Hybrid speciation through sorting of parental incompatibilities in Italian sparrows

JO S. HERMANSEN,* FREDRIK HAAS,*† CASSANDRA N. TRIER,*† RICHARD I. BAILEY,* ALEXANDER J. NEDERBRAGT,* ALFONSO MARZAL† and GLENN-PETER SÆTRE*

*Department of Biosciences, Centre for Ecological and Evolutionary Synthesis, University of Oslo, PO Box 1066, Blindern, N-0316 Oslo, Norway, †Department of Zoology, University of Extremadura, ES-06071 Badajoz, Spain

Abstract

Speciation by hybridization is emerging as a significant contributor to biological diversification. Yet, little is known about the relative contributions of (i) evolutionary novelty and (ii) sorting of pre-existing parental incompatibilities to the build-up of reproductive isolation under this mode of speciation. Few studies have addressed empirically whether hybrid animal taxa are intrinsically isolated from their parents, and no study has so far investigated by which of the two aforementioned routes intrinsic barriers evolve. Here, we show that sorting of pre-existing parental incompatibilities contributes to intrinsic isolation of a hybrid animal taxon. Using a genomic cline framework, we demonstrate that the sex-linked and mitonuclear incompatibilities isolating the homoploid hybrid Italian sparrow at its two geographically separated hybrid–parent boundaries represent a subset of those contributing to reproductive isolation between its parent species, house and Spanish sparrows. Should such a sorting mechanism prove to be pervasive, the circumstances promoting homoploid hybrid speciation may be broader than currently thought, and indeed, there may be many cryptic hybrid taxa separated from their parent species by sorted, inherited incompatibilities.

Keywords: birds, hybridization, introgression, Passer, reproductive barriers, speciation

Received 28 May 2014; revision received 29 August 2014; accepted 1 September 2014

Introduction

Natural hybridization is increasingly recognized as an important contributor to biological diversification (Abbott et al. 2013; Seehausen et al. 2014). Existing taxa exchange adaptations through introgression (Heliconius Genome Consortium 2012) and new species form—not only in the face of gene flow (Nosil 2012), but also as a direct consequence of interbreeding (Rieseberg 1997; Mallet 2007; Rieseberg & Willis 2007; Mavarez & Linares 2008; Abbott & Rieseberg 2012; Abbott et al. 2013). The latter process, known as hybrid speciation, encompasses the origin of genetically admixed taxa that are reproductively isolated from their parents. The development of reproductive barriers plays a central role in hybrid speciation, yet the nature and development of such barriers remain poorly understood (Rieseberg 1997; Mallet 2007; Abbott et al. 2013; Schumer et al. 2014). An important task in hybrid speciation research is therefore to unveil the evolution of the barriers that separate hybrid taxa from their parent species.

Both pre- and postzygotic barriers have been shown to act between hybrid taxa and their parents (Rieseberg 1997; Mallet 2007; Mavarez & Linares 2008; Abbott et al. 2013). A recent survey of the topic suggested that ecological divergence and geographic isolation are likely to be of greater importance for the establishment of hybrid taxa than are intrinsic postzygotic barriers, especially in animals (Abbott & Rieseberg 2012). The evolution of intrinsic reproductive isolation between hybrids and parents is poorly studied and understood, however (Buerkle et al. 2000; Coyne & Orr 2004; Jiggins et al. 2008; Schumer et al. 2014). The only direct empirical evidence comes from the Helianthus sunflower system,
where intrinsic isolation seems to result from a mixture of sorting of pre-existing chromosomal structural differences and structural rearrangements induced by recombination (Rieseberg et al. 1995, 2003; Lai et al. 2005). To what extent intrinsic barriers separating hybrids from their parents develop through sorting of pre-existing parental incompatibilities or through de-novo epistatic interactions in the hybrid genomes therefore remains an open question. This question can, however—as suggested by Rieseberg (1997)—be addressed by comparing the locations of genomic regions contributing to intrinsic reproductive barriers between the parent species with those isolating the hybrid species from their parents. If the genomic regions contributing to reproductive isolation in a hybrid taxon are a subset of those acting between its parents, then the sorting hypothesis would be supported.

