Phenotypic Divergence in a Hybrid Species: the roles of Isolation, Drift and Selection

Bernt Christian Helén

Master of Science Thesis

Centre for Ecological and Evolutionary Synthesis
Department of Biosciences
Faculty of Mathematical and Natural Sciences

UNIVERSITY OF OSLO, NORWAY

01.06.2016

2016

Phenotypic Divergence in a Hybrid Species: the roles of Isolation, Drift and Selection

Bernt Christian Helén

http://www.duo.uio.no/

Print: University Print Centre, University of Oslo

To my dearest family, Mari and Amalie – this work is more than anything a testimony to your unconditional love and support. Thank you immensely for putting up with me these past two years. It is something I will never be able to repay, but also something I shall never forget.

Affectionately yours.

“The more you know, the more you know you don't know.”

- Aristotle

Abstract

Hybrid speciation occurs when two diverged taxa hybridize and produce viable offspring that becomes reproductively isolated from their parent species. For non-hybrid species, most divergence and speciation events include some aspect of spatial isolation. How spatial isolation affects the divergence in hybrid species is, however, not well understood. Here I set out to investigate how isolation affects morphological divergence in the homoploid hybrid species *Passer italiae*. To do this I use isolated island populations and compare these to mainland populations which are likely exposed to clinal gene flow. First, I test if there is any difference in the extent of divergence between populations from island and mainland populations. Second, I use a P_{ST}-F_{ST} approach to investigate if divergence is likely to result from selection or reflect genetic divergence. Third, I use a model selection approach to find the ecological variables best explaining variation. Lastly, I compare variation within the Italian sparrow to that of the parent species to investigate if the phenotypic traits are transgressive, intermediate or mosaic and whether this differs in isolation. I use beak morphology, which is mainly subjected to natural selection and sexually selected male plumage traits as study traits. I found that both insular and mainland populations readily diverge, for many traits to a larger extent than expected based on genetic divergence. Island populations tend to be more diverged in beak shape and size than mainland populations, and the best models support a role for ecological factors such as $\delta^{13}C$ and precipitation regimes in driving this divergence. Italian sparrows were largely transgressive for beak size and intermediate for beak shape, and plumage colour was a mosaic with instances of transgression in back colour, cheek colour, and rump colour. Furthermore, island populations were transgressive across more traits than mainland. My data shows that hybrid populations resulting from a cross of the same two parent species can diverge phenotypically, and that this divergence can be adaptive. This adds to the growing body of evidence suggesting that hybridization is a non-negligible evolutionary process.

Preface

This thesis was written at the Centre for Ecological and Evolutionary Synthesis (CEES) at the Department of Biosciences, University of Oslo, under the supervision of Professor Glenn-Peter Sætre, Dr. Fabrice Eroukhmanoff and Dr. Anna Runemark.

Glenn, my great thanks for accepting me into the Sparrow group and giving me this great opportunity. You have been an inspiration and always taken my thoughts, questions and concerns seriously, no matter how mundane they surely must have felt like sometimes. Anna and Fabrice, I've truly no words to describe my appreciation of the way you have tutored me through this process. You have received me whenever I've had the need for guidance, and answered my sometimes tedious e-mails at all hours of the day and with great patience and understanding. You have followed me up actively and always made sure I never felt unwelcome. Fabrice, thank you so much for our great talks at your office – you provided me with the necessary perspectives to grapple these questions, and you helped me out when I was too dug down with details and made me appreciate the biological meaning of what I was dealing with. Anna, where to begin. Your patience and understanding it truly unique, even though I'm sure I've stretched those limits repeatedly. Your great knowledge of everything from programing to complex biological questions is awe-inspiring. I have learned so much from you these past two years and I am forever grateful for always taking the time to help and support me. Thank you so, so much!

And not least, to the rest of the group: thank you for your kindness and helpful support. I feel truly privileged to have been part of this great team of skillful scientists and researchers for the past two years. The amount of collective knowledge and experience this group has is mindboggling, and I stand in humble appreciation of the fact my modest contribution is barely scratching the surface of this exciting and innovative field of study.

Table of content

1 Introduction

Ever since Darwin, the importance of phenotypic variation in adaptive evolution has been widely recognized (Darwin, 1859; Schluter, 2000). How this variation within and between species arise lies at the heart of our understanding of why life here on earth is not just a coherent mass of biological entities, but diversify into specific groups that resemble each other. The formation of new species, speciation, typically proceeds in a *bifurcating* manner, where two different forms from one common ancestor ensue as a result of divergent selection or genetic drift (Coyne & Orr, 2004). The variants that are sorted by selection or drift originate from mutation. The data made available by the revolution of genome sequencing techniques has, however, revealed that the level of gene exchange in natural systems are higher than previously appreciated (Abbott et al., 2013; Mallet, 2005). Hybridization may thus be an important source of variation that could spur evolutionary change. Variation derived from hybridization differs from that originating from mutation as it allows variants already tested by selection to introgress, and can allow for the transfer of entire co-adapted gene complexes rather than just single base modifications or other mutational changes (Abbott et al., 2013). In addition, mutations are rare and often neutral (Kimura, 1984). Hence, the evolutionary potential of variation derived from hybridization is expected to differ from that of mutation accumulation.

Hybridization is believed to occur relative frequently in both plants and animals (Mallet, 2005). Previously considered a “dead-end” in evolution, we now know many evolutionary radiations where hybridization has promoted diversification, including the Galapagos finches (Grant et al., 2004) and the Rift lake cichlids (Wagner et al., 2013). Despite increasing recognition of hybridization as an important source of novel variation, we know very little of the evolutionary consequences of hybridization (Schumer et al., 2014). Most studies have only documented cases of genetic admixture, and we still lack an understanding of the processes involved (Losos et al., 2013). Hence, studies of the benefits and limitations for hybridization as a source of novel variation in phenotypic evolution are needed. Are hybrids from the same parental species restricted to a limited number of viable phenotypes, or can several variants readily persist? Does the standing genetic variation in hybrid genome make them able to rapidly achieve local adaptation?

Consequences of hybridization

Origination from a gradient or mixture of parental and viable hybrid individuals in sympatry, introgressive hybridization could result in both break-down of taxa to *hybrid swarms* (Bittner, 2010) and adaptive introgression of small parts of the genome (Macholán et al., 2007; Mavárez et al., 2006). Introgression (i.e. gene flow) between two divergent taxa through a hybrid population could spur novel gene combinations and thus open a potential for adaptation in the parent species (Anderson & Stebbins, 1954). Hybrid swarms have various outcomes depending on the fitness of the hybrid, and could result in hybrid speciation (Hermansen et al., 2011; Rieseberg et al., 1996; Rieseberg et al., 2003; Trier et al., 2014). Hybrid speciation requires the hybrid to evolve reproductive barriers preventing assimilation back into either parent. There are two main modes of hybrid speciation. *Polyploid* hybrid speciation occurs when two species mate to produce offspring with a number of chromosome sets different from their parents, and is more common in plants as extra chromosomes is putatively known to be more detrimental in animal fetal development. Conversely, *homoploid* hybrid speciation results in a hybrid without a change in chromosome number (Arnold, 1996). Without a change in ploidy, it is harder to establish the reproductive isolation needed for consolidating the hybrid as a distinct species. This mode of speciation has been considered rare, as hybrids often are ecological intermediates and fall between signaling preferences of either parental, thus having reduced fitness (Christophe & Baudoin, 1998). An additional issue is that hybrid species must escape the mass of unfit recombinants such as those arising from *Dobzhansky-Muller incompatibilities* when alleles from different parent species do not work in combination with each other and reduce fitness in the hybrids (Barton, 2001). However, homoploid speciation could occur under the right circumstances and seems to have the greatest potential for success under consequent genomic rearrangements, ecological divergence, and spatial isolation (Gross and Rieseberg, 2005).

The genetic admixture by hybridization can result in distinct phenotypic scenarios in a hybrid species. At the phenotypic level, one possibility is *mosaicism*, i.e. the hybrid will inherit a substantial contribution from each parental species and express either across various traits in different parts of its phenotype (Kunte et al., 2011). If both parent species have traits with a history of directional selection, the hybrid will likely show an intermediate phenotype in this trait (Paterson et al., 1991). An intermediate phenotype is a blend of the parent phenotype, instead of distinctly expressing either. A third possibility arises when the hybrid expresses an

extreme phenotype which extends outside the range of either parental (Bell & Travis, 2005). This *transgressive segregation* could result from epistasis between different interacting loci each having an effect on the same phenotype, provided these alleles are additive (Rieseberg et al., 1999a), and is often the result of complementary gene action (Vega & Frey, 1980). In addition, the exposure of recessive parental alleles in the hybrid has been linked to an increased chance of transgressive phenotypes (Rick & Smith, 1953). Transgressive segregation is positively related to the genetic distance between species, suggesting that more transgressive phenotypic variation will result from hybridization between diverged taxa (Stelkens & Seehausen, 2009). These genetic effects could lead to atypical phenotypes and increase the potential for divergence in the hybrid by new ecological opportunity (Rieseberg et al., 2003). In terms of the adaptive landscape analogy (Svensson & Calsbeek, 2012), this could allow for new fitness peaks, unavailable to both parents, to be reached by the hybrid, without waiting for a new mutational path to arise. Alternatively, the hybrid phenotype could be able to outcompete both parent species, leading to a total breakdown of the two taxa into a new species (Schierenbeck & Ellstrand, 2009).

Factors promoting population divergence

If a new hybrid species persists, it could be subject to intraspecific divergence as it spreads. Although behavioral isolation (Selz, 2014) can facilitate divergence, results from population genetics reveal that as little as one immigrant in a hundred (i.e. 1%) can be enough to prevent this from effectively occurring (Tufto, 2001). Maladaptive gene flow may hence prevent diversification and radiation. Moreover, most speciation events are associated with spatial divergence (Coyne & Orr 2004). Thus, spatial isolation can increase chances for differentiation within a species. Whether spatial isolation is as important for population divergence within a hybrid is not well known or understood. For instance, effects from hitchhiking of genetic incompatibilities in hybrid populations could fuel ecological population divergence in a non-classical way (Seehausen, 2013). One way to investigate the hypothesis that hybrid species divergence is spurred by isolation is to use spatially isolated populations, for instance, from different island habitats, and compare them to conspecific populations experiencing gene flow. Islands are notorious for harboring rapid divergence and evolution (e.g. Grant & Grant, 2008; Losos, 2009; Schluter, 2000). More specifically, stochasticity has a larger impact because of smaller population sizes, genetic isolation decreases maladaptive

gene flow, and potentially unique island habitats offer different ecological opportunity (c.f, Lomolino, 1985; Meiri & Dayan, 2003).

The role of selection and drift

The relative importance of genetic drift and selection in evolution is a matter of ongoing debate (c.f. Nei, 2013). However, both selection and genetic drift can change the frequency of variants in a population, potentially causing them to diverge (Nielsen & Slatkin, 2013). Drift is a random process that has a stronger effect in small populations where low frequency variants by chance could arise to high frequency (Nielsen & Slatkin, 2013). Subsequently, potentially favorable variants could be more easily lost from small populations due to stochasticity as there are fewer copies. Conversely, selection is expected to be more efficient in larger populations as the likelihood of favorable variation to arise is greater. These variants can then quickly be carried to high frequency through selective sweeps (Smith & Haigh, 1974). Identifying selection in natural populations can be achieved by investigating the discordance between phenotypic divergence from the expectations originating from genetic differentiation (Brommer, 2011). How this classical population genetics scenario plays out within a hybrid species is not well understood.

Can hybridization favor local adaptation?

If selection differs across a species' range, this could result in divergence due to local adaptation (c.f. Kawecki & Ebert, 2004). Contrasting selection pressures induced by abiotic and biotic ecological factors in their habitat may cause populations to diverge. Relative to their parent species, hybrids have increased additive genetic variation, which can predispose for rapid divergence and novel adaptations (Rieseberg, 1997). Local adaptation is expected to be higher for traits which are directly associated with survival, such as those involved in feeding and intrasexual - and interspecific competition (Williams, 1966). Conversely, traits which are sexually dimorphic are expected to mainly be subjected to sexual selection (Andersson, 1994). However, genetic and developmental constraints can put intrinsic limits on how much variation there is available for selection to act upon, which may be especially relevant to hybrids (Eroukhmanoff et al. in revision). In classical species, physical linkage and

epistatic fitness effects between loci contributing to phenotypic variation lead to genetic correlations, and hence constrain the independent evolution of traits (Chenoweth et al., 2010). If the hybrid inherits such genomic dependencies (i.e. contingencies), its trajectory may already be fixed on that of the parent species, which could delay or prevent divergence of populations. Little is known about how selection will shape divergence between hybrid populations, as most studies have so far focused on how hybrid species are well-adapted in comparison to their parents (but see e.g. Eroukhmanoff et al., 2013)

Here we set out to test various effects of isolation within a hybrid species using the homoploid hybrid Italian sparrow (*Passer italiae*) as a model system. Several studies have documented it as a distinct biological species with a hybrid origin from a cross between the house sparrow (*Passer domesticus*) and Spanish sparrow (*Passer hispaniolensis*) (Elgvin et al., 2011; Hermansen et al., 2011; Hermansen et al., 2014; Trier et al., 2014). Furthermore, it is one of two systems fulfilling all criteria for a being a homoploid hybrid species (c.f. Schumer, 2014). Although there is a lack of conclusive evidence, some have suggested that this hybridization could have occurred as late as 3600 years ago (Johnston, 1969). The Italian sparrow is mosaic in plumage traits, and sexually dimorphic with secondary sexual characteristics for males being more Spanish like by a chestnut crown. Previous studies have shown that the Italian sparrow can readily evolve phenotypic diversity across populations in both beak morphology and plumage colour, with respect to its parent species, and interspecifically across different island habitats (Eroukhmanoff et al., 2013; Piñeiro, 2014; Backe-Mathisen, 2014)

The Italian sparrow is thus an excellent model system to address these questions. It is a documented hybrid species, there is spatial variation in genetic composition and morphology, and it is distributed both on the mainland and geographically widely separated islands allowing for studies its patterns of dispersal and phenotypic diversity (Fig. 1). First, I assess if the inter-population variation between islands is different from the variation between mainland locations. Due to the reduced effects of gene flow, islands are expected to be more diverse. Second, I investigate if there is are different evolutionary processes driving divergence in island and mainland populations. Third, I investigate whether and which ecological factors that cause the observed divergence, and whether sexually and naturally selected traits show the same patterns of divergence. Lastly, I assess if there is any convention in parental resemblance between islands and mainland. I use beak morphology and plumage

colour as study traits as these are important for survival (Mallarino et al., 2010) and under sexual selection (Andersson, 1982) in passerine birds (Passeriformes).

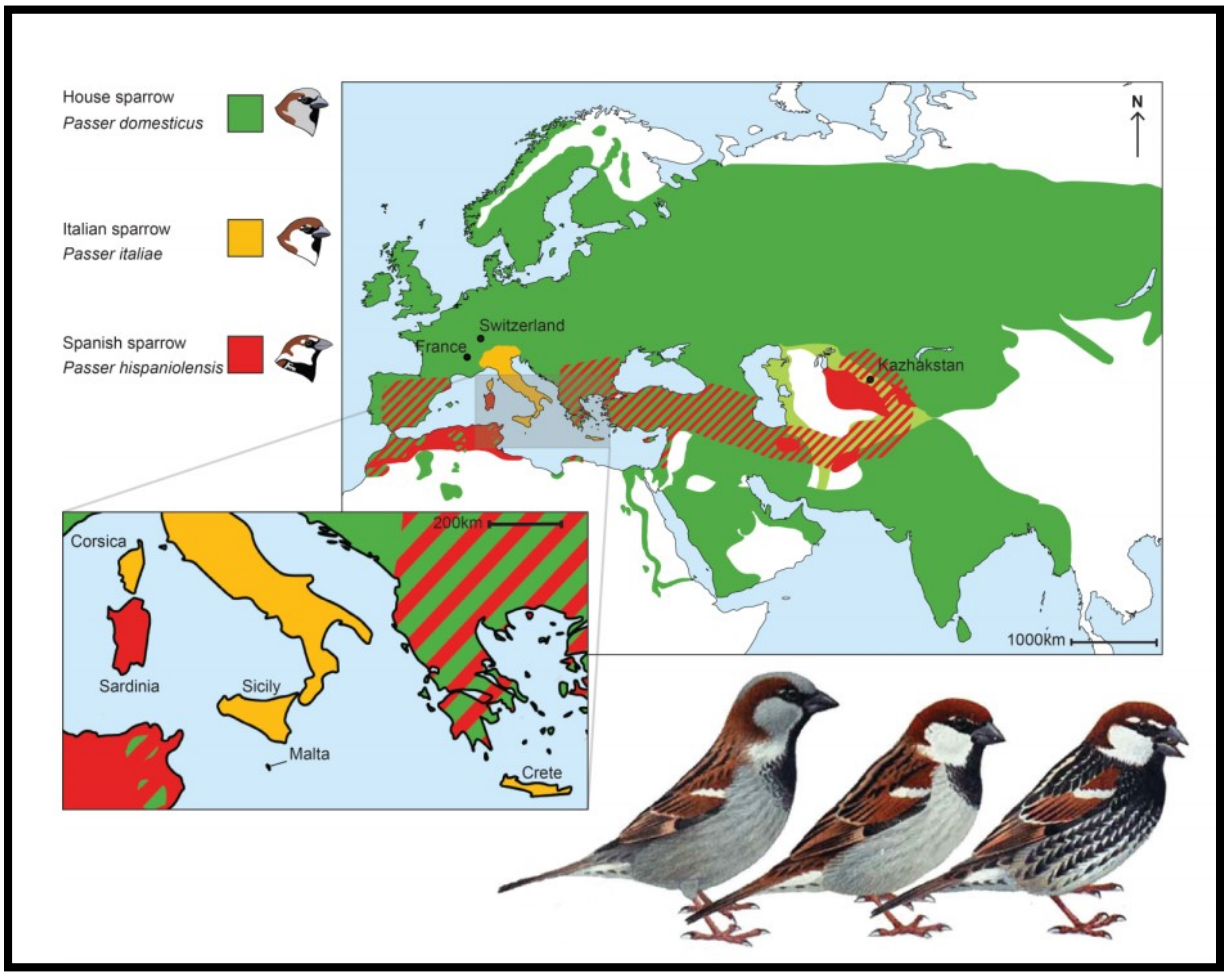

Figure 1. The distribution of species in the *Passer* hybrid complex. The inset map shows the Mediterranean region and the distribution of the Italian sparrow across mainland Italy, as well as the islands Corsica, Malta, Sicily and Crete. The large map shows the Eurasia region where the House sparrow and its subspecies *P.d.bactrianus* is largely dominating, with some regions of Spanish sparrow in sympatry (green/red). The three birds in the bottom-right picture is house sparrow (left), Italian sparrow (middle), and Spanish sparrow (right). The Italian sparrow is a mosaic with a chestnut coloured crown and brighter cheek from the Spanish sparrow, and is overall more house-like for the rest of the plumage.

2 Materials and methods

2.1 Data sampling

2.1.1 Field work

The Italian sparrow is dispersed throughout mainland Italy where it occurs both in allopatry and sympatry with its parental species, as well as on several Mediterranean islands (Fig.1). Like many of its closest relatives, it's a human commensal that feed mostly on seeds and insects. It is likewise also non-migratory although it wanders to some extent, especially during the breeding season (Gustin & Sorace, 2002).

We caught allopatric Italian sparrows from the Mediterranean islands of Corsica, Crete and Sicily in 2013 and three localities in mainland Italy (Crotone, Guglionesi and Rimini) in 2015 (Fig. 2; Appendix II, Table S1). We also caught house sparrows from Sales in France in June 2014, as well as Spanish sparrows near Manfredonia in Italy in 2011 (Fig. 2; Table S1). The locales ranged from farmsteads to rural communities, and to semi-populated areas (i.e. small villages, camping areas, and the like). All birds were caught using mist nets and released immediately after data had been collected. We extracted about 25 µl blood by venipuncture of a brachial vein for use in DNA-analysis. We photographed the birds' plumage colouration and beak morphology using a standardized setup described in detail by Tesaker (2014) (Nikon *D500*, 16.2 megapixels). Some specimen was categorized as *bushy* due to erect feathers on the head, which could change the colour composition as the skin can show through the feathers then. We gathered feather samples from each individual for stable isotope analysis ($\delta^{13}C$ and $\delta^{15}N$). We obtained all the necessary permissions from the appropriate authorities in order to carry out the work outlined in this thesis.

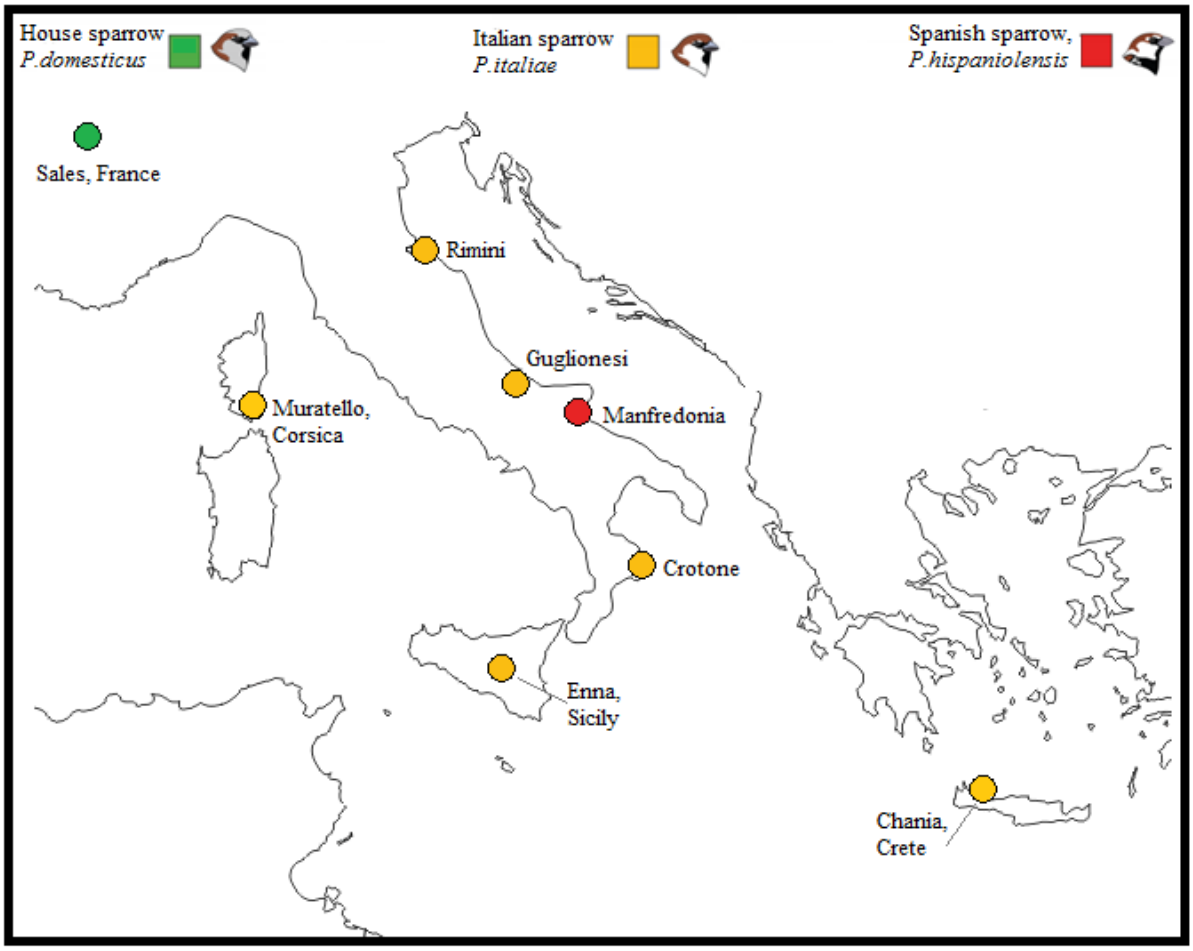

Figure 2. Sampling locations for *Passer* sparrows in the Mediterranean region. Approximate coordinates: Chania (35.5°N, 24.0°E), Crotone (39.1°N, 17.1°E), Enna (37.6°N, 14.3°E), Guglionesi (41.9°N, 14.9°E), Manfredonia (41.6°N, 15.9°E), Muratello (41.6°N, 9.2°E), Rimini (44.6°N, 12.5°E) and Sales (45.9°N, 6.0°E).

