

ORIGINAL ARTICLE

Genomic islands of differentiation in two songbird species reveal candidate genes for hybrid female sterility

Libor Mořkovský¹ | Václav Janoušek¹ | Jiří Reif² | Jakub Rídl³ | Jan Pačes³ | Lukáš Choleva^{4,5}  | Karel Janko⁴ | Michael W. Nachman⁶ | Radka Reifová¹ 

¹Department of Zoology, Faculty of Science, Charles University, Prague 2, Czech Republic

²Institute for Environmental Studies, Faculty of Science, Charles University, Prague 2, Czech Republic

³Laboratory of Genomics and Bioinformatics, Institute of Molecular Genetics, The Czech Academy of Sciences, Prague 4, Czech Republic

⁴Laboratory of Fish Genetics, Institute of Animal Physiology and Genetics, The Czech Academy of Sciences, Liběchov, Czech Republic

⁵Department of Biology and Ecology, Faculty of Science, University of Ostrava, Ostrava, Czech Republic

⁶Museum of Comparative Zoology and Integrative Biology, University of California, Berkeley, CA, USA

Correspondence

Radka Reifová, Department of Zoology, Faculty of Science, Charles University, Prague 2, Czech Republic.
Email: radka.reifova@natur.cuni.cz

Funding information

National Institutes of Health, Grant/Award Number: RO1 GM074245; The Czech Science Foundation, Grant/Award Number: 15-10884Y; ELIXIR CZ research infrastructure project, Grant/Award Number: LM2015047; Czech Academy of Sciences, Grant/Award Number: RVO 67985904; Ministry of Education, Youth and Sports of the Czech Republic, Grant/Award Number: EXCELLENCE CZ.02.1.01/0.0/0.0/15_003/0000460 OP RDE

Abstract

Hybrid sterility is a common first step in the evolution of postzygotic reproductive isolation. According to Haldane's Rule, it affects predominantly the heterogametic sex. While the genetic basis of hybrid male sterility in organisms with heterogametic males has been studied for decades, the genetic basis of hybrid female sterility in organisms with heterogametic females has received much less attention. We investigated the genetic basis of reproductive isolation in two closely related avian species, the common nightingale (*Luscinia megarhynchos*) and the thrush nightingale (*L. luscinia*), that hybridize in a secondary contact zone and produce viable hybrid progeny. In accordance with Haldane's Rule, hybrid females are sterile, while hybrid males are fertile, allowing gene flow to occur between the species. Using transcriptomic data from multiple individuals of both nightingale species, we identified genomic islands of high differentiation (F_{ST}) and of high divergence (D_{xy}), and we analysed gene content and patterns of molecular evolution within these islands. Interestingly, we found that these islands were enriched for genes related to female meiosis and metabolism. The islands of high differentiation and divergence were also characterized by higher levels of linkage disequilibrium than the rest of the genome in both species indicating that they might be situated in genomic regions of low recombination. This study provides one of the first insights into genetic basis of hybrid female sterility in organisms with heterogametic females.

KEYWORDS

birds, genomic islands of differentiation, hybrid female sterility, oogenesis, speciation

1 | INTRODUCTION

Understanding the genetic basis of reproductive isolation between incipient species is a major goal of evolutionary biology. Hybrid sterility is a common first step in the evolution of postzygotic reproductive isolation (Coyne & Orr, 2004). Haldane (1922) observed that

hybrid sterility occurs preferentially in the heterogametic sex, that is males in organisms with XY sex chromosomes, such as mammals and *Drosophila*, and females in organisms with ZW sex chromosomes, such as birds and butterflies. While the genetic basis of hybrid male sterility in organisms with heterogametic males has been studied for decades (e.g., Bhattacharyya et al., 2014; Good, Giger, Dean, &

Nachman, 2010; Masly & Presgraves, 2007; Mihola, Trachtulec, Vlček, Schimenti, & Forejt, 2009; Presgraves, 2010; Storchová et al., 2004; Zhang, Sun, Woldehellasie, Xiao, & Tao, 2015), the genetic basis of hybrid female sterility in organisms with heterogametic females has received much less attention. In fact, most studies on hybrid female sterility have been limited to anecdotal observations that hybrid females do not lay eggs (Reifová, Kverek, & Reif, 2011a; Stadie, 1991), that eggs of hybrid females fail to hatch nestlings (Gelter, Tegelström, & Gustafsson, 1992), or that hybrid females have atrophied ovaries (Eroukhanoff et al., 2016). One exception is the work on *Heliconius* butterflies showing that a large Z effect underlies hybrid female sterility, although specific genes have not been identified (Jiggins et al., 2001; Naisbit, Jiggins, Linares, Salazar, & Mallet, 2002).

In the last 15 years, technological advances in DNA sequencing have created new possibilities for studying the genetic basis of reproductive isolation in nonmodel organisms (Seehausen et al., 2014). One approach is to study patterns of genetic differentiation (F_{ST}) across the genome. In the presence of gene flow, genomic regions underlying reproductive isolation should show increased levels of genetic differentiation as selection will act against introgressed alleles in these regions (Via & West, 2008; Wu, 2001). These portions of the genome have been referred to as islands of high differentiation (Burri, 2017; Ellegren et al., 2012; Harr, 2006; Turner, Hahn, & Nuzhdin, 2005; Wolf & Ellegren, 2017). However, heterogeneous levels of differentiation can arise even in the absence of differential gene flow as a consequence of selection at linked sites leading to a local reduction in within-population diversity (Charlesworth, Morgan, & Charlesworth, 1993; Kaplan, Hudson, & Langley, 1989; Maynard Smith & Haigh, 1974). One way to distinguish between differential gene flow and selection at linked sites as causes for heterogeneous levels of differentiation is to take into account absolute divergence (D_{xy})—the average number of pairwise differences between alleles sampled from two populations (Nei, 1987). Selection at linked sites reduces variation within populations, and as F_{ST} (Wright, 1951) depends on the amount of within-population variation as well as the amount of between-population variation, selection at linked sites will lead to increased values of F_{ST} (Charlesworth, 1998). In contrast, absolute sequence divergence (D_{xy}) is not increased by selection at linked sites. In fact, D_{xy} may be reduced in such regions as selection in the ancestral population will reduce the coalescence time for alleles sampled in the two daughter populations. On the other hand, D_{xy} is expected to be higher in regions with low introgression as introgression reduces D_{xy} in a similar way as F_{ST} . Thus, genomic regions of high differentiation and high absolute divergence are likely to contain loci that are resistant to introgression (Cruickshank & Hahn, 2014; Nachman & Payseur, 2012; Noor & Bennett, 2009).

Analysis of islands of high differentiation and divergence thus seems to be a promising approach for identifying genomic regions involved in reproductive isolation. For example, this method has recently been used with two flycatcher species (*Ficedula hypoleuca* and *Ficedula albicollis*) (Burri et al., 2015; Ellegren et al., 2012). These taxa have been described as conforming to Haldane's Rule (e.g.,

Sætre et al., 2003), although much recent work suggests that hybrids of both sexes are sterile (Ålund, Immler, Rice, & Qvarnström, 2013). Moreover, genome-wide data to date have not revealed a single backcross individual in the hybrid zone (Kawakami et al., 2014), suggesting that postzygotic isolation is already complete in these species. Patterns of F_{ST} and D_{xy} across the flycatcher genome suggest that heterogeneity in levels of differentiation is due mostly to linked selection rather than to differences in levels of gene flow (Burri et al., 2015). Genome-wide comparisons of D_{xy} and F_{ST} have also been used in Swainson's thrush subspecies (*Catharus ustulatus ustulatus* and *C. u. swainsoni*) (Delmore et al., 2015), two cichlid ecomorphs that are in an early stage of ecological divergence (Malinsky et al., 2015) and subspecies of house mice (Phifer-Rixey, Bomhoff, & Nachman, 2014), among others.