Here, we apply the approach suggested by Rieseberg (1997) to investigate the evolution of intrinsic barriers during homoploid hybrid speciation in Passer sparrows. In this species complex, interbreeding between house P. domesticus and Spanish sparrows P. hispaniolensis has resulted in a taxon of hybrid origin, the Italian sparrow P. italica, which genetically and phenotypically is a mixture of its parent species (Meise 1936; Elgvin et al. 2011; Hermansen et al. 2011; Trier et al. 2014). The Italian sparrow is a small seed-eating bird occupying a human-commensal niche similar to that of the house sparrow (Summers-Smith 1988). It is distributed over the entire Italian Peninsula and meets the house sparrow in a narrow hybrid zone in the Alps. It lives sympatrically with the Spanish sparrow in restricted areas of mainland Italy (Summers-Smith 1988; Hermansen et al. 2011) and exchanges migrants with Spanish sparrows from Sardinia (Trier et al. 2014). The Italian sparrow’s parent species are themselves broadly sympatric across large parts of the Spanish sparrow range, maintaining phenotypic distinctness in all but a few locations (Summers-Smith 1988). This demonstrates the existence of reproductive barriers typically isolating house and Spanish sparrows.

Previous work has shown that the Italian sparrow has developed reproductive barriers against both parent species (Hermansen et al. 2011; Trier et al. 2014). Evidence suggests that intrinsic barriers consist mainly of sex-linked and mitonuclear incompatibilities (Trier et al. 2014). A set of seven candidate reproductive isolation (RI) loci have been identified, three of which exhibit steep clines at the Italian-house sparrow boundary and four which exhibit steep clines at the Italian–Spanish sparrow boundary. Among these candidate RI loci, sex chromosome (Z)-linkage is overrepresented, and the patterns of variation suggest a role for mitonuclear conflict in isolating Italian and Spanish sparrows (Trier et al. 2014).

Here, we take the next step in the study of reproductive isolation in the Passer system and investigate whether these intrinsic barriers have formed through sorting of pre-existing incompatibilities or through de-novo epistatic interactions in the hybrid genome. We accomplish this by replicating the hybrid–parents study of Trier et al. (2014) in an area of parental species sympatry and hybridization on the Iberian Peninsula.

Materials and methods

Focal population

Our focal population is situated on the Iberian Peninsula, close to Badajoz, Extremadura, Spain (38.64866N, –7.215453E, Fig. 1). This is an area where house and Spanish sparrows occur sympatrically, including breeding in the same stork nests. As in our previous hybrid–parents study (Trier et al. 2014), only male individuals, which are diploid for the Z chromosome, were analysed to avoid issues related to haplodiploidy of the Z chromosome. One hundred and sixty-three males (phenotypic house sparrows: N = 76, phenotypic Spanish sparrows: N = 87) were caught using mist nets, and a blood sample (20–50 μL) for use in genetic analyses was taken from each individual by puncturing a brachial vein. DNA was extracted from blood samples stored in standard buffer using Qiagen DNeasy 96 Blood and Tissue Kits (Qiagen N.V., Venlo, the Netherlands) according to the manufacturer’s instructions with the minor adjustment of adding 100 μL of blood/buffer in the initial step. Authorization to catch birds and take blood samples was obtained from the appropriate authorities, and all birds were released upon finished sampling.

Reference populations

As reference house sparrows, we used previously published data from two allopatric populations; Oslo, Norway (N = 58) and Hradec Králové, Czech Republic (N = 27) (Fig. 1; Trier et al. 2014). Population genetic analyses have shown that the house sparrow exhibits very little geographic structuring genetically (Sætre et al. 2012). This is likely a result of a recent, massive range and population expansion associated with the sparrows’ recent adaptation to human commensalism in connection with the rise and spread of agriculture (Sætre et al. 2012). House sparrows from Norway and Czech Republic are therefore suitable allopatric reference populations for use in comparative analyses with the sympatric area on the Iberian Peninsula.

As reference Spanish sparrows, we used a population on the Gargano Peninsula in east-central Italy (N = 52) (Fig. 1). This reference population was chosen based on
an assessment of hybrid indices from our available collection of samples. This assessment indicated that these birds, albeit locally sympatric with Italian sparrows, are genetically pure Spanish sparrows and hence should give good estimates of pure Spanish sparrow allele frequencies. In contrast, allopatric Spanish sparrows from Sardinia show evidence of some gene exchange with mainland Italian sparrows, through episodes of migration (Trier et al. 2014; see also Fig. 2A).

Hybrid–parent comparison

To obtain a direct comparison with hybrid–parent boundaries, we reanalysed the data set utilized by Trier et al. (2014) using the exact same marker set and parental reference populations as used when analysing the focal sympatric parental population. The hybrid–parents data set consists of 57 populations of Italian sparrows, Italian-house hybrids and parapatric house sparrows from the Italian peninsula and the Alps (N = 385), as well as parapatric Spanish sparrows from Sardinia (N = 48).

