Extensive transgressive gene expression in a homoploid hybrid species

Homa Papoli Yazdi[1], Melissah Rowe[2], Mark Ravinet[3], Glenn-Peter Sætre[4], Caroline Øien Guldvog[4], Alfonso Marzal[5], Anna Runemark[1]

[1] Department of Biology, Lund University, SE-223 62 Lund, Sweden

[2] Department of Animal Ecology, Netherlands Institute of Ecology (NIOO-KNAW), 6700 AB, Wageningen, The Netherlands

[3] School of Life Sciences, University of Nottingham, Nottingham, UK

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

[5] Departamento de Anatomía, Biología Celular y Zoología, Universidad de Extremadura, E-506071, Badajoz, Spain

Abstract

Hybridization is increasingly recognized as an important evolutionary force contributing novel variation for selection to act on. While mis-expression in F1 hybrids is well documented, how gene expression evolves in stabilized hybrid taxa and contributes novel variation remains an open question, especially for hybrid species without an increase in ploidy. As gene expression evolves in a stabilizing manner, break-up of co-evolved *cis*- and *trans*-regulatory elements could lead to transgressive patterns of gene expression in hybrids. Here, we address to what extent gonad gene expression has evolved in an old homoploid hybrid, Italian sparrow, *Passer italiae*. Through comparing the gene expression of parental species and F1 hybrids to that of the Italian sparrow, we find evidence for strongly transgressive expression in the Italian sparrow, with 22% of the testis genes exhibiting expression patterns outside the range of both parent species, compared to only 0.37% in the F1s. In contrast, Italian sparrow ovary expression was similar to that of one parent species, the house sparrow (*P. domesticus*). Moreover, the Italian sparrow testis transcriptome is 26.2 -26.6 times as diverged from those of the parent species compared to how divergent their transcriptomes are in spite of it being genetically intermediate. This highlights the potential for regulation of gene expression to produce novel variation following hybridization. Genes involved in energy production and protein synthesis are enriched in the subset that is over-dominantly expressed in Italian sparrow testis, suggesting that selection on key functions have molded the hybrid transcriptome.

Introduction

Hybridization is an important contributor to novel genetic variation, providing the raw material that natural selection can act on (1-4). While divergence in parental genomes brings novel allelic combinations in the hybrid, which may lead to a reduction in hybrid fitness (5, 6), it is now clear that hybridization is frequent, rendering phylogenetic tree to appear more network like than previously recognized (7). Hence, we can no longer study hybridization as an isolated phenomenon, and the processes at play in hybrids should be integrated in our general understanding of how novel variation can arise. Changes in gene expression is increasingly being recognized as an important process for establishing reproductive isolation (8-10). Patterns of gene expression has been widely investigated in F1 hybrids between mostly evolutionarily strongly diverged lineages. However, we lack an understanding of how gene expression contributes to novel phenotypes and adaptation in old, stabilized hybrid taxa, in particular for homoploid hybrid species, with no increase in ploidy level.

Patterns of gene expression divergence between the parental species might not be a good indicator for expression pattern in the hybrids. The reason for that is gene expression is under stabilizing selection, hence, regulatory elements diverge in a way that keep expression levels at an optimum level for physiological functions (11-13). The impact of the divergence in these *cis*-(proximate) and t*rans*-(distal) regulatory elements can then only be revealed if these elements come together from different parental species in the hybrid. Combination of these regulatory elements in the hybrids might lead to transgressive patterns of expression where the gene expression in the hybrid is either lower or higher than that of both parent species. Indeed, transgressive gene expression has been reported in F1- and F2 hybrids in several taxa, and the underlying regulatory divergence for transgressive expression has been associated

with divergence in *cis*- and *trans*-regulatory elements in the parental species (9, 14-17). Patterns of gene expression in post-F1 generation hybrids show a higher level of transgressive expression due to uncoupling of regulatory elements in the process of recombination. In Lake whitefish, the F2 hybrids showed a higher level of nonadditive inheritance in gene expression compared to F1 hybrids (9). In interpopulation hybrids at F2 or later generation of copepod species, *Tigriopus californicus*, gene expression showed transgressive inheritance only in about 1.2% of the transcriptome with many genes involved in mito-nuclear interactions (18). Moreover, recent studies on allopolyploid plant taxa have shown preferential expression of one of the parental genomes may arise (19), and that this dominance may vary with the tissue investigated (20). This suggests that gene expression will not be proportional to the parental contributions to the genome and could hence cause transgressive gene expression from intermediate genome composition, but whether transgressive patterns of gene expression may arise in homoploid hybrid taxa is not known.

Moreover, most F1-hybrids studied stem from crosses of evolutionarily strongly diverged parental species, with a large number of fixed differences. Such fixed differences have been extensively investigated as candidate incompatibility loci. However, transgressive expression patterns in the hybrid of recently diverged species pairs can either arise due to polymorphic incompatible loci or physiological response to hybrid dysfunction (18, 21). For instance, a recent study on the F1 hybrid between the pied and collared flycatchers, showed a tissue-specific pattern of mis-expression with testis showing the highest divergence between the parental species but the least amount of mis-expression in F1 hybrids (22). To which extent this pattern holds up across species, and how sorting of segregating polymorphisms may result in novel patterns of gene expression remains a challenge for researchers to address.

Here, we leverage a unique study system to address how gene expression evolves in a homoploid hybrid species, the Italian sparrow (*Passer italiae*) (Figure 1) through combining data from the old, stabilized hybrid taxon, the parental species and experimental F1-hybrids. This enables us to test to what extent the Italian sparrow,a homoploid hybrid species that has arisen from hybridization between the house sparrow (*P. domesticus*) and Spanish sparrow (*P. hispaniolensis*) (23-26) has diverged in gene expression from the parent species. We compared the gene expression profiles of the gonads of the Italian sparrow to gene expression of the parental species and lab bred Experimental F1 hybrids to address a) if Italian sparrows exhibit transgressive gene expression compared to the parent species and b) whether inheritance mode and misexpression in F1s predict which genes are transgressively expressed in Italian sparrows. Conserved expression level between parents controlled by compensatory variation in parental species is often associated with misregulation in their hybrids (27-29). By using sites in experimental F_1 hybrids that contain both parental alleles, we measured allele specific expression, and compared that to expression levels in the parental species to infer whether the divergence is due to *cis*-, *trans*- or compensatory evolution.

Materials and methods

Sampling, RNA extraction, library preparation, and sequencing

We compared the gene expression of Italian sparrows, which originated ca. 5800 years ago, most likely as the house sparrow expanded its range into Europe and hybridized with the Spanish sparrow (30), to its two parent species which diverged approximately 0.85 million years ago (30) and experimental F1-hybrids with house sparrow mothers and Spanish sparrow fathers. As the species differ slightly in phenology, birds were sampled during their peak reproductive periods at their respective locations. Spanish sparrows and F1-hybrids were sampled near Olivenza, Spain (38º 40' 56"N, 7º 11' 17"W) in March 2016 and 2017, house sparrows in Oslo, Norway (59º 55' 2"N, 10º 46' 11"E) in May 2016 and Italian sparrows from Montanari in the Gargano peninsula (41º 54' 36.8"N, 15º 51' 13.0"E) in May 2017 (Fig 1).

Experimental F1 hybrids were produced using house sparrow females and Spanish sparrow males. While the Experimental F1 hybrids were bred in outdoor cages, all birds were trapped using mist nets outside or within cages. Birds were immediately euthanized by cervical dislocation, and gonads (ovary and left testis) were harvested directly after death. Sampling was performed between 10:00 and 16:00, with individuals from each of the species being sampled over most of the day, in order to avoid biases in sampling time. The gonads were stored in RNAlater, and immediately stored at -80 degrees, with extractions later performed at the University of Oslo using the Qiagen miRNeasy micro kit. Samples were prepped with TruSeq® Stranded mRNA library prep from Illumina on an automated Perkin Elmer Sciclone NGSx liquid handler and sequenced in the Norwegian Sequencing Center in Oslo, as paired-end (2x150 bp) on the Illumina Hiseq4000. The number of reads obtained per each sample is

reported in Table S1 As F1-crosses were bred in an area where house- and Spanish sparrows are sympatric (Olivenza, Spain; see Fig 1) and correct identification of females can be challenging, we performed a post-hoc genomic analysis based on mitochondrial and genomic DNA to ensure that all individuals clustered with the expected species. (Supplementary materials and methods, Figure S1, S2).