2.1.2 Digitization of beak morphology

As beak morphology has been found to be phenotypically diverse in the Italian sparrow (Eroukhmanoff et al., 2013; Piñeiro, 2015), it provides a sound trait for a study into the degree of population divergence between island and mainland habitats. I gathered beak data using landmark-based methods (Zelditch et al., 2012). This is advantageous compared to more traditional techniques as the use of homologous morphological loci decrease bias, improves consistency across specimen, and makes size measurements less ambiguous. Moreover, it is easier to correct for variation not related to size, such as isometry (Zelditch et al., 2012). I digitized the beaks using five homologous landmarks and seven equidistant semi-landmarks along curves outlying the upper and lower mandible respectively (Fig. S1). The semi-landmarks are used for quantifying morphology in areas where it is difficult to specify any

clear homologous points. I used TPS-based (*thin-plate spline*) programs developed by Rohlf (1998) available at http://life.bio.sunysb.edu/morph/. *tpsUTIL* was used for utility and file conversions, *tpsDIG2* for digitization of landmarks, and *tpsRELW* for extraction of partial and relative warps, and centroid size. *TpsRELW* uses a generalized least-squared procrustes superimposition to remove scale and orientate images and produces a consensus configuration as a reference to visualize shape differences between geometric objects, and to calculate centroid sizes and relative warps. The latter are the principal component vectors of the W-matrix, a matrix of the partial warp scores, and thus the approximation of form. The minimum bending energy sliding method was selected because of its ability to prevent the displacement of semi-landmarks (Gunz & Mitteroecker, 2013). As recommended for studies without any strong a priori opinion about the relative importance of individual landmarks, I set the *alpha-level* to zero in order to weigh all landmarks equally (Rohlf, 1998).

I used the residuals from a regression of relative warps onto centroid size, which represents the shape-components corrected for allometric effects of size. I extracted relative warp scores (n = 28) and centroid sizes for 175 Italian sparrows with at least 25 individuals from each population, and 50 individuals of the parent species house and Spanish sparrow respectively (Table S2). I selected an approximately equal number of males and females in all populations. I used the *broken-stick* distribution criteria (Jackson, 1993) to select relative warps (n = 5) explaining roughly 90% of the variation, based on the eigenvalues (Table S2). To ensure sampling consistency I estimated the repeatability of the digitalization for thirty specimens and only started digitalizing the data set when *R-square* (R^2) for two data sets of these specimens was above 0.9. To obtain a measure of overall shape distance, I extracted pairwise Euclidean distances between all individuals based on all relative warps (n = 28). Euclidean distances are the shortest way ("as the crow flies") between two data points and are calculated as square roots of the sum of the squares of the differences between corresponding values. I included a resampling procedure to alleviate the problem of pseudo-replication due to the pair-wise character of the estimates using the R-package *resample* (Hesterberg, 2015).

2.1.3 Colour quantification

I digitized the colour on the back, cheek, crown and rump plumage of each individual as these traits are likely to be important for species recognition (Bailey et al., 2015). Moreover, these traits may be subjected to sexual selection as they are sexually dimorphic (Summers-Smith,

1963). The Italian sparrow has a fairly uniform colored crown and cheek in brown and white respectively. However, as the back and rump have patchy coloration, a method that measures the colour of the entire trait rather than a point estimate, and that takes spatial variance into account, was needed. Spectrometers typically sample a point, and estimates average wavelength of the reflected light of this specific patch. Alternatively, there are methods that measure the amount of red, green and blue (RGB) for each pixel of an image. We used the Chromatic Spatial Variance Toolbox that implements an algorithm from Brydegaard et al., (2012), available at http://www.models.life.ku.dk/ChromatricSpatialVarianceToolbox which takes the spatial correlation between R, G and B in the pixels into account to incorporate information on patchiness. This software is implemented in *MATLAB* R2014a (version 8.2.0.701; http://se.mathworks.com/products/matlab/). It also uses an X-rite colour checker to standardize each image, which corrects for differences in illumination between the photographs. The method includes a *singular value decomposition* (SVD), which produce principal component scores for each individual. These can take the form of eigenvectors, eigenfields, or eigenplanes, depending on the number of dimensions one wishes to include in the decomposition of colour variation. Following Bache-Mathiesen (2014), I used a two-dimensional distribution script (i.e. RGB *without* measuring reflection) for all plumage traits except cheek colour (Appendix I, Fig. S2a; Fig. S2c; Fig.S2d). For cheek colour, where the variation lies in how light the cheek is, the three dimensional decomposition (i.e. RGB *with* a measure of reflection) was more successful in capturing variation (Fig. S2b). Regardless of dimensional decomposition, these eigenvalues will in unison be referred to as principal components (PC). A total of 89 principal components were calculated for each plumage trait, and I again used the broken-stick distribution to select which principal components to include in any further analysis (Table S2).

2.1.4 Genetics

We extracted DNA from the blood samples using a Qiagen DNeasy Blood and Tissue Kits (*Qiagen corp*., Valencia, CA) according to the manufacturer's protocol. DNA was then sent to McGill (McGill University and Génome Québec Innovation Centre, Montréal (Québec), Canada) where they performed whole genome re-sequencing to 8-12 times coverage of 10 individuals from each of the study populations (with the exception of Corsica for which 21 individuals were re-sequenced). We aligned the raw reads obtained to the house sparrow reference genome (Trier et al., in prep.). We used the Broad Institute Best Practices workflow

to call single nucleotide polymorphisms (SNPs) (DePristo et al., 2011; Van der Auwera et al., 2013). First, we aligned the reads to the reference genome using *bwaMem* (Li et al., 2010), then we sorted the file, marked and removed duplicates and realigned around indels with *samtools* (Li et al., 2009). We then called variants with *haplotypecaller* (DePristo et al., 2011) and subsequently merged the 'gvcfs' (i.e. the file format) and jointly genotyped all individuals using *gatk3.2.2* (McKenna et al., 2010). We first filtered according to the strict filtering recommendations from gatk, and then further used *vcftools* (Danecel et al., 2011) to remove inserts and deletions (indels), as well as all SNPs with a read depth lower than 3 or higher than 65, and SNPs with genotype quality lower than 20. I used vcftools to calculate the mean weighted F_{ST} (Weir and Cockerham, 1984) for all sites on 29 autosomal chromosomes and the Z-chromosome between all population pairs (n = 6). I calculated average F_{ST} per chromosome as the mean of all SNPs on that chromosome; the number of SNPs spanned from approximately 20 to several thousands. The average F_{ST}-estimates per chromosome were then used to investigate if phenotypic divergence departed from that expected based this genetic differentiation.

2.2 Population divergence in different biogeographic settings

To estimate the extent of variation between populations I first performed a canonical variates analysis (CVA) using the R-package *Morpho* (Schlager, 2016). A common challenge with CVA in morphometrics is that the number of variables is high relative to the sample size, thus restricting full-rank covariance matrix. For shape the sample size (n = 175) was large enough to include all 32 procrustes coordinates. Plumage samples (n = 89) were too small to avoid matrix singularities based on the raw data and subsequently the first 10 principal components were used for each plumage trait separately. I ran a jackknife cross-validation procedure with 1000 iterations to investigate if groups were significantly differentiated. This mitigates overestimation of group separation (Mitteroecker & Bookstein, 2011). I finally calculated 80% inclusion ellipsoids and assessed statistical significance of the CV-axes using Wilk's Lambda (λ).

While CVA is useful to visually asses the degree of separation between the populations, it is not as robust in evaluating which populations are statistically different from one another and hence significantly biologically diverged. I therefor followed up with an *omnibus* step-wise

analysis of variance (ANOVA) procedure using MANOVAs, ANOVAs, and finally a Tukey's honest significance test (Tukey, 1949). Only significant MANOVAs are followed by ANOVAs with adjusted *p*-values to the number of subsequent univariate components (i.e. principal components) within each multivariate trait (Dunn-Sidak approach: Dunnet, 1964). Combined, this approach minimizes the risk of inflation error (Type-I) and takes into account the combined effects of all trait components in explaining the entire phenotype. Only the principal components selected based on the broken-stick distribution were used for both beak shape and plumage colouration (Table S2). The chi-square (χ^2) and Pillai-Bartlett trace (V) test statistics were used for the univariate and multivariate analysis respectively, and a Student's *t*-test (t) was used for the post-hoc to assess statistical significance between each of the population means. I visually inspected residual plots to ensure there were no deviations from homoscedasticity or normality, and I checked various multivariate assumptions using several different test statistics (e.g. Box's - M and Levene's test). All analyses were done using the default commands in *R* (R-Team, 2010).

To determine if inter-population variation was significantly different between biogeographic settings, and to augment the analysis with a different methodology, I performed a variance decomposition analysis (VC) using MCMC posterior sampling implemented in the R package, *MCMCglmm* (Hadfield, 2010). I placed a joint probability distribution over the parameters (i.e. populations) to get the posterior modes of coefficients and their corresponding confidence intervals. The size of these coefficients are an expression of how much of the variation is explained by grouping individuals into population categories compared to the variation between individuals (i.e. the residuals from the analysis). I performed the analysis on each multivariate trait using the PCs selected based on the broken stick criteria.

When the number of groups are low (i.e. < 6) the posterior distribution of the variance becomes increasingly *tail-heavy* rendering the results unreliable as the mixing of the MCMC-chain is poor and gets stuck close to zero. As I only had three groups I needed to consider how this could affect the results. Fitting a stronger prior can be helpful although you run the risk of placing to much bias in the analysis. To mitigate this, I used a technique called *parameter expansion* (Hadfield, 2010) to the MCMCglmm algorithm to speed up the rate of convergence in the MCMC-chain. It uses information from a run with an uninformative prior on the same data to choose proper values for the prior means and prior covariance matrix

(*alpha* mean and variance) to be specified in the parameter expanded run. I then used the *Cauchy* probability distribution as a prior, as recommended for the parameter expanded run (Hadfield, 2010), with the alpha variance set to the square of the standard deviation (i.e. the variance) in the posterior distribution from the uninformative prior. The posterior sampling was run for 200 000 iterations with a *burn-in* of 40 000 and a *thinning* of 100. The MCMC-chain was plotted and inspected for proper mixing, and autocorrelation remained low (< 0.1) between successive samples in the chain.

2.3 The roles of drift and selection

2.3.1 The relationship between genetic structure and phenotypic divergence

To investigate the potential role of selection and genetic drift as causes of population differentiation in beak and plumage colouration I performed $P_{ST} - F_{ST}$ comparisons. This is analogous to the more traditional $Q_{ST} - F_{ST}$ analysis but uses phenotype variation instead of additive genetic variance when a pedigree is not possible to obtain (Brommer, 2011). As with Q_{ST}, P_{ST} measures the (phenotypic) variation among populations relative to the total variation. If P_{ST} exceeds the F_{ST} distribution, this indicates that *divergent selection* is acting on the trait, if it is lower than F_{ST} it indicates *stabilizing selection*, and if it approximately equals P_{ST} *genetic drift* cannot be ruled out as an explanation to the divergence. I used the average F_{ST} score across all chromosomes from an F_{ST} distribution consisting of one F_{ST} estimate per chromosome as the proxy for genetic divergence. The P_{ST} equation also includes a component (h^2) representing phenotypic variation as a proportion of additive genetic effects within populations and a component (c) representing variation as a proportion of additive genetic effects between populations.

I did not have data to accurately calculate c and h^2, and the $\frac{c}{h^2}$ ratio strongly influence the phenotypic divergence estimate (Brommer, 2011). Subsequently, I used a conservative approach with low value for $\frac{c}{h^2}$ (< .4) for instances where traits appeared to be under divergent selection and high $\frac{c}{h^2}$ value (> 1.7) where traits appeared to be under stabilizing selection or canalization. I calculated P_{ST}-scores between all possible populations combinations between island and mainland populations, respectively (n = 6), using procrustes

coordinates for shape (n = 32), centroid size for beak size, and broken-stick PCs for plumage (n = 3 or 4, depending on the trait; Table S2). The sums of squares (SS) from an ANOVA using population pairs as independent variables is equivalent to the within population component, and the SS of the residuals as the between population component. To obtain a .95 confidence interval I added a bootstrapping loop of 10 000 iterations in the sums of squares calculation using the adjusted bootstrap percentile (BCa) method through the *R*-package *boot* (Canty & Ripley, 2015).

2.3.2 Factors explaining morphological variation

As climate and diet are known to affect beak shape (Grant et al., 2004), and as plumage is under strong sexual selection (Bailey et al., 2015; Shultz & Burns, 2013), I wanted to examine to what extent these factors where important in explaining phenotypic variation for these different traits. Hence, I estimated the effects of available proxies for diet, stable isotope ratio for carbon ($\delta^{13}C$) and nitrogen ($\delta^{15}N$) respectively, and the climate variables mean annual temperature (AT), mean annual precipitation (AP), temperature seasonality (ST) and precipitation seasonality (SP). I used these measures as independent variables in ordinary least square (OLS) linear mixed effects regression models using the *lme4* package in R (Bates, 2015). Possible interactions between fixed effects were also considered with the exception of between climate variables. I tested if including the interaction terms significantly improved the model by applying an ANOVA to a reduced (i.e. without the interaction) and a full model and only included the interaction in the final model if that model was significantly better ($\alpha = .05$). I included sex (for beak only as this is the only trait for which both sexes were sampled) as a categorical fixed effect while population and bushy (crown plumage only) were modelled as random factors. Population was also added as a sole fixed effect for general population structuring. Following Eroukhmanoff (2013), climate data was collected from the *Worldclim* database (Hijmans et al., 2005). The data was extracted using the *R* packages *raster* (Hijmans, 2014), *rgdal* (Bivand et al., 2014) and *foreach* (Weston, 2014).

I used both a maximum-likelihood (ML) and a Bayesian model selection (BMS) scheme to find the most plausible hypotheses explaining my data. The ML approach started with a *step-wise* regression of the maximized model (i.e. all factors and significant interactions included) to reduce it in a step-wise manner by removing interactions and or the main effects (i.e. the main independent variables) one at a time. Each step was ranked using the second-order

Akaike's information criteria (AIC_C) to penalize over parameterization more severely. Models with a deviance in AIC_C by less than two from the best-fitted model was regarded as not statistically worse in explain the variation. If more models fell within this interval, I selected the most parsimonious model as the best model. I used the maximum-likelihood (ML) option as opposed to restricted likelihood (REML) in *lme4*. This reduces the estimating power of the random effects but is necessary when comparing models that include different fixed effects. The final model was then re-fitted with REML to model the random factors more accurately.

To overcome caveats with a step-wise regression (c.f. Whittingham et al., 2006), I also performed a Bayesian model selection (BMS). Using an independent and different methodology evaluating the same data enables stronger inference. Bayesian statistics is conceptually different from maximizing likelihoods as it generates a probability distribution of models instead of estimating a single best model. The result is not one model, but many models with different *posterior model probabilities* (PMP). The Bayesian approach requires the researcher to specify a prior belief that represents the prior distribution of model parameters (i.e. similar to the MCMCglmm procedure described previously). As I wanted to penalize over-parameterization, I modelled with an informative prior biased towards a low number of factors. If the best model selected by the AIC_C criteria still appear with high a *posterior model probability* (PMP) even after a low-factor biased prior, one can be more confident that the step-wise procedure was sound. I also used Bayesian model selection to decide between more models in instances where several equally parsimonious models (i.e. the same number of independent variables) fell within 2 AIC_C of each other. I used the package *BMS* in *R* (Zeugner & Feldkircher, 2015). As the number of factors in the models were lower than 15, the entire model-space was evaluated as recommended by the authors (Zeugner & Feldkircher, 2015). This also eliminated the need for posterior distribution sampling methods (e.g. MCMC).

Linear mixed effect models (LMEMs) are analytically powerful but lack reliable *goodness-of-fit* criteria associated with the fixed effect-only models (e.g. the coefficient of determination, R^2). In addition, *p*-values are not included in the *lme4* package. As recommended by the authors of *lme4* (Bates, 2010) I instead used an ANOVA to compare the fitted ML-model against a model representing only the intercept and random factors as a null-hypotheses, with $\alpha = .05$. For a coefficient of determination equivalent to R^2, the Omega^2 – criteria (Ω^2) is reliable for LMEMs (Xu, 2003). Due to the risk of marginality in the presence of interactions

(i.e. erroneously interpreting the main effect or interactions) when calculating *p*-values for the model factors, I included a parametric bootstrap to obtain these *p*-values using the package *afex* (Singmann et al., 2016). I used an ANOVA 'Type-II' calculation of the sums of squares with a F-distribution (*F*) for models without interactions *and* random factors, and a Chi-square (χ^2) distribution for models without interactions but *with* random factors. Both ANOVAs was performed using the *car* package in R (Fox & Weisberg, 2011).

To complement null-hypotheses testing I included a measure of standardized *effect sizes* to infer the magnitude of a factor on the dependent variable by using Cohen's *d* (Rice & Harris, 2005) on all independent variables. This is similar to the *standardized regression coefficient*, beta (β). However, interpreting the comparison between binary categorical variables (i.e. sex) with numerical variables in terms of standard deviations is problematic. To overcome this, I used the 'standardize' function in the *arm* package (Gelman & Yu-Sung, 2015) which corrects numerical variables by two-times their standard deviation which is equivalent to going from one extreme to another. The impacts of random factors were tested using the package *RLRsim* using 1000 iterations (Scheipl, 2008). This checks if adding a random factor improves the model or not.

2.4 Hybridization and phenotypic diversity

Italian Sparrows has previously been found to show both mosaic (Elgvin et al., 2011) and transgressive phenotypes (Bache-Mathiesen, 2015) with respect to their parent species. To examine if there are any differences in how island and mainland populations combine trait values from the parent species and transgressive features, I added house and Spanish sparrow populations to the analyses. I used a combination of different discriminant functions (i.e. CVA for more than two groups and LDA for only two groups) to find the maximum axis of parental variation, and subsequently placed the Italian Sparrow populations in this ordinance space. This was formally achieved by performing a linear discriminant function analyses (Fischer, 1938) on the two parental populations to get the axis of maximal separation. I then added the first canonical variates (i.e. CV1) from a CVA using the six Italian sparrow populations and classified where on the parental axes the six Italian Sparrow populations fell. I used the first 10 principal components for each plumage trait separately, the 32 allometrically corrected procrustes coordinates for beak shape, and centroid sizes for beak size.

A second CVA using an island-mainland grouping combined with the two parental populations (i.e. 4 groups) was used to investigate the proportion of correct posterior classifications. A higher proportion of correctly classified specimens reflect the ability of the CVA to assign individuals to the posterior grouping, and is thus a reflection of overall divergence between groups. I again used the first 10 PCs for each plumage trait separately and the 32 allometrically corrected procrustes coordinates for beak shape. Being univariate, this procedure was inappropriate for beak size. I also used the Mahalanobis distances based on all significant CVs from this analysis to see how close each group where from each other while taking into account their group covariance, and performed a 1000 round permutation test to see if the different groups differed statistically. All analysis was performed using the package *morpho* in R (Schlager, 2016)

3 Results

3.1 Population divergence in different biogeographic settings

Islands tend to have more inter-population variation in beak morphology. Plumage colour show a more mixed pattern as there are examples of certain plumage traits being more diverged between island populations, as well as between mainland populations. This holds true both for the combined effects of all individual trait components (Table S3), and along the various axes of variation (i.e. principle components) (Table S4). The proportion of variation explained by grouping individuals by population of origin (Fig. 3), and the extent of pairwise differences in means between populations (Table S5), lends further support towards more extensive phenotypic divergence between island populations.

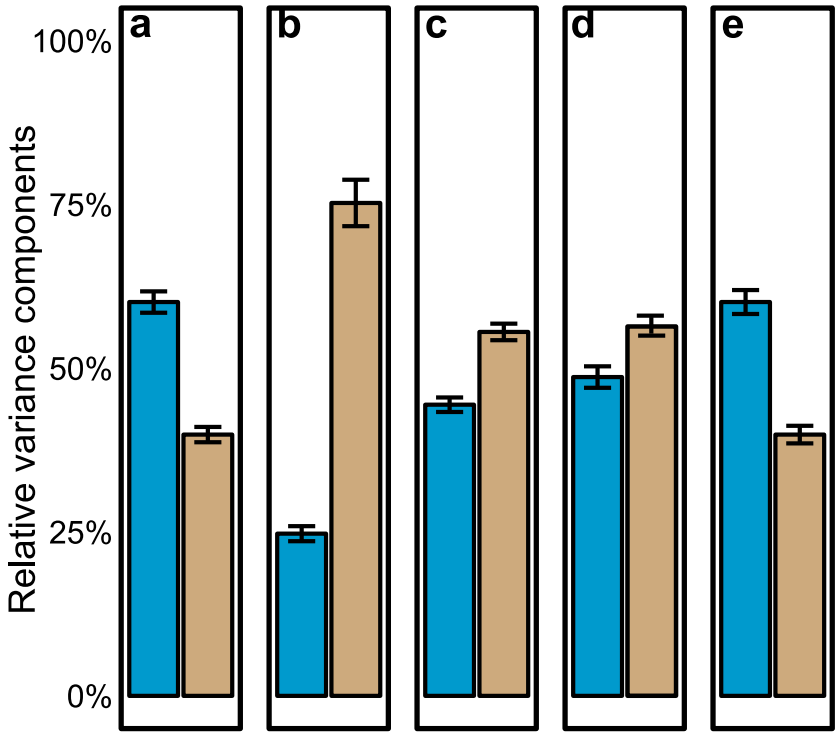

Figure 3. The relative sizes between island (blue bars) and mainland (brown bars) variance components for each trait, with corresponding .95 confidence intervals. Data derived from a variance decomposition using principle components selected by the broken stick criteria for each multivariate trait: (**a**) back colour, **(b**) cheek colour, (**c**) crown colour, (**d**) rump colour, and (**e**) beak shape. Shows the amount of variation that is explained by grouping individuals by population of origin. within either island or mainland.

Beak morphology

In general, island populations have *smaller beaks* compared to mainland populations ($F(1, 169) = 10.00$, $p = .002$). This is somewhat counteracted by a highly varying Crete population which has some of the largest beaks (Fig. S3). The two other island populations (Corsica and Sicily) are statistically similar (Table S5). The mainland populations are all statistically differentiated (Table S5) but the inter-populations variation relative to the within population variation is not as extensive compared to islands (islands: $F(2,87) = 41.69$; mainland: $F(2,78) = 31.98$; Table S4).

Island-mainland differences in population variation in *beak shape* are more ambiguous by inspection of the CVA plots (Fig. 4). All three ellipsoids for both islands and mainland groups are encompassing all the population means (cross). The difference in ellipsoid orientation is more conspicuous for the island group indicative of unequal patterns of variation. Overall classification accuracy is also relatively poor with slightly better posterior grouping of island populations (islands: 58%, mainland: 53%). The statistical test shows that population structuring within islands contribute more to the discriminant function (islands: $\lambda(2, N = 89) = .7$, $p = < .001$; mainland: $\lambda(2, N = 80) = .8$, $p = < .001$).