Marker set

We utilized 77 species-informative SNP markers (Table S1, Supporting information) from 75 genes developed through 454 transcriptome sequencing of cDNA from heart, liver and brain tissues (described in Trier et al. 2014). To obtain species-informative SNPs from 454 transcriptome data of six house and six Spanish sparrows lacking reference genomes, we mapped against the most closely related reference genome available, that of the zebra finch Taeniopygia guttata (Warren et al. 2010). First, we mapped all cDNA reads from the house sparrows against the zebra finch transcriptome and used the consensus called on the mapped reads as reference for mapping all cDNA reads from both house and Spanish sparrows. Then, we singled out the bases where all reads from one species exhibited a different
base than that of the other species, that is putatively species-diagnostic SNPs. These steps were repeated a second time starting with the cDNA reads from the Spanish sparrows (for detailed description of the species-diagnostic SNP detection method, see https://github.com/lexnederbragt/454_transcriptome_snps). Based on the species-diagnostic SNPs from the 12 samples, multiplex sets of PCR primers were designed and all genotyped at each of 77 SNP loci (with the exception of five loci where genotyping failed for 44 of the Spanish sparrow reference individuals, see Table S1, Supporting information) using the Sequenom MassARRAY platform at CIGENE, Norwegian University of Life Sciences, As, Norway. Genes were annotated by blasting against the zebra finch and chicken Gallus gallus (International Chicken Genome Sequencing Consortium 2004) genomes. Two exceptions were SNPs within CHD1Z and ND2 genes, which were genotyped using existing primers (see Elgvin et al. 2011). The CHD1Z SNP in this set is located within an intron.

We note that the six Spanish sparrows used for transcriptome sequencing and SNP detection were sampled from the same sympatric locality on the Iberian Peninsula as the focal sparrows in this study, whereas the six house sparrows used for transcriptome sequencing and SNP detection were sampled in Oslo, Norway. Thus, all else being equal, one might expect to observe more introgression into house sparrows than Spanish sparrows in the area of parental sympatry, as the house sparrows in the parental sympatric population were not among the populations used for identification of species-informative markers. This potential problem should not, however, adversely affect the comparison of locus-specific introgression between hybrid–parents and parent–parent systems, that is the comparative genomic clines analyses (see below).

Comparative investigation of introgression patterns

We used Bayesian admixture analysis as implemented in STRUCTURE (Pritchard et al. 2000; Falush et al. 2003) to compare the genetic makeup of individuals in the focal sympatric area in Spain (house: \( N = 76 \), Spanish: \( N = 87 \)) to that of other populations of house and Spanish sparrows and hence to assess patterns of introgression. As representative allopatric house sparrow
populations, birds from Oslo, Norway (N = 58) and Hradec Králové, Czech Republic (N = 27) were used. As representative Spanish sparrow populations, birds from the Gargano Peninsula, Italy (N = 52) and Sardina, Italy (N = 48) were used. Known FI avian hybrids between house (Oslo, Norway) and Spanish (Badajoz, Spain) sparrows (N = 6) were also included to verify that STRUCTURE assigned admixed individuals correctly. In STRUCTURE, the admixture model was used and we assumed two groups (K = 2) with correlated allele frequencies and ran 1 000 000 iterations after a burn-in of 500 000 iterations.

**Detection of SNPs associated with reproductive isolation**

To test for loci potentially associated with reproductive isolation, we applied a genomic cline approach (Payseur 2010; Gompert & Buerkle 2011; Fitzpatrick 2013). This approach takes advantage of the fact that introgression can vary across the genome due to recombination in admixed individuals. Genomic cline analysis quantifies locus-specific introgression relative to a genome-wide average as estimated by hybrid index or average genome-wide ancestry. We used genomic cline analysis to identify loci exhibiting restricted introgression relative to the genome-wide average, as such loci are candidates for being associated with reproductive isolation. The basic assumption is that alleles at loci responsible for barriers to gene flow and neutral loci linked to such selected loci will exhibit reduced introgression into the foreign genomic background, due to selection against unfit admixed individuals. This will result in steep and possibly shifted genomic clines. By contrast, gene flow at unlinked, neutral loci is not expected to be influenced by such barriers. Neutral loci therefore exhibit shallower clines that do not change more sharply with changing hybrid index than a neutral expectation.