Here, we explore the genomic landscape of differentiation and divergence in two closely related songbirds, the common nightingale (*L. megarhynchos*) and the thrush nightingale (*L. luscinia*). These two species diverged 1.8 Mya and currently hybridize in a zone of secondary contact across Central Europe (Storchová, Reif, & Nachman, 2010). Despite relatively strong assortative mating, F_1 as well as backcross hybrids occur in sympatric populations (Reifová, Reif, Antczak, & Nachman, 2011b; Reifová et al., 2011a; Vokurková et al., 2013) indicating that postzygotic isolation is still not complete. Crosses of the two species in captivity as well as observations in nature show that F_1 hybrid females are sterile, while F_1 hybrid males are fertile, consistent with Haldane's Rule (Reifová et al., 2011a; Stadie, 1991). The two nightingale species are very similar morphologically and ecologically, and for that reason, strong competition occurs between the species in sympatry (Reif et al., 2015). Interspecific competition has resulted in ecological character displacement in beak size (Reifová et al., 2011b), which might enhance reproductive isolation and thus also contribute to speciation. In previous work, using coalescent models of isolation with migration, we showed that gene flow has occurred between the species and that gene flow has been higher on the autosomes than on the Z chromosome, consistent with a large Z effect (Storchová et al., 2010).

Here, we expand on that work by comparing loci across the nightingale genomes. Using transcriptomic data from multiple allopatric individuals of both nightingale species, we identified (i) genomic islands of high differentiation and (ii) genomic islands of both high differentiation and divergence. We then analysed gene content within these islands. In addition, we compared levels of linkage disequilibrium and patterns of molecular evolution using McDonald–Kreitman tests in genomic regions inside and outside these islands. Interestingly, we found significant enrichment of genes related to female meiosis and metabolism in islands of high differentiation. The same functional categories were also enriched in islands of high differentiation and divergence although in this case the enrichment was not significant after correction for multiple tests. Islands of high differentiation and divergence were also characterized by increased levels of linkage disequilibrium in both species, suggesting that these regions might be situated in genomic regions with reduced recombination. Our study confirms that hybrid female sterility is an

important reproductive barrier between the two nightingale species and provides a first look at the genetic basis of hybrid female sterility in organisms with heterogametic males.

2 | MATERIALS AND METHODS

2.1 | Samples

Birds were sampled in allopatric regions close to the contact zone, in southwestern Poland for *L. megarhynchos* and in northeastern Poland for *L. luscinia*. Within each species, birds were captured at different localities 10–30 km apart. The whole sampling area spanned approximately 100 km in both species. Eight male *L. megarhynchos* and seven male *L. luscinia* were collected. In addition, we sampled one male individual of the bluethroat (*L. svecica*) as an outgroup. *L. svecica* diverged from the other two species approximately 12 Mya (Jetz, Thomas, Joy, Hartmann, & Mooers, 2012). Sampling was performed in 2010 at the beginning of the breeding season (early May). The birds were euthanized, and a piece of liver was removed and placed into RNA later. All work with live animals was done in accordance with the Local Ethics Committee for Scientific Experiments on Animals in Poznan, Poland (permission no. 36/2010).

2.2 | Transcriptome sequencing

Total RNA from liver was isolated using TRIzol Reagent (Ambion) according to the manufacturer's protocol. The integrity of the RNA was verified using an Experion automated electrophoresis system (Bio-Rad). cDNA synthesis and normalization were performed by the Evrogen Lab (Moscow, Russia). Total RNA samples were used for ds cDNA synthesis using a SMART approach (Zhu, Machleder, Chenchik, Li, & Siebert, 2001). SMART-prepared cDNAs were then normalized using the duplex-specific nuclease (DSN) normalization method (Zhulidov et al., 2004). The normalization procedure included cDNA denaturation/reassociation, treatment by DSN (Shagin et al., 2002) and amplification of the normalized fraction by PCR. One microgram of each normalized cDNA sample was used to generate a tagged 454 sequencing library using MID adaptors following the Rapid Library Preparation Method from Roche. The libraries were sequenced on a GS FLX instrument using Titanium Chemistry. Most libraries were sequenced in pairs in which two different libraries were pooled and loaded onto one sequencing region.

2.3 | Transcriptome assembly and variant calling

The sequences were split according to MIDs using the sfffile tool and converted to fastq. Transcriptome assembly was performed with NEWBLER 2.6 (Margulies et al., 2005). Resulting contigs were mapped onto the zebra finch (*Taeniopygia guttata*) reference genome (TAE-GUT1, WUSTL version 3.2.4). The assembly produced 61,753 unique contigs belonging to 9,332 unique genes in the reference genome. The average length of the contigs was 1,264 bp and the median length 1,023 bp. The chromosome and position in the zebra finch

genome were used to build a contig scaffold by concatenating the contigs in the order of their zebra finch homologs. All reads were mapped onto the contig scaffold with SMALT (<http://www.sanger.ac.uk/science/tools/smalt-0>), and variants were called using SAMTOOLS MPILEUP (Li et al., 2009). The variant detection pipeline closely followed one used in Mořkovský, Pačes, Řídl, and Reifová (2015). In brief, we applied the default filtering criteria used by vcfutils varFilter and additionally filtered for Phred quality score above 10. In addition, we excluded from the analysis all variants from contigs that spanned more than 50 kb in the zebra finch genome after the mapping as this probably reflects incorrect mapping. We identified 300,433 SNPs which met the filtering criteria. The average coverage of each variant that was used in analyses was 42.5.

Nightingales and zebra finch diverged approximately 45 million years ago (Jetz et al., 2012). However, large-scale synteny is observed among bird genomes separated by even greater evolutionary distances (Backström et al., 2008). Nonetheless, using the zebra finch genome may result in incorrect ordering of some genes in nightingales if rearrangements have occurred. This is likely to introduce some noise into downstream analyses, but should not result in systematic bias affecting the gene content or patterns of molecular evolution in islands of high differentiation and divergence.

2.4 | Analysis of genetic differentiation (F_{ST}) and divergence (D_{xy})

Per site unweighted F_{ST} (Wright, 1951) was calculated for all SNP variants that passed our quality filters using VCFTOOLS (Danecek et al., 2011). We calculated smoothed F_{ST} values by averaging per site F_{ST} in sliding windows along chromosomes using positions in the zebra finch genome. In cases where the nightingale sequence mapped perfectly to the zebra finch genome, the exact genomic position for SNPs within this sequence was taken. In cases where there was no exact mapping, for example in less conserved introns (note that a small number of introns are captured in unspliced transcripts), the positions of SNP variants were taken within the range of the gene to which the intron belonged (this occurred in 23,819 of 410,245 SNP variants, or 5.8%). The size of the sliding window was 1 Mb, and the step size was 100 kb.