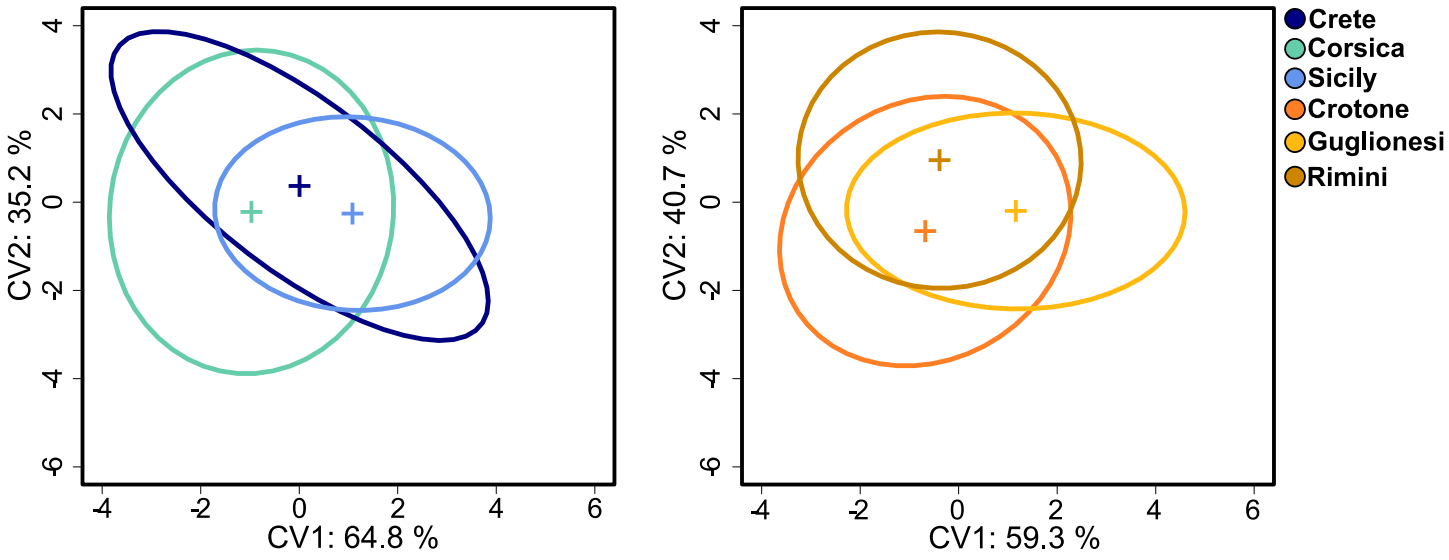

Figure 4. Population differentiation in beak shape for islands (left) and mainland (right). Data is calculated from a CVA based on 32 allometrically corrected procrustes coordinates. Ellipses show 80% inclusion quantiles and the crosses are group means.

Total trait variation based on the combined effects of all principal components (n = 5) is also more apparent for islands (islands: $V(2, 87) = .49$, $p < .001$; mainland: $V(2,78) = .41$, $p < .001$; Table S3). This finding is mirrored by the individual trait components (Table S4). Along the axis of maximum variation (PC1) there is only significant differences between island populations (F (2, 87) = 11.70, p = < .001; Table S4). This axis of maximum variation relates mostly to the thickness of the lower mandible and the position of landmark #4 (the junction between the par mandibularis and the malar region of the head: Fig. S1; Fig. S3). The second axis of variation, PC 2, is distinguishable within both groups but more prominently between island populations (island: F (2, 87) = 11.79 > mainland: F (2, 78) = 5.63; Table S4). This axis mostly mediates the thickness of the upper mandible, posteriorly (Fig. S3). Although mainland shows statistical significant divergence for PC5, islands are as a whole more differentiated along the axes of variation where the variation between individuals is largest.

Plumage coloration

Starting with *back coloration* for plumage trait, island population ellipsoids are less entangled compared to mainland, and ellipsoid orientation is more dissimilar (Fig. 5a). This is reflected in overall classification accuracy, where islands are more easily differentiated (78%) compared to mainland (68%). Total trait variation is significantly different for both islands and mainland, but more extensive for islands (island: $V(2, 40) = 1.04$, p = < .001, mainland: $V(2, 40) = .69$, p = < .001; Table S3). Islands are statistically differentiated for all axes of variation (PC1-3), and mainland for PC1 and PC2 (Table S4). The primary axis of variation in back colour reflects mainly brown hues, with some abundance of red hues as well (Fig. S4). The second axis of variation represents grey to bluish hues, and the third marks a transition towards yellowish hues (Fig. S4).

Cheek coloration has an obvious island-mainland difference in inter-population variation as island populations are virtually indistinguishable, while mainland ellipsoids show some separation (Fig. 5b). There are also large differences in classification accuracy (islands= 68%; mainland = 25%). These findings are mirrored by the variance analysis (islands: $V(2, 40) = .14$, p = .491, mainland: $V(2, 40) = .72$, p = < .001; Table S3). As island populations were not significantly different in the multivariate analysis, I only proceeded with univariate analysis for mainland populations. Mainland populations were differentiated for the primary and third

axis of variation (i.e. PC1 and PC3) (Table S4). The eigenfields of these axes represent reflection of grey and the transition between greyish towards darker reflection, respectively (Fig. S5).

Inter-population variation for *crown colour* appears less discernible between islands and mainland than cheek colour (Fig. 5c). The contrast in ellipsoid orientation indicate that populations at both islands and mainland have differences in which canonical component they vary for, but they are equally discernible (islands= 78%; mainland = 78%). Both geographical regions have significant overall trait differentiation, however the degree of between population variation is slightly larger for mainland (island: $V(2, 40) = .93$, $p = < .001$, mainland: $V(2, 40) = .94$, $p = < .001$; Table S3). In addition, mainland populations are more differentiated where most of the variation between individuals resides (i.e. PC 1), while islands are most varied in the second and fourth axis of variation (i.e. PC2 and 4; Table S4). The primary axis of variation in crown colour relates to grey with a transition towards bluish hues, while the second axis of variation reflects brown and some abundance in red (Fig. S6). The fourth axis marks a transition towards yellow hues (Fig. S6)

Rump coloration also falls into the pattern of subtle differences in variation, however there is more separation between island populations than mainland populations (islands = 58%; mainland= 51%; Fig. 5d). Overall trait differences are significant for both the island and mainland groups, but more pronounced between island populations (islands: $V(2, 40) = 0.75$, $p = < .001$, mainland: $V(2, 40) = .63$, $p = < .001$; Table S3). Islands as more variation in the second axis of variation (Table S4), which renders mostly towards greenish hues (Fig. S7). Mainland has more variation in the primary axis of variation (Table S4), which represents mostly grey hues (Fig. S7).

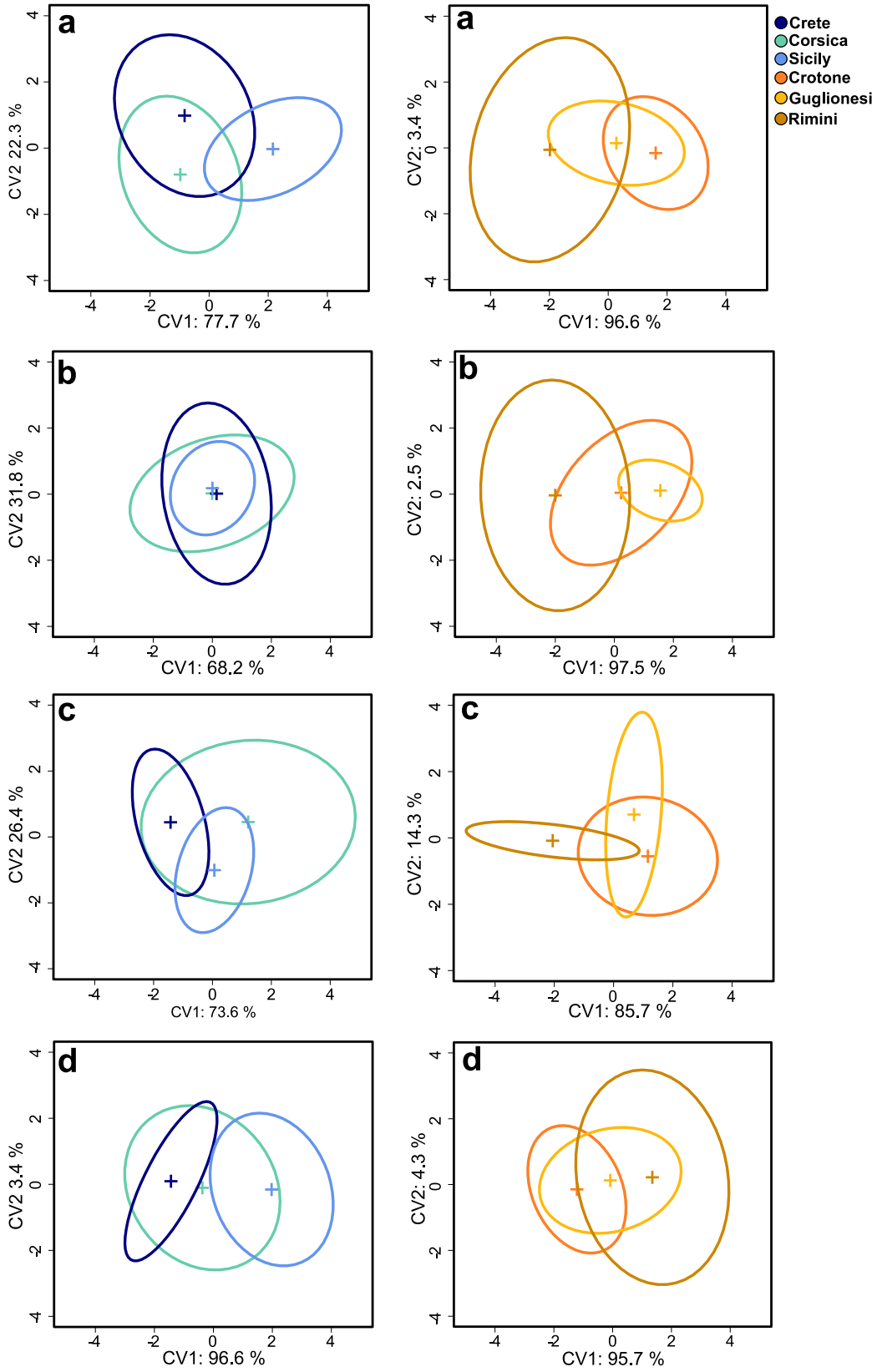

Figure 5. Population divergence for the island (left) and mainland (right) group for each plumage trait: (**a**) back colour, (**b**) cheek colour, (**c**) crown colour, (**d**) rump colour. Data is calculated from a CVA based on the first 10 principal components for each trait. Ellipses show 80% inclusion quantiles and the crosses are group means.

3.2 The roles of drift and selection

3.2.1 The relationship between genetic structure and phenotypic divergence

Genetic divergence is higher between island populations than mainland populations (Table S6). Hence, island populations have more frequent robust P_{ST} (e.g. using the conservative threshold – sensu: Brommer, 2011) estimates below the neutral expectation (i.e. stabilizing selection) (Table 1). There is little difference between island and mainland in the proportion of divergence explained by selection (i.e. $P_{ST} > F_{ST}$) (Table 1). Among traits, divergent selection most frequently occurs for beak size, followed in turn by plumage colour and beak shape (Table 1).

Table 1. Summary of the robust P_{ST}/F_{ST} analysis for all traits, derived from centroid sizes for beak size, the 32 allometrically corrected procrustes coordinates (n = 32) for beak shape, and the principle components selected by the broken stick distribution for plumage traits (n = 3 or 4, depending on the trait). The '*Divergent ~ Non-divergent*' main columns shows occurrences (n) of: statistically significant ~ insignificant cases of phenotypic divergence; robust $P_{ST} > F_{ST} \sim P_{ST} < F_{ST}$; robust $P_{ST} = F_{ST}$. The number in parentheses is the relative proportion of total occurrences in each P_{ST}/F_{ST} scenario[1].

		Divergent			~	**Non-divergent**		
Trait	**Group**	n	n($P_{ST} > F_{ST}$)	n($P_{ST} = F_{ST}$)	~	n	n($P_{ST} < F_{ST}$)	n($P_{ST} = F_{ST}$)
Beak-size	island†	2	2 (1.00)	-		1	-	1 (1.00)
	mainland	3	3 (1.00)	-		-	-	-
Beak-shape	island†	26	2 (.08)	24 (.92)		70	23 (.33)	47 (.67)
	mainland	13	4 (.31)	9 (.69)		83	6 (.07)	77 (.93)
Back-colour	island†	4	3 (.75)	1 (.25)		5	1 (.20)	4 (.80)
	mainland	3	2 (.67)	1 (.33)		6	0 (.00)	6 (1.00)
Cheek-colour	island	-	-	-		9	3 (.33)	6 (.67)
	mainland†	5	3 (.60)	2 (.40)		4	1 (.25)	3 (.74)
Crown-colour	island	2	2 (1.00)	-		7	1 (.14)	6 (.86)
	mainland†	5	2 (.40)	3 (.60)		4	-	4 (1.00)
Rump-colour	island†	2	2 (1.00)	0 (.00)		7	1 (.14)	6 (.86)
	mainland	5	5 (1.00)	-		4	-	4 (1.00)

† The group which was most phenotypically diverged of island or mainland

[1] See Fig. S8, Appendix I, for examples of how the ratio of $\frac{c}{h^2}$ affects the P_{ST} estimate and how the different outcomes (i.e. directional – and stabilizing selection vs. drift) is presented.

Beak morphology

Beak size is unique compared other traits as it has no instances of stabilizing selection (Table 1). Selection is equally important in causing divergence for both island and mainland (islands = 100%, mainland = 100%; Table 1). Phenotypic divergence between mainland populations are all above the upper F_{ST} confidence interval when $\frac{c}{h^2}$ = .2 (lower P_{ST} = .027, .079, .032 > upper F_{ST} = .009, .019, .012). $P_{ST} > F_{ST}$ is associated with Crete in islands, also when $\frac{c}{h^2}$ = .2 (lower P_{ST} = .085, .080 > upper F_{ST}= .082, .059). The Guglionesi-Rimini pair has a large P_{ST} confidence interval rendering it undifferentiated from neutral already at $\frac{c}{h^2}$ = .7 (upper P_{ST} = .037 < lower F_{ST} = .033).

In contrast to size, *beak shape* has the highest frequency of stabilizing selection (islands: 33.0%, mainland: 7.0%; Table 1). Crete and Corsica among the island populations and Crotone and Guglionesi among the mainland populations are involved in the combinations where stabilizing selection is frequent (Fig. 6). However, the majority of trait components have robust P_{ST}-estimates that are overlapping the genetic differentiation distribution, which make beak shape the trait that departs the least from the neutral expectation that morphological divergence follows genetic divergence (islands: 74.0%, mainland: 89.6%). Of the areas of the beak that are diverging, selection is more important for mainland (31%; Table 1) compared to islands (8%; Table 1). Islands have instances of divergent selection in both the *long-to-short* (i.e., X-axis; reflecting changes in beak length) direction and the *wide-to-narrow* direction (i.e., Y-axis; reflecting changes in beak width), while mainland only show divergent selection in the wide-to-narrow direction (Fig. 6).

Of notice is the landmark ‘Y4’, representing the wide-to-narrow position of the junction of the par mandibularis and the malar region of the head (Fig. S1; Fig. S3). It is the only landmark which is under directional selection ($P_{ST} > F_{ST}$) for both islands and mainland (CoSi: lower P_{ST} = .097 > upper F_{ST} = .033; CrGu: lower P_{ST} = .030 > upper F_{ST} = .009; GuRi: lower P_{ST} = .023 > upper F_{ST} =.013; Fig. 6). Interestingly, the landmark ‘X4’ (same position as ‘Y4’, but in the long-to-short direction: Fig. S1; Fig. S3) has robust stabilizing selection between Corsica and Crete (upper P_{ST} =.056 < lower F_{ST} =.059) and between Crotone and Guglionesi (upper P_{ST} = .001 < lower F_{ST} =.009). It appears that different directions of selection can act over a very confined area at the same time.

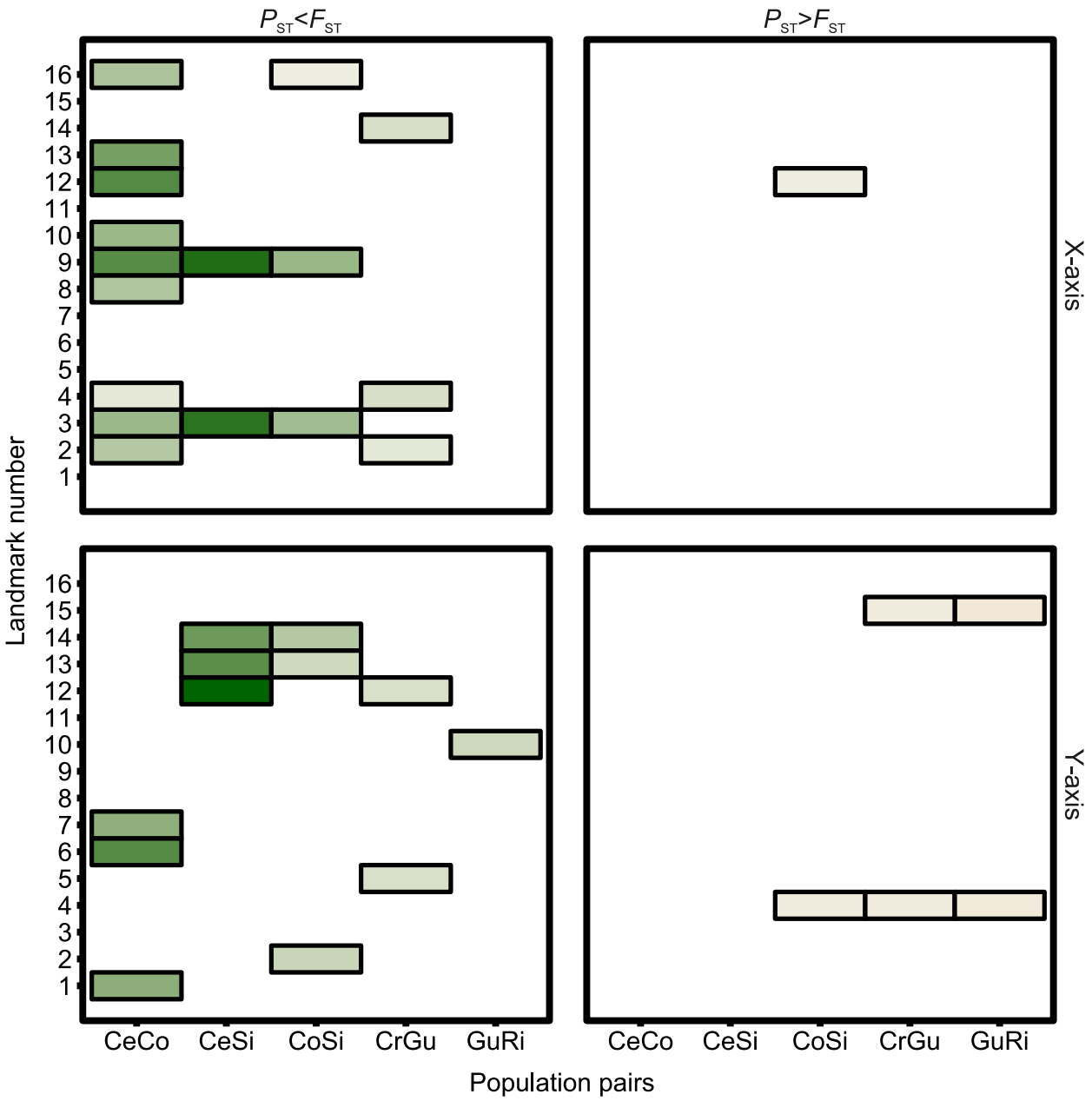

Figure 6. Summary of the P_{ST}/F_{ST} analysis on beak shape for each population pair between islands and mainland, respectively. Calculation of phenotypic divergence was performed using conservative estimates on each landmark (n = 32) for both the *long-to-short* direction (i.e. X-axis) and *wide-to-narrow* direction (Y-axis). Enclosed rectangles represent instances where $P_{ST} \neq P_{ST}$. They are colour coded by the extent of difference between phenotypic divergence and genetic differentiation (i.e. $\Delta P_{ST}/F_{ST}$): below F_{ST} (greener), above F_{ST} (redder). Missing squares represents instances of robust P_{ST}/F_{ST} overlap (note: Crotone-Rimini population pair is not depicted as it had only P_{ST}/F_{ST} overlaps). Population abbreviations: *Ce* = Crete, *Co* = Corsica, *Cr* = Crotone, *Gu* = Guglionesi, *Ri* = Rimini, *Si* = Sicily.

Plumage coloration

Island individual divergence in *back colour* is more frequently caused by selection compared to the frequency in mainland (island: 75%, mainland: 67%; Table 1). However, mainland individuals show more divergent selection along the axis of maximum variation (Fig. 7). Sicily is consistently under divergent selection, especially against Crete (CeSi) (PC1: lower P_{ST} = .119 > upper F_{ST} = .083; PC2: lower P_{ST} = .109 > upper F_{ST} = .083; Fig. 7). The same is true for Rimini among the mainland populations (GuRi: lower P_{ST} = .027 > upper F_{ST} = .013; CrRi: lower P_{ST} = .074 > upper F_{ST} = .019; Fig. 7). Only one instance of stabilizing selection is found, between the island populations on Crete and Corsica (CeCo) (PC3: upper P_{ST} = .019 < lower F_{ST} = .059; Fig. 7).

Island-mainland difference in *cheek colour* is more prominent than in any other plumage trait as islands experience stabilizing selection between all populations pairs (upper P_{ST} = .062, .002, .010 < lower F_{ST} = .082, .033, .059; Fig. 7). This accounts for 33% of undifferentiated traits between population at islands (Table 1). Mainland divergent selection is driven by divergent cheek color in Guglionesi (lower P_{ST} = .095, .070, .010 > upper F_{ST} = .012, .012, .009; Fig. 7). Interestingly, Guglionesi and Rimini (GuRi), which have most cheek color PCs with divergent selection, have one instance of stabilizing selection (Cheek PC2: upper P_{ST} = .002 < lower F_{ST} = .012; Fig. 7): selection is acting in different directions at different axes of variation for the same trait.

Divergence in *crown colour* is more often the result of divergent selection for island individuals (islands: 100%, mainland: 40%; Table 1). Among mainland individuals, divergent selection mainly along the axis of maximum variation (PC1) and is driven by divergent crown color in the Rimini population already at $\frac{c}{h^2}$ = .1 (GuRi: lower P_{ST} = .061 > upper F_{ST} = .012, CrRi: lower P_{ST} =. 027 > upper F_{ST} = .019; Fig. 7). Again, there is an instance where selection is acting in different directions on different PCs, this time between Crete and Sicily (CeSi). While there is stabilizing selection on PC1 (upper P_{ST} = .026 < lower F_{ST} = .082; Fig. 7) there is divergent selection on PC2 (lower P_{ST} = .122 > upper F_{ST} = .082; Fig. 7).

Patterns of divergence in *rump colour* deviate from those expected based on genetic differentiation for both islands and mainland (Table 1). The divergent phenotype on Sicily is causes divergent selection with respect to both the other island populations in PC 2(CeSi:

lower P_{ST} = .086 > upper F_{ST} = .082, CoSi: lower P_{ST} = .059 > upper F_{ST} = .033; Fig. 7). The same is true for Rimini within mainland, especially against Crotone as it covers almost the entire trait (CrRi: lower P_{ST} = .036, .028, .035 > upper F_{ST} = 0.019; Fig. 7). Rump colour also has instances of both divergent and stabilizing selection, between Corsica and Sicily (CoSi) (PC1: upper P_{ST} = .004 < lower F_{ST} = .033, PC2: lower P_{ST} = .059 > upper F_{ST} = .033; Fig. 7).