Read mapping, quality check and counting

We mapped reads from samples of all four groups (house, Spanish, F1 hybrids and Italian sparrow) to the house sparrow reference genome using STAR2 v2.7.2b (31) with default parameters and the house sparrow gene annotation as reported in (24). Over 90% of the reads mapped successfully (see Table S1). The house sparrow general feature format (GFF) file contained 14,734 annotated genes of which 92.52% were anchored to chromosomes and 7.48% were located on unanchored scaffolds. Of the genes anchored to chromosomes, 12,595 genes were annotated on autosomes and 598 genes were on Z chromosome (Table S2). To ensure homogeneity in quality across our samples, we used RSeQC (32) to measure RNA integrity at the transcriptome level for each sample by calculating Transcript Integrity Number (TIN). The median TIN scores across samples ranged between 73.7 and 85.9 (Table S3). We counted the number of reads mapping to the house sparrow gene features using HTSeq v0.9.1 (33). Read counting was performed for all reads with minimum quality score of 30 and it was configured to handle reverse-stranded sequencing data, with the parameter controlling for overlapping gene features set to union. We created a dendrogram to show sample relationships. The main split clearly between tissue types, i.e. testes and ovaries (Figure S3).

Differential gene expression analysis

We analyzed the raw read counts obtained from htseq-count using the R package DESeq2 (34, 35). We first tested genes for differential expression between house and Spanish samples (house (testis: n = 5, ovary: n = 5) *vs.* Spanish (testis: n = 5, ovary: N = 3)). We then compared each of the parents to the experimental F1 hybrid (testis: n = 5, ovary: n = 8) and the Italian sparrow (testis: n = 5, ovary: n = 5). We pre-filtered the data to a minimum of total read counts of 10 across at least half of the samples in each pairwise comparison. To generate more accurate estimates of log2 fold change (LFC) for genes with low count number or large dispersion, we shrank LFC estimates. We considered genes to be differentially expressed if they showed False Discovery Rate (FDR) Padj< 0.05 and shrunken LFC > 0.32.

Classification of gene expression inheritance in experimental F1 hybrids and Italian sparrows

The mode of inheritance for differentially expressed genes was determined following (36). We normalized gene expression using median of ratios method in DESeq2 (35). Genes whose total expression in hybrids deviated significantly more than 1.25-fold (LFC > 0.32) from that of either parent were considered to have non-conserved inheritance, and were classified as having additive, dominant, under-dominant, or over-dominant inheritance, based on the magnitude of the difference between total expression in the experimental F1 hybrid / Italian sparrow and in each parental species. Genes for which expression in the experimental F1 hybrid / Italian sparrow was less than for house sparrows and greater than for Spanish sparrow (or vice versa) were classified as additive; genes for which expression in the experimental F1 hybrid / Italian sparrow was similar to one of the parents were classified as dominant; and genes for which expression in the experimental F1 hybrid / Italian sparrow was either greater than or less than both parent species were classified as over-dominant and

under-dominant, respectively (Fig 4A and B).

Allele-specific expression and *cis*- and *trans*- regulatory divergence assignment

Analysis of allele-specific expression (ASE) in hybrids is a powerful approach to study regulatory basis of gene expression evolution. A typical approach for the study of ASE is to select the sites for which the parental species are fixed for two different alleles and study the ASE in heterozygote hybrids (37). House and Spanish sparrows diverged only about 0.85 MYA with most of polymorphisms still shared between the two species. We therefore took an approach similar to the one used in other studies with species pairs containing high number of shared polymorphisms and few fixed differences (22). We performed the SNP calling and filtering as described in the section "Genetic Characterization of Samples" (Supplementary Materials and Methods) with the difference of filtering for DP of minimum 10, base quality of 20 and mapping quality of 60. We used the R package MBASED (38) in the unphased mode to identify genes with allele-specific expression. We filtered the MBASED results for the FDR < 0.05 and major allele frequency of minimum 0.7. To balance the power, we chose 5 individuals for each house, Spanish, F1 and Italian sparrow in testis and 3 individuals of each group for ovary. To identify allele-specific expression at the population level, we required genes to show significant allele-specific expression in at least two individuals. Following (39), we divided the regulatory categories as follow: Cis-regulatory divergence was attributed to genes that are differentially expressed between parents and show allele-specific expression in hybrids. Trans-regulatory divergence was assigned to genes that are differentially expressed between parents and do not show allele-specific expression in hybrids. Putative compensatory regulatory variation was determined if a gene 1) did not show differential expression between parents, 2) showed significant allele-specific expression in hybrids, and 3) did not show allele-specific expression in parents.

Functional Annotation and Enrichment Analysis

Gene ontology (GO) analysis was done for differentially expressed genes in each pairwise comparison. GO functional annotations and gene descriptions were obtained for protein sequences from the entire house sparrow protein set as described in (40) using PANNZER (41). Functional enrichment of GO terms present in the set of differentially expressed genes relative to the background consisting of all genes expressed and tested for differential expression was performed using the clusterProfiler (42) and significant enrichment was determined at FDR < 0.05. Interactions among proteins with significant GO terms were predicted using STRING-v11 (43) using "Experiment", "Databases" and "Co-expression" as interaction sources. We set the minimum required interaction score to the highest confidence (0.9) and used only the query proteins to build the interaction network after removing the disconnected nodes in the network. The obtained network was then exported to Cytoscape (44) and major representations of biological processes were detected using ClueGO (45) with Padj< 0.01.

Results

Differential expression between the parental species

Gene expression divergence in gonads between the parental species was strongly asymmetric with testis showing a more conserved pattern of expression compared to ovary (Table 1). In testis, 135 genes (1.16%) and in ovary, 1382 genes (11.65%) were significantly differentially expressed. Of the differentially expressed genes, 60.7% of genes in testis and 69.3% of genes in ovary were upregulated in the Spanish sparrow. A significantly higher proportion of genes that were differentially expressed in testis were located on the Z chromosome (24 genes, hypergeometric test, *P*-value < 1.82-08). We found no evidence for functional enrichment among differentially expressed genes in parental species for neither testis nor ovary.

Differential expression in Italian sparrows compared to the parental species

In spite of intermediate genome composition between the parent species, the Italian sparrow differed more in comparison to the parental species than the parent species differed to each other for testis (3536 and 3581 genes vs 135 genes that differed between parent species). About one-third of the testis transcriptome was differentially expressed compared to both parental species, whereas the ovary expression was similar to that of the house sparrow (Table 1, Figs. 2-3). In the comparisons to house and Spanish sparrow testis, 3536 genes (30.45%) and 3581 genes (30.9%) were differentially expressed, respectively. In ovary, only 22 genes (0.18%) differed from the house sparrow whereas 1508 genes (12.63%) were differentially expressed compared to the Spanish sparrow. Significant over-representation of Z-linked genes among the differentially expressed genes was detected in testis but only in the comparison to Spanish sparrow (196 of the 3581 genes, hypergeometric test, *p* value = 0.006). Most genes in both ovary and testis were up-regulated compared to house and down-

regulated compared to Spanish sparrow (Chi-squared test *p* value = 1.05e-06 in testis and 0.001 in ovary).

