 2013 (Fig. 2A, Table S1). On each island we sampled individuals from three geographically separated populations (Fig. 2B, Fig. 3C). Population sample size varied between 16 and 30 (Table S1). 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); (Table S1). All birds were caught using mist nets, 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, and library preparation, sequencing, demultiplex 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).

Data processing and variant calling

First, the quality of all RAD sequencing was checked using FASTQC [58]. Raw reads were filtered using the module process_radtag from the software Stacks [59]. 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 (99.9% 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) [60] to map the reads to the house sparrow reference genome [5] using default parameters. We re-aligned indels realignment with GATKs (v 3.7) RealignerTargetCreator and IndelRealigner [61,62] and called the variants using HaplotypeCaller [62]. For a detailed description of the variant pipeline, see [32]. We filtered SNPs using Vcftools v. 0.1.14 [63], setting the filter parameters to --max-missing 0.8, --minDP 10.00, --minGQ 20.00 and --maf 0.02. We pruned linked sites using PLINK v. 1.9 [64] with 100 kb windows, 25 bp step and an R^2 threshold of 0.1. VCF-files containing different set of individuals were generated to suit the different analyses (Table S2).

464

465 **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 [65]. 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 [66]. 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 (Table S2). After filtering for LD, the remaining 2224 SNPs were included in the PCA and Admixture analyses. 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. To characterize potential variation in genomic diversity we also estimated nucleotide diversity in sliding windows of 100kb in size with 25-kb steps for each population using vcftools v. 0.1.14 [63]. 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 pairwise F_{ST} among islands in 100kb windows using vcftools v. 0.1.14 [63]. The window size was selected as linkage disequilibrium in sparrows is known to decay within this distance [5]. 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 [67,68]. We assessed significance by a randomization of population assignments using a Monte Carlo test with 1000 permutations implemented in the randtest() function from the ADE4 R-package [69]. See supporting methods for additional analyses assessing the sensitivity to missing data.

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 [37,40]. 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 [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.

517

518 **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) [70], 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 correlations between each island pair of the

integrated haplotype homozygosity score (iHS) [71] 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 [72,73] and converted the resulting VCF-file using the vcfr R-package [74]. We then used the functions `data2haplohh()`, `ihh2ihs()` and `ies2xpehh()` from the rehh R- package [75,76] to prepare the data, estimate the integrated haplotype homozygosity score (iHS) and estimate Extended Haplotype Homozygosity (XP-EHH), respectively.

IV) Distribution 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 pair-wise comparison, applying Bonferroni corrections for multiple testing.

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) using within-island F_{ST} as the response variable: $F_{ST} = \text{ancestry proportion (hybrid index) + recombination rate + island + island to house sparrow differentiation } (F_{ST}) + \text{island to Spanish sparrow differentiation } (F_{ST}) + \text{parental differentiation (house-Spanish } F_{ST})$. We also evaluated how these factors affected the probability of a locus to belong 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}).

Finally, to address if variation in minor parent ancestry affects within-island differentiation, we tested the correlation between genomic differentiation and hybrid index (HI) reflecting the relative contribution of each parent species. We also tested if highly differentiated loci were found in blocks with high allele frequencies from major- (less than 35% minor parent alleles) or minor parent ancestry (less than 35% major parent alleles) more frequently than expected by chance 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. Hybrid index (HI) was estimated using whole genome data from [22] for Crete, Corsica and Sicily, including parental species from [52] and [5]. We phased data using SHAPEIT/v2.r837 [72,73] and estimated ancestry using LOTER [77] and translated this into a per locus hybrid index, where 0 corresponds to only Spanish ancestry and 1 corresponds to pure house sparrow ancestry. We estimated the hybrid index separately for each island.

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), Consejería de Industria, Energía y Medio Ambiente (Spain), Norwegian Food Safety Authority (Norway), and Bundesamt für Umwelt BAFU, Abteilung (Switzerland)) and Ministry of Education and Science (Republic of Kazakhstan).