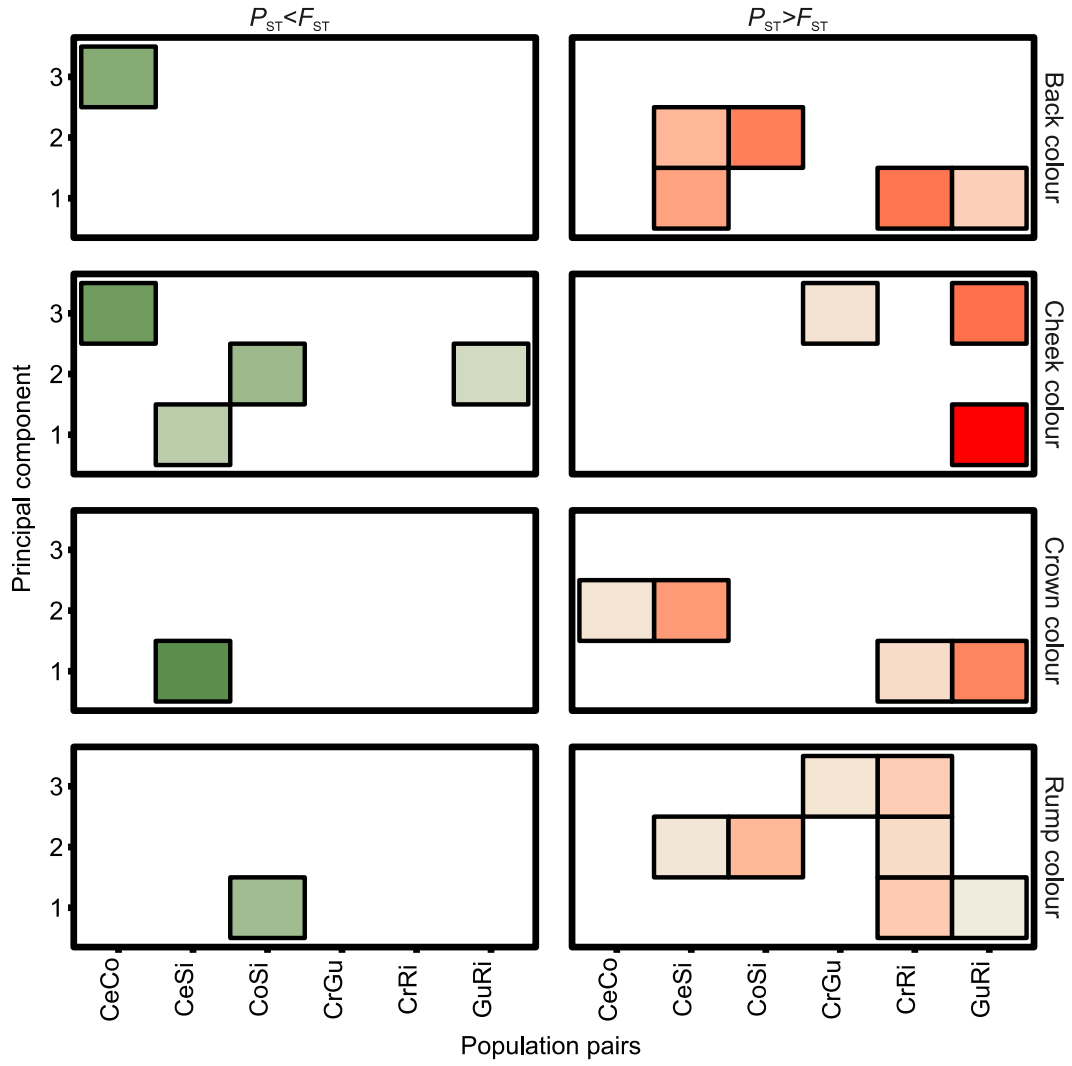

Figure 7. Summary of the P_{ST}/F_{ST} analysis on all plumage traits for each population pair between islands and mainland, respectively. Calculation of phenotypic divergence was performed using conservative estimates on the principle components selected by the broke-stick criteria. Enclosed rectangles represent instances where robust $P_{ST} \neq F_{ST}$. They are colour coded by the extent of difference between phenotypic divergence and genetic differentiation (i.e. $\Delta P_{ST}/F_{ST}$): below F_{ST} (greener), above F_{ST} (redder). Missing squares represents instances of robust P_{ST}/F_{ST} overlap (note: PC 4 for crown and rump colour is not shown as it had only P_{ST}/F_{ST} overlaps). Population abbreviations: *Ce* = Crete, *Co* = Corsica, *Cr* = Crotone, *Gu* = Guglionesi, *Ri* = Rimini, *Si* = Sicily.

3.2.2 Factors explaining morphological variation

Beak morphology

The model that best explains *beak size* variation between island individuals includes several ecological factors and their interactions (AP, $\delta^{13}C$, AP**:**$\delta^{13}C$ and the interaction between 'Sex' and $\delta^{13}C$ (Sex**:**$\delta^{13}C$) (Table 2). There were many models within an AIC_C of 2 from the best island model (Table S8a), and the best model was chosen because it was the one with the lowest number of factors (i.e. the most parsimonious). Overall model fit was good for the best island model (islands: $\Omega^2 = .579$; Table 2). However, there are indications of overfitting in the island model, as the main effect of $\delta^{13}C$ has a small effect-size on beak size ($\beta = -.005$; Table S7) and a reduced Bayesian model without $\delta^{13}C$ increase the posterior probability (PMP = .016, - C = .681; Table S7). Still, the AIC_C score did not improve after such a refit.

Annual precipitation had the largest effect on island individual variation in beak size (change in two standard deviations in mean annual precipitation is associated with a change of .21 standard deviations in beak centroid size). Other factors in the best model for island include the main effects $\delta^{13}C$, 'Sex', and their interactions with 'AP' (Table 2). Males and females seem to have different diets, as size correlates negatively with the amount of $\delta^{13}C$ in males and positively for females for island individuals (PB: $p = .002$, $\beta = -.147$; Table S7). The interaction between annual precipitation and $\delta^{13}C$ also has a relatively large effect (PB: $p = .002$, $\beta = .169$; Table S7). A re-fit with population of origin as a random factor did not significantly improve the fit of the best island model ($p >$ log lik. = .821; Table S7).

Population structure is the only factor in the model that best explains beak size variation between mainland individuals (Table 2). However, there were a large number of models with many ecological variables and their interactions within an AIC_C of 2 (i.e. they are not statistically worse at explaining the variation) from the population model for mainland (Table S8a). Hence, we cannot rule out that ecological factors may influence beak size variation between mainland individuals. The overall fit for the best mainland model was strong (mainland: $R^2 = .444$; Table 2).

Table 2. Summary of all models (i.e. both those for the axis of maximum divergence and multidimensional distances) to show the importance of ecological factors in explaining variation between both island and mainland individuals. Display which factors are present in the best model, the average effect size (β), and the model fit (i.e. R^2 or Ω^2). A 'dash' represent non-applicable data (i.e. there is only one independent variable in the best model). '**:**' is the symbol for interaction between two main effects.

Trait	Group	Data type	Significant independent variables	avg. β	Model fit[1]
Beak Size	island	Centroid size	Annual prec + $\delta^{13}C$ + Sex + Annual prec**:**$\delta^{13}C$ + Annual prec**:**Sex	0.13	0.58
"	mainland	Centroid size	Population	-	0.44
Beak shape	island	PC1	Population	-	0.21
"	island	Euclidean	$\delta^{13}C$ + $\delta^{15}N$ + Annual prec**:**$\delta^{13}C$ + Annual temp**:**$\delta^{13}C$ + Annual temp**:**Sex + Sex**:**$\delta^{15}N$	0.29	0.12
Back colour	island	PC1	Annual prec + $\delta^{15}N$	0.46	0.50
"	mainland	PC1	Temp.seasonality	-	0.62
"	island	Euclidean	Annual temp + $\delta^{13}C$ + $\delta^{15}N$ + Temp.seasonality + $\delta^{13}C$**:**$\delta^{15}N$	0.41	0.28
"	mainland	Euclidean	Annual prec + Annual prec**:**$\delta^{13}C$	0.63	0.21
Cheek colour	mainland	PC1	Population	-	0.46
"	island	Euclidean	$\delta^{15}N$ + Temp.seasonality + Prec.seasonality**:**$\delta^{15}N$	0.20	0.12
"	mainland	Euclidean	Population	-	0.06
Crown colour	island	PC1	Population	-	0.21
"	mainland	PC1	Annual prec + $\delta^{13}C$	0.59	0.63
"	island	Euclidean	Population	-	0.14
"	mainland	Euclidean	$\delta^{13}C$ + $\delta^{15}N$	0.27	0.32
Rump colour	island	PC1	Annual prec + $\delta^{15}N$	0.37	0.27
"	mainland	PC1	$\delta^{13}C$ + Temp.seasonality	0.50	0.45
"	island	Euclidean	$\delta^{15}N$ + Annual prec**:**$\delta^{13}C$ + Annual temp**:**$\delta^{15}N$	0.44	0.12
"	mainland	Euclidean	$\delta^{15}N$ + Temp.seasonality	0.30	0.24

Population structure was the only factor in the model that best explain *beak shape* variation between island individuals along the axis of largest variation (i.e. PC1: Table 2). There were many models with ecological factors in them within 2 AIC_C of the best island model (i.e. they were not significantly worse at explaining the variation), and the current model was selected as it was the most parsimonious one (i.e. contained the least variables) (Table S8a). Overall fit for the best island model was good (R^2 = .209; Table 2). Annual precipitation (AP) is the only factor in the model that best explains beak shape variation between mainland individuals along the axis of largest variation (Table 2). The best model for mainland was chosen because of better Bayesian support than a model with population of origin (Population) within its AIC_C interval (Table S8a). Still, the model fit for mainland remained low (Ω^2 = .061; Table 2), and did not improve after a re-fit with population of origin as the only independent variable.

Ecological factors are more important in the best model explaining *multidimensional beak shape divergence* (i.e. Euclidean distances) between island individuals compared to the best model explaining variation along the major axis of divergence (Table 2). Annual precipitation and its interaction with $\delta^{13}C$ (AP**:**$\delta^{13}C$) has the largest effect on multidimensional beak shape variation between island individuals (PB: p = .001, β = .78; Table S7). The other factors in the best model explaining this variation were $\delta^{13}C$, $\delta^{15}N$,'AT: $\delta^{13}C$' and 'AT**:**Sex' (Table 2). Interestingly, there is evidence that the sparrows of the different sexes in islands exploit trophic levels in their diet differently (Sex**:** $\delta^{15}N$; Table 2). This effect is relatively small compared to that of the 'Sex**:**$\delta^{13}C$' interactions explaining beak size in islands (PB: p = .001, β = .01; Table S7). However, this result is not statistically strong as its prior inclusion probability from the Bayesian model selection is very low for this variable (PIP: .02).

The overall fit of the best model explaining multidimensional distances for island was low (Ω^2 = .117; Table 2). This is consistent with the results from the Bayesian analysis, which favors fewer factors (islands: PMP = .002; Table 2). Removing the main effect of 'Sex' and its interaction with $\delta^{15}N$ for islands increase PMP (islands: PMP = .371; Table 2). However, such a re-fit did not improve AIC_C. This could indicate that the models are too complicated for the amount of variation present in the data. There was one model that was not statistically worse than the best model for islands (i.e. within an AIC_C of 2) that puts seasonal precipitation (SP) in lieu of annual temperature (Table S8b). The current best model explaining multidimensional distances between island individuals was chosen as it had the best Bayesian

support. However, caution should be exercised as to what degree annual temperature (AT) is more important than seasonal precipitation in this instance (ST). The best model for explaining variation in multidimensional distances between mainland individuals was not significantly better than the null-model hypothesis (p = .121). and neither were any of the other models within an AIC_C of 2 (Table S7; Table S8b).

Plumage coloration

Variation between island individuals along the *axis of major back colour divergence* (i.e. PC1) includes the two variables annual precipitation and $\delta^{15}N$ (AP: β = .653, $\chi^2(1, N = 41)$ = 10.08, p = .002; $\delta^{15}N$: β = .273, $\chi^2(1, N = 41) = 5.52$, p = .019; Table 2; Table S7). The best model for back colour variation between mainland individuals includes only temperature seasonality (ST: $\chi^2(1, N = 41) = 12.22$, p = .014; Table 2; Table S7). The model that best explains mainland variation has better explanatory power compared to the best island model (island: Ω^2 = .496, mainland: R^2 = .620; Table 2). Neither models for variation between island nor mainland individuals show evidence of model over-parameterization (Table S7). Population of origin was included as a random factor in both island and mainland models, as it significantly improved model fits (p > log lik. < .001: Table S7).

Models for *multidimensional divergence* in back colour variation between both island and mainland individuals fit the data well, but worse compared to the main axis of back color variation (island: Ω^2 = .280, mainland: Ω^2 = .209; Table 2). The best model for island individual variation has a better fit as the Bayesian support is considerable higher for islands than for mainland (mainland PMP < .001, island: PMP = .600: Table S7). In the best model for island variation in multidimensional distances both main effects of diet ($\delta^{13}C$, $\delta^{15}N$) and climate variables (AT, ST), as well as diet-diet interactions ($\delta^{13}C$**:**$\delta^{15}N$), are significantly influencing back colour variation between island individuals (Table 2). Of these, temperature seasonality (ST) has the largest effect-size (ST: β = .680, PB: p < .001); Table S7). Multidimensional divergence between mainland individuals is best explained by the main effect of annual precipitation (AP) and its interaction with $\delta^{13}C$ (AP**:**$\delta^{13}C$), as these two factors are the only significant ones ('AP**:**$\delta^{13}C$': β = .788, PB: p < .001; Table S7). This is suggesting that precipitation regimes variation between individuals in back colour. Population of origin was included as a random factor as this significantly improved both island and

mainland model fits (p > log lik. < .001: Table S7).

Population of origin was the only factor in the best model for explaining variation along the axis of maximum variation in *cheek colour* between both island and mainland individuals (Table 2). Variation in cheek colour along the axis of maximum variation is similar within and between islands (Population: $F(2, 40) = .76$, $p = .475$; Table S7), and the overall model fit is poor ($R^2 = 0.037$, $p = 0.32$). In contrast, the variation between mainland individuals shows a relatively good fit as at least some mainland populations differ in variation of cheek colour along this axis (Population: $F(2, 40) = 17.10$, $p < .001$, $R^2 = .461$; Table S7). The population only model for both island and mainland, did however, not fit the data significantly better than other models model containing various climate variables (Table S8a). The population only models was chosen as a lower number of factors is a more parsimonious explanation than a higher if the model fit does not differ significantly. This makes an effect of ecological factors difficult to rule out as explanations for variation between both island and mainland individuals along the axis of maximum variation.

Ecological factors influence variation between island individuals in *multidimensional cheek colour divergence*, as the best model has our proxy for trophic levels ($\delta^{15}N$), a temperature seasonality (ST), and a climate-diet interaction (SP**:**$\delta^{15}N$) in its top model (Table 2). Of these factors in the best model for islands, temperature seasonality (ST) has the largest effect on variation between island individuals (ST: $\beta = -.341$, $\chi^2(1, N = 41) = 9.76$, $p = .002$; Table S7). The total closeness of fit for the island model is acceptable, with good Bayesian support ($\Omega^2 = .12$, PMP = .40; Table S7). As suggested by the Bayesian model selection, the exclusion of the interaction 'SP**:**$\delta^{15}N$' in the best island model could improve fit (Table S7), but the AIC_C did not increase from such a re-fit. Population of origin was included as a random factor as this significantly improved island model fit (p > log lik. < .001: Table S7).

Variation in multidimensional cheek colour between mainland individuals is once again best explained by population of origin only (Table 2), but as opposed to the model for the axis of largest variation which showed relatively good overall model fit, the explanatory potential is very poor ($R^2 = .058$). The other model which was not statistically different from the population-only model for mainland (i.e. within 2 AIC_C) did not improve R^2 or differ significantly from the null-model hypothesis (i.e. intercept only model) (Table S8b).

Variation in *crown colour* along the *axis of maximum variation* between island individuals is best explained by population structure only (Population: $\chi^2(2, N = 40) = 9.32$, $p = .010$; Table 2; Table S7). The best model for between mainland individual variation in crown colour consists of annual precipitation (AP) and our proxy for different diets, $\delta^{13}C$ (Table 2). Of these two, annual precipitation (AP) has the largest effect on variation (AP: $\beta = .850$, $\chi^2(1, N = 41) = 10.76$, $p = .001$; Table S7). The best model fits for crown colour are good for both island and mainland (island: $R^2 = .21$, mainland: $\Omega^2 = .63$; Table 2). A correction for the effect of erected feathers on variation only improve model fit for mainland significantly (p >log lik. = .004), but there is little influence of population structure (p >log lik. = .222).

For *multidimensional crown colour divergence*, the variation between island individuals is best explained by population of origin only (Population: $\chi^2(2, N = 41) = 75.25$, $p < .001$; Table 2; Table S7), and model fit was acceptable ($\Omega^2 = .14$). The best model for multidimensional variation between mainland individuals for crown colour consist of our two proxies for diet, $\delta^{13}C$ and $\delta^{15}N$, where the former has the largest effect ($\delta^{13}C$: $\beta = .330$, $\chi^2(1, N = 41) = 74.21$, $p < .001$; Table 2; Table S7). There is also good overall support for this model ($\Omega^2 = .322$, PMP = .833; Table S7), and correcting both for bushiness (p > log lik. < .001) and for population structure (p >log lik. < .001) through adding these as random variables improves model fit. However, we cannot rule out that climate variables are affecting variation between mainland individuals, as there were many models containing climate variables that were not a significantly worse fit (Table S8b). The best model was chosen based on parsimony (e.g. the lowest number of variables).

The best model explaining *rump colour* variation between island individuals along the *axis of maximum variation* includes annual precipitation (AP) and our proxy for trophic levels $\delta^{15}N$ ($\delta^{15}N$) only (Table 2). These two independent variables have comparable effects on variation between island individuals (i.e., AP: $\beta = .336$, $\delta^{15}N$: $\beta = .398$; Table S7). The best model explaining rump colour variation between mainland individuals along the axis of maximum variation includes our proxy for the type of plant diet ($\delta^{13}C$) and temperature seasonality (ST) only (Table 2). Of these two independent variables, temperature seasonality has the largest effect on rump colour variation between mainland individuals (ST: $\beta = .704$, $\chi^2(1, N = 41) = 10.41$, $p < .001$; Table S7).

Model fit for mainland is better than model fit explaining variation along the axis of maximum variation between island individuals (island: $\Omega^2 = .270$, mainland: $\Omega^2 = .450$; Table

2). Both the model for island individual variation and that for mainland individual variation have good Bayesian support (island: PMP = .360, mainland: PMP = .380; Table S7), and adding population as a random factor did not improve model fit for either group (island: p >log lik. = .092, mainland: p >log lik = .072; Table S7). Annual temperature (AT) in lieu of annual precipitation (AP) was in a model within 2 AIC_C of the best island model (Table S8a). The current model was selected as best model as it had better Bayesian support. Several models were also insignificantly worse at explaining rump colour variation between mainland individuals (Table S8a). The current model was selected as it was the most parsimonious.

The model for rump colour explaining variation between island individuals for *multidimensional divergence* has many ecological variables in its best model but the parametric bootstrap renders most main effects statistically unreliable (Table 2; Table S7). Hence, island variation is best explained by remaining climate-diet interactions (AP**:**$\delta^{13}C$, AT**:**$\delta^{15}N$) and the main effect of $\delta^{15}N$ (Table 2). All of these variables have comparable effects on explaining island variation, the largest being the interaction between annual precipitation (AP) and $\delta^{13}C$ ('AP**:**$\delta^{13}C$': β = - .50, PB: p = .002; Table S7). Temperature seasonality (ST) and $\delta^{15}N$ are the only two variables in the best model explaining variation in multidimensional distances between mainland individuals, where temperature seasonality has the largest effect (ST: β = .490, $\chi^2(1, N = 41) = 15.02$, $p < .001$; Table S7).

The mainland model explains data better than does the island model for multidimensional distances (island: Ω^2 = .120, mainland: Ω^2 = .580; Table 2). Overfitting is penalized more by the Bayesian approach in the best island model for multidimensional distances, as expected from the high number of factors. Removing the main effect of $\delta^{13}C$ increase Bayesian support effectively (islands: PMP = .180 < - $\delta^{13}C$ = .772; Table S7). There were several other models that were insignificantly worse at explaining variation in multidimensional distances between both island and mainland individuals (Table S8b). The best island model was selected on grounds of better Bayesian support, and the best mainland model was selected as it was the simplest one (i.e. the most parsimonious). Adding population of origin as a random factor improves model fit in multidimensional distances for the best island model (p > log lik. < .001), but not for the best mainland model (p > log lik. = .076).

3.3 Hybridization and phenotypic diversity

Overall, there is evidence for the occurrence of a variety of phenotypic parental combinations: beak morphology is transgressive or intermediate, and plumage colour is mosaic with both transgressive and intermediate traits (Table 3; Fig. 8). However, the occurrence of transgressions is both more frequent and extensive for island populations. Islands and mainland individuals tend to have fixated different parental phenotypes for multidimensional distances (D^2) (Table 3). For the axis of maximum parental separation (CV1), both islands and mainland individuals show some diversity between traits in terms of parental resemblance, being not solely restricted to neither Spanish nor house sparrows (Fig. 3).

Table 3. Multidimensional distance estimates (Mahalanobis distance, D^2) from a CVA using the four groups: 'House parental' (house sparrow), 'Spanish parental' (Spanish sparrow),'island' (all individuals from island populations of Italian sparrows), and 'mainland' (all individuals from mainland populations of Italian sparrows). The CVA was performed for each plumage trait separately, using principle components (n = 10) for plumage traits, and allometrically corrected procrustes coordinates (n = 32) for beak shape. The number in parentheses are the results from a 1000 round permutation to check if the groups were significantly differentiated.

Trait	Group	House parental		Spanish parental		island	
Beak shape	Spanish parental	3.12	(0.001)				
	island	1.98	(0.001)	2.37	(0.001)		
	mainland	2.01	(0.001)	1.79	(0.001)	1.33	(0.001)
Plumage, all	Spanish parental	10.16	(0.001)				
	island	9.18	(0.001)	4.13	(0.001)		
	mainland	9.79	(0.001)	4.62	(0.001)	4.32	(0.001)
Back colour	Spanish parental	4.62	(0.001)				
	island	1.24	(0.395)	4.36	(0.001)		
	mainland	2.49	(0.001)	4.23	(0.001)	2.16	(0.001)
Cheek colour	Spanish parental	4.08	(0.001)				
	island	2.03	(0.002)	3.01	(0.001)		
	mainland	4.41	(0.001)	1.29	(0.380)	3.59	(0.001)
Crown colour	Spanish parental	10.91	(0.001)				
	island	11.01	(0.001)	1.37	(0.999)		
	mainland	11.11	(0.001)	2.03	(0.535)	1.34	(0.999)
Rump colour	Spanish parental	4.27	(0.001)				
	island	2.16	(0.001)	3.16	(0.001)		
	mainland	3.93	(0.001)	2.37	(0.001)	3.29	(0.001)

Beak morphology

Beak size has the most transgressive phenotype, as both island and mainland have larger beaks than both parent species (Fig. 8). The Crotone and Crete populations has the most transgressive individuals, while Sicily and Rimini has the least transgressive (Fig.8). Interestingly, the two parental populations are statistically undifferentiated (t (98) = 2.07, p = .594).

Overall, *beak shape* had the lowest global classification accuracy of all phenotypes (48%). Island and mainland individuals are closer to one another than to either parent species (Table 3). With house and Spanish sparrows considered independently, islands are closer to house sparrows (island; D^2 = 1.98, p = .001; Table 3) and mainland is closer to Spanish sparrows (mainland: D^2 = 1.79, p = .001; Table 3). However, considering the axis of maximal separation explaining 42% of the variation, this pattern is reversed so that islands are now closer to Spanish sparrows and mainland closer to house sparrows (Fig. 8). Both island and mainland populations show considerable overlap along the axis of maximum separation, and the phenotype is intermediate (Fig. 8).

Plumage coloration

Back colour has an overall classification accuracy of 66%. For multidimensional distances both islands and mainland are closer to house sparrows, with Spanish sparrows as an outgroup (Table 3). Island populations are more similar to house sparrows than are mainland populations (islands-house: D^2 = 1.24, p = .395, mainland-s: D^2 = 2.49, p = .001: Table 3). Parent species resemblance along the axis of maximum separation does, however, vary strongly between island populations, with Crete and Corsica resembling house sparrows and Sicily Spanish sparrows (Fig. 8). Corsica and Crete have both instances of transgressive individuals (Fig. 8: red ‘dots’). Mainland populations are also located at different positions along the main parental axis, where Rimini individuals are more similar to house sparrows compared to the others (Fig. 8.).