Genes that were differentially expressed in Italian sparrow testis were enriched for similar biological processes in comparison to both parent species. Primarily, functions involving protein synthesis and mitochondrial gene expression and function were overrepresented. In the comparison to house sparrow, 27 significant GO categories were found (Table S4), with top biological processes involved in viral transcription, SRP-dependent co-translational protein targeting to membrane, translation, ribosome and mitochondrial respiratory chain complex I assembly. In comparison to the Spanish sparrow, the 15 enriched GO terms reflected similar biological processes, with ribosome as the top GO term (Table S5). In ovary, 6 GO terms were detected in the comparison to house sparrow with sodium ion transmembrane transporter activity as the top term (Table S6) and 1 GO term, negative regulation of protein phosphorylation detected in the comparison to Spanish sparrow (Table S7).

Differential expression in F1 hybrids compared to the parental species

Both testis and ovary showed a more conserved pattern of expression between the F1 hybrid and the house sparrow than between F1 hybrids and the Spanish sparrow (Table 1, Figure 2 and 3). The overall magnitude of differences to the parent species was much lower than for the Italian sparrow. In testis, 106 genes (0.9%) and 263 genes (2.25%) were differentially expressed compared to house and Spanish sparrows, respectively. In ovary, 30 genes (0.25%) and 140 genes (1.15%) showed a significant difference in expression compared to the house and Spanish sparrows, respectively. Similar to Italian sparrow, the pattern of up- and down-regulation was consistent in both testis and ovary in the comparison to each of the parental

species. A larger number of genes were up-regulated in F1 hybrid in the comparison to house sparrow and down-regulated in the comparison to Spanish sparrow (Chi-squared test *P*-value = 1.96-05 in testis and 0.01 in ovary). As in Italian sparrow, there was an over-representation of Z-linked among genes differentially expressed between F1 hybrids and Spanish sparrows (31 of the 263 differentially expressed genes, hypergeometric test, *P*-value = 2.42-06). The enriched functions were different to those enriched for Italian sparrows. Testis genes differentially expressed in F1 showed functional enrichment for genes involved in primarily collagen catabolic process. In the comparison to house sparrow, 46 GO terms were significantly over-represented. Among these GO terms, 20 GO terms had an $P_{adj} < 0.01$, whereas there were few significant GO terms compared to Spanish sparrows with top GO term common with the comparison to house sparrow (Tables S8 and S9). The top 5 GO terms were enriched in proteinaceous extracellular matrix, extracellular matrix structural constituent, extracellular vesicle, collagen trimer and collagen catabolic process. In ovary, while there was no functionally significant term associated with differentially expressed genes in F1 compared to the house sparrow, there were 28 significant GO terms in the comparison to Spanish sparrow with the top category to be response to vitamin D (Table S10).

Classification of inheritance patterns of gene expression

Gene expression was generally conserved in both Italian sparrow and F1 hybrid (Table 2, Figure 4), but in the testis of Italian sparrow, 2611 genes (22.71%) showed a non-conserved pattern of inheritance with the house-dominant and transgressive expression being the two largest categories. Transgressively expressed genes in Italian sparrow comprised 22% of the testis transcriptome tested for inheritance. This means more than a fifth of the expressed genes in testis are expressed at either a higher or lower level compared to both parent species.

In contrast, in F1 testis only 0.37% of the genes were transgressively expressed. The largest category of inheritance in Italian sparrow ovary was the house-dominant pattern of expression with over 96% of genes with non-conserved pattern of inheritance showing a pattern of inheritance similar to house sparrow. Despite the strong transgressive inheritance pattern observed in Italian sparrow testis, only 3 genes (0.028%) were transgressively expressed in ovaries. Gene expression pattern in F1 ovary was more conserved compared to Italian sparrow. Out of 376 genes with non-conserved pattern of inheritance, the two largest categories of inheritance were additive and house-dominant inheritance pattern.

Overdominant genes in Italian sparrow testis were enriched for functional categories involved in protein synthesis, mitochondrial protein complex and gene expression and binding of sperm to zona pellucida. We found a gene network with significant protein-protein interaction (PPI Enrichment: 1.0E-16) among the overdominant genes in Italian sparrow (Figure 5). A different set of genes were underdominant in Italian sparrow, with functions including regulation of cellular component size and negative regulation of neuron project development. Overdominant genes in F1 testis were functionally enriched for different GO-terms, including collagen catabolic process and proteinaceous extracellular matrix. The only significant GO term for underdominant genes in F1 testis was voltage-gated potassium channel complex.

Comparison of gene expression divergence in respect to parent species between Italian sparrows and F1

To evaluate how gene expression evolves over time asking whether mis-expression in early generation hybrids predicts differential expression in Italian sparrows, we tested if the same genes or gene functions that were differentially expressed in comparison to parental species

in F1 hybrids also were differentially expressed in the Italian sparrow (Fig 2C and Fig 3C). In testis, 1.28% of the genes that were differentially expressed in Italian sparrows compared to house sparrow were also differentially expressed in experimental F1 hybrids, whereas in comparisons to Spanish sparrows it was 3.9%. In ovary, only 1 gene (2%) was differentially expressed in F1 and Italian sparrow compared to house sparrow, whereas 5.9% were differentially expressed in F1 hybrids and Italian sparrows in comparison to Spanish sparrows. The majority of the commonly differentially expressed genes were differentially expressed in the same direction (i.e., up- or down-regulated) in both testis and ovary, both in F1 hybrids and Italian sparrows, as reflected by a significant correlation of Log Fold Change (Pearson's R = 0.89, P-value < 2.2e-16 for house sparrow testis, R = 0.94, P-value < 2.2e-16 for Spanish sparrow testis R = 0.95, P-value < 2.2e-16 for Spanish ovary). Testis genes commonly differentially expressed in F1 hybrid and Italian sparrow compared to house sparrow were enriched in extracellular vesicle, insulin like growth factor binding, extracellular matrix structural constituent and extracellular matrix component. In the comparison to Spanish sparrow testis, only one significant GO term, cAMP binding was detected.

Mode of regulatory divergence and prediction of transgressive expression in Italian sparrows

To address if the regulatory basis can predict transgressive gene expression in hybrid species, we classified genes dependent on their regulatory basis (Table 3). We expect genes under *cis-trans* compensatory regulation and trans-regulated genes to more easily produce transgressive patterns of gene expression in later generation hybrids than *cis*-regulated genes, through de-coupling of *cis*- and *trans*- or different trans-elements in the sorting of parental variation in the hybrid lineage. Therefore, we checked the percentage of a gene to be transgressively

expressed in the Italian sparrow depending on the mode of regulatory divergence. We found that 16/130 genes (12%) of the *trans*-regulated testis genes were transgressively expressed in Italian sparrow, and 28/159 genes (18%) of the genes under *cis*-trans compensatory regulation were transgressively expressed in Italian sparrows. Finally, we tested whether genes with a non-additive inheritance, that were transgressively expressed in F1 hybrids, were more likely to be transgressively expressed in the Italian sparrow and found that 20 out of 45 transgressively inherited genes in F1 hybrid (44.4%) showed a transgressive pattern of inheritance in Italian sparrow testis.

Discussion

In this study, we aimed to document the evolution of gene expression in the gonads of a homoploid hybrid species, the Italian sparrow. An extensive level of transgressive expression was found in Italian sparrow testis, with an over-expression of genes involved in functions related to mitochondrial respiratory chain, cytosolic and mitochondrial ribosomal proteins and proteins involved in protein folding and the proteasome complex. In contrast to testis, the gene expression pattern in the ovary was asymmetric compared to parental species with a house-dominant expression in the Italian sparrow, potentially due to its house sparrow mitochondria. Experimental F1 hybrids had a much more conserved pattern of gene expression. In the following sections, we discuss the potential implications of these results and consider the limitations of the data.