We used genomic cline analysis as implemented in the BGC software (Gompert & Buerkle 2012). This approach allows for the use of markers that are not fixed in each parent species. BGC uses Markov Chain Monte Carlo (MCMC) methods to obtain Bayesian estimates of two parameters that describe the bias and rate of locus-specific introgression into foreign genomic background, based on ancestry (Gompert & Buerkle 2011, 2012). For a given locus, cline parameter \( \alpha \) specifies the probability of ancestry from one of the parental species where an increase in probability produces a positive parameter value and a decrease, a negative parameter value. Cline parameter \( \beta \), on the other hand, specifies the rate of transition from low to high probability of ancestry from the same parental species as a function of hybrid index, with an increased rate yielding a positive parameter value and a decreased rate, a negative parameter value (Gompert & Buerkle 2012). \( \alpha \) is therefore analogous to cline centre in geographic cline analysis, whereas \( \beta \) is analogous to cline steepness. Locus-specific introgression differs from the genome-wide average when credibility intervals for the posterior probability distributions of \( \alpha \) and \( \beta \) do not overlap with zero (Gompert & Buerkle 2011, 2012).

As parental reference populations in the BGC analyses, and hence to indicate parental allele frequencies, we used house sparrows from Oslo, Norway (N = 58) and Hradec Králové, Czech Republic (N = 27), and Spanish sparrows from the Gargano Peninsula, Italy (N = 52). For the BGC analyses in parental sympatry, all individuals from the sympatric parental population were pooled into one admixed test population. Similarly, for the reanalysis of the hybrid–parents data set, all individuals from the Alps, the Italian peninsula (excluding Spanish sparrows from the south-east Italian sympatric zone), Sicily and Sardinia were pooled into one admixed test population.

Genotypes at the mitochondrial locus ND2 were coded as diploid homozygotes as BGC failed to run with a haploid marker included. Ten independent runs of 100 000 iterations each were run with the first 25 000 iterations discarded as burn-in, MCMC samples thinned by recording every fifth value, while the rest of the BGC settings were as default. We verified that the runs gave qualitatively similar results, and for each system, we present the results from the run with the best fit, that is the run with the lowest mean negative log-likelihood. SNPs were identified as significantly deviating from genome-wide expectations when the 95% credibility intervals of the cline parameters \( \alpha \) and \( \beta \) did not cross zero. We also estimated the quantiles of the cline parameters in the genome-wide distribution of introgression, namely \( \alpha _q \) and \( \beta _q \). These quantiles represent the position of each SNP on the posterior probability distribution of the genome-wide average for each of \( \alpha \) and \( \beta \), respectively.

The genomic cline model implemented in BGC assumes no drift and that all loci have the same effective population size. As Z-linked loci have a lower effective population size than autosomal markers (3/4), this could lead to an overrepresentation of Z linkage of steep clines simply as an effect of stronger drift at these loci. To test for this, we also analysed the Z-linked and autosomal loci in parental sympatry separately using the same procedure as when analysing the full data set.

To compare the patterns of introgression between the parent species in sympatry on the Iberian peninsula with the introgression patterns between the Italian sparrow and its parent species in Italy, we investigated which markers exhibited significant excess ancestry.
(significant $\alpha$ parameter) and/or steep clines (significant, positive $\beta$ parameter) in either or both systems. We also estimated Pearson's product–moment correlation coefficients among systems for both cline parameters ($\alpha$ and $\beta$) for the autosomal markers, for the Z-linked markers, for the candidate intraspecific incompatibilities in the Italian sparrow and for the putative reproductive isolation markers (see below). The correlations were performed on point estimates based on the median of the posterior distribution for $\alpha$ and $\beta$. Finally, we plotted the quantiles of the cline parameters in the genome-wide distribution of introgression, $q_\alpha$ and $q_\beta$, to visualize where in the genome-wide distribution of locus-specific introgression, the markers were situated.