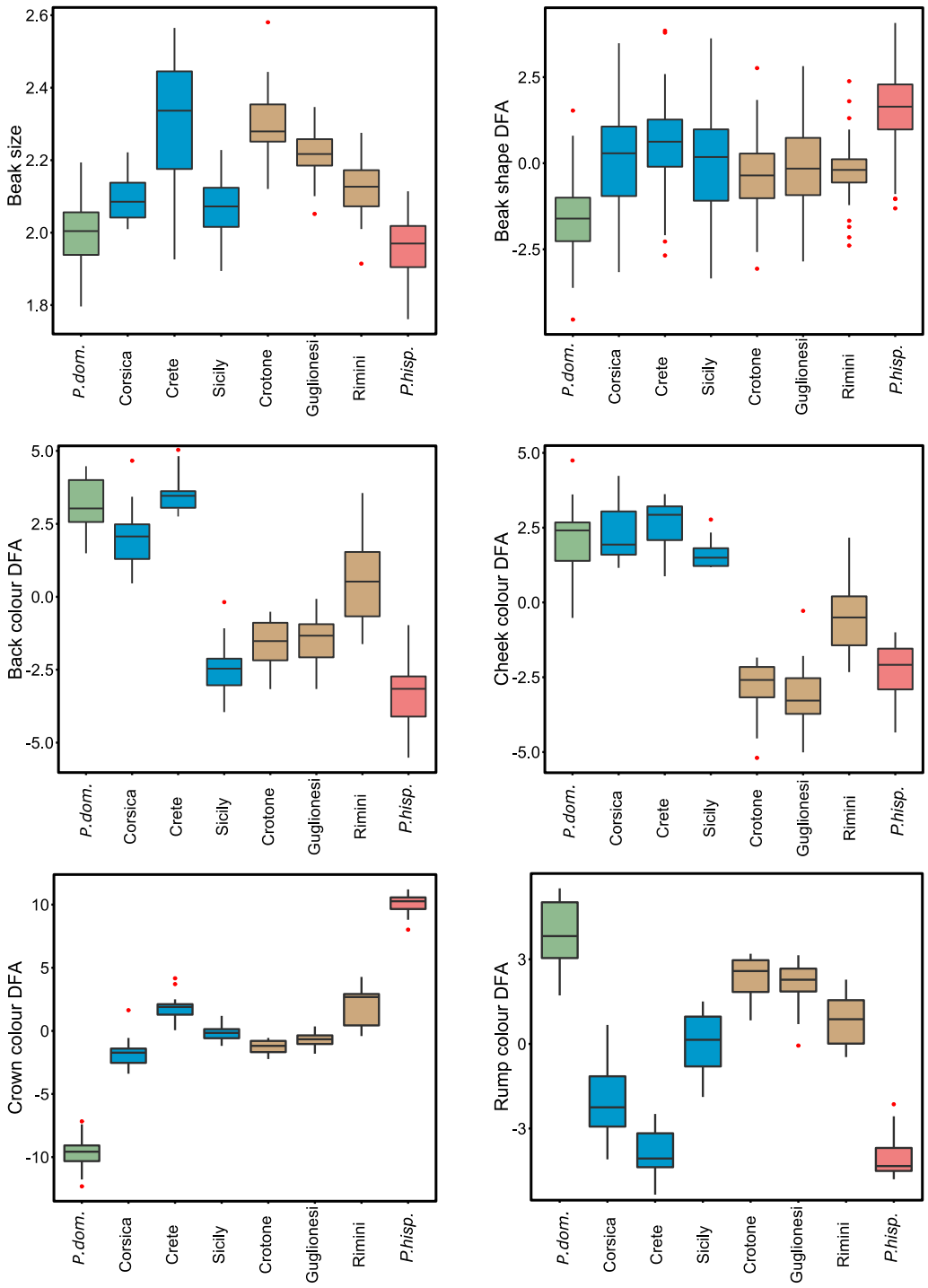

Figure 8. DFAs on all populations of Italian sparrows (n = 6; blue = islands, brown = mainland) and the two parent species populations, house sparrows (n = 1) and Spanish sparrows (n = 1). The data for beak size are the raw centroid sizes. Plumage trait scores for Italian sparrows are derived from the first CV of a CVA using the first 10 principal components for each trait separately. Beak shape scores are derived from the first CV of a CVA using all allometrically corrected procrustes coordinates (n =32). House sparrows and Spanish sparrow scores are derived from an LDA using the same data as the Italian sparrow. Red dots represent outliers.

Mainland populations resemble Spanish sparrows for multidimensional *cheek colour* distances, whereas islands are more similar to house sparrows (islands-house: $D^2 = 2.03$, $p = .002$; mainland-Spanish: $D^2 = 1.29$, $p = .380$; Table 3). The island populations are more similar to house sparrows than to the mainland populations, and mainland populations more similar to Spanish sparrows than to the island populations compared to the resemblance between island and mainland groups themselves ($D^2 = 3.59$, $p = .001$; Table 3). The CVA could relatively easily distinguish between groups (classification accuracy: 72%). This pattern of resemblance is repeated along the axis of maximal separation (Fig. 8). Both island and mainland groups have transgressive individuals, especially the mainland populations Guglionesi and Crotone (Fig. 8).

Classification accuracy for *crown colour* was very good (85%). House sparrows form a distant outgroup, and both islands and mainland strongly resemble Spanish sparrows to the point where they are statistically indistinguishable (island-Spanish: $D^2 = 1.37$, $p \approx 1$; mainland-Spanish: $D^2 = 2.03$, $p = .52$; Table 3). Along the axis of maximum separation, Crown plumage is clearly intermediate (Fig. 8). However, the individual populations within both islands and mainland vary between themselves (Fig. 8).

The different groups were not easily predicted by the DFA for *rump colour* (classification accuracy = 56%). Island populations resemble house sparrows in rump coloration, whereas mainland populations resemble Spanish sparrows (island-house: $D^2 = 2.16$, $p = .001$; mainland-Spanish: $D^2 = 2.37$, $p = .001$; Table 3). This pattern is not repeated along the axis of parental separation. Instead, island populations are more similar to Spanish sparrows, and mainland populations resemble house sparrows along this axis (Fig. 8). Island populations show signs of transgressive individuals for Crete (i.e. boxplot lower end extend outside the range of the Spanish sparrow; Fig. 8).

4 Discussion

I found that hybrid populations of Italian sparrows readily diverge from each other in beak size, beak shape and plumage coloration, independent of spatial isolation. It hence is likely that variation arising from hybrid speciation can provide the raw material for evolution to occur over relatively short time spans. Furthermore, I found widespread decoupling of phenotypic divergence from genetic differentiation, which suggests that adaptive divergence is taking place. Although directional selection appears to be widespread and affect several traits, I also found evidence consistent with stabilizing selection for other traits. It is likely that differences in climate or diet provide some of the selection pressures that are driving diversification. The Italian sparrow can combine parental phenotypes differently across traits, showing mosaicism, intermediacy and also signs of transgressive segregation at the various traits. However, there are important differences between the traits, where some always resemble the trait values of one parent whereas others differ between populations.

I found that island populations are more phenotypically diverged from each other than are mainland populations, especially in traits putatively important for local adaptation (beak morphology). The genetic divergence in the Italian sparrow is substantial for island populations, but low of mainland populations. A meta study found that parent species of hybrids typically have an average F_{ST} of below .03 (Chapman & Burke, 2007) supporting that the divergence between the island populations is relatively high. Although divergence by selection was equally common for both islands and mainland, my results suggest that ecological differences account for more of the variation between individuals across islands. Consistent with expectation from non-hybrid speciation (Coyne & Orr, 2004; Price, 2008), this suggests that spatial isolation and island habitats favor more rapid local adaptation also in a hybrid species. For instance, the absence of gene flow from populations adapted to other conditions could increase the opportunity for divergence. Furthermore, island populations show evidence of transgression in more traits, and exhibited greater variability among populations as to which parent species they resembled.

The observed divergence could either be due to sorting of parental alleles causing allele frequency differences that result in evolutionary change, or due to phenotypic plasticity. Phenotypic plasticity is the genetic capacity to produce a variety of phenotypes in response to contemporary changes in abiotic and biotic factors (Pigliucci, 2001). Studies on beak shape

and size in the Darwin finches have, however, demonstrated a strong heritable component controlled by few genes (Grant & Grant, 2002), so potentially beak shape divergence could be genetic in this system too. Baily (2015) also found evidence suggesting that crown plumage colour is largely governed by few genes. Hence, it is plausible that some of the variation in the Italian sparrow is due to adaptive genetic change. To disentangle of the role of plasticity common garden experiments should be employed (see, for instance; Wikelski et al., 2003; c.f. Pigliucci, 2001) to investigate to what degree the Italian sparrow is able to change its contemporary phenotypic expression in the face of changing environmental factors.

Decoupling of phenotypic divergence from genotypic differentiation

Genomic and phenotypic evolution following hybridization events is not well understood (Schumer et al., 2014). Whether hybrids are shaped by historical contingencies and stochastic processes as opposed to adaptive divergence are important questions (Eroukhmanoff, 2013). This study provides evidence for phenotypic divergence that exceeds that expected from genetic differentiation, suggesting that adaptive divergence takes place in this hybrid species. Hence, variation arising from hybridization can allow for rapid local adaptation in some traits. As there are indications of stabilizing selection for some traits, selection may also oppose phenotypic diversification across populations for other traits.

I found instances where phenotypic divergence differed from the expectation based on genetic differentiation for all traits. This suggests that populations have diverged by selection, even in the presence of gene flow, in line with the findings of Eroukhmanoff (2013). Hybrid speciation thus seems to generate variation that directional selection can act on and cause divergent forms. This is an interesting addition to our understanding of how a hybrid species is formed, complementing our knowledge of how selection against unfit recombinants has molded the hybrid genome of the Italian sparrow (Trier et al., 2014).

Although less frequent than divergent selection, stabilizing selection appears to have kept some traits or their components more similar than expected from genetic divergence. However, there is also the possibility that constraints imposed by historical or genetic contingencies could limit divergence and explain these patterns (c.f. Schluter, 1996). If traits are co-adapted, future alternatives may be contingent on the prior fixation of other traits in the population, reflecting the process occurring at a genetic level (c.f. Trier et al., 2014). There is

also a possibility that an erroneous inference of stabilizing selection could result from using a genome wide estimation of F_{ST}. Future studies should aim to identify neutral regions to ascertain a more accurate assessment.

Isolation promotes population divergence

In several independent statistical tests in this study the Italian sparrow inhabiting the Mediterranean islands show greater inter-population variation compared to mainland populations. The proportion of variation explained by grouping individuals by population of origin suggests that this difference in population divergence is significant. It is more plausible that the standing genetic variation arising from hybridization combined with spatial isolation has allowed for the island populations to diverge more than the mainland populations. The time it takes for mutations to generate favorable variants for selection or drift to act upon, happens on a large geological time scale (Uyeda et al., 2011). The house sparrow spread with the flow of agriculture from Mesopotamia, and came into contact with the non-commensal Spanish sparrow around 6000-10 000 years ago (Summers-Smith, 1963). It is therefore unlikely that considerable proportions of this variation is the result of new point mutations. Conspicuous divergence between isolated populations is a well-known phenomenon (Mayr, 1954). Still, the extent of the difference between islands and mainland is not as substantial as anticipated, as mainland populations also readily diversify. This suggests a strong role for local ecology rather than genetic differentiation in shaping hybrid species phenotypes.

Factors explaining divergence – the roles of drift and selection

P_{ST}/F_{ST} comparisons suggested that significant divergent phenotypic selection was equally frequent for island and mainland populations. Interestingly island population divergence seems to less frequently be consistent with stochasticity than that of mainland population, contrary to the expectation. Genetic drift is expected to be stronger on islands as the population sizes there usually are smaller (Slatkin, 1977). For a hybrid species, certain combinations of parental alleles may be more likely lost by chance in smaller populations. Hence, an overall stronger phenotypic divergence between island populations could be expected merely because of drift acting on genotypic and phenotypic traits, but the observed patterns of stronger departure from the neutral expectation in phenotypic divergence are likely

to reflect selection. Stronger differences in ecology between the islands combined with spatial isolation would, however, be consistent with the observed patterns of phenotypic divergence.

In general, differences in ecological variables were more important in explaining variation between island individuals than between mainland individuals. Potentially, this could be due to stronger differences in ecology between island populations than between mainland populations, but this remains to be tested. Interestingly, Eroukhmanoff et.al (2013) found that mainland populations readily diverge by selection driven by climate factors. I found that precipitation regimes and variation in diet have a relatively large effect on beak size and back colour compared to other ecological variables. Moreover, these factors explained the variation more reliably for island populations than for mainland populations. It is thus plausible that more unique island habitats have resulted in strong adaptive divergence.

The prevalence of stabilizing selection was higher between island populations compared to mainland populations. Mainland populations have few instances where morphological divergence is significantly lower the genome wide expectation, probably as a result of the overall lower genetic divergence on the mainland. Interestingly, along the axis of maximum variation (i.e. PC1), the traits that show signs of stabilizing selection are frequently best explained by population structure only. Moreover, the models for the multidimensional distances have relatively poor fit. This could indicate that these traits are constrained to a small range of phenotypic values, either due to a lack of variation or due to strong selection for values within this range. For instance, a specific suit of sexually selected plumage characteristics could have been required for pre-mating isolation against its parent species (c.f. Mavárez et al., 2006).

Parental resemblance: phenotypic mosasicm, intermediacy and transgression

I found that both island and mainland populations showed diversity in patterns of parental resemblance. Individuals readily combine different parental traits in a mosaic pattern. However, island population show transgressive segregation for more traits, including back and rump colour. In addition, island populations vary more in which parent they resemble along the axis of maximum parental separation. For multidimensional divergence, island individuals tend to be more house sparrow-like while mainland individuals are more similar to Spanish sparrows.

For some traits, patterns of divergence differ with population, or dependent on island or mainland geography. For instance, mainland populations resemble Spanish sparrows with respect to cheek colour, whereas the island populations resemble house sparrows. Both island and mainland populations vary more in which parent species they resemble for back and rump colour, whereas crown is fixed for the Spanish sparrow phenotype for all Italian populations.

Phenotypic composition in a hybrid can be dependent on the level of fixation of alleles in the parents and the degree of differentiation between them (Stelkens & Seehausen, 2009). For instance, when the parental species have similar phenotypes and when there is stabilizing selection in the parent species, transgression in hybrids increases (Riesberg et al., 1999). It has been suggested that transgression is relatively common (Rieseberg et al., 1999). In the Italian sparrow, there were individuals with transgressive values for over half of the traits. Exploitation of new ecological niches aided by transgressive features could aid homoploid hybrid speciation to succeed (Abbot et al., 2013; Rieseberg et al., 2003). Interestingly, island populations have increased levels of transgression compared to mainland populations, potentially reflecting that the island environment is less similar to that of the sampled populations of the parent species.

Patterns of divergence and selection regimes differ with study trait

A reoccurring pattern in this study is how the effects of isolation appear to differ across traits. Island populations have more inter-population variation, but not for all traits. Island population appears to evolve more by non-random evolution, but dependent on which trait. This is not surprising. These are traits putatively under selection as they are involved in social interactions and feeding. Much research into bird biodiversity have shown different adaptive roles for beak morphology and plumage colour (c.f. Price, 2008). It is thus tempting to speculate that the differences found in this study are somehow linked to their adaptive roles. However, life-history trade-offs can distort a broad categorization into purely sexual traits or traits adapted for survivability. Beak morphology is mainly associated with feeding behaviour and expected to be subjected to strong natural selection (Mallarino et al., 2012). In the Italian sparrow, Eroukhmanoff et al. (2013) found that climate variables were important in explaining beak size, and Piñeiro (2014) a strong effect of diet on beak shape. However, beak morphology is also important for species recognition and sexual selection (Huber and Podos, 2006; Kimball, 1996). Plumage colour in the Italian sparrow has been shown to be important

for species recognition (Bailey et al., 2015), and in intraspecific competition (Sætre et al., in preparation). Traits subjected to natural selection could be expected to diverge more between localities with different local ecologies. However, environmental factors may also influence sexually selected traits. For instance, colouration used for display in social interactions must be conspicuous enough to shrouding vegetation so the recipients can perceive the signal efficiently (Endler & Thery, 1996). I will thus proceed with caution, as my data suggests a difference between beak morphology and plumage colour in several aspects of the above discussion. In addition, there is variation between different plumage colour traits that could tempt further speculation.

Beak size –divergence, selection, and strong influence of ecology

Beak size is the most diverged with most occurrences of inferred divergent selection. Interestingly, ecological factors explain beak size variation in most models. Furthermore, beak size is a highly transgressive. These patterns are consistent with strong ecological divergent selection on beak size. Interestingly, island individuals have smaller beaks compared to mainland individuals, which is not expected as the "island-syndrome" states that larger traits are expected in small species because of the availability of ecological niches due to low levels of interspecific competition (Adler & Levins, 1994). This is the case some for passerine birds too (Grant, 1965).

Models with $\delta^{13}C$ and its additive effects with both annual precipitation and sex explained more of the beak size variation between individuals on islands than in mainland populations. The carbon isotopic signature of plants is transplanted through the food chain and deposits traces in organic tissue (Peterson & Fry, 1987). The ratio of the stable isotope of carbon in plants varies depending on how carbon fixation occurs. The ratio increases from C_3 plants (grains and fruit), through CAM (Crassulacean acid metabolic) plants, and is highest for C_4 plants (e.g. corn) (Hobson, 1999). Its interaction with climate may reflect that climate affect which plants are cultivated. Hence, stronger plant composition divergence between islands could explain why diet is a better predictor of beak shape in island populations. Interestingly, the sexes have different feeding patterns that influence beak size (i.e. there is a significant interaction between $\delta^{13}C$ and sex). This is consistent with how sexual dimorphism is expected to increase during ecological niche divergence (e.g. Selander's rule; Selander 1966).

Beak size is also strongly transgressive, and this could have aided speciation by new ecological opportunity in the Italian sparrow, both for islands and mainland populations.

Transgression is expected to be more frequent when there is no divergence between parental species in a particular trait (Rieseberg et al., 1999, Bailey et al., 2015). I found that the parent species had similar beak sizes. All Italian sparrow populations show transgression in beak size, with Crete having the most transgressive individuals. Together, this suggests that ecological opportunity offered by island habitats can favour local adaptation.

Beak shape – stabilizing selection with an interesting exception

Beak shape shows limited but in some cases significant differentiation, and some aspects frequently exhibit patterns consistent with stabilizing selection. Beak shape is consistently intermediate between parent species. Interestingly, one landmark, associated with the intersection between the par mandibularis and the malar region of the head experience divergent selection both among island and mainland individuals. The position of this landmark is changing considerably both vertically and horizontally. This has an influence on how tapered the lower mandible is towards the posterior part of the beak. The shape of the lower mandible is, especially for seed eaters, important for the ability to crack open seed casings (Bowman, 1961). Hence, depending on where in the beak crushing takes place, there could be a need to alter the tapering to accommodate different seed types and feeding habits. A study of how the Italian sparrow handles the seed is needed to investigate if this could be driving the shape divergence.

Plumage colour – a mixed picture

I found that back and rump colour, as well as cheek and crown colour consistently grouped together when comparing plumage traits. Support for this can be found from all parts of my analysis. Firstly, back and rump colour were more diverged between island populations, and crown and cheek colour were more diverged between mainland populations. Second, counting the number of occurrences of divergence and the lack of it, islands have more non-random modes of evolution in back colour. However, mainland showed more divergent selection in cheek colour and the same number of instances in rump colour. Third, ecological factors accounts for more of the variation between individuals, independent of island or mainland setting, for back and rump colour. Conversely, cheek and crown colorations is to a larger extent better explained by population of origin. And lastly, results from the parental comparisons show that both island and mainland population vary more in which parent species they resemble for back and rump colour, and that cheek and crown is fixed for either parental phenotype. The most immediate interpretation is that there is a distinction between

the two groups of traits in which adaptive roles they play in the Italian sparrow. However, there are findings to suggest that such categorization is problematic.

The sexual dimorphism in plumage in Italian sparrows is a strong indication that male plumage traits are under sexual selection (Andersson, 1994). If sexual selection causes reduced divergence in islands for cheek (Galván, & Sanz, 2008) and crown (Bailey et al., 2015) colour, phenotypic divergence estimates below genetic differentiation would be expected and the role of ecology in explaining divergence should be small. Cheek colour is the least differentiated trait, and shows the greatest signs of stabilizing selection/constraints for islands. Furthermore, there are very few of the best cheek color models containing any ecological factors. However, mainland individuals are highly divergent in cheek colour, and patterns are consistent with strong divergent selection. This is likely not driven primarily by ecological factors as there are none in the best model for neither the axis of maximum variation nor multidimensional distances. Baily (2015) did not find cheek colour to act as a sexual signal in the Alps hybrid zone hence promoting a different role in this trait, but as Bache-Mathiesen points out, this could be due to methodological problems (Bache-Mathiesen 2014). Crown colour is significantly diverged between both island and mainland groups, and models with ecological factors have good support. Furthermore, there selection causing this divergence. Potentially, presence of winter plumage in the crowns could be a confounding factor (c.f. Tesaker, 2014).

Hence, the categorization certain plumage traits as purely adapted to meet social demands is perhaps masking a more complex role. A study in male blue tits found that there were divergent selection acting on the blue crown colour due to alternative mating strategies (Delhey, 2007). Sometimes, dependency on other sexual signals could diminish the need to maintain others. Studies in finches has found that more elaborate vocalization patterns were linked to a decrease in plumage ornamentation (Badyaev et al., 2002). Other life-history trade-offs like the link between testosterone levels needed to build secondary sexual plumage characteristics and immosuppressive effects could spur selection in either direction (Roberts & Peters, 2009).

In summary, there are some indications that traits mainly subjected to natural selection, such as beak size diverge more between island populations and the best models explaining divergence include ecological factors, whereas traits expected to be under sexual selection such as crown and cheek color are less diverged in both island and mainland individuals and

the patterns of divergence in the latter traits are not as well explained by ecological factors. However, divergence and ecology for sexual traits cannot be ruled out. Future studies need to investigate the role of other social traits, e.g. vocalization, to grasp the fuller picture. Life-history studies could also help to identify how the Italian sparrow might invest differently in sexual ornamentation due to ecological selection pressures.

This study is to my knowledge the first to investigate how spatial isolation affect phenotypic and genetic diversity within a homoploid hybrid species, and partakes in the growing body of research mapping the role of hybridization in adaptive evolution. I have shown that island populations are more phenotypically diverged compared to their mainland conspecifics, likely driven by greater ecological differences and local adaptation. Moreover, I have found reliable evidence for selection acting on almost all the traits examined here, irrespective of the degree of spatial isolation. Although there is reason to believe that ecology has played a part in molding beak morphology, more thorough investigation is needed before any firm conclusions can be drawn as to how the trade-off between survival and reproduction influence the evolutionary patterns in plumage traits. Future studies might include comparison of insular and mainland species of house and Spanish sparrows to investigate whether the degree of local adaptation and how this is affected by spatial isolation differ between hybrid and non-hybrid species.

References

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*(5689), 1462-14 65.

Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N. & Butlin (2013). Hybridization and speciation. *Journal of Evolutionary Biology*, *26*(2), 229-246.

Adler, G. H., & Levins, R. (1994). The island syndrome in rodent populations.*Quarterly Review of Biology*, 473-490.

Anderson, E., & Stebbins Jr, G. L. (1954). Hybridization as an evolutionary stimulus. *Evolution*, 378-388.

Andersson, M. (1982). Female choice selects for extreme tail length in a widowbird. *Nature*, *299*(5886), 818-820.

Andersson, M. B. (1994). *Sexual selection*. Princeton University Press.

Bache-Mathiesen, L. (2015). *The Evolutionary Potential of Male Plumage Colour in a Hybrid Sparrow Species* (Master thesis). University of Oslo, Oslo.