Extensive transgressive expression in Italian sparrow testis

About one-fifth of testis transcriptome in Italian sparrow showed transgressive expression pattern, while this level was less than 1% in the F1 hybrid. A higher level of transgressive expression despite a conserved gene expression profile between parents is expected to occur in later generation hybrids due to the uncoupling of co-evolved *cis*- and *trans*- regulatory elements through recombination (46, 47). There are few studies of gene expression in post F1-hybrids. In Lake whitefish, F2 hybrids showed higher misexpression and percentage of genes with nonadditive inheritance compared to F1 hybrids (9). Transgressive expression is commonly referred to "mis-expression" in early generation hybrids while Italian sparrow is a hybrid taxon formed within 5800 years ago (30). Here, we show that transgressive expression is found not only as mis-expression in early generation hybrids, but remains in well-functioning homoploid hybrid taxa long after hybridization. Interestingly, the Italian

sparrow transcriptomes are generally much more divergent from both parent species than are the transcriptomes of the F1 hybrids, with the exception for the house sparrow like expression in the Italian sparrow ovary. Hence, evolution of the regulation of gene expression following hybridization may produce novel expression from intermediate genomes. In hybrid lineages, transgressive segregation which leads to extreme phenotypes has been considered as a major mechanism by which extreme or novel adaptations observed in new hybrid ecotypes or species are thought to arise (48), but sorting of regulatory elements have not formally been suggested to contribute to this. This research suggests that break-up of co-evolved regulatory compounds, resulting in transgressive patterns of gene expression partly can explain the generation of transgressive phenotypes in hybrid taxa.

Transgressive expression arising in spite of low parental divergence

Hybrid incompatibilities arisen due to bringing together novel genetic combinations have been mainly considered when the parental species are fixed for alleles potentially contributing to reproductive isolation (21). House and Spanish sparrows diverged less than one million years ago. Transgressive genes in Italian sparrow has a genome-wide distribution, hence, it is unlikely that the observed transgressive inheritance is driven by bringing together fixed differences in the Italian sparrow. In recently diverged species, reproductive incompatibility may arise due to polymorphic loci within each lineage before fixation for the incompatibility alleles (21). Therefore, if transgressive expression is driven in the Italian sparrow by bringing together *cis*- and *trans*-regulatory variants not yet co-adapted, this must either be by bringing together polymorphic loci in each of the parental species or novel mutations that have occurred in Italian sparrow since isolation from the parental species. Each of these hypotheses can be evaluated in future work by studying the F1 and F2 hybrids and the direct parents of the cross in order to find *cis* and *trans*-regulatory variations.

Interestingly, during ca. 5.800 years of evolution, Italian sparrows have achieved 26 of the evolution of gene expression the parents have accumulated during 0.68 million years, in spite of being intermediate genomically (24)

Tissue- and parent specific patterns of differentiation in gene expression

We find that the degree of difference in gene expression are tissue- and parent specific. How conserved gene expression was in the Italian sparrow differed between the gonads, with much more conserved patterns in ovary compared to testis. The ovary transcriptome of the Italian sparrow showed a house sparrow like pattern of expression. We generally observed fewer differentially expressed genes relative to house sparrow than relative to the Spanish sparrow, both for Italian sparrows and F1 hybrids, and especially for ovaries. This is likely partly explained by the fact that both F1 hybrids and Italian sparrows have house sparrow mitochondria. Mito-nuclear interactions can strongly affect gene expression, and even result in misexpression (49). An additional explanation could be a contribution of the W chromosome. While we can assume that F1 hybrids inherit house sparrow W chromosomes, there is not yet an assembly of the W chromosome of any of the three sparrow species. Investigating how the W chromosome affects gene expression is an interesting future line of inquiry. Finally, we cannot exclude that a technical artefact contributes to the house-dominant pattern of expression. Since only 3 Spanish sparrows were used in the analysis of ovary transcriptome, the result might be driven by large within-species variance in gene expression.

Enrichment of over-dominant genes in Italian sparrow for protein synthesis and respiratory chain enzymes

A large set of genes involved in protein synthesis were over-expressed in Italian sparrow testis compared to both parents. These include a set of ribonucleoproteins present in both cytosolic and mitochondrial ribosomes, eukaryote translation initiation factor such as *EIF1AY* required for maximal rate of protein synthesis and SRP (signal recognition particle), a ribonucleoprotein that recognizes and targets specific proteins to the endoplasmic reticulum in eukaryotes. Several genes coding for subunits of T-complex protein Ring Complex (TRiC) (*CCT4, CCT5, CCT6A, TCP1*) were also among the overdominant genes. TRiC complex aids in the folding of about 10% of the proteome. These genes also contribute to binding of sperm to zona pellucida and have shown evidence for positive selection among the seminal fluid genes between house and Spanish sparrows (40). Another category of gene networks which were upregulated in Italian sparrow included genes encoding for components of each of the four mitochondrial respiratory chain enzymatic complexes. These included genes involved in the NADH dehydrogenase, cytochrome *b-c1* and cytochrome oxidase complex and the last component of the electron transport chain, the enzyme which drives ATP synthesis through oxidative phosphorylation, the ATP synthase. Additionally, genes encoding for chaperone proteins involved in the assembly of these enzymatic complexes such as *TIMMDC1* were also over-expressed.

Overexpression of cytosolic ribonucleoproteins and OXPHOS process has also been observed in later generation hybrids (F3+) in copepod species *Tigriopus californicus* and has been linked to cellular responses to physiological dysfunction (18). In Italian sparrow testis, we observe an increase in gene expression in both cytosolic and mitochondrial ribonucleoproteins. Mito-nuclear genes have been shown to be important in generating

reproductive isolation in the Italian sparrow (25), and mito-nuclear sites have been under selection to inherit the house sparrow allele in Italian sparrows (50). Although mitochondrion has its own genome, more than 98% of its proteins are encoded in the nuclear genes. An increase in genes involved in protein synthesis might imply a response from mitochondria to changes in the physiological environment of the cell.

This study is the first in investigating evolution of gene expression in a homoploid hybrid species. Two important aspects of our study were comparing patterns of gene expression in homoploid hybrid individuals that contain mixed genetic background to those seen in laboratory crosses and the fact that alleles contributing to reproductive isolation may be polymorphic rather than fixed. Future studies should evaluate the current results in more tissues and by using pedigrees of parents and hybrid individuals, identify the regulatory underpinnings of gene expression divergence.

Acknowledgements:

We thank Fabrice Eroukhmanoff and Angelica Cuevas for assistance with sampling. The computations were enabled by resources in project SNIC2020-6-222 provided by the Swedish National Infrastructure for Computing (SNIC) at UPPMAX, partially funded by the Swedish Research Council through grant agreement no. 2018-05973. This work was financed by a Norwegian Research council grant to G-P. S. and A. R. and a Wenner-Gren and a Swedish Research Council grant to A.R.

Animal Ethics

In Spain, all trapping and sampling of birds was conducted in accordance with Spanish Animal Protection Regulation RD53/2013, and all methods were approved by the

Institutional Commission of Bioethics at the University of Extremadura (CBUE 49/2011). Birds were sampled in Oslo with permission from Miljø-direktoratet (2016/2225). In Italy, all birds were sampled with permission from ISPRA (protocol number 12404) and with permission (no. 305) from regione Puglia. All samples taken from Italy to Oslo in accordance with the Nagoya protocols.

Data access

The RNA-seq data generated in this study will be submitted to the NCBI BioProject database under accession number XXX upon the acceptance of the manuscript.

Scripts for the analyses

All scripts for the manuscripts are deposited in github under

github.com/Homap/Expression_sparrow

Author contributions:

A.R. and H.P.Y. conceived of the study. A.M. raised the hybrids and M. Ro., M. Ra., A.R., C.O. and G-P. S. performed field work and sampling, C.O. performed the molecular lab work, and H.P.Y. designed and performed the bioinformatical analysis based on discussions with A.R. H.P.Y. drafted the manuscript, with revisions based on the input from A.R., M.Ra and M.Ro.

Table 1. Number of differentially expressed genes and log2 fold change (LFC) with Padj< 0.05 and LFC* > 0.32.