We used a permutation analysis to identify islands of elevated F_{ST} . A distribution of potential F_{ST} values in each window was determined by shuffling the real F_{ST} values among the genomic positions of SNP variants—this kind of resampling procedure serves to establish a distribution of F_{ST} values as expected if genome position had no effect on F_{ST} . We sampled the distributions with 25,000 realizations of the permutation. Then, the upper 1% quantile of the distribution was used as a threshold to identify windows with elevated F_{ST} . To account for the different mode of inheritance and effective population size of the sex chromosomes and their effect on F_{ST} , permutations were conducted separately for the Z chromosome and for all autosomes. Windows of elevated F_{ST} less than 1 Mbp apart were pooled, and the whole genomic region containing these windows was designated as an island of high differentiation.

D_{xy} , defined as the average number of pairwise differences between alleles sampled from two populations, was calculated as follows. First, for each site, we calculated the average number of pairwise differences as $p_1q_2 + p_2q_1$, where p and q represent the frequencies of alternative alleles, and subscripts represent the species of origin. Second, the average pairwise differences for individual SNPs were summed and divided by the total length of the contig including all variable and invariant sites that passed the filtering criteria. The average D_{xy} in the F_{ST} islands was calculated. A subset of islands of high differentiation with D_{xy} higher than the genomic average was then identified for further analyses. We refer to these islands as islands of high differentiation and divergence.

2.5 | Gene set enrichment analysis

To assess whether the genes in the islands of high differentiation and islands of high differentiation and divergence were associated with specific functions, we carried out a test for over-represented gene groups. We took all ENSEMBL (Yates et al., 2016) genes found (i) in the islands of high differentiation and (ii) in the islands of high differentiation and divergence and assigned them to KEGG pathways (Kanehisa & Goto, 2000; Kanehisa, Sato, Kawashima, Furumichi, & Tanabe, 2016) using the GENOMENET REST API and KEGG REST API. For each pathway, we counted the genes found in the islands and calculated the enrichment score using all ENSEMBL genes as a background (17,488 genes). We used Fisher's exact tests and binomial tests followed by a correction for false discovery rate using the procedure of Benjamini and Hochberg (1995). To highlight the candidate genes in the oocyte meiosis pathway (tgu04114), we used CYTOSCAPE 3.3.0 (Shannon et al., 2003; Smoot, Ono, Ruschinski, Wang, & Ideker, 2011).

2.6 | Analysis of linkage disequilibrium

Genotypic linkage disequilibrium was assessed inside and outside (i) islands of high differentiation and (ii) islands of high differentiation and divergence separately for the two nightingale species. We calculated the composite correlation coefficient (r^2) between all pairs of SNPs within 1-Mb sliding windows using the VCFTOOLS software (Danecek et al., 2011). From these pairwise comparisons, we calculated the average value of r^2 for each window. SNP positions were based on the zebra finch genome. Only SNPs having a minor allele frequency (MAF) higher than 0.15 were considered. Significance was assessed using Mann–Whitney U tests. The analysis of linkage disequilibrium was performed for the entire data set including the Z chromosome as well as for the data set excluding the Z chromosome. Results were nearly the same in both analyses, and therefore, we present results only for the entire data set including the Z chromosome.

2.7 | Patterns of molecular evolution

We conducted McDonald–Kreitman (MK) tests (McDonald & Kreitman, 1991) and calculated neutrality indexes (NI; Rand & Kann,

1996) inside and outside (i) islands of high differentiation and (ii) islands of high differentiation and divergence (Table S1). The MK test and the NI are based on comparisons of the amount of polymorphism within species and substitutions between species for synonymous and nonsynonymous mutations. Calculations were made separately for the two nightingale species. As the two nightingale species are closely related and there is only limited number of substitutions between them, we determined substitutions for each species relative to the outgroup, *L. svecica*. Sequences were then compared to two other more distantly related outgroups, collared flycatcher (*Ficedula albicollis*) and zebra finch, to assign substitutions either to *L. svecica* or to *L. megarhynchos/L. luscinia* lineage. In our analysis, we considered only those substitutions that occurred in the lineage to *L. megarhynchos/L. luscinia*. Analyses were performed for the entire data set including the Z chromosome as well as for the data set excluding the Z chromosome. Results were nearly identical for both data sets, and therefore, we present results only for the entire data set including the Z chromosome.

To identify synonymous and nonsynonymous mutations, we used the EXONERATE program (Slater & Birney, 2005) and custom Perl and Shell scripts. For gene prediction, we used protein sequences from the collared flycatcher and zebra finch obtained from ENSEMBL (Version 81, July 2015). Only protein coding sequences of genes exhibiting a one-to-one ortholog relationship between the zebra finch and flycatcher genomes were included (based on the ENSEMBL COMPARA database; Herrero et al., 2016; Vilella et al., 2009). For each gene, only the best prediction of the two reference species was considered (based on the blast score provided by the EXONERATE program). We also considered predictions for which only one species provided a match. As the genomes of the flycatcher and zebra finch are fairly diverged, we aligned their coding sequences with the coding sequences created for individual nightingale samples. Regions of poor alignment containing gaps, unknown bases, stop codons or multiple substitutions/polymorphisms per codon were removed. The final data set consisted of 4,657 protein-coding genes. The number of within-species polymorphisms associated with coding sequence was 3,186 for *L. megarhynchos* and 3,987 for *L. luscinia*. From these 588 were shared between the species. The number of substitutions was 3,505 for the *L. megarhynchos* lineage and 2,704 for the *L. luscinia* lineage. From these 2,129 were shared between the two lineages (i.e., occurred between the common ancestor of both nightingale species with *L. svecica* and the ancestor of *L. megarhynchos* and *L. luscinia*).

To compare patterns of molecular evolution inside and outside (i) islands of high differentiation and (ii) islands of high differentiation and divergence, we constructed a generalized linear model (glm) with terms for the type of mutation (synonymous or nonsynonymous), type of variant (polymorphism or substitution), genomic location (inside or outside islands) and the three-way interaction between these terms. This three-way interaction tests for the difference in MK tests inside and outside islands. To assess its significance, we compared the full model containing this interaction to a simpler model without

this interaction using a log-likelihood ratio test (ΔLL). The analysis was carried out in R (R Core Team 2016).

3 | RESULTS

3.1 | Patterns of differentiation and divergence along the genome

We found 136 islands of elevated differentiation (F_{ST}) covering ~319 Mbp. The mean island size was ~2.3 Mbp (maximum 13 Mb; minimum 1 Mbp) (Figure 1). Overall, the Z chromosome exhibited higher average F_{ST} values (0.088) than the autosomes (0.063). This is consistent with previous work that found lower levels of gene flow on the Z chromosome compared to the autosomes (average $2Nm$ was 0.422 on the autosomes and 0.018 on the Z chromosome; Storchová et al., 2010). However, we note that at equilibrium with respect to migration and drift, $F_{ST\text{-autosomes}} = 1/(4Nm + 1)$ and $F_{ST\text{-Z chromosome}} = 1/(3Nm + 1)$, assuming an equal sex ratio and no

sex-biased migration. The observed values of F_{ST} for the Z chromosome and for the autosomes are very close to these theoretical expectations, suggesting that these differences in F_{ST} can largely be explained by differences in the effective population size of the Z chromosome and autosomes. Within the Z chromosome, islands of high differentiation were not more common than on the autosomes. However, as the permutation tests were conducted separately for the Z chromosome and autosomes, Z-linked islands of differentiation were generally more differentiated than on the autosomes.