ACKNOWLEDGEMENTS

We thank Jo S. Hermansen and Maria Tesaker for help with fieldwork. This research was funded by a Wenner-Gren fellowship to A.R., a Norwegian Research Council grant to G-P.S. and A.R., and a Marie-Curie Fellowship and a University of Oslo, Faculty of Sciences grant to F.E.

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.
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.
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.
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>
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>
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.
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>
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.
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.
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>
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>
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>
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>
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.
15. Rieseberg, L.H. (2003). Major Ecological Transitions in Wild Sunflowers Facilitated by Hybridization. *Science* **301**: 1211–1216.
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>
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.
18. Mallet, J. (2007). Hybrid speciation. *Nature*, 446(7133), 279–283.
<https://doi.org/10.1038/nature05706>
19. Runemark, A., Vallejo-Marín, M. & Meier, J.I. (2019). Eukaryote hybrid genomes. *PLoS Genet* **15**: e1008404.
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.
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.
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. Mayr, E. (1963) Animal Species and Evolution, *Harvard University Press*.
24. 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.
25. Coyne, J.A. & Orr, H.A. (1989). Patterns of Speciation in *Drosophila*. *Evolution* **43**: 362.
26. Haldane, J.B.S. (1922). Sex ratio and unisexual sterility in hybrid animals. *Journal of*
Genetics **12**: 101–109.
27. 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>
28. Hill, G.E. (2016). Mitonuclear coevolution as the genesis of speciation and the
mitochondrial DNA barcode gap. *Ecol Evol* **6**: 5831–5842.
29. 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.
30. 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.

31. 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.
32. 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>
33. 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.
34. 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>
35. 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.
36. 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.
37. 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>
38. 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>
39. 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>
40. 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>
41. 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/10.1101/297531>
42. 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>.
43. 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.
44. 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.
45. 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.
46. 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>

47. 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>.
48. Summers-Smith, J. D. (1988). The Sparrows: a study of the genus *Passer*. T & AD Poyser.
49. 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.
50. Consortium, T.H.G. (2012). Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. **487**: 94–98.
51. Marques, D.A., Meier, J.I. & Seehausen, O. (2019). A Combinatorial View on Speciation and Adaptive Radiation. *Trends in Ecology & Evolution* **34**: 531–544.
52. Ravinet, M., Elgvin, T.O., Trier, C., Aliabadian, M., Gavrillov, A. & Sætre, G.-P. (2018b). Signatures of human-commensalism in the house sparrow genome. *Proc. Biol. Sci.* **285**: 20181246.
53. 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>
54. Stelkens, R. & Seehausen, O. (2009). Genetic distance between species predicts novel trait expression in their hybrids. *Evolution* **63**: 884–897.
55. 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.
56. 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>
57. 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>
58. Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
59. 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>
60. 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>
61. 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>.
62. Mckenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytzky, 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.107524.110.20>
63. 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>
64. 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>
65. Jombart, T. (2008). Adegnet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24(11), 1403–1405. <https://doi.org/10.1093/bioinformatics/btn129>
66. 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>
67. 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>
68. 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>
69. 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>
70. 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>
71. Voight, B. F., Kudravalli, 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>
72. 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>
73. 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>

74. 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>
75. Gautier, M., Klassmann, A., & Vitalis, R. (2017). rehh 2.0: a reimplement 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>
76. 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>
77. 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>

SUPPORTING INFORMATION

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

Table S2. VCF files.

Table S3. Per-island population genomic statistics. Mean values of π and within-island genomic differentiation (F_{ST}). t-test for pairwise comparison between genome wide within island F_{ST} . Mean values of Tajima's D per population within each island.