	Testis			Ovary		
Comparison	**Significant**	**LFC > 0**	**LFC < 0**	**Significant**	**LFC > 0**	**LFC < 0**
Spain - House	135	82	53	1382	958	424
F_1 - House	106	77	29	30	16	14
F_1 - Spain	263	125	138	140	39	101
Italy - House	3536	1962	1574	22	13	9
Italy - Spain	3581	1779	1802	1508	394	1114

* LFC: Log Fold Change

Table 2. Number and percentage of genes in each inheritance category.

	Testis		Ovary	
Inheritance	**F_1**	**Italy**	**F_1**	**Italy**
Conserved	11515 (99.14%)	8883 (77.30%)	11600 (96.9%)	10889 (92.26%)
Additive	6 (0.05%)	2 (0.02%)	272 (2.27%)	19 (0.16%)
house dominant	45 (0.39%)	54 (0.47%)	94 (0.78%)	895 (7.58%)
Spain dominant	4 (0.03%)	25 (0.22%)	6 (0.05%)	7 (0.06%)
Underdominant	12 (0.1%)	1205 (10.47%)	2 (0.17%)	3 (0.02%)
Overdominant	33 (0.27%)	1325 (11.53%)	2 (0.17%)	1 (0.008%)

Table 3. Counts of genes with ASE and *cis*- , *trans*- and CE*-evolution

	Testis				Ovary			
Comparison	**ASE**	**cis**	**trans**	**CE**	**ASE**	**cis**	**trans**	**CE**
House	261	-	-	-	72	-	-	-
Spain	456	-	-	-	118	-	-	-
F_1	164	5	130	159	54	9	1373	45
Italy	151	4	131	147	53	10	1372	43

* Compensatory evolution

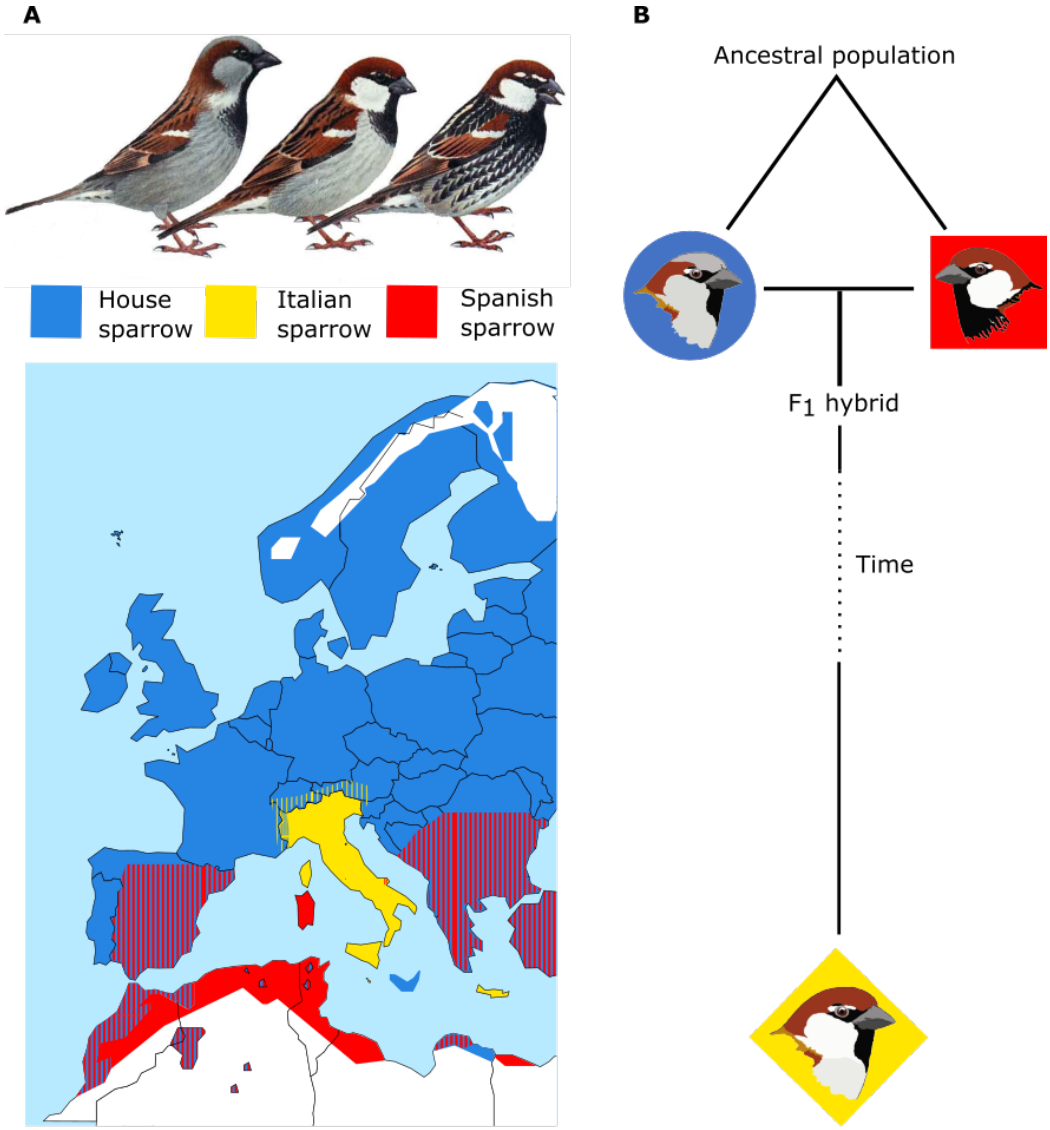

Figure 1. **Species distribution and F_1 hybrid breeding design.** (A) Top: Illustrations of male plumage patterns in house, Italian, and Spanish sparrows modified from (51) Bottom: A distribution map of house, Italian, and Spanish sparrows throughout Europe and northern Africa (52). This picture is modified from (24). (B) Pedigree of the F_1 hybrids with house sparrow mother and Spanish sparrow father.