The average absolute divergence (D_{xy}) was slightly lower within islands of high differentiation ($D_{xy} = 0.00210$) than outside islands of high differentiation ($D_{xy} = 0.00211$), but the difference was not significant ($p = .6098$; Welch's two-sample t test). To identify genomic regions possibly involved in reproductive isolation, we selected all islands of high differentiation having D_{xy} higher than the genomic average—we call these regions islands of high differentiation and divergence. There were 67 such islands harbouring a total of 2,007 genes.

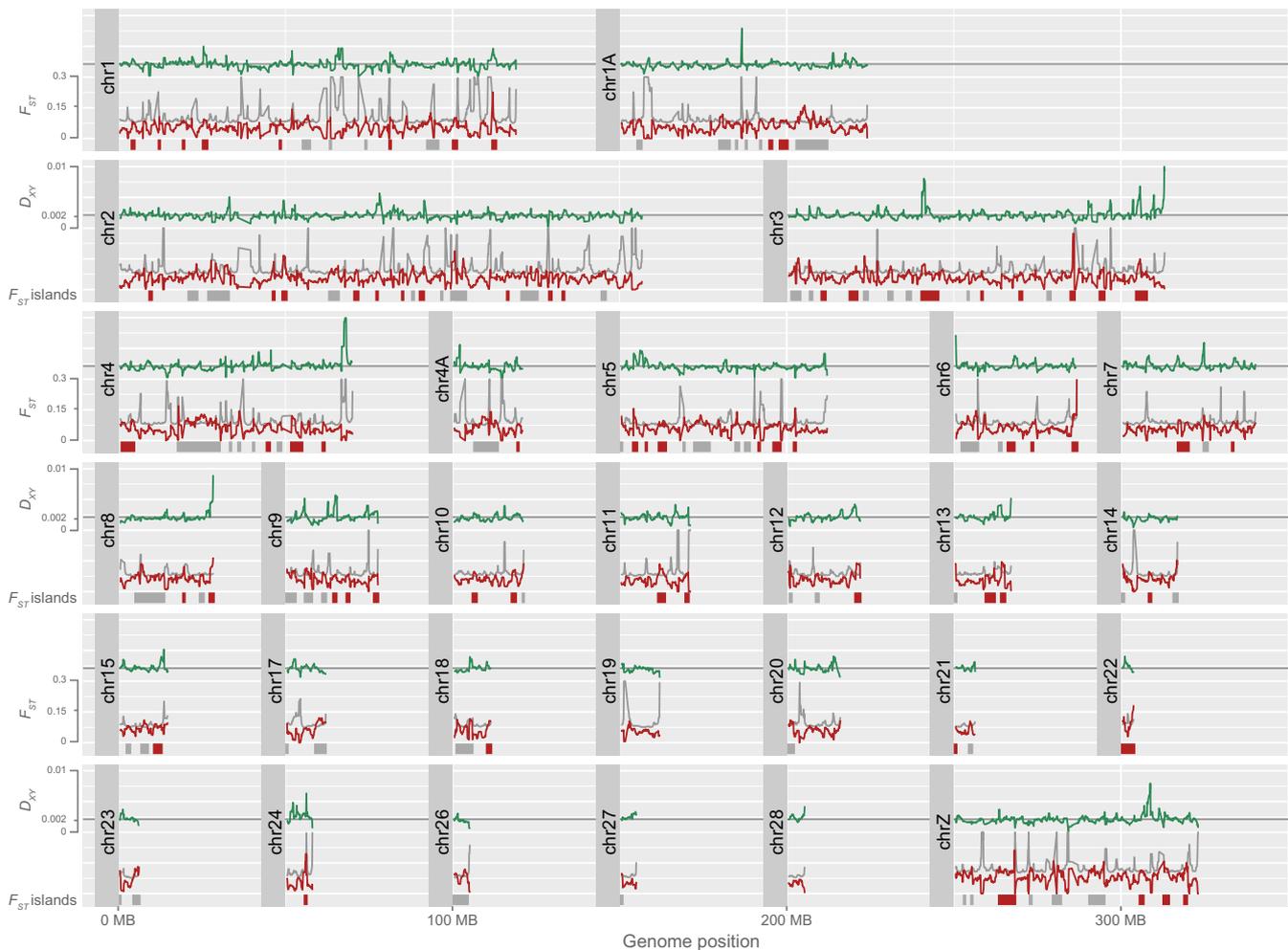


FIGURE 1 F_{ST} and D_{xy} values along the chromosomes. F_{ST} is drawn by red line and D_{xy} by green line. Grey lines represent the genomic average for D_{xy} and bootstrap values for F_{ST} . Genomic position is based on the mapping of nightingale contigs to the zebra finch genome. Genomic islands with significantly elevated F_{ST} values are depicted as grey or red bars. Red bars indicate genomic islands of elevated F_{ST} with average D_{xy} higher than the genome average

3.2 | Functional analysis of genes in islands of high differentiation and divergence

To better understand the nature of reproductive isolation between the nightingale species and to identify candidate speciation genes, we performed functional analyses of genes located in (i) islands of high differentiation and (ii) islands of high differentiation and divergence. We used KEGG pathways as a proxy for protein function and conducted gene set enrichment analyses. Altogether, we found three KEGG pathways significantly enriched in islands of high differentiation after FDR correction (Table S2A). These pathways included metabolic pathway (tgu01100), lysosome pathway (tgu04142) and oocyte meiosis pathway (tgu04114). When we focused the analysis on islands of high differentiation and divergence, no pathway was enriched after FDR correction. However, when we considered pathways significantly enriched in these islands without FDR correction (excluding pathways with fewer than 10 genes in islands to reduce the effect of false positives), we identified four enriched pathways (Table S2B). Interestingly, these included metabolic pathway (tgu01100), oocyte meiosis pathway (tgu04114), lysosome pathway

(tgu04142) and purine metabolism pathway (tgu00230). We were further interested whether any of the three pathways that were significantly enriched in islands of high differentiation as well as (without FDR correction) in islands of both high differentiation and divergence were also enriched in islands of high differentiation and low divergence (i.e., with D_{xy} lower than the genomic average). We found that metabolic and lysosome pathways were significantly (without FDR correction) enriched in these islands. Interestingly, oocyte meiosis pathway was not enriched in these islands (Table S2C), suggesting that the enrichment of this pathway in islands of high differentiation is driven mainly by higher density of genes related to this pathway in islands of high differentiation and divergence.

There were 17 genes belonging to the oocyte meiosis pathway in islands of high differentiation and divergence (Table 1). These genes represent candidate hybrid female sterility genes and are located on several autosomes. None of them are located on the Z chromosome. We also looked at their position in the oocyte meiosis pathway (Figure S1). The genes function at different stages of oogenesis from prophase I of meiosis I until meiosis II.