Badyaev, A. V., Hill, G. E., & Weckworth, B. V. (2002). Species divergence in sexually selected traits: increase in song elaboration is related to decrease in plumage ornamentation in finches. *Evolution*, *56*(2), 412-419.

Bailey, R. I., Tesaker, M. R., Trier, C. N., & Sætre, G. P. (2015). Strong selection on male plumage in a hybrid zone between a hybrid bird species and one of its parents. *Journal of evolutionary biology*, *28*(6), 1257-1269.

Barton, N. H. (2001). The role of hybridization in evolution. *Molecular Ecology*, *10*(3), 551-568.

Bates, D. M. (2010). lme4: Mixed-effects modeling with R. *URL http://lme4. r-forge. r-project. org/book*.

Bates, D. M., Maechler, N., Bolker, B., Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67 (1), 1-48. doi:10.18637/jss.v067.i01.

Bell, M. A., & Travis, M. P. (2005). Hybridization, transgressive segregation, genetic covariation, and adaptive radiation. *Trends in ecology & evolution*,*20*(7), 358-361.

Bittner, D., Excoffier, L., & Largiader, C. R. (2010). Patterns of morphological changes and hybridization between sympatric whitefish morphs (Coregonus spp.) in a Swiss lake: a role for eutrophication? *Molecular ecology*, *19*(10), 2152-2167.

Bivand, R., Keitt, T., Rowlingson, B. (2016). rgdal: Bindings for the Geospatial Data Abstraction Library. R package version 1.1-10. https://CRAN.R-project.org/package=rgdal

Bowman, R. I. (1961). Morphological differentiation and adaptation in the galápagos finches. diferenciación morfológica y adaptación en los pinzones de las galápagos. *University of California Publications in Zoology*, *58*, 1-302.

Brommer, J. E. (2011). Whither PST? The approximation of QST by PST in evolutionary and conservation biology. *Journal of evolutionary biology*, *24*(6), 1160-1168.

Campàs, O., Mallarino, R., Herrel, A., Abzhanov, A., & Brenner, M. P. (2010). Scaling and shear transformations capture beak shape variation in Darwin's finches. *Proceedings of the National Academy of Sciences*, *107*(8), 3356-3360.

Canty, A. & Ripley.B. (2015). boot: Bootstrap R (S-Plus) Functions. R package version 1.3-17.

Chapman, M. A., & Burke, J. M. (2007). Genetic divergence and hybrid speciation. *Evolution*, *61*(7), 1773-1780.

Charles Darwin., *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*, (London: John Murray) 1859; Reprinted in Ernst Mayr, ed., *On the Origin Of Species: A Facsimile of the First Edition*, Cambridge, Massachusetts: Harvard University Press, 1964, 495 pp.

Chenoweth, S. F., Rundle, H. D., & Blows, M. W. (2010). The contribution of selection and genetic constraints to phenotypic divergence. *The American Naturalist*, *175*(2), 186-196.

Christophe, N., & Baudoin, C. (1998). Olfactory preferences in two strains of wild mice, Mus musculus musculusandMus musculus domesticus, and their hybrids. *Animal behaviour*, *56*(2), 365-369.

Coyne, J. A., & Orr, H. A. (2004). *Speciation* (Vol. 37). Sunderland, MA: Sinauer Associates.

Danecek, P., Adam Auton, Goncalo Abecasis, Cornelis A. Albers, Eric Banks, Mark A. DePristo, Robert Handsaker, Gerton Lunter, Gabor Marth, Stephen T. Sherry, Gilean McVean, Richard Durbin and 1000 Genomes Project Analysis Group. The Variant Call Format and VCFtools. *Bioinfromatics*, 2011

Delhey, K., Johnsen, A., Peters, A., Andersson, S., & Kempenaers, B. (2003). Paternity analysis reveals opposing selection pressures on crown coloration in the blue tit (Parus caeruleus). *Proceedings of the Royal Society of London B: Biological Sciences*, *270*(1528), 2057-2063.

DePristo, M., Banks E, Poplin R, Garimella K, Maguire J, Hartl C, Philippakis A, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell T, Kernytsky A, Sivachenko A, Cibulskis K, Gabriel S, Altshuler D, Daly M, 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *NATURE GENETICS 43:491-498*

Dunnett, C. W. (1964). New tables for multiple comparisons with a control. *Biometrics*, *20*(3), 482-49.

Elgvin, T. O., Hermansen, J. S., Fijarczyk, A., Bonnet, T., Borge, T., Saether, S. A., ... & SÆTRE, G. P. (2011). Hybrid speciation in sparrows II: a role for sex chromosomes? *Molecular ecology*, *20*(18), 3823-3837.

Endler, J. A., & Thery, M. (1996). Interacting effects of lek placement, display behavior, ambient light, and color patterns in three neotropical forest-dwelling birds. *American Naturalist*, 421-452.

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.

Fisher, R. A. (1938). The statistical utilization of multiple measurements. *Annals of eugenics*, *8*(4), 376-386.

Fox, J. & Weisberg, S. (2011). An {R} Companion to Applied Regression, Second Edition. Thousand Oaks CA: Sage. URL: http://socserv.socsci.mcmaster.ca/jfox/Books/Companion

Galván, I., & Sanz, J. J. (2008). The cheek plumage patch is an amplifier of dominance in great tits. *Biology Letters*, *4*(1), 12-15.

Gelman, A. & Yu-Sung Su (2015). arm: Data Analysis Using Regression and Multilevel/Hierarchical Models. R package version 1.8-6. https://CRAN.R-project.org/package=arm

Grant, P. R. (1965). The adaptive significance of some size trends in island birds. *Evolution*, 355-367.

Grant, P. R., & Grant, B. R. (2002). Unpredictable evolution in a 30-year study of Darwin's finches. *Science*, *296*(5568), 707-711.

Grant, P. R., Grant, B. R., Markert, J. A., Keller, L. F., & Petren, K. (2004). Convergent evolution of Darwin's finches caused by introgressive hybridization and selection. *Evolution*, *58*(7), 1588-1599.

Grant, P. R., & Grant, B. R. (2011). *How and why species multiply: the radiation of Darwin's finches*. Princeton University Press.

Gross, B. L., & Rieseberg, L. H. (2005). The ecological genetics of homoploid hybrid speciation. *Journal of heredity*, *96*(3), 241-252.

Gunz, P., & Mitteroecker, P. (2013). Semilandmarks: a method for quantifying curves and surfaces. *Hystrix, the Italian Journal of Mammalogy*,*24*(1), 103-109.

Gustin, M., & Sorace, A. (2002). Autumn movements of Italian Sparrows Passer italiae in central Italy. *Ringing & Migration*, *21*(1), 1-4.

Hadfield, J. D. (2010). MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. *Journal of Statistical Software*, 33(2), 1-22. URL http://www.jstatsoft.org/v33/i02/.

Hermansen, J. S., Saether, 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.

Hesterberg, T. (2015). resample: Resampling Functions. R package version 0.4. https://CRAN.R-project.org/package=resample

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International journal of climatology*, *25*(15), 1965-1978.

Hijmans, R. J. (2015). raster: Geographic Data Analysis and Modeling. R package version 2.5-2. https://CRAN.R-project.org/package=raster

Hobson, K. A. (1999). Tracing origins and migration of wildlife using stable isotopes: a review. *Oecologia*, *120*(3), 314-326.

Hodges, S. A., Burke, J. M., & Arnold, M. L. (1996). Natural formation of Iris hybrids: experimental evidence on the establishment of hybrid zones.*Evolution*, *50*(6), 2504-2509.

Huber, S. K., & Podos, J. (2006). Beak morphology and song features covary in a population of Darwin's finches (Geospiza fortis). *Biological Journal of the Linnean Society*, *88*(3), 489-498.

Jackson, D. A. (1993). Stopping rules in principal components analysis: a comparison of heuristical and statistical approaches. *Ecology*, 2204-2214.

Johnston, R. F. (1969). Taxonomy of house sparrows and their allies in the Mediterranean basin. *The Condor*, *71*(2), 129-139.

Kawecki, T. J., & Ebert, D. (2004). Conceptual issues in local adaptation.*Ecology letters*, *7*(12), 1225-1241.

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

Kimball, R. T. (1996). Female choice for male morphological traits in house sparrows, Passer domesticus. *Ethology*, *102*(4), 639-648.

Kimura, M. (1984). *The neutral theory of molecular evolution*. Cambridge University Press.

Kunte, K., Shea, C., Aardema, M. L., Scriber, J. M., Juenger, T. E., Gilbert, L. E., & Kronforst, M. R. (2011). Sex chromosome mosaicism and hybrid speciation among tiger swallowtail butterflies. *PLoS Genet*, *7*(9), e1002274.

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. and 1000 Genome Project Data Processing Subgroup (2009) The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics*, 25, 2078-9. [PMID: 19505943]

Li H. and Durbin R. (2010) Fast and accurate long-read alignment with Burrows-Wheeler Transform. *Bioinformatics*, Epub. [PMID: 20080505]

Lomolino, M. V. (1985). Body size of mammals on islands: the island rule reexamined. *The American Naturalist*, *125*(2), 310-316.

Losos, J. B. (2009). *Lizards in an evolutionary tree: ecology and adaptive radiation of anoles* (Vol. 10). Univ of California Press.

Losos JB, Arnold SJ, Bejerano G et al., (2013) Evolutionary biology for the 21st century. Plos Biol 11, e1001466.

Macholán, M., Munclinger, P., Šugerková, M., Dufková, P., Bímová, B., Božíková, E., ... & Piálek, J. (2007). GENETIC ANALYSIS OF AUTOSOMAL AND X-LINKED MARKERS ACROSS A MOUSE HYBRID ZONE. *Evolution*, *61*(4), 746-771.

Mallarino, R., Campàs, O., Fritz, J.A., et al., 2012. Closely related bird species demonstrate flexibility between beak morphology and underlying developmental programs. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, pp.16222–16227.

Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, *20*(5), 229-237.

Mavárez, J., Salazar, C. A., Bermingham, E., Salcedo, C., Jiggins, C. D., & Linares, M. (2006). Speciation by hybridization in Heliconius butterflies. *Nature*, *441*(7095), 868-871.

MATLAB and Statistics Toolbox Release 2014a, The MathWorks, Inc., Natick, Massachusetts, United States.

Mavárez, J., & Linares, M. (2008). Homoploid hybrid speciation in animals. *Molecular Ecology*, *17*(19), 4181-4185.

Mayr, E. (1954). Geographic speciation in tropical echinoids. *Evolution*, 1-18.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA, 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *GENOME RESEARCH 20:1297-303*

Meiri, S., & Dayan, T. (2003). On the validity of Bergmann's rule. *Journal of biogeography*, *30*(3), 331-351.

Mitteroecker, P., & Bookstein, F. (2011). Linear discrimination, ordination, and the visualization of selection gradients in modern morphometrics.*Evolutionary Biology*, *38*(1), 100-114.

Nei, M. (2013). *Mutation-driven evolution*. OUP Oxford.

Nielsen, R., & Slatkin, M. (2013). *An introduction to population genetics: theory and applications*. Sunderland, MA: Sinauer Associates.

Peterson, B. J., & Fry, B. (1987). Stable isotopes in ecosystem studies.*Annual review of ecology and systematics*, 293-320.

Pigliucci, M. *Phenotypic Plasticity* (Johns Hopkins University Press, 2001)

Piñeiro, L. (2015). *Beak shape variation in a hybrid species: effects of diet, insularity and species interactions* (Master thesis.). University of Oslo, Oslo.

Price, T. (2008). *Speciation in birds*. Roberts and Co.

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC Area, Cohen's d, and r. *Law and human behavior*, *29*(5), 615.

Rieseberg, L. H., Van Fossen, C., & Desrochers, A. M. (1995). Hybrid speciation accompanied by genomic reorganization in wild sunflowers.*Nature*, *375*(6529), 313-316.

Rieseberg, L. H. (1997). Hybrid origins of plant species. *Annual review of Ecology and Systematics*, 359-389.

Rieseberg, L. H., Archer, M. A., & Wayne, R. K. (1999). Transgressive segregation, adaptation and speciation. *Heredity*, *83*(4), 363-372.

Rieseberg, L. H., Raymond, O., Rosenthal, D. M., Lai, Z., Livingstone, K., Nakazato, T., ... & Lexer, C. (2003). Major ecological transitions in wild sunflowers facilitated by hybridization. *Science*, *301*(5637), 1211-1216.

Roberts, M., & Peters, A. (2009). Is testosterone immunosuppressive in a condition-dependent manner? An experimental test in blue tits. *Journal of Experimental Biology*, *212*(12), 1811-1818.

Rohlf, F. J. (1998). On applications of geometric morphometrics to studies of ontogeny and phylogeny. *Systematic Biology*, *47*(1), 147-158.

Runemark, A., Wellenreuther, M., Jayaweera, H. H., Svanberg, S., & Brydegaard, M. (2012). Rare events in remote dark-field spectroscopy: an ecological case study of insects. *Selected Topics in Quantum Electronics, IEEE Journal of*, *18*(5), 1573-1582.

Scheipl, F., Greven, S. and Kuechenhoff, H. (2008) Size and power of tests for a zero random effect variance or polynomial regression in additive and linear mixed models. Computational Statistics & Data Analysis, 52(7):3283--3299.

Schierenbeck, K. A., & Ellstrand, N. C. (2009). Hybridization and the evolution of invasiveness in plants and other organisms. *Biological invasions*, *11*(5), 1093-1105

Schlager, S. (2016). Morpho: Calculations and Visualisations Related to Geometric Morphometrics. R package version 2.3.1.1. https://CRAN.R-project.org/package=Morpho

Schluter, D. (1996). Adaptive radiation along genetic lines of least resistance. *Evolution*, 1766-1774.

Schluter, D. (2000). *The ecology of adaptive radiation*. OUP Oxford.

Schumer, M., Rosenthal, G. G., & Andolfatto, P. (2014). How common is homoploid hybrid speciation?. *Evolution*, *68*(6), 1553-1560.

Seehausen, O. (2013). Conditions when hybridization might predispose populations for adaptive radiation. *Journal of Evolutionary Biology*, *26*(2), 279-281.

Selander, R. K. (1966). Sexual dimorphism and differential niche utilization in birds. *The Condor*, *68*(2), 113-151.

Selz, O. M., Thommen, R., Maan, M. E., & Seehausen, O. (2014). Behavioural isolation may facilitate homoploid hybrid speciation in cichlid fish. *Journal of evolutionary biology*, *27*(2), 275-289.

Shultz, A. J., & Burns, K. J. (2013). Plumage evolution in relation to light environment in a novel clade of Neotropical tanagers. *Molecular phylogenetics and evolution*, *66*(1), 112-125.

Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2016). afex: Analysis of Factorial Experiments. R package version 0.16-1. https://CRAN.R-project.org/package=afex

Slatkin, M. (1977). Gene flow and genetic drift in a species subject to frequent local extinctions. *Theoretical population biology*, *12*(3), 253-262.

Smith, J. M., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetical research*, *23*(01), 23-35.

Stelkens, R., & Seehausen, O. (2009). Genetic distance between species predicts novel trait expression in their hybrids. *Evolution*, *63*(4), 884-897.

Summers-Smith, J. D. (1963). *The house sparrow* (pp. 149-159). London: Collins.

Svensson, E., & Calsbeek, R. (Eds.). (2012). *The adaptive landscape in evolutionary biology*. OUP Oxford.

Tesaker, L. (2014). *Male plumage variation and its role in reproductive isolation between house sparrows (Passer domesticus) and Italian sparrows (P. italiae) & A new method for quantifying colours of Passer sparrows using digital imaging in the field* (Master thesis.). University of Oslo, Oslo.

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 Genet*, *10*(1), e1004075.

Tufto, J. (2001). Effects of releasing maladapted individuals: a demographic-evolutionary model. *The American Naturalist*, *158*(4), 331-340.

Tukey, J. W. (1949). Comparing individual means in the analysis of variance.*Biometrics*, 99-114.

Uyeda, J. C., Hansen, T. F., Arnold, S. J., & Pienaar, J. (2011). The million-year wait for macroevolutionary bursts. *Proceedings of the National Academy of Sciences*, *108*(38), 15908-15913.

Van der Auwera GA, Carneiro M, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, Banks E, Garimella K, Altshuler D, Gabriel S, DePristo M, 2013. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *CURRENT PROTOCOLS IN BIOINFORMATICS 43:11.10.1-11.10.33*

Vega, U., & Frey, K. J. (1980). Transgressive segregation in inter and intraspecific crosses of barley. *Euphytica*, *29*(3), 585-594.

Weston, S. (2014). Using The foreach Package.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *evolution*, 1358-1370.

Whittingham, M. J., Stephens, P. A., Bradbury, R. B., & Freckleton, R. P. (2006). Why do we still use stepwise modelling in ecology and behaviour?.*Journal of animal ecology*, *75*(5), 1182-1189.

Wikelski, M., Spinney, L., Schelsky, W., Scheuerlein, A., & Gwinner, E. (2003). Slow pace of life in tropical sedentary birds: a common-garden experiment on four stonechat populations from different latitudes.*Proceedings of the Royal Society of London B: Biological Sciences*,*270*(1531), 2383-2388.

Xu, R. (2003). Measuring explained variation in linear mixed effects models.*Statistics in medicine*, *22*(22), 3527-3541.

Zelditch, M. L., Swiderski, D. L., & Sheets, H. D. (2012). *Geometric morphometrics for biologists: a primer*. Academic Press.

Zeugner, S., Feldkircher, M. (2015). Bayesian Model Averaging Employing Fixed and Flexible Priors: The BMS Package for R. Journal of Statistical Software, 68(4), 1-37. doi:10.18637/jss.v068.i04

List of abbreviations

AIC$_C$: Akaike information criterion (corrected for finite sample size)

ANOVA: analysis of variance

AP: annual precipitation

AT: annual temperature

BMS: Bayesian model selection

CVA: canonical variance analysis

DFA: discriminant function analysis

DNA: deoxyribonucleic acid

F: Fisher – Snedecor distribution

***F*$_{ST}$**: fixation index

LDA: linear discriminant function analysis

LMEM: linear mixed effects models

Log-lik: logarithmic likelihood

MANOVA: multivariate analysis of variance

MCMC: marcov chain monte carlo

ML: maximum likelihood

OLS: ordinary least squares

PB: parametric bootstrap

PC: principal component

PIP: prior inclusion probability

PMP: posterior model probability

***P*$_{ST}$**: quantitative fixation index, estimate

REML: restricted likelihood

RGB: red, green, blue

RW: relative warp

R^2: R-squared

SNP: single nucleotide polymorphism

SP: precipitation seasonality

SS: sums of squares

ST: temperature seasonality

SVD: singular value decomposition

TPS: thin – plate spline

λ = Wilk's lambda

χ^2 = Chi-square

Ω^2 = Omega squared

t = Student's t-test

V = Pillai-Bartlett trace

c = additive genetic effects between populations

h^2 = additive genetic effects within populations

$\delta^{13}C$ = stable environmental isotope of carbon

$\delta^{15}N$ = stable environmental isotope of nitrogen

β = standardized regression coefficient, or effect size

Appendix I

Supplementary figures

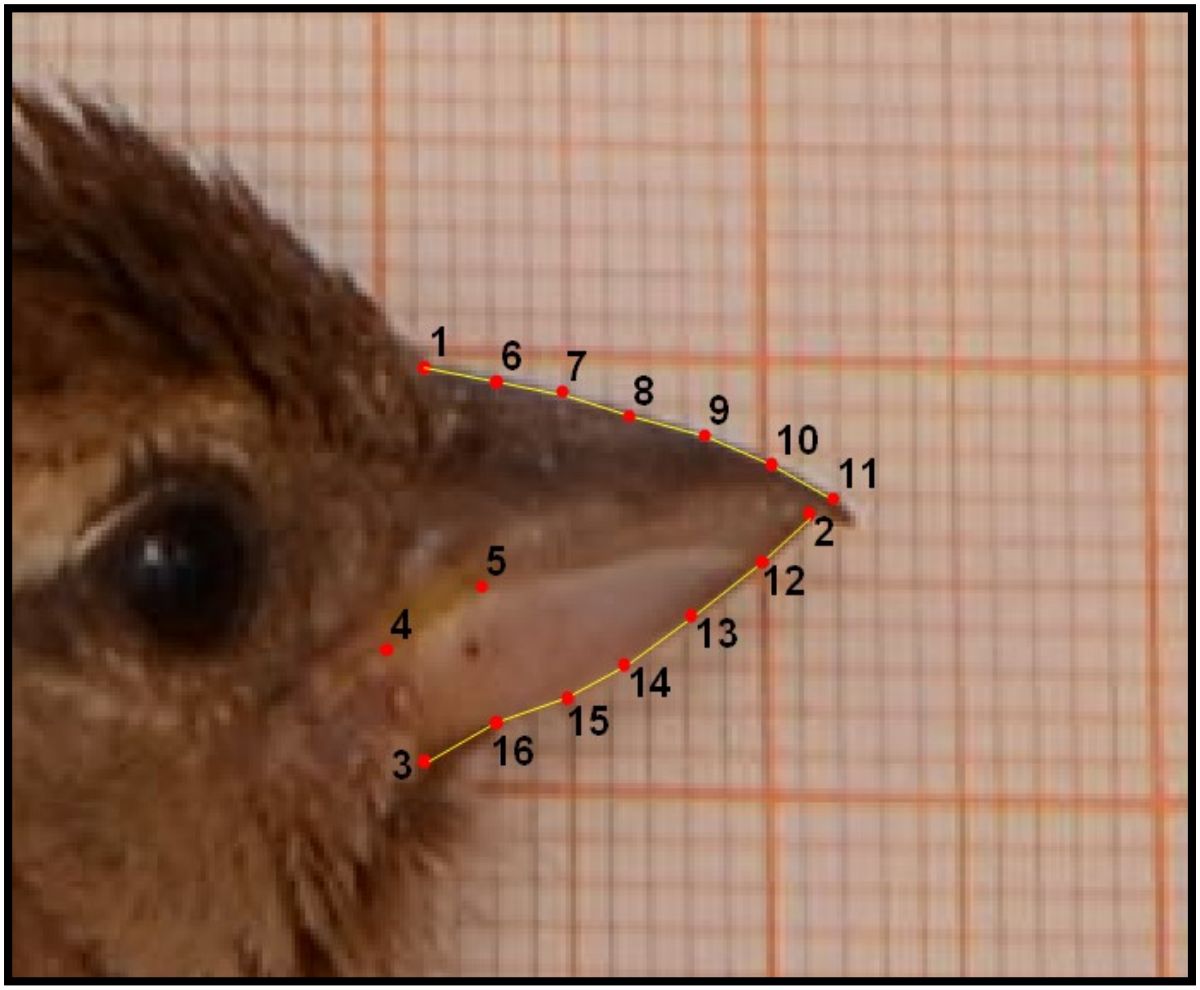

Figure S1. Example of digitization of beak shape (*P.italiae*, female). Shows landmarks (1-5) and 7 equidistant semi-landmarks (6-16). The yellow line is a "best-fit" curve for the upper and lower mandible, respectively. (The individuals were placed on millimetre paper for comparison, i.e. one square enclosed by bold lines equals 1 cm^2).