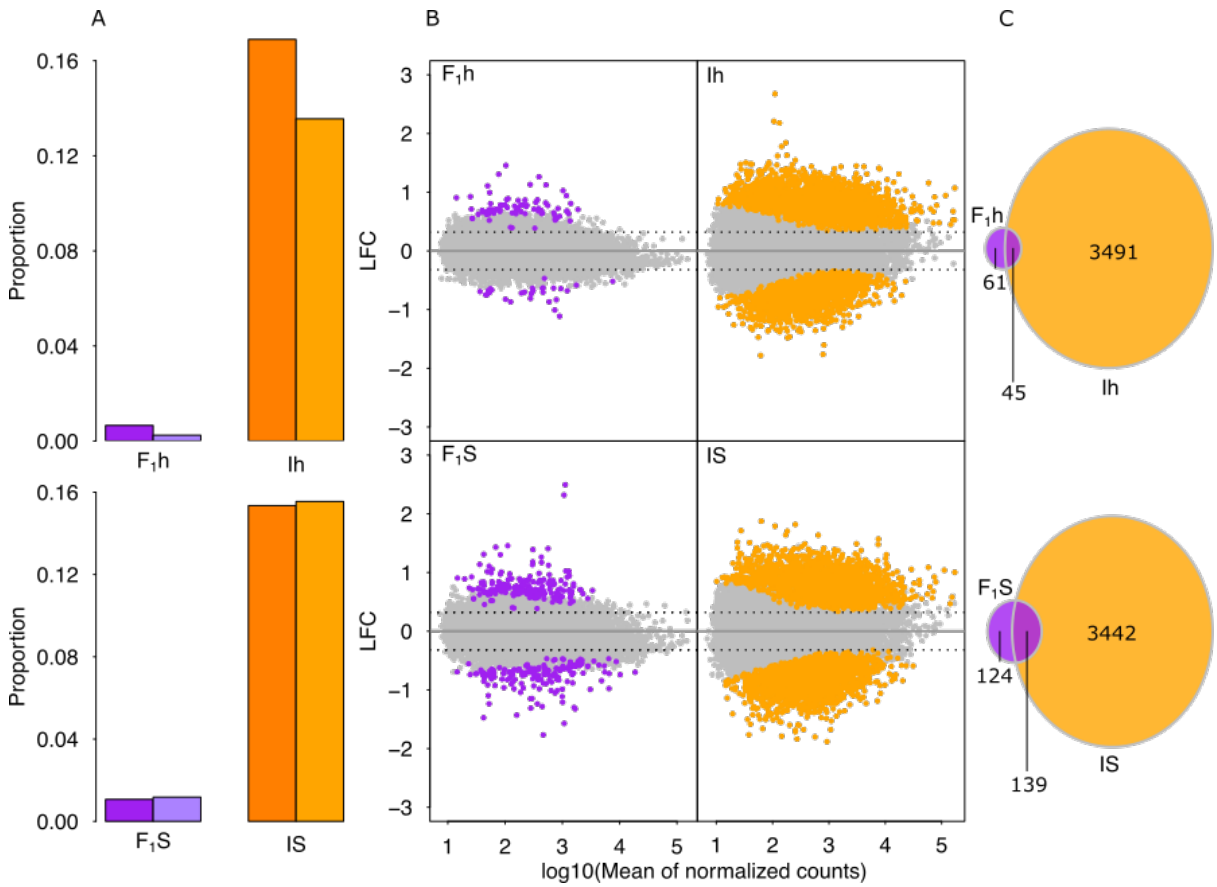

Figure 2. **Gene expression in experimental F1 hybrid and Italian sparrow in comparison with parental species in testes.** A) Proportion of up- and down-regulated genes. Dark color: up-regulated, light color: down-regulated. B) Log2 fold change (LFC) as a function of mean of normalized counts. C) Venn diagram depicting the intersection between differentially expressed genes between the experimental F1 hybrid and Italian sparrow with each of the parental species. F_1h: experimental F1 hybrid – house sparrow; Ih: Italian sparrow – house sparrow; F_1S: experimental F1 hybrid – Spanish sparrow; IS: Italian sparrow – Spanish sparrow. Purple: F1 hybrids Orange: Italian sparrows.

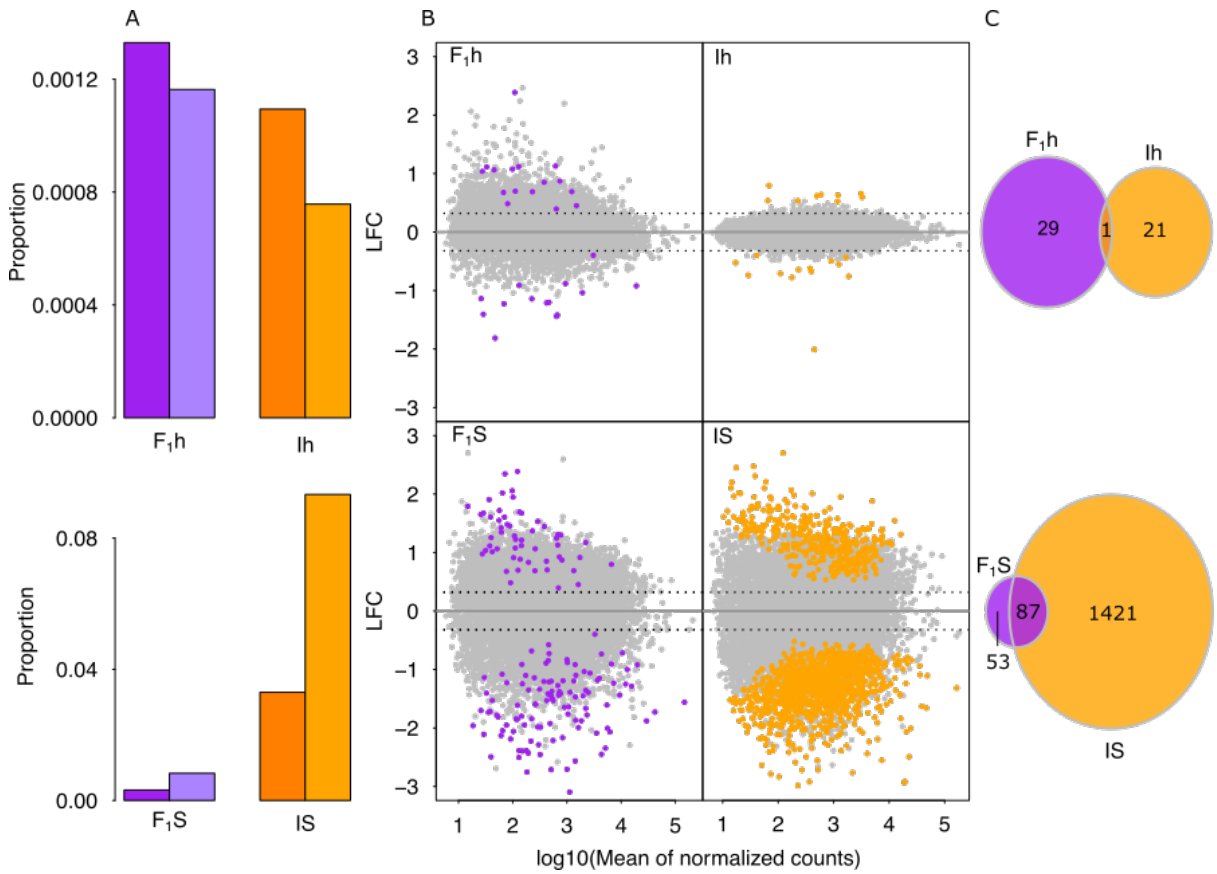

Figure 3**. Gene expression in experimental F1 hybrid and Italian sparrow in comparison with parental species for ovaries.** A) Proportion of up- and down-regulated genes. Dark color: up-regulated, light color: down-regulated. B) Log2 fold change (LFC) as a function of mean of normalized counts. C) Venn diagram depicting the intersection between differentially expressed genes between the experimental F1 hybrid and Italian sparrow with each of the parental species. F_1h: experimental F1 hybrid – house sparrow; Ih: Italian sparrow – house sparrow; F_1S: experimental F1 hybrid – Spanish sparrow; IS: Italian sparrow – Spanish sparrow. Purple: F1 hybrids Orange: Italian sparrows.