Results

Patterns of introgression in parental sympatry

Our Bayesian admixture analysis revealed a pattern of asymmetric introgression in the area of parental sympatry. The majority of individuals with house sparrow phenotype showed signs of introgression when compared to allopatric house sparrow populations (Fig. 2A). Spanish sparrows, on the other hand, did not exhibit signs of extensive introgression. They showed less introgression than the insular Spanish sparrow population on the Mediterranean island of Sardinia previously shown to undergo gene exchange with Italian sparrow migrants from mainland Italy (Fig. 2A; Trier et al. 2014). With the aforementioned SNP ascertainment bias caveat in mind, however, this apparent asymmetry in introgression must be interpreted cautiously as it may to some degree reflect an artefact of the SNP detection strategy used in this study. Nonetheless, the relative difference in introgression between allopatric and sympatric house sparrow populations as evident in Fig. 2A is a signal of significant ongoing introgression in sympatry. Moreover, the distribution of hybrid indices estimated using the Bayesian algorithm implemented in BGC provided an informative basis for investigating locus-specific introgression using genomic clines analysis (Fig. 2B, C).

Genomic clines analysis in parental sympatry

Ten SNPs exhibited significant excess house sparrow ancestry (negative $\alpha$), whereas eight SNPs exhibited excess Spanish sparrow ancestry (positive $\alpha$) in parental sympatry (Fig. S1, Supporting information). This indicates that locus-specific introgression does not show the same overall bias towards introgression into house sparrows as previously described for admixture proportion results from STRUCTURE. Furthermore, 19 SNPs exhibited significantly steep clines (positive $\beta$), indicating reduced introgression and hence a potential association with reproductive isolation, whereas 10 SNPs exhibited significantly shallow clines (negative $\beta$) (Fig. S1, Supporting information). Of the 19 loci exhibiting steeper clines than expected given the marker set average, 15 were Z-linked: a significant overrepresentation of sex-linkage compared to the genome-wide expectation. This remains a significant overrepresentation of sex-linkage also when compared to the proportion of sex-linked markers in our marker set (Two-tailed binomial test: null probability based on proportion of sex-linked marker in marker set $= 0.273$, successes $= 15$, trials $= 19$, $P = 4.10 \times 10^{-6}$). One could argue that this overrepresentation of Z-linked loci may result from stronger drift due to lower effective population size of Z-linked loci compared to autosomal loci, and we cannot exclude the possibility that differences in effective population size between Z and autosomes may contribute to the observed pattern. However, our independent analyses of Z-linked and autosomal loci gave qualitatively the same results for the $\beta$ parameter as when all markers were analysed together (Z-linked loci: $N = 21$, $r = 0.964$, $P = 2.15 \times 10^{-12}$, autosomal loci: $N = 55$, $r = 0.954$, $P = 2.20.0 \times 10^{-16}$, Fig. S2, Supporting information). This indicates that there is a real overrepresentation of Z linkage among the candidate RI loci.

Genomic clines reanalysis of hybrid–parent data set

In our reanalysis of the hybrid–parents data set using a different, larger, parental Spanish reference set than previously, 26 SNPs exhibited significant excess house sparrow ancestry, whereas 29 SNPs exhibited excess Spanish sparrow ancestry. Furthermore, 13 SNPs exhibited significantly steep clines, whereas 17 SNPs exhibited significantly shallow clines (Fig. S3, Supporting information). All Z-linked loci exhibiting steep clines at the hybrid–parent range boundaries in Trier et al. (2014) also exhibited significantly steep and shifted genomic clines when reanalysing the hybrid–parent data set (Figs 3 and 4; Fig. S3, Supporting information). That is, the Z-linked candidate reproductive isolation loci remained unchanged: CETN3 and CHD1Z at the Italian/house boundary and HSDL2, MCC2 and GTF2H2 at the Italian/Spanish boundary. Autosomal candidate marker RPS4 did not exhibit a significantly steep cline in the reanalysis, however, but was the marker exhibiting the strongest excess Spanish sparrow ancestry (Figs 3 and 4; Fig. S3, Table S2, Supporting information). In our previous study of hybrid–parent barriers,
Fig. 3 Comparison of locus-specific patterns of introgression between parental sympatry and hybrid–parent range boundaries. Displayed are cline estimates for the 77 markers used in our BGC analyses. Each symbol denotes the point estimate of the given parameter based on the median of the posterior distribution. White symbols denote autosomal loci, black symbols denote Z-linked loci, grey symbols denote putative internal incompatibilities in the Italian sparrow, red symbols denote the putative nuclear reproductive isolation genes situated at hybrid–parent range boundaries, and the yellow symbol denotes mitochondrial marker ND2. The shape of the symbol denotes the significance level in each comparison based on 95% credibility intervals of the posterior distribution. Squares denote comparisons where the parameter estimates are significant in both parental sympatry and in the hybrid–parent analysis, triangles denote comparisons where the parameter estimates are significant only in parental sympatry, diamonds denote comparisons where estimates are significant only in the hybrid–parent analysis, circles denote comparisons where the parameter estimates do not deviate from neutral expectations in any of the two systems, and finally, crosses denote where the estimates are significant in both systems but in opposite direction. Numbers refer to the different putative RI markers as specified in panel A. In all panels, parameters from Italy (hybrid–parents) are plotted against parameters from Spain (parental sympatry) (A) $\alpha$ (bias) parameter: positive and negative values indicate bias in favour of Spanish sparrow and house sparrow alleles, respectively. Note different scale on axes. (B) $\beta$ (rate) parameter: positive values indicate restricted introgression, and negative values indicate elevated introgression. Note different scale on axes. (C) $q_\alpha$ parameter: values closer to 0 and 1 indicate increasing evidence for excess house or Spanish sparrows ancestry, respectively. (D) $q_\beta$ parameter: values closer to 0 and 1 indicate increasing evidence for shallow and steep clines, respectively.
F$_{ST}$ outlier analysis revealed *RPS4* to be the strongest candidate for being under directional selection in the Italian-house hybrid zone in the Alps (Trier et al. 2014). Moreover, analysis of spatial genetic variation using GENELAND (Guillot et al. 2005a,b) showed that *RPS4* exhibited a sharp shift from house to Spanish sparrow genotypes in the Alps hybrid zone (Trier et al. 2014).