TABLE 1 List of 17 genes of the oocyte meiosis pathway that were found in the genomic islands of high differentiation and divergence

Ensembl gene ID	Gene name	Gene description	Location in the pathway	Chr	Gene start	Gene end	Strand
ENSTGUG00000009440	CDC16	Cell division cycle 16	APC/C-Cdc20 complex	1	26,694,321	26,715,128	+
ENSTGUG00000007142	ADCY1	Adenylate cyclase 1	AC	2	77,509,308	77,671,457	+
ENSTGUG00000011936	CCNE2	Cyclin E2	CycE-Cdk2 complex	2	132,836,273	132,856,217	-
ENSTGUG00000004938	PPP2R5D	Protein phosphatase 2 regulatory subunit B, delta	PP2A-B56 phosphatase	3	19,981,662	20,006,850	-
ENSTGUG00000005486	SPDYA	Speedy/RINGO cell cycle regulator family member A	Ringo-Cdc2	3	21,135,408	21,139,927	-
ENSTGUG00000001762	PPP2CB	Protein phosphatase 2 catalytic subunit beta	PP2A, PP2A-B56 phosphatase	4	2,417,725	2,421,587	-
ENSTGUG00000001917	MAD2L1	Mitotic arrest deficient 2-like 1	Mad1/2	4	4,749,027	4,751,057	-
ENSTGUG00000009328	ANAPC4	Anaphase promoting complex subunit 4	APC/C-Cdc20 complex	4	52,684,835	52,701,909	-
ENSTGUG00000009326	INS	Insulin	INS	5	13,622,537	13,625,762	+
ENSTGUG00000005982	CCNB2	Cyclin B2	CyclinB2/B5-Cdc2 complex (MPF)	10	6,469,106	6,472,886	-
ENSTGUG00000009363	MAP2K1	Mitogen-activated protein kinase kinase 1	MEK	10	19,125,593	19,183,532	+
ENSTGUG00000008504	ADCY7	Adenylate cyclase 7	AC	11	12,630,250	12,660,343	-
ENSTGUG00000001287	PPP2CA	Protein phosphatase 2 catalytic subunit alpha	PP2A, PP2A-B56 phosphatase	13	10,125,175	10,140,010	+
ENSTGUG00000001291	SKP1	S-phase kinase-associated protein 1	SCF-BTRC complex	13	10,154,467	10,160,268	+
ENSTGUG00000001417	PTTG1	Pituitary tumour-transforming 1	Securin	13	11,608,459	11,612,472	+
ENSTGUG00000002334	MAD2L2	Taeniopygia guttata MAD2 homolog (LOC100226442)	Mad1/2	21	907,423	914,246	-
ENSTGUG00000003464	CAMK2B	Calcium/calmodulin-dependent protein kinase II beta	CamKII	22	112,126	120,186	+

Location of these genes in the oocyte meiosis pathway is indicated (see also Figure S1). Gene information was retrieved using the ENSEMBL BIOMART, ENSEMBL Version 81, July 2015. Physical location of genes in zebra finch genome is provided.

3.3 | Linkage disequilibrium and patterns of molecular evolution in the islands of high differentiation and divergence

We estimated linkage disequilibrium and performed MK tests for genomic regions inside and outside (i) all islands of high differentiation and (ii) all islands of high differentiation and divergence separately in both nightingale species. In both species, linkage disequilibrium (measured as r^2) was significantly higher inside than outside islands for both islands of high differentiation and islands of high differentiation and divergence (Table 2). All MK tests except one (*L. luscinia*, inside islands) were statistically significant (Table S1). The NI revealed that the ratio of nonsynonymous to synonymous mutations was higher for polymorphisms than for substitutions. Importantly, when the amount of synonymous and nonsynonymous variation within and between species was compared inside and outside islands using glm models (see Section 2), we found a statistically significant difference in patterns of molecular evolution for islands of high differentiation in both species (*L. luscinia*: $\Delta LL = -4.4178$, p -value = .03557; *L. megarhynchos*: $\Delta LL = -6.9914$, p -value = .0082, Table S1), but not for islands of high differentiation and divergence (Table S1). Islands of high differentiation were characterized by an excess of nonsynonymous polymorphisms relative to nonsynonymous substitutions (Table S1).

4 | DISCUSSION

We conducted a genome-wide survey of differentiation and divergence to study the genetic basis of reproductive isolation between *L. megarhynchos* and *L. luscinia*, two closely related species that diverged 1.8 Mya (Storchová et al., 2010) and are separated by incomplete prezygotic isolation and female-limited hybrid sterility. Using transcriptomic data from multiple individuals of both species, we identified 136 regions of elevated differentiation (F_{ST}) across the genome, 67 of which had absolute divergence (D_{xy}) higher than the

TABLE 2 Linkage disequilibrium (measured as r^2) inside and outside (i) islands of high differentiation (F_{ST}) and (ii) islands of high differentiation and divergence (F_{ST} & D_{xy}) in both species (LL: *L. luscinia*, LM: *L. megarhynchos*)

Type of island	Species	Position	r^2	U	p -value
F_{ST}	LL	Inside	.1667		
		Outside	.1406	27,970	.0327
	LM	Inside	.4032		
		Outside	.2984	21,977	.0176
F_{ST} & D_{xy}	LL	Inside	.1667		
		Outside	.1432	26,106	.0256
	LM	Inside	.4145		
		Outside	.2954	22,142	.0002

Difference between r^2 inside and outside the islands was tested by Mann–Whitney U test.

genomic average. This subset of islands with high differentiation and divergence identifies candidate genomic regions involved in maintaining reproductive isolation between the nightingale species. To identify specific candidate genes and to gain insight into the molecular mechanisms underlying reproductive isolation, we analysed gene content and studied patterns of molecular evolution in these regions.

Three pathways (metabolic pathway, lysosome pathway and oocyte meiosis pathway) were significantly enriched in islands of high differentiation (Table S2). The same three pathways together with the purine metabolism pathway were also enriched in islands of high differentiation and divergence, although in this case the enrichment was not significant after FDR correction for multiple comparisons (Table S2). Intriguingly, the oocyte meiosis pathway could be related to female sterility, which is the main reproductive barrier observed between the nightingale species. We identified 17 genes located in islands of high differentiation and divergence that were associated with the oocyte meiosis pathway (Table 1). These are among the first candidate genes for female-limited hybrid sterility in taxa with heterogametic females.

Some of the candidate genes operate in the early stages of meiosis. This agrees with our current understanding of hybrid female sterility in nightingales. Experimental crosses of both species in captivity (Stadie, 1991) and rare observations of hybrid females in nature (Reifová et al., 2011a) show that hybrid females do not lay eggs, suggesting that sterility may be caused by an early meiotic arrest. In addition, hybrid females show impaired sexual behaviour and a reduced tendency to mate which might indicate reduced levels of sex hormones produced by the ovaries (Reifová et al., 2011a; Stadie, 1991). More detailed analysis of the ovaries of F_1 females is needed to better understand the underlying mechanisms of hybrid female sterility in these species.

The sex chromosomes are expected to play a crucial role in hybrid female sterility and Haldane's Rule (Presgraves, 2008). We previously showed that levels of between-species differentiation are higher and that levels of interspecific gene flow are lower on the Z chromosome compared to the autosomes (Storchová et al., 2010). The present results are consistent in showing higher average levels of differentiation on the Z chromosome compared to the autosomes. However, islands of high differentiation were not markedly enriched on the Z chromosome, and none of the candidate genes for hybrid female sterility was located on the Z chromosome. It is difficult to draw strong inferences from this observation, as it likely reflects the conservative nature of the analysis in which significance for each genomic region on the Z chromosome was assessed only in relation to the distribution of regions within this chromosome. As the Z chromosome showed higher levels of differentiation overall, it was more difficult to detect significant outliers. Direct mapping experiments would be useful for determining whether QTL for female sterility are located on the Z chromosome.