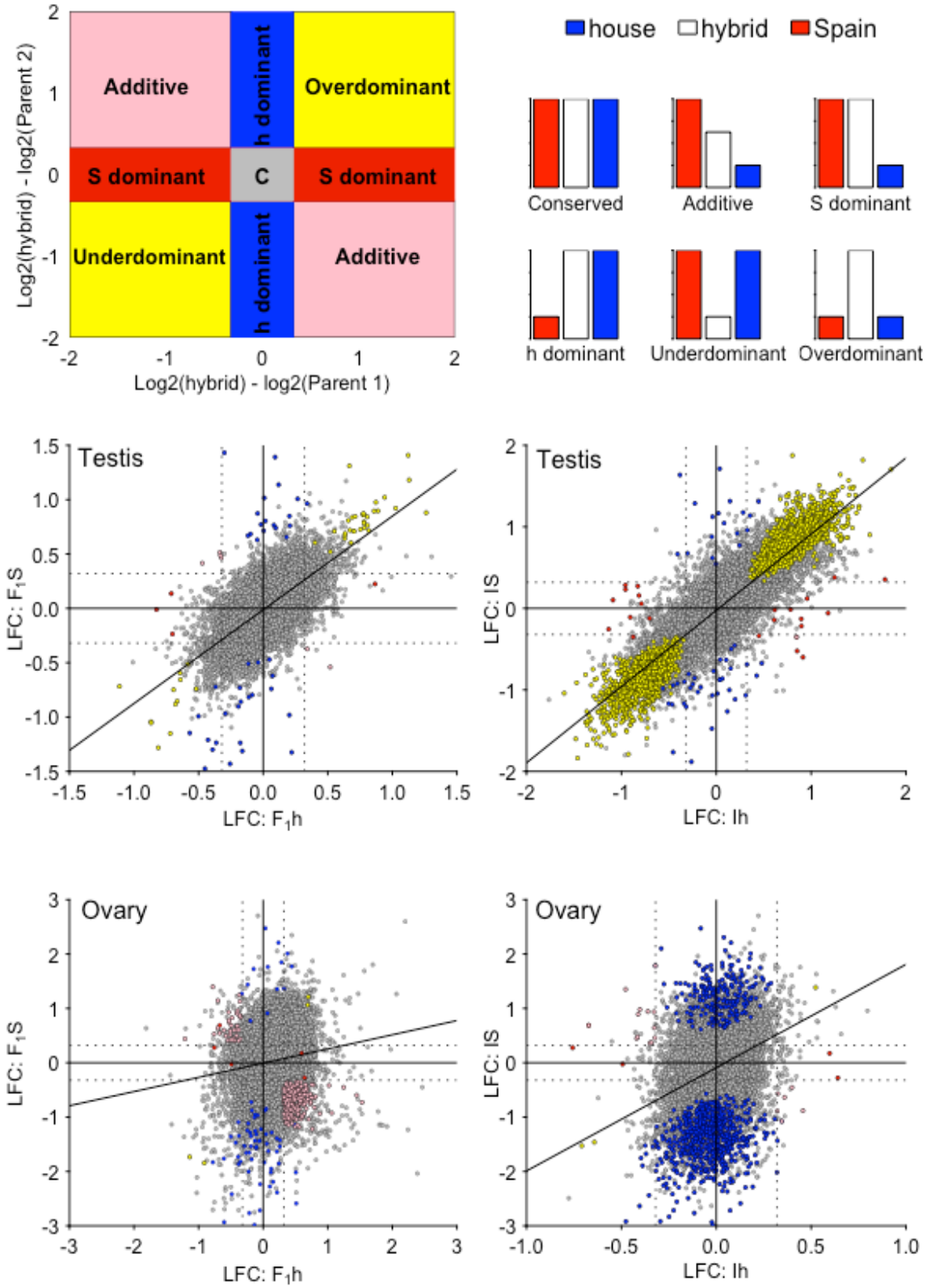

Figure 4. **Inheritance pattern of gene expression in testes and ovaries in experimental F1 hybrids and Italian sparrows.** Top: Schematic figure representing the distribution of inheritance patterns. Middle: Scatter plots show shrunken log2 fold change (LFC) between

the experimental F1 hybrid/Italian sparrow with the house sparrow on the *x*-axis and with Spanish sparrow on the *y*-axis. Grey points in each graph depicts the total number of genes studies for gene expression with those colored representing the ones significantly different from parental species to be divided into each of the inheritance category. Grey dotted lines indicate the fold-change threshold of 1.25 used for classification of inheritance pattern. Regression line represents the major variation in log2 fold change. F_1h: experimental F1 hybrid – house sparrow; Ih: Italian sparrow – house sparrow; F_1S: experimental F1 hybrid – Spanish sparrow; IS: Italian sparrow – Spanish sparrow.

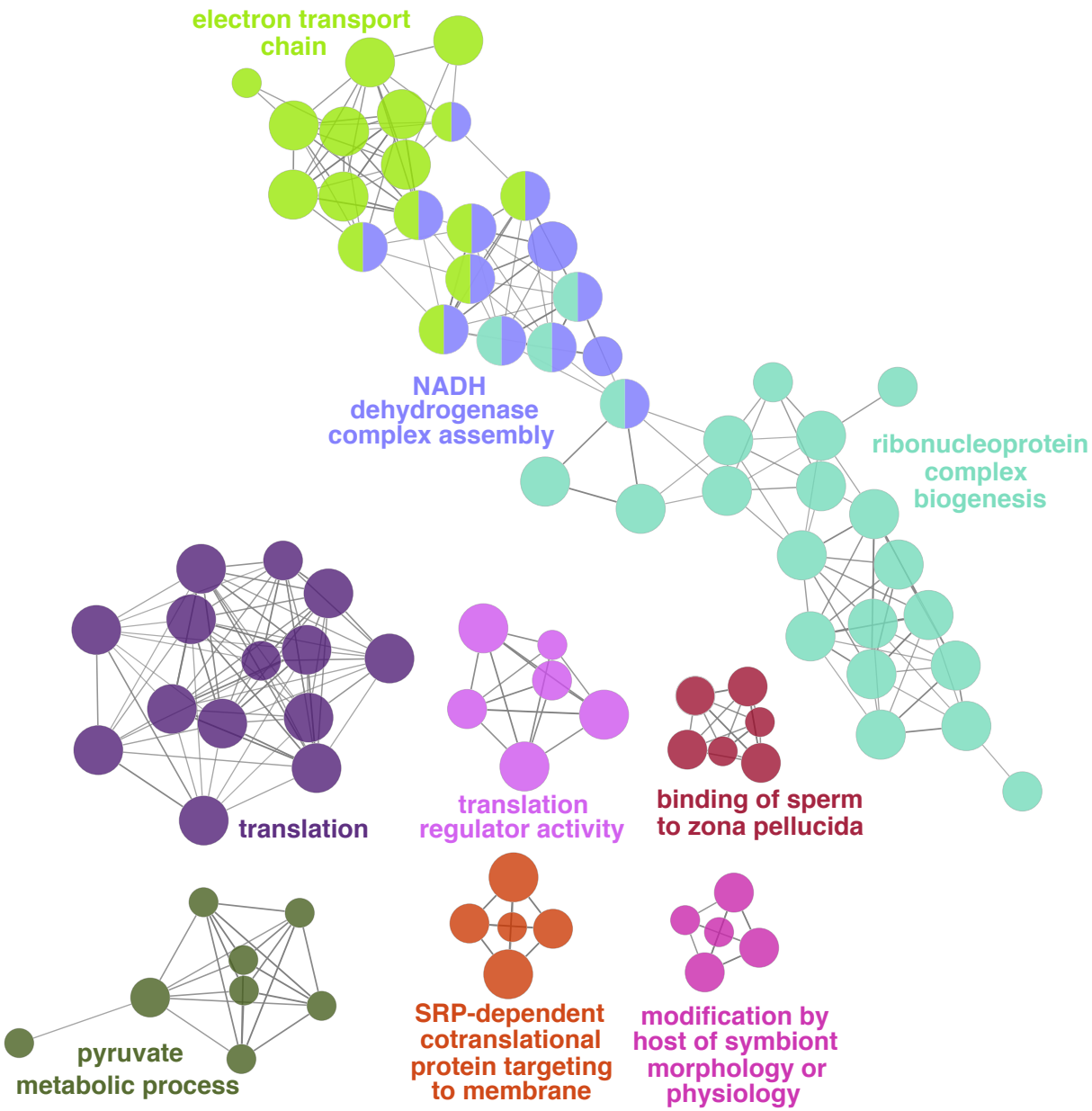

Figure 5. **Gene ontology network represented in overdominant genes in testis of Italian sparrow.**