We thus retain *RPS4* as a candidate RI locus between Italian and Spanish sparrows when plotting the results from the comparative genomic cline analysis as well as in our correlation comparison of candidate hybrid–parent RI loci. Furthermore, in our reanalysis, six loci exhibited steep clines not situated at either hybrid–parent boundary, compared with 15 such loci reported by Trier et al. (2014). These six markers were, however, a subset of the 15 loci that Trier et al. (2014) interpreted as candidates for being unpurged intraspecific incompatibilities within the Italian sparrow.

Hybrid–parents intrinsic barriers represent a subset of the barriers isolating the parent species.

All Z-linked candidate hybrid–parent RI loci (*CETN3, CHD1Z, HSDL2, MCCC2 and GTF2H2*) were among the 19 loci exhibiting significantly steep clines in parental sympatry (Figs 3 and 4; Figs S1 and S3, Supporting information). Indeed, there was strong concordance between the two systems in patterns of locus-specific introgression for the candidate nuclear RI loci (Figs 3 and 4; Table 1, Supporting information, correlation among systems for the $\alpha$ parameter for the nuclear putative RI markers: $N = 6, r = 0.986, P = 2.77 \times 10^{-6}$; correlation among systems for the $\beta$ parameter for the nuclear putative RI markers $\beta$: $N = 6, r = 0.942, P = 4.94 \times 10^{-3}$).

Of the six candidate intraspecific incompatibility loci in the Italian sparrow, three exhibited restricted introgression also in parental sympatry (Fig. 3; Figs S1 and S3, Table S2, Supporting information). Furthermore, more
Table 1 Pearson’s product–moment correlations between cline parameters in different marker groups. The correlations were performed on point estimates based on the median of the posterior distribution for \( \alpha \) and \( \beta \).

<table>
<thead>
<tr>
<th>Marker group</th>
<th>( \alpha )-Parameter</th>
<th>( \beta )-Parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( N ) ( r ) ( P )</td>
<td>( r ) ( P )</td>
</tr>
<tr>
<td>RI (nuclear)</td>
<td>6 0.986 2.77 ( \times 10^{-4} )</td>
<td>0.942 4.94 ( \times 10^{-3} )</td>
</tr>
<tr>
<td>II</td>
<td>6 0.617 0.192</td>
<td>0.030 0.956</td>
</tr>
<tr>
<td>Z-linked</td>
<td>21 0.558 8.56 ( \times 10^{-3} )</td>
<td>0.240 0.295</td>
</tr>
<tr>
<td>Autosomal</td>
<td>55 0.426 1.17 ( \times 10^{-3} )</td>
<td>0.485 1.77 ( \times 10^{-4} )</td>
</tr>
</tbody>
</table>

RI refers to candidate reproductive isolation loci and II refers to candidate unpurged intraspecific incompatibilities within the Italian sparrow.

loki appeared to be involved in reproductive isolation in parental sympatry than between hybrid and either parent as ten loci exhibited steep clines only in parental sympatry (Fig. 3; Figs S1 and S3, Supporting information).