825 **Table S4. Intercept, slope and confidence intervals of the slope of individual linear regression**
 826 **of *within*-island genomic differentiation and recombination rate as well as parent-parent**
 827 **differentiation and recombination rate.**

828

829 **Table S5. Evaluating the effect that the interaction between recombination rate and the type**
 830 **of comparison (parental differentiation (house-Spanish), which is the null model, and within-**
 831 **island differentiation) has on genomic differentiation (F_{ST}). Individual linear models per island**
 832 **were run to test if there is a significant interaction between recombination rate and comparison,**
 833 **as expected if the relationship between recombination rate and differentiation differs between**
 834 **parent species and the hybrid Italian sparrow (Fig. 1A).**

835

836 **Table S6. Logistic regressions per island, on the probability of being a local F_{ST} outlier within**
 837 **island.**

838

839 **Table S7. Generalized linear model on within-island F_{ST} .**

840

841 **Table S8. Generalized linear models per island on within-island F_{ST} .**

842

843 **Table S9. Concordance of within-island selection (iHS) and divergent selection (xp-EHH with**
 844 **genomic differentiation.**

845

846 **Table S10. Number and percentage of *within*-island outlier loci shared between islands. Chi-**
 847 **squared denote tests for overrepresentation compared to the genome wide average.**

848

849 **Table S11. Number and percentage of within-island outlier loci identical to between-island**
 850 **outliers.** Chi-squared denote tests for overrepresentation compared to the genome wide average

851

852 **Figure S1. AMOVA significance - Randomization via permutation.** Monte Carlo test with 1000
 853 permutations implemented in the randtest() function from the ADE4 R-package to evaluate
 854 significance. Black line denotes the observed values of Sigma (Variance in each hierarchical level).

855

856 **Figure S2. Concordance of patterns of selection and genomic differentiation.** Correlations of
 857 the integrated haplotype homozygosity score (iHS) and genomic differentiation (within-island
 858 F_{ST}).

859

860 **Figure S3. Distribution of Tajima's D per population in each island.**

861

862 **FigureS4. Correlation of within-islands differentiation across the three Mediterranean**
 863 **islands.** Bonferroni corrections of the p-values are reported.

864

865 **FigureS5. Correlation of within-islands differentiation v.s. between-islands divergence.**
 866 Adjusted p-values after resampling and Bonferroni corrections.

867

868 **FigureS6. Correlation of within-islands differentiation and the parental species.** Adjusted p-
 869 values after Bonferroni corrections.

870

871 **Figure S7. Correlations of within-island differentiation and within-parent differentiation**
 872 **(within-house or and within-Spanish sparrow).**

873

874 **Figure S8. A.** Intraspecific genomic differentiation in the parental species for the within-island F_{ST}
 875 outlier loci. Dash lines represent the within-parent F_{ST} global mean. **B.** t-tests evaluating whether
 876 within-island outlier loci present higher/lower values than expected by chance in the within-parent
 877 differentiation

878

879 **Figure S9. A.** Relation between within-island F_{ST} and ancestry (hybrid index - HI). Results of
 880 linear regression reported. Dashed lines depict the 1% outliers threshold. **B.** Frequency proportion
 881 of outlier loci found in regions of mainly house ancestry ($0.65 < HI$) and mainly Spanish ancestry
 882 ($HI < 0.35$) (minor-major parental ancestry). Distribution of 10.000 random resampling draws of 8
 883 outlier loci.

884

885 **DATA ACCESSIBILITY:** Genomic data will be deposit at the NCBI Sequence Read Archive or
 886 at the European Nucleotide Archive (ENA) at [http://doi.org/\[doi\]](http://doi.org/[doi]), reference number [reference
 887 number]. Other data will be deposited in the Dryad Digital Repository at [accession numbers or
 888 DOI will be provided once the data is uploaded]. Data will be uploaded upon acceptance of the
 889 manuscript.

890

891 **AUTHORS' CONTRIBUTIONS:** A.R. conceived and designed the study with advice from G-
 892 P.S. and F.E.; A.R., F.E. and G-P.S. collected field data; A.R. and A.C. wrote the manuscript; A.C.
 893 conducted the laboratory work; A.C. developed the RAD-data analysis pipeline with input from
 894 M.R.; A.C. and A.R. designed the analyses; A.C. analysed the data with contributions from F.E.,
 895 M.R. and A.R.; all authors discussed the results and commented on earlier drafts of the manuscript.

896

897 **COMPETING INTERESTS**

898 The authors declare that they have no competing interests.