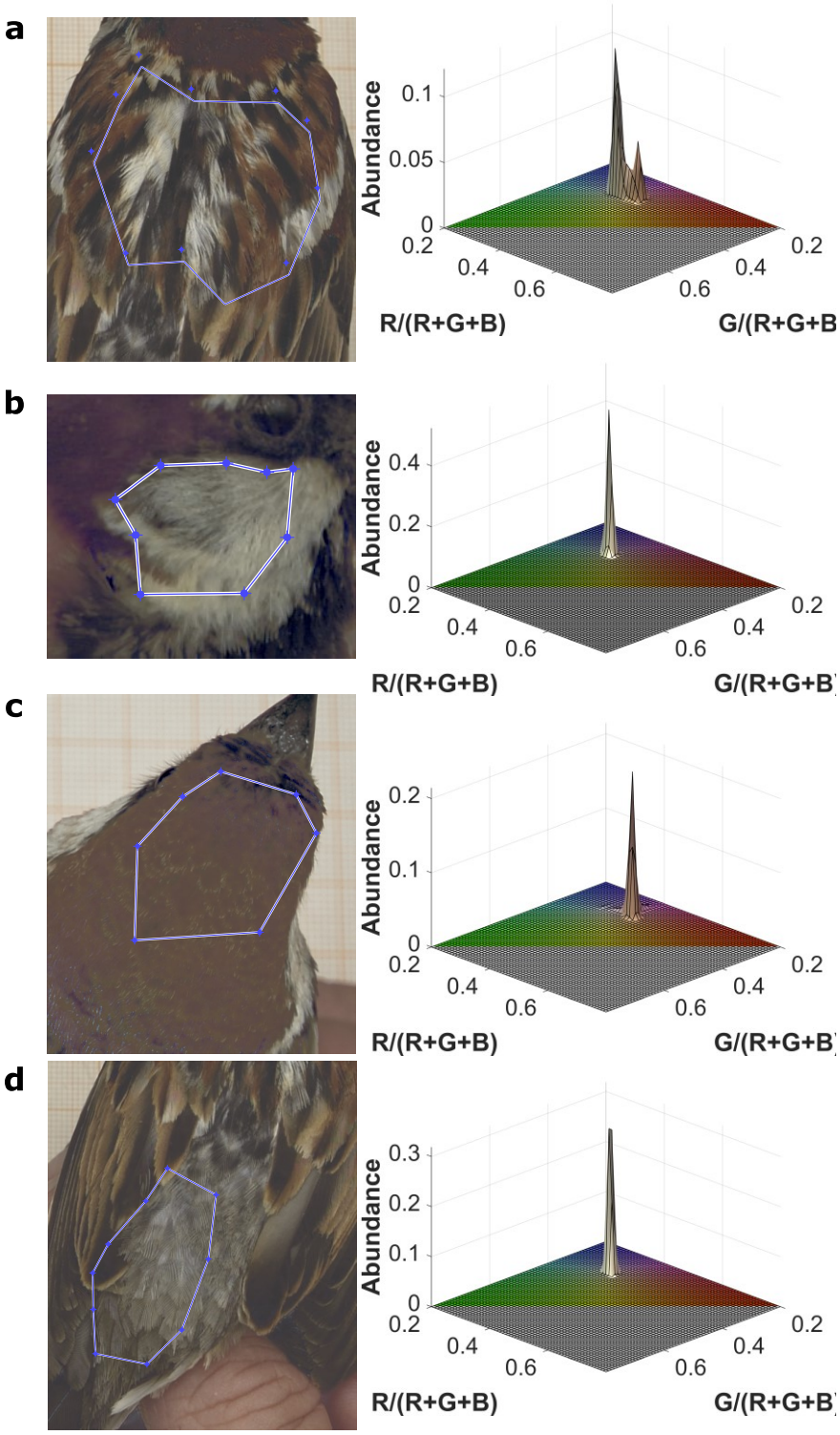

Figure S2. Examples of plumage colour quantification for each trait: (**a**) back colour, (**b**) cheek colour, (**c**) crown colour, and (**d**) rump colour. The encircled regions in the pictures to the left are represented by the abundance of different colour in the chromaticity histograms to the right.

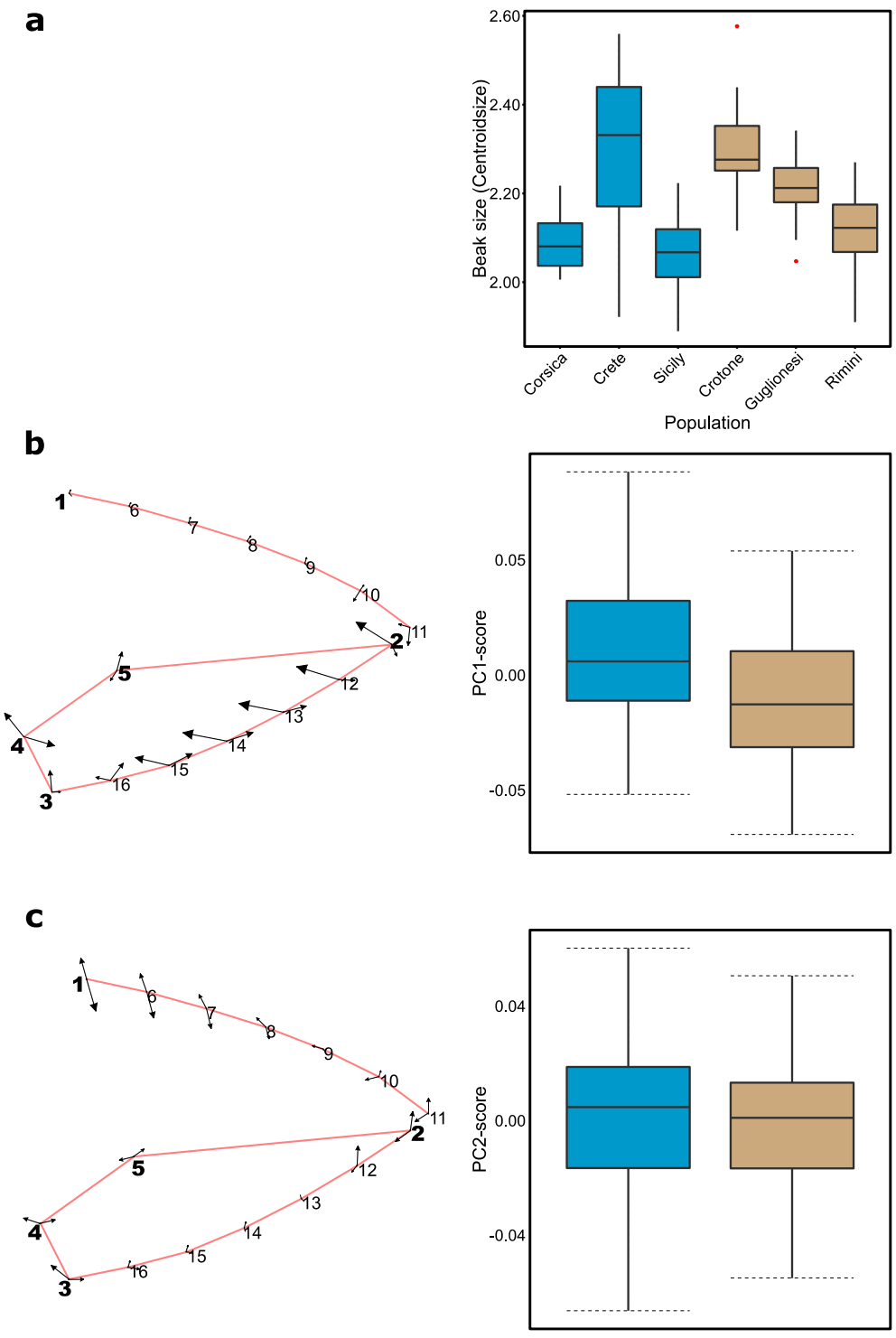

Figure S3. (**a**) The amount of variation in beak size estimates derived from centroid sizes for each population. (**b**)-(**c**): Examples of how the beak changes along two axis of variation (i.e. PC1 and 2, respectively). Variation in principal component score for RW1/PC1 (**b**) and PC2 (**c**), for all individuals within island (blue bars) and mainland (brown bars). The beak outlines to the left depicts each landmark (bold) and semi-landmarks position with vectors representing the result of movement along the respective principal component axis, from one extreme to the other. (note: PCs 3-5 are not depicted).

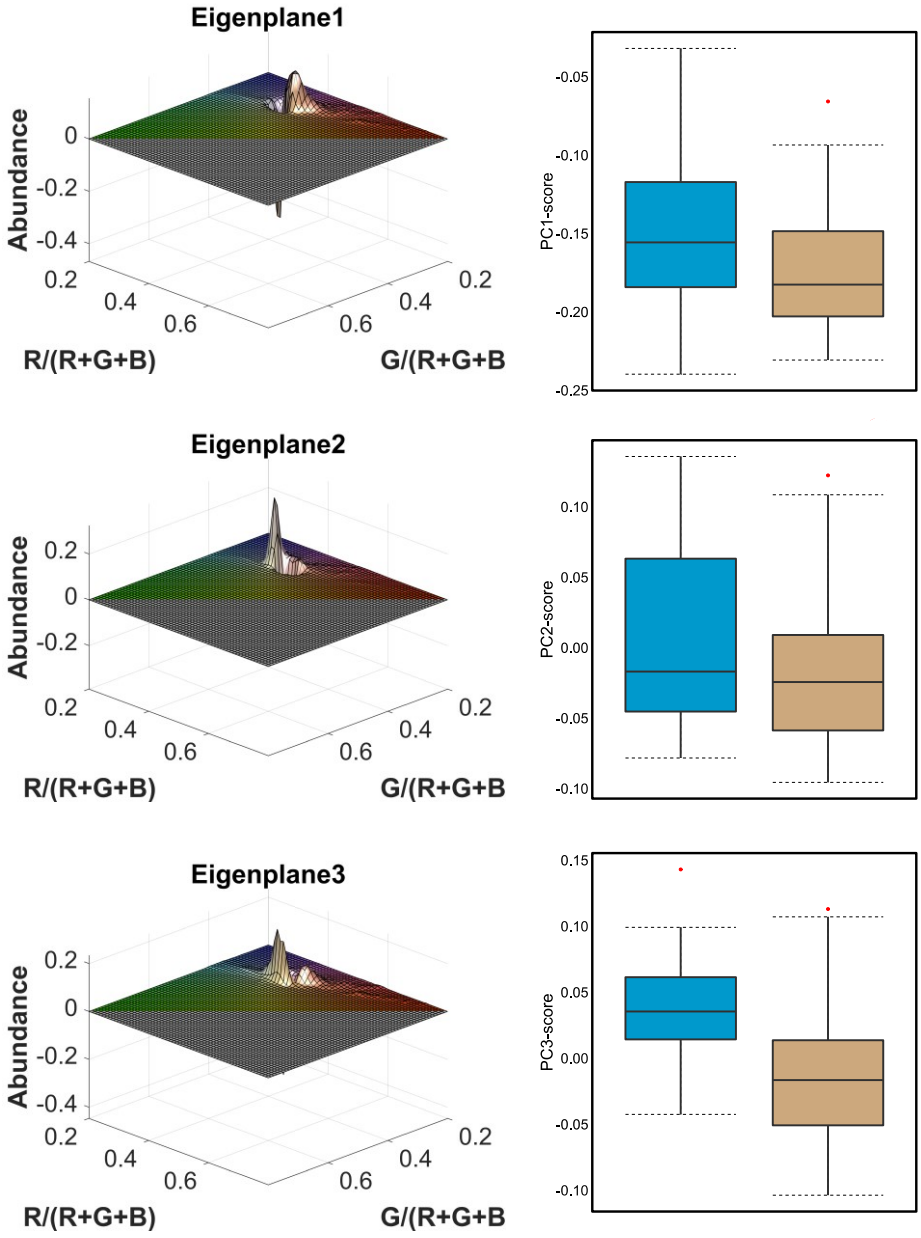

Figure S4. Back colour chromaticity histograms (left) for eigenplanes 1-3 from the SVD of RGB-data derived from plumage colour digitization. (right) Variation between island (blue bars) and mainland (brown bars) individuals in principle component 1-3.

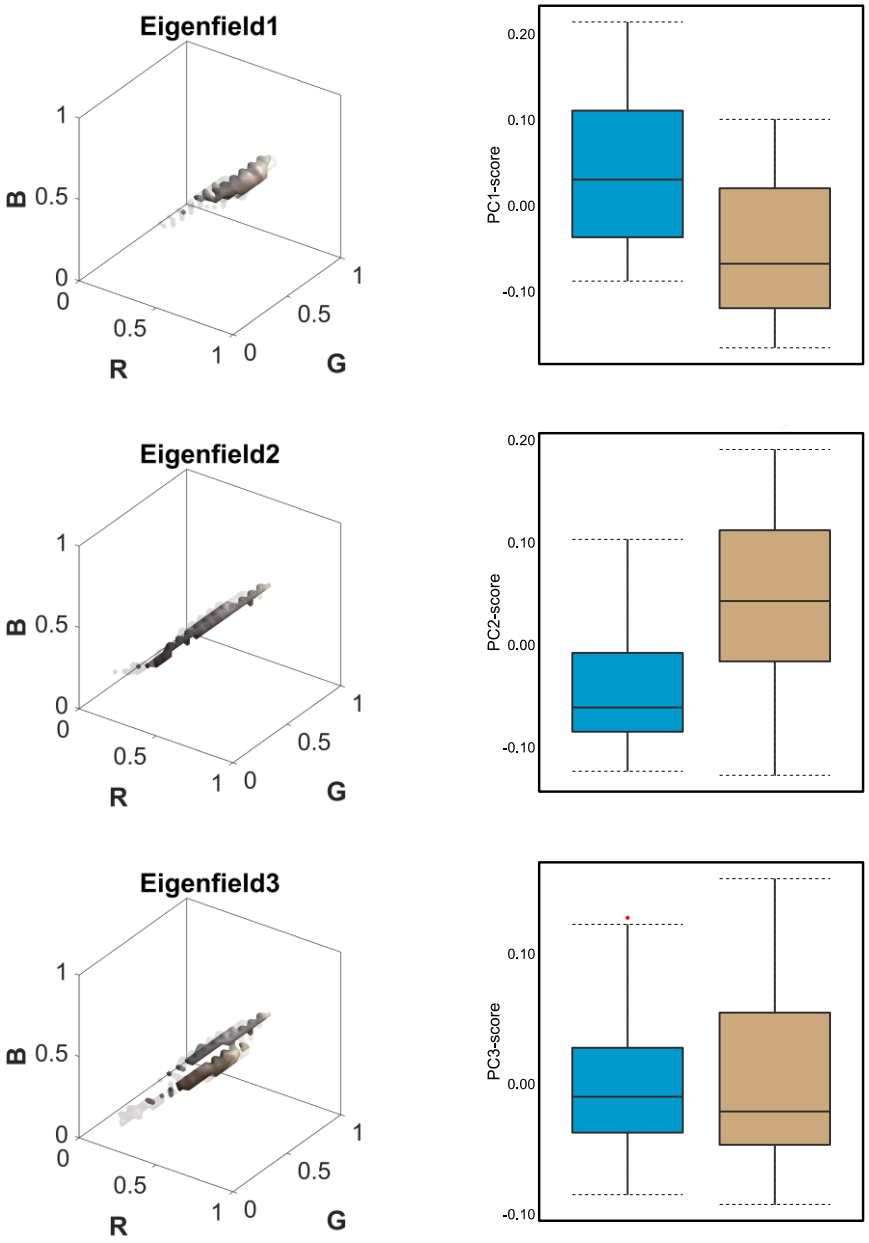

Figure S5. Cheek colour chromaticity histograms (left) for eigenfields 1-3 from the SVD of RGB-data derived from plumage colour digitization. (right) Variation between island (blue bars) and mainland (brown bars) individuals in principle component 1-3.

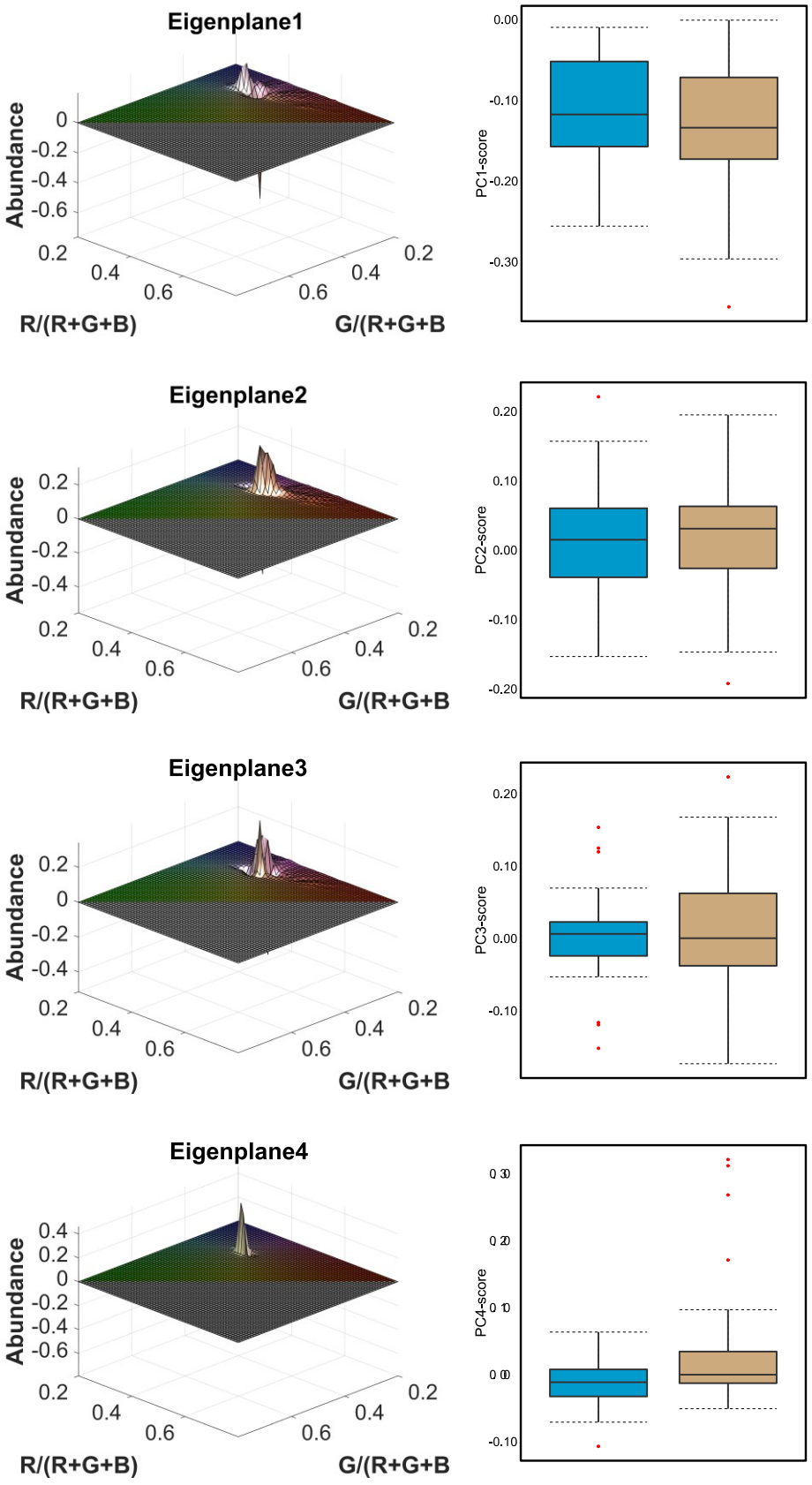

Figure S6. Crown colour chromaticity histograms (left) for eigenplanes 1-4 from the SVD of RGB-data derived from plumage colour digitization. (right) Variation between island (blue bars) and mainland (brown bars) individuals in principle component 1-4.

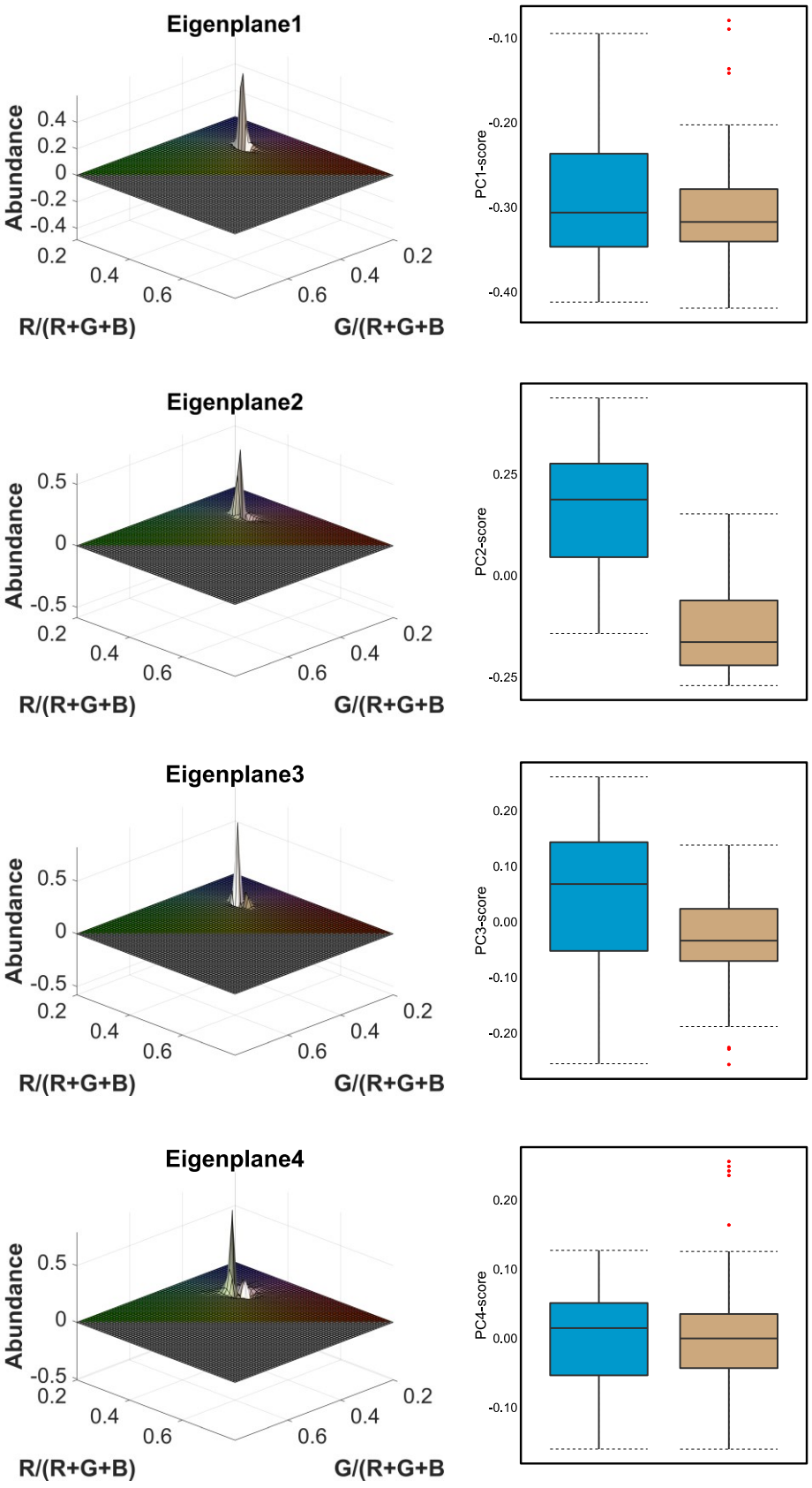

Figure S7. Rump colour chromaticity histograms (left) for eigenplanes 1-4 from the SVD of RGB-data derived from plumage colour digitization. (right) Variation between island (blue bars) and mainland (brown bars) individuals in principle component 1-4.