In parental sympatry, mitochondrial marker ND2 exhibited the same steep cline and excess house sparrow ancestry as in Italian sparrows (Figs 3 and 4; Figs S1 and S3, Supporting information). Similarly, Z-linked nuclear-encoded mitochondrial markers HSDL2 and MCCCI2 also exhibited the same steep clines and excess house sparrow ancestry in parental sympatry as at the Italian-Spanish boundary (Figs 3 and 4; Figs S1 and S3; Table S2, Supporting information).

Discussion

In this study, we investigated the evolution of intrinsic reproductive isolation in a hybrid species system. We compared loci exhibiting restricted introgression at the range boundaries between the hybrid Italian sparrow and its parent species, house and Spanish sparrows, to loci exhibiting restricted introgression in an area where the parent species co-occur and hybridize. The rationale behind our approach is that if the hybrid–parent barriers have arisen through sorting of pre-existing parental incompatibilities, the same loci should also act as barriers between the parents when they hybridize in sympatry. If the hybrid–parent barriers have arisen through \textit{de-novo} interactions in the hybrid genome, no such concordance is expected (Rieseberg 1997).

Hybrid–parent intrinsic barriers represent a subset of those isolating the parent species

Consistent with the main prediction from the sorting hypothesis, all five Z-linked candidate RI loci between the Italian sparrow and either parent species were among the 19 loci exhibiting restricted introgression in parental sympatry. Moreover, these five loci all exhibited similar patterns of excess ancestry in parental sympatry as in the hybrid taxon, indicating that the sorting of these incompatibilities in the Italian sparrow was driven by a deterministic process. Autosomal marker RPS4 did not exhibit a significantly steep cline in parental sympatry, however, suggesting that this locus may be involved in a novel barrier that has developed between the hybrid Italian sparrow and the house sparrow.

Our comparative analysis further revealed a set of ten autosomal loci to exhibit restricted introgression only in parental sympatry. This supports the prediction from the sorting hypothesis that hybrid–parent RI genes should represent a subset of parent–parent RI genes. It further suggests that the association of these loci with reproductive isolation is caused by epistasis through linkage disequilibrium, which has been broken down by recombination in the hybrid taxon. In sum, our results support the hypothesis that the previously reported intrinsic barriers between the Italian sparrow and its parents have arisen through the sorting of pre-existing parental incompatibilities. Allelic mosaicism of hybrid species genomes has previously been demonstrated in \textit{Helianthus} sunflowers (Rieseberg et al. 1995, 2003) as well as in \textit{Lycaenides} (Gompert et al. 2006; Nice et al. 2013) and \textit{Papilio} butterflies (Kunte et al. 2011) and cichlids (Keller et al. 2013), but this is the first time mosaicism of parental intrinsic incompatibility alleles is demonstrated.

Not all narrow clines in the hybrid Italian sparrow are located at hybrid–parent range boundaries, indicating a possible presence of unpurged intraspecific incompatibilities (Trier et al. 2014). Such incompatibilities may become coupled with extrinsic barriers and this can set the stage for further diversification of the hybrid taxon (Bierne et al. 2011). Findings consistent with isolation-by-adaptation have previously been reported from the \textit{Passer} system, so such a coupling mechanism may be at work (Eroukhmanoff et al. 2013). In our comparative genomic cline analysis, three of the six candidate intraspecific incompatibilities exhibited restricted introgression also in parental sympatry. This suggests that a subset of the candidate intraspecific incompatibilities in the Italian sparrow may represent unpurged parental incompatibilities, albeit further careful investigation is needed to unveil their potential role in further diversification.

The set of markers associated with restricted introgression in parental sympatry exhibited a strong over-representation of Z chromosome linkage. This finding must, however, be interpreted with caution. We cannot fully control for the lower effective population size of Z-linked markers in our analyses, and hence this
finding may to some degree be influenced by elevated drift at Z-linked loci. However, the finding is consistent with a growing number of studies demonstrating that sex chromosome linkage plays a disproportionate role in reproductive barriers between species with chromosomal sex determination systems (Qvarnström & Bailey 2009), including in the Passer sparrows (Elvgren et al. 2011; Trier et al. 2014). The disproportionate role for sex chromosomes in reproductive isolation is thought to result from the higher mutation rate in the male germ line, combined with exposure of recessive alleles to selection in the heterogametic sex, an overrepresentation of genes with sexual function on the sex chromosomes, high average linkage between the sex-linked genes involved in reproductive isolation, and stronger drift due to a lower effective population size of sex-linked markers (Charlesworth et al. 1987; Coyne & Orr 2004; Qvarnström & Bailey 2009; Mank et al. 2010; Sætre & Saether 2010; Ellegren et al. 2012).