References

1. Runemark A, Vallejo-Marin M, Meier JI. Eukaryote hybrid genomes. Plos Genet. 2019;15(11):e1008404.
2. Taylor SA, Larson EL. Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. Nat Ecol Evol. 2019;3(2):170-7.
3. Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJ, Bierne N, et al. Hybridization and speciation. J Evol Biol. 2013;26(2):229-46.
4. Mallet J. Hybridization as an invasion of the genome. Trends Ecol Evol. 2005;20(5):229-37.
5. Dobzhansky T. Studies on Hybrid Sterility. II. Localization of Sterility Factors in Drosophila Pseudoobscura Hybrids. Genetics. 1936;21(2):113-35.
6. Muller HJ. Isolating mechanisms, evolution, and temperature. Biol Symp. 1942;6:71-125.
7. Pennisi E. A shortcut to a species. Science. 2016;354(6314):818.
8. Ortiz-Barrientos D, Counterman BA, Noor MA. Gene expression divergence and the origin of hybrid dysfunctions. Genetica. 2007;129(1):71-81.
9. Renaut S, Nolte AW, Bernatchez L. Gene expression divergence and hybrid misexpression between lake whitefish species pairs (Coregonus spp. Salmonidae). Mol Biol Evol. 2009;26(4):925-36.
10. Nowick K, Carneiro M, Faria R. A prominent role of KRAB-ZNF transcription factors in mammalian speciation? Trends Genet. 2013;29(3):130-9.
11. Hodgins-Davis A, Rice DP, Townsend JP. Gene Expression Evolves under a House-of-Cards Model of Stabilizing Selection. Mol Biol Evol. 2015;32(8):2130-40.
12. Gilad Y, Oshlack A, Rifkin SA. Natural selection on gene expression. Trends Genet. 2006;22(8):456-61.
13. Coolon JD, McManus CJ, Stevenson KR, Graveley BR, Wittkopp PJ. Tempo and mode of regulatory evolution in Drosophila. Genome Res. 2014;24(5):797-808.
14. Ranz JM, Namgyal K, Gibson G, Hartl DL. Anomalies in the expression profile of interspecific hybrids of Drosophila melanogaster and Drosophila simulans. Genome Res. 2004;14(3):373-9.
15. Rockman MV, Kruglyak L. Genetics of global gene expression. Nat Rev Genet. 2006;7(11):862-72.
16. Malone JH, Michalak P. Gene expression analysis of the ovary of hybrid females of Xenopus laevis and X. muelleri. Bmc Evol Biol. 2008;8:82.
17. Haerty W, Singh RS. Gene regulation divergence is a major contributor to the evolution of Dobzhansky-Muller incompatibilities between species of Drosophila. Mol Biol Evol. 2006;23(9):1707-14.
18. Barreto FS, Pereira RJ, Burton RS. Hybrid dysfunction and physiological compensation in gene expression. Mol Biol Evol. 2015;32(3):613-22.
19. Edger PP, Smith R, McKain MR, Cooley AM, Vallejo-Marin M, Yuan Y, et al. Subgenome Dominance in an Interspecific Hybrid, Synthetic Allopolyploid, and a 140-Year-Old Naturally Established Neo-Allopolyploid Monkeyflower. Plant Cell. 2017;29(9):2150-67.
20. Kryvokhyzha D, Salcedo A, Eriksson MC, Duan T, Tawari N, Chen J, et al. Parental legacy, demography, and admixture influenced the evolution of the two subgenomes of the tetraploid Capsella bursa-pastoris (Brassicaceae). Plos Genet. 2019;15(2):e1007949.
21. Cutter AD. The polymorphic prelude to Bateson-Dobzhansky-Muller incompatibilities. Trends Ecol Evol. 2012;27(4):209-18.

22. Mugal CF, Wang M, Backstrom N, Wheatcroft D, Alund M, Semon M, et al. Tissue-specific patterns of regulatory changes underlying gene expression differences among Ficedula flycatchers and their naturally occurring F1 hybrids. Genome Res. 2020;30(12):1727-39.
23. Hermansen JS, Saether SA, Elgvin TO, Borge T, Hjelle E, Saetre GP. Hybrid speciation in sparrows I: phenotypic intermediacy, genetic admixture and barriers to gene flow. Molecular Ecology. 2011;20(18):3812-22.
24. Elgvin TO, Trier CN, Torresen OK, Hagen IJ, Lien S, Nederbragt AJ, et al. The genomic mosaicism of hybrid speciation. Sci Adv. 2017;3(6):e1602996.
25. Trier CN, Hermansen JS, Saetre GP, Bailey RI. Evidence for mito-nuclear and sex-linked reproductive barriers between the hybrid Italian sparrow and its parent species. Plos Genet. 2014;10(1):e1004075.
26. Hermansen JS, Haas F, Trier CN, Bailey RI, Nederbragt AJ, Marzal A, et al. Hybrid speciation through sorting of parental incompatibilities in Italian sparrows. Mol Ecol. 2014;23(23):5831-42.
27. Denver DR, Morris K, Streelman JT, Kim SK, Lynch M, Thomas WK. The transcriptional consequences of mutation and natural selection in Caenorhabditis elegans. Nat Genet. 2005;37(5):544-8.
28. Goncalves A, Leigh-Brown S, Thybert D, Stefflova K, Turro E, Flicek P, et al. Extensive compensatory cis-trans regulation in the evolution of mouse gene expression. Genome Res. 2012;22(12):2376-84.
29. Landry CR, Wittkopp PJ, Taubes CH, Ranz JM, Clark AG, Hartl DL. Compensatory cis-trans evolution and the dysregulation of gene expression in interspecific hybrids of Drosophila. Genetics. 2005;171(4):1813-22.
30. Ravinet M, Elgvin TO, Trier C, Aliabadian M, Gavrilov A, Saetre GP. Signatures of human-commensalism in the house sparrow genome. Proc Biol Sci. 2018;285(1884).
31. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15-21.
32. Wang L, Wang S, Li W. RSeQC: quality control of RNA-seq experiments. Bioinformatics. 2012;28(16):2184-5.
33. Anders S, Pyl PT, Huber W. HTSeq-a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31(2):166-9.
34. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biology. 2010;11(10).
35. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology. 2014;15(12).
36. McManus CJ, Coolon JD, Duff MO, Eipper-Mains J, Graveley BR, Wittkopp PJ. Regulatory divergence in Drosophila revealed by mRNA-seq. Genome Res. 2010;20(6):816-25.
37. Wittkopp PJ, Haerum BK, Clark AG. Evolutionary changes in cis and trans gene regulation. Nature. 2004;430(6995):85-8.
38. Mayba O, Gilbert HN, Liu J, Haverty PM, Jhunjhunwala S, Jiang Z, et al. MBASED: allele-specific expression detection in cancer tissues and cell lines. Genome Biol. 2014;15(8):405.
39. McGirr JA, Martin CH. Hybrid gene misregulation in multiple developing tissues within a recent adaptive radiation of Cyprinodon pupfishes. Plos One. 2019;14(7):e0218899.
40. Rowe M, Whittington E, Borziak K, Ravinet M, Eroukhmanoff F, Saetre GP, et al. Molecular Diversification of the Seminal Fluid Proteome in a Recently Diverged Passerine Species Pair. Mol Biol Evol. 2020;37(2):488-506.

41. Koskinen P, Toronen P, Nokso-Koivisto J, Holm L. PANNZER: high-throughput functional annotation of uncharacterized proteins in an error-prone environment. Bioinformatics. 2015;31(10):1544-52.
42. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16(5):284-7.
43. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019;47(D1):D607-D13.
44. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498-504.
45. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics. 2009;25(8):1091-3.
46. Rieseberg LH, Raymond O, Rosenthal DM, Lai Z, Livingstone K, Nakazato T, et al. Major ecological transitions in wild sunflowers facilitated by hybridization. Science. 2003;301(5637):1211-6.
47. Mallet J. Hybrid speciation. Nature. 2007;446(7133):279-83.
48. Rieseberg LH, Archer MA, Wayne RK. Transgressive segregation, adaptation and speciation. Heredity (Edinb). 1999;83 (Pt 4):363-72.
49. Runemark A, Eroukhmanoff F, Nava-Bolanos A, Hermansen JS, Meier JI. Hybridization, sex-specific genomic architecture and local adaptation. Philos Trans R Soc Lond B Biol Sci. 2018;373(1757).
50. Runemark A, Trier CN, Eroukhmanoff F, Hermansen JS, Matschiner M, Ravinet M, et al. Variation and constraints in hybrid genome formation. Nat Ecol Evol. 2018;2(3):549-56.
51. Svensson L, Grant PJ, Mullarney K, Zetterstroem D. Gyldendals store fugleguide: Europas og middelhavsområdets fugler i felt: Gyldendal; 1999.
52. Summers-Smith JD. The Sparrows: A Study of the Genus Passer: T & AD Poyser; 1988.