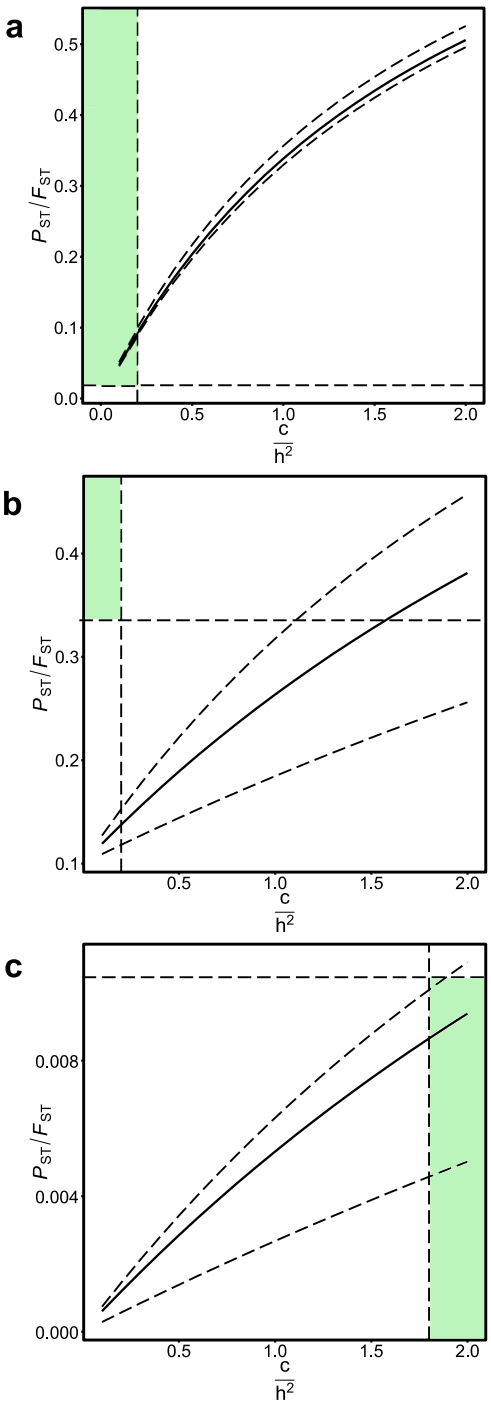

Figure S8. Examples of three different scenarios from the P_{ST}/F_{ST} analysis. Green areas are conservative regions (i.e. a c/h^2 ratio of < .3 and > 1.7) and the solid lines is the P_{ST} value with the 'robust' estimate expanded by .95 confidence limits. The horizontal stapled line is the F_{ST} estimate. (**a**) Evidence for divergent selection. The robust P_{ST} estimate is in the green are and does not overlap F_{ST} when the c/h^2 -ratio is <. 3. (**b**) Possible stabilizing selection, but the robust estimate crosses F_{ST} at a c/h^2 ratio of approx. 1. (**c**) Evidence for stabilizing selection as the upper P_{ST} confidence interval is in the green area before it crosses F_{ST}.

Appendix II

Supplementary tables

Table S1. Number of individuals (n) digitized from each location (i.e. populations) of *Passer* sparrows. Check-mark indicate that the type of data has been collected, and the 'line' indicate that no data collected.

Location	Species	Group	n[1]	Beak	Plumage	Isotopes	Climate
Corsica	*P.italiae*	islands	30(13)	✓	✓	✓	✓
Crete	*P.italiae*	islands	30(14)	✓	✓	✓	✓
Crotone	*P.italiae*	mainland	30(13)	✓	✓	✓	✓
Guglionesi	*P.italiae*	mainland	30(15)	✓	✓	✓	✓
Manfredonia	*P.hispaniolensis*	mainland	50(13)	✓	✓	-	-
Rimini	*P.italiae*	mainland	25(12)	✓	✓	✓	✓
Sales	*P.domesticus*	mainland	50(25)	✓	✓	-	-
Sicily	*P.italiae*	islands	30(25)	✓	✓	✓	✓

1 Total number of individuals and males in parentheses

Table S2. The number of sampled principle components for each trait, and the number of selected principle components by the broken-stick distribution criteria. Shows the explained variation for each of the selected components and the cumulative variation from all the selected principal components.

Trait	Principal components	Selected Principal components	explained variance	Cum.%
Beak shape	28	1	34%	34%
		2	24%	57%
		3	15%	73%
		4	8%	80%
		5	7%	87%
Back colour	89	1	68%	68%
		2	9%	77%
		3	7%	84%
Cheek colour	89	1	62%	62%
		2	20%	82%
		3	7%	89%
Crown colour	89	1	42%	42%
		2	10%	51%
		3	9%	60%
		4	7%	67%
Rump colour	89	1	53%	53%
		2	22%	76%
		3	9%	85%
		4	4%	89%

Table S3. MANOVA (multivariate analysis of variance) on each multivariate trait, using principle components selected by the broken-stick distribution. *P*-values below .001 are not displayed precisely.

Trait	Group	Df[1]	V[2]	F[3]	num. df[4]	den. df[5]	*p*[6]
Beak shape	island	2	0.49	5.38	10	168	< 0.001*
	mainland	2	0.41	3.91	10	150	< 0.001*
Back colour	island	2	1.04	12.90	6	72	< 0.001*
	mainland	2	0.69	6.44	6	74	< 0.001*
Cheek colour	island	2	0.14	0.91	6	72	0.491
	mainland	2	0.72	6.93	6	74	< 0.001*
Crown colour	island	2	0.93	7.66	8	70	< 0.001*
	mainland	2	0.94	7.99	8	72	< 0.001*
Rump colour	island	2	0.75	5.27	8	70	<0.001*
	mainland	2	0.63	4.17	8	72	<0.001*

[1]Degrees of freedom
[2]Pillai-Bartlett trace
[3]Approximate *F*-value
[4]Numerator degrees of freedom for the *F*-value calculation
[5]Denominator degrees of freedom for the *F*-value calculation
[6]*p*-value with $\alpha = .05$

Table S4. ANOVA (analysis of variance) on all univariate traits, or trait components from a significant MANOVA. The trait component is shown in parentheses (i.e. principal component, PC).

Trait	Group	Df[1]	SS[2]	Mean SS[3]	F[4]	*p*[5]
Beak size	island	2	1.0379	0.5189	41.69	< 0.001*
	mainland	2	0.4431	0.2216	31.98	< 0.001*
Beak shape (PC1)	island	2	0.0162	0.0081	11.70	< 0.001*
	mainland	2	0.0042	0.0021	2.98	0.057
Beak shape (PC2)	island	2	0.0132	0.0066	11.79	< 0.001*
	mainland	2	0.0047	0.0024	5.63	0.005*
Beak shape (PC3)	island	2	0.0027	0.0014	3.16	0.047
	mainland	2	0.0008	0.0004	1.07	0.347
Beak shape (PC4)	island	2	0.0008	0.0004	1.97	0.145
	mainland	2	0.0003	0.0001	0.67	0.515
Beak shape (PC5)	island	2	0.0009	0.0005	2.07	0.132
	mainland	2	0.0029	0.0015	13.89	< 0.001*
Back colour (PC1)	island	2	0.0449	0.0225	13.63	< 0.001*
	mainland	2	0.0438	0.0219	30.63	< 0.001*
Back colour (PC2)	island	2	0.1033	0.0517	31.14	< 0.001*
	mainland	2	0.0153	0.0077	2.64	0.085

Back colour (PC3)	island	2	0.0140	0.0070	6.24	0.005*
	mainland	2	0.0264	0.0132	5.63	0.007*
Cheek colour (PC1)	mainland	2	0.1054	0.0527	17.10	< 0.001*
Cheek colour (PC2)	mainland	2	0.0019	0.0010	0.15	0.859
Cheek colour (PC3)	mainland	2	0.0829	0.0414	18.12	< 0.001*
Crown colour (PC1)	island	2	0.0365	0.0183	4.85	0.014
	mainland	2	0.1368	0.0684	16.26	< 0.001*
Crown colour (PC2)	island	2	0.1156	0.0578	15.96	< 0.001*
	mainland	2	0.0189	0.0095	1.50	0.237
Crown colour (PC3)	island	2	0.0364	0.0182	6.89	0.003*
	mainland	2	0.0794	0.0397	6.36	0.004*
Crown colour (PC4)	island	2	0.0264	0.0132	23.20	< 0.001*
	mainland	2	0.1064	0.0532	9.53	< 0.001*
Rump colour (PC1)	island	2	0.0279	0.0140	2.43	0.102
	mainland	2	0.0920	0.0460	11.33	< 0.001*
Rump colour (PC2)	island	2	0.5398	0.2699	22.19	< 0.001*
	mainland	2	0.1333	0.0666	7.78	0.002*
Rump colour (PC3)	island	2	0.0500	0.0250	1.19	0.316
	mainland	2	0.1039	0.0519	9.07	0.001*
Rump colour (PC4)	island	2	0.0032	0.0016	0.28	0.761
	mainland	2	0.0933	0.0467	5.54	0.008*

[1]Degrees of freedom (n -1)

[2]Type I calculation of sums of squares

[3]Mean sums of squares

[4]Calculated from *F*-distribution

[5]$\alpha = .05$, adjusted according to the Dunn-Sidak principle

Table S5. Post-hoc Tukey's on each significant univariate trait or trait component from the ANOVA. Shows each population pair within the island or mainland group, their difference in mean and if this is statistically reliable (* = significant *p*-value).

Trait	Trait component	Group	Population pair	diff.[1]	*p*[2]
Beak size	Centroid size	island	Crete-Corsica	0.216	< 0.001*
"	Centroid size	island	Sicily-Corsica	-0.022	0.956
"	Centroid size	island	Sicily-Crete	-0.238	< 0.001*
"	Centroid size	mainland	Guglionesi-Crotone	-0.089	0.010*
"	Centroid size	mainland	Rimini-Crotone	-0.186	< 0.001*
"	Centroid size	mainland	Rimini-Guglionesi	-0.096	0.008*
Beak shape	PC1	island	Crete-Corsica	0.007	0.578
"	PC1	island	Sicily-Corsica	-0.024	0.002*
"	PC1	island	Sicily-Crete	-0.031	< 0.001*
"	PC2	island	Crete-Corsica	-0.004	0.984
"	PC2	island	Sicily-Corsica	0.024	0.001*
"	PC2	island	Sicily-Crete	0.027	< 0.001*
"	PC2	mainland	Guglionesi-Crotone	-0.001	0.999
"	PC2	mainland	Rimini-Crotone	0.017	0.088
"	PC2	mainland	Rimini-Guglionesi	0.017	0.064
"	PC5	mainland	Guglionesi-Crotone	0.011	< 0.001*
"	PC5	mainland	Rimini-Crotone	-0.002	0.737
"	PC5	mainland	Rimini-Guglionesi	-0.014	< 0.001*
Back colour	PC1	island	Crete-Corsica	0.061	<0.001*
"	PC1	island	Sicily-Corsica	-0.019	0.700
"	PC1	island	Sicily-Crete	-0.08	< 0.001*
"	PC1	mainland	Guglionesi-Crotone	0.021	0.579
"	PC1	mainland	Rimini-Crotone	0.078	< 0.001*
"	PC1	mainland	Rimini-Guglionesi	0.057	0.001*
"	PC2	island	Crete-Corsica	0.02	0.413
"	PC2	island	Sicily-Corsica	0.119	< 0.001*
"	PC2	island	Sicily-Crete	0.099	< 0.001*
"	PC2	island	Crete-Corsica	0.004	0.999
"	PC2	island	Sicily-Corsica	-0.039	0.171
"	PC2	island	Sicily-Crete	-0.043	0.120
"	PC2	mainland	Guglionesi-Crotone	0.025	0.617
"	PC2	mainland	Rimini-Crotone	0.062	0.003*
"	PC2	mainland	Rimini-Guglionesi	0.037	0.195
Cheek color	PC1	mainland	Guglionesi-Crotone	-0.05	0.059
"	PC1	mainland	Rimini-Crotone	0.075	0.003*
"	PC1	mainland	Rimini-Guglionesi	0.124	< 0.001*
"	PC2	mainland	Guglionesi-Crotone	-0.054	0.013*
"	PC2	mainland	Rimini-Crotone	0.057	0.011*
"	PC2	mainland	Rimini-Guglionesi	0.111	< 0.001*
Crown color	PC1	mainland	Guglionesi-Crotone	0.007	0.961
"	PC1	mainland	Rimini-Crotone	0.127	< 0.001*
"	PC1	mainland	Rimini-Guglionesi	0.121	< 0.001*
"	PC2	island	Crete-Corsica	-0.121	< 0.001*
"	PC2	island	Sicily-Corsica	-0.016	0.762
"	PC2	island	Sicily-Crete	0.105	< 0.001*
"	PC2	island	Crete-Corsica	-0.056	0.248
"	PC2	island	Sicily-Corsica	-0.068	0.105
"	PC2	island	Sicily-Crete	-0.012	0.998
"	PC2	mainland	Guglionesi-Crotone	-0.106	0.001*
"	PC2	mainland	Rimini-Crotone	-0.058	0.229
"	PC2	mainland	Rimini-Guglionesi	0.049	0.419
"	PC5	island	Crete-Corsica	0.059	0.069
"	PC5	island	Sicily-Corsica	0.012	0.994

"	PC5	island	Sicily-Crete	-0.047	0.286
"	PC5	mainland	Guglionesi-Crotone	0.019	0.950
"	PC5	mainland	Rimini-Crotone	0.117	< 0.001*
"	PC5	mainland	Rimini-Guglionesi	0.099	< 0.001*
Rump colour	PC1	mainland	Guglionesi-Crotone	0.029	0.463
"	PC1	mainland	Rimini-Crotone	0.113	< 0.001*
"	PC1	mainland	Rimini-Guglionesi	0.084	0.004*
"	PC2	island	Crete-Corsica	0.050	0.781
"	PC2	island	Sicily-Corsica	-0.226	< 0.001*
"	PC2	island	Sicily-Crete	-0.276	< 0.001*
"	PC2	mainland	Guglionesi-Crotone	0.047	0.818
"	PC2	mainland	Rimini-Crotone	0.139	0.009*
"	PC2	mainland	Rimini-Guglionesi	0.091	0.194
"	PC2	mainland	Guglionesi-Crotone	-0.079	0.024*
"	PC2	mainland	Rimini-Crotone	-0.122	< 0.001*
"	PC2	mainland	Rimini-Guglionesi	-0.043	0.314
"	PC5	mainland	Guglionesi-Crotone	0.059	0.215
"	PC5	mainland	Rimini-Crotone	0.118	0.005*
"	PC5	mainland	Rimini-Guglionesi	0.058	0.237

[1]Difference in mean

[2]α = .05, corrected for family-wise error

Table S6. Genome wide mean (weighted) F_{ST} estimate of genetic differentiation calculated from 30 genes between each population within the island and mainland group, respectively. There were thousands of base pairs in each locus, and the confidence interval did not affect the F_{ST} value beyond the 3rd decimal. Hence, it is not shown.

Group	Population pair	F_{ST}
islands	Crete-Corsica	0.059
islands	Crete-Sicily	0.082
islands	Corsica-Sicily	0.033
mainland	Crotone-Guglionesi	0.009
mainland	Crotone-Rimini	0.019
mainland	Guglionesi-Rimini	0.012

Table S7. The best model that explain variation between island and mainland individuals, respectively, for each trait. Selected by an AIC_C criteria (i.e. lowest AIC_C value) after a step-wise and Bayesian model selection scheme, divided by the main columns: PC 1 (*axis of maximum variation*) and Euclidean distances (*multidimensional divergence*). The value for independent variables are standardized effect sizes (β) or the Chi-square or *F*-value for models with a sole fixed effect. The value for the random variables is the *p*-score from a logarithmic likelihood simulation. All significant values are marked with '*' (α = .05, corrected for family-wise error, and the precise *p*-value is not shown).

	Axis of maximum variation		*Multidimensional divergence*	
	Island	**Mainland**	**Island**	**Mainland**
Beak size				
AIC_C		-167.42	-	-
PMP†	0.16 (*0.86*)	-	-	-
R^2/Ω^2	0.58*	0.44*	-	-
Independent variables				
AP	0.210*	-	-	-
$\delta^{13}C$ †	-0.005*	-	-	-
Sex	0.011*	-	-	-
Population	-	31.980*	-	-
AP**:**C		-	-	-
Sex**:**C	-0.147*	-	-	-
Random variables				
Population	0.820	-	-	-
Beak shape				
AIC_C	-394.02	-325.42	-12884.95	-10701.10
PMP†‡	-	-	< 0.01 (*0.37*)	< 0.01 (*0.57*)
R^2/Ω^2	0.21*	0.06	0.12*	0.05
Independent variables				
AP	-	9.210*	0.879	0.251
AT‡	-	-	-0.699	0.010
$\delta^{13}C$	-	-	0.263*	0.089*
$\delta^{15}N$ ‡	-	-	0.172*	0.144*
Population	11.700*	-	-	-
Sex†‡	-	-	0.035	-0.018*
AT**:**$\delta^{13}C$	-	-	0.494*	-
AT**:**Sex‡	-	-	-0.025*	0.009*
AP**:**$\delta^{13}C$	-	-	0.779*	-
AP**:**$\delta^{15}N$ ‡	-	-	-	-0.192*
Sex**:**$\delta^{15}N$ †	-	-	0.013*	-
Random variables				
Population	-	0.100	< 0.001*	0.085
Back colour				
AIC_C	-104.00	-153.00	-1858.90	-1833.30
PMP‡	0.67	-	0.599	< 0.01 (*0.46*)
R^2/Ω^2	0.50*	0.62*	0.28*	0.21*
Independent variables				
AP	0.653*	-	-	-0.473*
AT	-	-	0.436*	-
$\delta^{13}C$	-	-	0.260*	-0.061

$\delta^{15}N$	0.273*	-	0.198*	-
Population	-	-	-	-
ST‡	-	12.221*	0.679*	0.703
AP:$\delta^{13}C$ ‡	-	-	-	0.788*
$\delta^{13}C$:$\delta^{15}N$	-	-	-0.476*	-
ST:$\delta^{13}C$	-	-	-	-0.509
Random variables				
Population	< 0.001*	< 0.001*	< 0.001*	< 0.001*
Cheek colour				
AIC_C	-71.71	-114.74	-12884.90	-1383.10
PMP†	-	-	0.40 (0.55)	-
R^2/Ω^2	0.04	0.46*	0.12*	0.06*
Independent variables				
$\delta^{15}N$	-	-	0.151*	-
Population	0.760	17.100*	-	17.170*
SP	-	-	0.322	-
ST	-	-	-0.341*	-
SP:$\delta^{15}N$ †	-	-	0.102*	-
Random variables				
Population	-	-	< 0.001*	-
Crown colour				
AIC_C	-79.81	-78.30	-1078.10	-1104.60
PMP	-	0.64	-	0.83
R^2/Ω^2	0.21*	0.63*	0.14*	0.32*
Independent variables				
AP	-	0.854*	-	-
$\delta^{13}C$	-	-0.323*	-	0.330*
N		-	-	0.211*
Population	9.320*	-	75.280*	-
Random variables				
Bushy	0.156	0.004*	0.115	< 0.001*
Population	-	0.222	-	< 0.001*
Rump colour				
AIC_C	-59.70	-79.20	-689.12	-871.70
PMP†	0.36	0.38	0.18 (0.77)	0.58
R^2/Ω^2	0.27*	0.45*	0.12*	0.24*
Independent variables				
AP	0.336*	-	0.863	-
AT	-	-	-0.992	-
$\delta^{13}C$ †	-	-0.291*	0.250	-
$\delta^{15}N$	0.398*	-	-0.381*	-0.103*
ST	-	0.704*	-	0.488*
AP:$\delta^{13}C$	-	-	-0.503*	-
AT:$\delta^{15}N$	-	-	0.436*	-
Random variables				
Population	0.089	0.082	< 0.001*	0.076

†/‡ Indicates a better Bayesian model estimation by removing this variable, with the new PMP shown in parentheses. For island (†) and mainland (‡), respectively

Table S8. Models that were not significantly worse than the best model in explaining variation between individuals within island and mainland (i.e. within an AIC_C of 2). Each column represents and independent variable with an associated '+' if it is included in a model. Bold letters signifies the best model AIC_C score (ML-fit) with its independent variables. Both models for the axis of maximum variation (**a**) and multidimensional distances (**b**) are shown. Missing traits indicate that there were no models within an AIC_C of 2 from the best model (*Note 1: the AIC_C of the best model in Table S7 might have a different value than the one listed here due to the REML re-fit after model selection. Note 2: $\delta^{15}N$ is depicted as 'N' and $\delta^{13}C$ is depicted as 'C'*)

a. Axis of maximum variation

Beak size, island

	AP	AT	**C**	N	**Sex**	**AP:C**	AT:C	AT:N	**Sex:C**	SP	ST:N	AIC_C
	+		+	+	+	+			+			-143.64
		+	+	+	+		+	+	+			-143.55
	+	+	+	+	+	+		+	+			-142.10
	+		**+**		**+**	**+**			**+**			**-142.03**
	+		+	+	+	+			+	+	+	-141.94

Beak size, mainland

Pop	AP	AT	C	N	Sex	AP:C	AT:C	C:N	ST	ST:C	SP	SP:C	AIC_C
+													**-167.42**
			+	+	+				+	+			-167.13
			+		+				+	+			-166.96
			+						+	+			-166.71
	+				+				+				-166.61
			+	+					+	+			-166.59
	+		+		+				+	+			-166.20
	+		+		+						+	+	-166.18
	+		+		+	+			+				-165.93
	+		+						+	+			-165.86
	+		+			+			+				-165.63
			+						+		+	+	-165.51
	+	+	+				+						-165.50
					+				+				-165.47
			+	+	+			+	+	+			-165.45

Beak shape, island

	Pop	AP	AT	C	N	Sex	Sex:N	ST	SP	Sex:SP	AIC_C
		+			+	+	+		+	+	-394.40
	+										**-394.02**
		+		+	+	+	+		+	+	-393.04
		+			+	+			+	+	-392.66
		+				+			+	+	-392.50

Beak shape, mainland

					Pop	**AP**	N	ST	AIC_C
					+				-351.73
						+			**-350.95**
						+	+		-350.34
								+	-349.95

Back colour, island

				AP	**N**	ST	AIC_C
				+	**+**		**-174.64**
				+	+	+	-174.37
					+	+	-174.16

Cheek colour, island

					Pop	C	ST	SP	AIC_C
					+				**-71.71**
					+		+		-71.66
					+			+	-71.63
						+		+	-71.57
					+	+			-70.68

Cheek colour, mainland

Pop	AP	AT	C	N	AT:N	ST	SP	N:SP	AIC_C
+									**-114.74**
		+		+	+		+		-113.89
		+		+			+	+	-113.88
		+	+	+	+	+			-113.60
		+	+	+			+	+	-113.57

Rump colour, island

					AP	AT	**N**	AIC_C
					+		**+**	**-91.69**
						+	+	-90.97

Rump colour, mainland

	AP	AT	**C**	AP:C	AT:C	**ST**	C:ST	AIC_C
			+			+	+	-107.54
	+		+	+				-106.45
	+	+	+		+			-106.15
			+			**+**		**-106.08**

b. Multidimensional distances

Beak shape, mainland

AP	**AT**	**C**	**N**	**Sex**	**AP:N**	**Sex:AT**	SP	Sex:SP	AIC_C
+	+	+	+	+	+	+			**-10852.3**
+		+	+	+	+		+	+	-10852.2

Back colour, mainland

AP	**C**	N	**AP:C**	AP:N	**ST**	**ST:C**	ST:N	AIC_C
+	+	+	+		+	+		-1908.72
+	**+**		**+**		**+**	**+**		**-1907.81**
+	+	+	+		+	+	+	-1907.38
+	+	+	+	+	+	+		-1907.01

Cheek colour, mainland

Pop	AP	N	ST	AIC_C
	+	+	+	-1105.38
+				**-1104.23**

Crown colour, mainland

	AP	AT	**C**	**N**	AT:N	ST	SP	SP:N	AIC_C
		+	+	+	+	+			-1132.07
		+	+	+			+	+	-1132.06
		+	+	+		+			-1131.85
		+	+	+	+				-1131.78
			+	+			+	+	-1131.74
			+	**+**					**-1131.07**
		+	+	+					-1131.05
			+	+			+		-1131.01
	+		+	+					-1130.79
		+	+	+	+		+	+	-1130.24

Rump colour, island

AP	**AT**	**C**	**N**	**AP:C**	AT:C	**AT:N**	C:N	SP	SP:N	AIC_C
+	**+**	**+**	**+**	**+**		**+**				**-770.24**
+	+	+	+	+	+	+				-770.12
+		+	+	+				+	+	-769.05
+	+	+	+	+		+	+			-768.85

Rump colour, mainland

			AP	AT	C	**N**	**ST**	SP	AIC_C
						+	**+**		**-902.11**
					+	+	+		-900.99
			+			+	+		-900.19
				+		+		+	-900.19
				+		+	+		-900.18
			+			+		+	-900.18
						+	+	+	-900.18