Mother’s curse influences isolation both between the parents and between hybrid and parent

‘Mother’s curse’ is the phenomenon by which selection in males has no direct effect on mitochondrial fitness due to the maternal inheritance of mitochondria, favouring female (irrespective of male) interests, consequently creating a selective sieve allowing for the build-up of male-detrimental mutations (Frank & Hurst 1996; Gemmel et al. 2004). This in turn selects for suppressor alleles that restore male fitness. In female-heterogametic taxa (ZZ/ZW), Z-linked genes spend two-thirds of their time in the male lineage, and suppressor alleles are therefore expected to be disproportionately situated on the Z chromosome (Rice 1984; Trier et al. 2014). Mismatches between mitochondrial alleles and nuclear suppressor alleles lead to detrimental effects in hybrids whose parents differ in mitonuclear system (Frank & Hurst 1996; Gemmel et al. 2004).

In parental sympatry, we found both mitochondrial marker ND2 and Z-linked nuclear-encoded mitochondrial markers HSDL2 and MCCC2 to exhibit the same steep clines and shifts in favour of the house sparrow allele as in Italian sparrows at the Italian–Spanish boundary (Trier et al. 2014). The finding of strong house sparrow excess ancestry for mitochondrial DNA also in parental sympathy suggests that the fixation of house rather than Spanish sparrow mitochondrial DNA in the Italian sparrow was driven by deterministic rather than stochastic processes. The mechanism favouring house sparrow mitochondria remains unknown, but the candidacy of HSDL2 and MCCC2 as Z-linked mother’s curse suppressors is strengthened as these genes exhibit among the steepest clines both at the Italian-Spanish sparrow boundary and in parental sympathy, as well as exhibiting significant shifts in the same direction as mtDNA in both systems.

Conclusions

Our findings suggest that intrinsic barriers isolating the Italian sparrow from its parent species have mainly developed through the sorting of pre-existing sex-linked parental incompatibilities and that isolation in both cases is driven in part by mitonuclear conflict involving the Z chromosome. This represents the first evidence that the sorting process contributes to the persistence of a hybrid animal taxon. It remains to be further investigated whether and to what extent novel evolution in the hybrid Italian sparrow also contributes to reproductive isolation against its parent species, but autosomal marker RPS4 suggests that de-novo epistatic interactions may also play a role in this system. Should a sorting mechanism similar to the one described here prove to be pervasive, the circumstances promoting homoploid hybrid speciation may be broader than currently suspected, and indeed, there may be many cryptic hybrid taxa separated at two boundaries by sorted, inherited incompatibilities, as in the Italian sparrow.

Acknowledgements

We thank M. H. Tu, M. Moan, P. Munclinger, M. F. G. Rojas, S. A. Saether as well as numerous field assistants for help with acquiring the data, B. Dogan for help with laboratory work and N. H. Barton, A. Runemark and anonymous referees for helpful comments on a previous draft of the manuscript. This work was supported by The Research Council of Norway, Molecular Life Science (MLS), University of Oslo, and Centre for Ecological and Evolutionary Synthesis (CEES), University of Oslo.

References


The study was conceived and designed by J.S.H. and G.P.S. Fieldwork was carried out by J.S.H., F.H., C.N.T., A.M. and G.P.S. Bioinformatics work was carried out by A.J.N. The data were analysed by J.S.H. with contributions from C.N.T., R.I.B. and F.H. The study was written by J.S.H. with contributions from R.I.B., G.P.S. and the other authors.
Data accessibility

Genotype data for all individuals used in the presented analyses, estimated cline parameters for all markers, hybrid indices, BGC input files and code used for running BGC are available through Dryad (doi:10.5061/dryad.v6f4d).

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 Individual Bayesian genomic clines for all 77 markers in parental sympatry analysis.

Fig. S2 Comparison of estimates of genomic cline steepness in parental sympatry for Z-linked and autosomal loci analysed separately and together with all markers.

Fig. S3 Individual Bayesian genomic clines for all 77 markers in hybrid–parents analysis.

Table S1 Details of SNPs used in analyses.

Table S2 Estimates of $\alpha$-parameter ('cline shift') and $\beta$-parameter ('cline steepness') in hybrid–parents and parental sympatry analyses for candidate RI markers and putative unpurged intraspecific incompatibilities in the Italian sparrow.