The two nightingale species diverged 1.8 Mya and although hybrid female sterility is an important reproductive barrier separating the two nightingale species, there are also other forms of reproductive isolation. Competition between the two species has resulted in

partial habitat segregation (Reif, Reifová, Skoracka, & Kuczyński, 2018) and ecological character displacement in bill size (Reifová et al., 2011b), which might enhance prezygotic as well as extrinsic postzygotic isolation in sympatric populations. There is also evidence for marked differences in sperm length between the species, which might contribute to postmating prezygotic isolation (T. Albrecht, K. Opletalová, J. Reif, V. Janoušek, L. Piálek, E.R.A. Cramer, A. Johnsen, & R. Reifová, unpublished results). In addition, the two nightingale species differ in breeding distribution; *L. megarhynchos* has more southern breeding distribution, while *L. luscinia* is distributed more northerly. In principle, this could be associated with differences in metabolic rates between the two species and could result in disrupted metabolism in hybrids (Qvarnström, Ålund, McFarlane, & Sirkkiä, 2016) as has been demonstrated in two flycatcher species (McFarlane, Sirkkiä, Ålund, & Qvarnström, 2016). Possible adaptations to different climates in the two nightingale species could provide an explanation for why the metabolic pathway is enriched in islands of high differentiation and islands of high differentiation and divergence. The lysosome pathway could be also related to different climatic adaptations as it is involved in energy metabolism (Settembre, Fraldi, Medina, & Ballabio, 2013), and in fact, there is some overlap between the metabolic and lysosome pathways.

Analysis of linkage disequilibrium revealed significantly higher levels of linkage disequilibrium inside than outside islands of high differentiation as well as islands of high differentiation and divergence in both nightingale species. In addition, islands of high differentiation were characterized by a significant excess of nonsynonymous polymorphism relative to nonsynonymous substitutions in both species. This suggests that purifying selection might be less efficient in these regions, although it should be noted that differences in linkage disequilibrium and effective population size in these islands could bias the results of MK tests (Eyre-Walker, 2002). Nevertheless, together, these observations suggest that some loci involved in reproductive isolation may lie in regions of locally reduced recombination. In such regions, selection at linked sites can lead to a locally reduced effective population size and less efficient purifying selection (Charlesworth et al., 1993; Maynard Smith & Haigh, 1974). This observation is consistent with theory that predicts an association between regions of low recombination and regions contributing to reproductive isolation (Butlin, 2005; Nachman & Payseur, 2012). It is also consistent with recent empirical studies in other species (Carneiro, Ferrand, & Nachman, 2009; Gerald, Basset, Smith, & Nachman, 2011; Janoušek, Munclinger, Wang, Teeter, & Tucker, 2015), suggesting that variation in recombination rate plays an important role in the evolution of reproductive isolation.

Our results together with previous studies (Delmore et al., 2015; Malinsky et al., 2015) demonstrate that analysis of the genomic landscape of differentiation and divergence is a promising approach for identifying genomic regions and candidate genes involved in speciation. The two nightingale species represent an especially suitable model system for this approach. These species diverged a relatively long time ago (1.8 Mya) yet show relatively high postdivergence interspecific gene flow (average $2Nm$ on autosomes is 0.422,

Storchová et al., 2010). This situation facilitates the identification of islands of high differentiation and divergence. Nightingales also provide a useful model system for studying mechanisms that underlie female-limited hybrid sterility. Further studies, including analysis of the ovary transcriptome of both nightingale species and hybrids, together with a detailed inspection of the ovaries of hybrid females, will shed more insight into female-limited hybrid sterility in organisms with heterogametic females.

ACKNOWLEDGEMENTS

We thank late Marcin Antczak for the assistance in the field and Megan Phifer-Rixey for useful discussion. The study was supported by a junior grant of the Czech Science Foundation (15-10884Y) to RR and an NIH grant (RO1 GM074245) to MWN. This work was also supported by ELIXIR CZ research infrastructure project (MEYS Grant No: LM2015047) including access to computing and storage facilities. We are grateful for financial support from the Czech Academy of Sciences (grant no. RVO 67985904) and the Ministry of Education, Youth and Sports of the Czech Republic (grant no. EXCELLENCE CZ.02.1.01/0.0/0.0/15_003/0000460 OP RDE).

DATA ACCESSIBILITY

DNA sequences, demultiplexed FASTQ files and accompanying sample identification are available from Dryad repository: <https://doi.org/10.5061/dryad.2p4t3>, <https://doi.org/10.5061/dryad.41ng6>. The analysis scripts are publicly available (<https://github.com/libor-m/islalnds-paper>).

AUTHOR CONTRIBUTION

The study was designed by R.R., L.M. and M.W.N. The samples were collected by R.R. and J.Re. The samples for sequencing were prepared by L.C. and K.J. The samples were sequenced by J.Ri. The data were analysed by L.M., V.J. and J.P. The manuscript was written by R.R., L.M., V.J. and M.W.N. with contribution of all coauthors.

ORCID

Lukáš Choleva  <http://orcid.org/0000-0002-3603-0266>

Radka Reifová  <http://orcid.org/0000-0001-5852-5174>

REFERENCES

- Ålund, M., Immler, S., Rice, A. M., & Qvarnström, A. (2013). Low fertility of wild hybrid male flycatchers despite recent divergence. *Biology Letters*, *9*, 20130169. <https://doi.org/10.1098/rsbl.2013.0169>
- Backström, N., Karaiskou, N., Leder, E. H., Gustafsson, L., Primmer, C. R., Qvarnström, A., & Ellegren, H. (2008). A gene-based genetic linkage map of the collared flycatcher (*Ficedula albicollis*) reveals extensive synteny and gene-order conservation during 100 million years of avian evolution. *Genetics*, *179*, 1479–1495. <https://doi.org/10.1534/genetics.108.088195>

- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57, 289–300.
- Bhattacharyya, T., Reifová, R., Gregorová, S., Simecek, P., Gergelits, V., Mistrik, M., ... Forejt, J. (2014). X chromosome control of meiotic chromosome synapsis in mouse inter-subspecific hybrids. *PLoS Genetics*, 10, e1004088. <https://doi.org/10.1371/journal.pgen.1004088>
- Burri, R. (2017). Interpreting differentiation landscapes in the light of long-term linked selection. *Evolution Letters*, 1, 118–131. <https://doi.org/10.1002/evl3.14>
- Burri, R., Nater, A., Kawakami, T., Mugal, C. F., Olason, P. I., Smeds, L., ... Hogner, S. (2015). Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. *Genome Research*, 25, 1656–1665. <https://doi.org/10.1101/gr.196485.115>
- Butlin, R. K. (2005). Recombination and speciation. *Molecular Ecology*, 14, 2621–2635. <https://doi.org/10.1111/j.1365-294X.2005.02617.x>
- Carneiro, M., Ferrand, N., & Nachman, M. W. (2009). Recombination and speciation: Loci near centromeres are more differentiated than loci near telomeres between subspecies of the European rabbit (*Oryctolagus cuniculus*). *Genetics*, 181, 593–606.
- Charlesworth, B. (1998). Measures of divergence between populations and the effect of forces that reduce variability. *Molecular Biology and Evolution*, 15, 538–543. <https://doi.org/10.1093/oxfordjournals.molbev.a025953>
- Charlesworth, B., Morgan, M. T., & Charlesworth, D. (1993). The effect of deleterious mutations on neutral molecular variation. *Genetics*, 134, 1289–1303.
- Coyne, J. A., & Orr, H. A. (2004). *Speciation*. Sunderland, MA: Sinauer Associates.
- Cruikshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, 23, 3133–3157. <https://doi.org/10.1111/mec.12796>
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... McVean, G. (2011). The variant call format and VCFtools. *Bioinformatics*, 27, 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Delmore, K. E., Hübner, S., Kane, N. C., Schuster, R., Andrew, R. L., Câmara, F., ... Irwin, D. E. (2015). Genomic analysis of a migratory divide reveals candidate genes for migration and implicates selective sweeps in generating islands of differentiation. *Molecular Ecology*, 24, 1873–1888. <https://doi.org/10.1111/mec.13150>
- Ellegren, H., Smeds, L., Burri, R., Olason, P. I., Backström, N., Kawakami, T., ... Uebbing, S. (2012). The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature*, 491, 756–760.
- Eroukhanoff, F., Rowe, M., Cramer, E. R., Haas, F., Hermansen, J. S., Runemark, A., ... Sætre, G. P. (2016). Experimental evidence for ovarian hypofunction in sparrow hybrids. *Avian Research*, 7, 3. <https://doi.org/10.1186/s40657-016-0038-1>
- Eyre-Walker, A. (2002). Changing effective population size and the McDonald-Kreitman test. *Genetics*, 162, 2017–2024.
- Gelter, H. P., Tegelström, H., & Gustafsson, L. (1992). Evidence from hatching success and DNA fingerprinting for the fertility of hybrid Pied × Collared Flycatchers *Ficedula hypoleuca* × *albicollis*. *Ibis*, 134, 62–68.
- Geraldes, A., Basset, P., Smith, K. L., & Nachman, M. W. (2011). Higher differentiation among subspecies of the house mouse (*Mus musculus*) in genomic regions with low recombination. *Molecular Ecology*, 20, 4722–4736. <https://doi.org/10.1111/j.1365-294X.2011.05285.x>
- Good, J. M., Giger, T., Dean, M. D., & Nachman, M. W. (2010). Widespread over-expression of the X chromosome in sterile F₁ hybrid mice. *PLoS Genetics*, 6, e1001148. <https://doi.org/10.1371/journal.pgen.1001148>
- Haldane, J. B. S. (1922). Sex-ratio and unisexual sterility in hybrid animals. *Journal of Genetics*, 12, 101–109. <https://doi.org/10.1007/BF02983075>
- Harr, B. (2006). Genomic islands of differentiation between house mouse subspecies. *Genome Research*, 16, 730–737. <https://doi.org/10.1101/gr.5045006>
- Herrero, J., Muffato, M., Beal, K., Fitzgerald, S., Gordon, L., Pignatelli, M., ... Flicek, P. (2016). Ensembl comparative genomics resources. *Database: The Journal of Biological Databases and Curation*, 2016, bav096. <https://doi.org/10.1093/database/bav096>
- Janoušek, V., Munclinger, P., Wang, L., Teeter, K. C., & Tucker, P. K. (2015). Functional organization of the genome may shape the species boundary in the house mouse. *Molecular Biology and Evolution*, 32, 1208–1220. <https://doi.org/10.1093/molbev/msv011>
- Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K., & Mooers, A. O. (2012). The global diversity of birds in space and time. *Nature*, 491, 444–448. <https://doi.org/10.1038/nature11631>
- Jiggins, C. D., Linares, M., Naisbit, R. E., Salazar, C., Yang, Z. H., & Mallet, J. (2001). Sex-linked hybrid sterility in a butterfly. *Evolution*, 55, 1631–1638. <https://doi.org/10.1111/j.0014-3820.2001.tb00682.x>
- Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 28, 27–30. <https://doi.org/10.1093/nar/28.1.27>
- Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*, 44, D457–D462. <https://doi.org/10.1093/nar/gkv1070>
- Kaplan, N. L., Hudson, R. R., & Langley, C. H. (1989). The “hitchhiking effect” revisited. *Genetics*, 123, 887–899.
- Kawakami, T., Backström, N., Burri, R., Husby, A., Olason, P., Rice, A. M., ... Ellegren, H. (2014). Estimation of linkage disequilibrium and inter-specific gene flow in *Ficedula* flycatchers by a newly developed 50k single-nucleotide polymorphism array. *Molecular Ecology Resources*, 14, 1248–1260. <https://doi.org/10.1111/1755-0998.12270>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... Durbin, R. (2009). The sequence alignment/map (SAM) format and SAMtools. *Bioinformatics*, 25, 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Malinsky, M., Challis, R. J., Tyers, A. M., Schifffels, S., Terai, Y., Ngatunga, B. P., ... Turner, G. F. (2015). Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science*, 350, 1493–1498. <https://doi.org/10.1126/science.aac9927>
- Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., ... Dewell, S. B. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437, 376–380. <https://doi.org/10.1038/nature03959>
- Masly, J. P., & Presgraves, D. C. (2007). High-resolution genome-wide dissection of the two rules of speciation in *Drosophila*. *PLoS Biology*, 5, e243. <https://doi.org/10.1371/journal.pbio.0050243>
- Maynard Smith, J., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetics Research*, 23, 23–35. <https://doi.org/10.1017/S0016672300014634>
- McDonald, J. H., & Kreitman, M. (1991). Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature*, 351, 652–654. <https://doi.org/10.1038/351652a0>
- McFarlane, S. E., Sirkiä, P. M., Ålund, M., & Qvarnström, A. (2016). Hybrid dysfunction expressed as elevated metabolic rate in male *Ficedula* Flycatchers. *PLoS ONE*, 11, e0161547. <https://doi.org/10.1371/journal.pone.0161547>
- Mihola, O., Trachtulec, Z., Vlček, C., Schimenti, J. C., & Forejt, J. (2009). A mouse speciation gene encodes a meiotic histone H3 methyltransferase. *Science*, 323, 373–375. <https://doi.org/10.1126/science.1163601>
- MorŤovský, L., Pačes, J., Rídl, J., & Reifová, R. (2015). Scrimmer: Designing primers from transcriptome data. *Molecular Ecology Resources*, 15, 1415–1420.
- Nachman, M. W., & Payseur, B. A. (2012). Recombination rate variation and speciation: Theoretical predictions and empirical results from rabbits and mice. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367, 409–421. <https://doi.org/10.1098/rstb.2011.0249>

- Naisbit, R. E., Jiggins, C. D., Linares, M., Salazar, C., & Mallet, J. (2002). Hybrid sterility, Haldane's rule and speciation in *Heliconius cydno* and *H. melpomene*. *Genetics*, *161*, 1517–1526.
- Nei, M. (1987). *Molecular evolutionary genetics*. New York, NY: Columbia University Press.
- Noor, M. A., & Bennett, S. M. (2009). Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity*, *103*, 439–444. <https://doi.org/10.1038/hdy.2009.151>
- Phifer-Rixey, M., Bomhoff, M., & Nachman, M. W. (2014). Genome-wide patterns of differentiation among house mouse subspecies. *Genetics*, *198*, 283–297. <https://doi.org/10.1534/genetics.114.166827>
- Presgraves, D. C. (2008). Sex chromosomes and speciation in *Drosophila*. *Trends in Genetics*, *24*, 336–343. <https://doi.org/10.1016/j.tig.2008.04.007>
- Presgraves, D. C. (2010). The molecular evolutionary basis of species formation. *Nature Reviews Genetics*, *11*, 175–180. <https://doi.org/10.1038/nrg2718>
- Qvarnström, A., Ålund, M., McFarlane, S. E., & Sirkkiä, P. M. (2016). Climate adaptation and speciation: Particular focus on reproductive barriers in *Ficedula* flycatchers. *Evolutionary Applications*, *9*, 119–134. <https://doi.org/10.1111/eva.12276>
- R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rand, D. M., & Kann, L. M. (1996). Excess amino acid polymorphism in mitochondrial DNA: Contrasts among genes from *Drosophila*, mice, and humans. *Molecular Biology and Evolution*, *13*, 735–748. <https://doi.org/10.1093/oxfordjournals.molbev.a025634>
- Reif, J., Jiran, M., Reifová, R., Vokurková, J., Dolata, P. T., Petrusek, A., & Petrusková, T. (2015). Interspecific territoriality in two songbird species: Potential role of song convergence in male aggressive interactions. *Animal Behaviour*, *104*, 131–136. <https://doi.org/10.1016/j.anbehav.2015.03.016>
- Reif, J., Reifová, R., Skoracka, A., & Kuczyński, L. (2018). Competition-driven niche segregation on a landscape scale: Evidence for escaping from sympatry toward allopatry in two coexisting sibling passerine species. *Journal of Animal Ecology*. <https://doi.org/10.1111/1365-2656.12808>
- Reifová, R., Kverek, P., & Reif, J. (2011a). The first record of a female hybrid between the Common Nightingale (*Luscinia megarhynchos*) and the Thrush Nightingale (*Luscinia luscinia*) in nature. *Journal of Ornithology*, *152*, 1063–1068. <https://doi.org/10.1007/s10336-011-0700-7>
- Reifová, R., Reif, J., Antczak, M., & Nachman, M. W. (2011b). Ecological character displacement in the face of gene flow: Evidence from two species of nightingales. *BMC Evolutionary Biology*, *11*, 138. <https://doi.org/10.1186/1471-2148-11-138>
- Sætre, G. P., Borge, T., Lindroos, K., Haavie, J., Sheldon, B. C., Primmer, C., & Syvänen, A. C. (2003). Sex chromosome evolution and speciation in *Ficedula* flycatchers. *Proceedings of the Royal Society B: Biological Sciences*, *270*, 53–59. <https://doi.org/10.1098/rspb.2002.2204>
- Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., ... Brelsford, A. (2014). Genomics and the origin of species. *Nature Reviews Genetics*, *15*, 176–192. <https://doi.org/10.1038/nrg3644>
- Settembre, C., Fraldi, A., Medina, D. L., & Ballabio, A. (2013). Signals from the lysosome: A control centre for cellular clearance and energy metabolism. *Nature Reviews Molecular Cell Biology*, *14*, 283–296. <https://doi.org/10.1038/nrm3565>
- Shagin, D. A., Rebrikov, D. V., Kozhemyako, V. B., Altshuler, I. M., Shcheglov, A. S., Zhulidov, P. A., ... Lukyanov, S. (2002). A novel method for SNP detection using a new duplex-specific nuclease from crab hepatopancreas. *Genome Research*, *12*, 1935–1942. <https://doi.org/10.1101/gr.547002>
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., ... Ideker, T. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research*, *13*, 2498–2504. <https://doi.org/10.1101/gr.1239303>
- Slater, G. S., & Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*, *6*, 31. <https://doi.org/10.1186/1471-2105-6-31>
- Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P. L., & Ideker, T. (2011). Cytoscape 2.8: New features for data integration and network visualization. *Bioinformatics*, *27*, 431–432. <https://doi.org/10.1093/bioinformatics/btq675>
- Stadie, C. (1991). Erdsanger I; Nachtigall und Sprosser. *Europäische Vogelwelt*, *3*, 130–189.
- Storchová, R., Gregorová, S., Buckiová, D., Kyselova, V., Divina, P., & Forejt, J. (2004). Genetic analysis of X-linked hybrid sterility in the house mouse. *Mammalian Genome*, *15*, 515–524.
- Storchová, R., Reif, J., & Nachman, M. W. (2010). Female heterogamety and speciation: Reduced introgression of the Z chromosome between two species of nightingales. *Evolution*, *64*, 456–471. <https://doi.org/10.1111/j.1558-5646.2009.00841.x>
- Turner, T. L., Hahn, M. W., & Nuzhdin, S. V. (2005). Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biology*, *3*, e285. <https://doi.org/10.1371/journal.pbio.0030285>
- Via, S., & West, J. (2008). The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular Ecology*, *17*, 4334–4345. <https://doi.org/10.1111/j.1365-294X.2008.03921.x>
- Vilella, A. J., Severin, J., Ureta-Vidal, A., Heng, L., Durbin, R., & Birney, E. (2009). EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. *Genome Research*, *19*, 327–335.
- Vokurková, J., Petrusková, T., Reifová, R., Kozman, A., Mořkovský, L., Kipper, S., ... Petrusek, A. (2013). The causes and evolutionary consequences of mixed singing in two hybridizing songbird species (*Luscinia* spp.). *PLoS ONE*, *8*, e60172. <https://doi.org/10.1371/journal.pone.0060172>
- Wolf, J. B., & Ellegren, H. (2017). Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics*, *18*, 87–100. <https://doi.org/10.1038/nrg.2016.133>
- Wright, S. (1951). The genetical structure of populations. *Annals of Eugenics*, *15*, 323–354.
- Wu, C. I. (2001). The genic view of the process of speciation. *Journal of Evolutionary Biology*, *14*, 851–865. <https://doi.org/10.1046/j.1420-9101.2001.00335.x>
- Yates, A., Akanni, W., Amode, M. R., Barrell, D., Billis, K., Carvalho-Silva, D., ... Girón, C. G. (2016). Ensembl 2016. *Nucleic Acids Research*, *44*, D710–D716. <https://doi.org/10.1093/nar/gkv1157>
- Zhang, L., Sun, T., Woldesellassie, F., Xiao, H., & Tao, Y. (2015). Sex ratio meiotic drive as a plausible evolutionary mechanism for hybrid male sterility. *PLoS Genetics*, *11*, e1005073. <https://doi.org/10.1371/journal.pgen.1005073>
- Zhu, Y. Y., Machleder, E. M., Chenchik, A., Li, R., & Siebert, P. D. (2001). Reverse transcriptase template switching: A SMART approach for full-length cDNA library construction. *BioTechniques*, *30*, 892–897.
- Zhulidov, P. A., Bogdanova, E. A., Shcheglov, A. S., Vagner, L. L., Khaspekov, G. L., Kozhemyako, V. B., ... Shagin, D. A. (2004). Simple cDNA normalization using kamchatka crab duplex-specific nuclease. *Nucleic Acids Research*, *32*, e37. <https://doi.org/10.1093/nar/gnh031>

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Mořkovský L, Janoušek V, Reif J, et al. Genomic islands of differentiation in two songbird species reveal candidate genes for hybrid female sterility. *Mol Ecol*. 2018;27:949–958. <https://doi.org/10.1111/mec.14479>