



OPEN

Comparative transcriptomics between *Drosophila mojavensis* and *D. arizonae* reveals transgressive gene expression and underexpression of spermatogenesis-related genes in hybrid testes

Cecilia A. Banho^{1,2}, Vincent Mérel², Thiago Y. K. Oliveira³, Claudia M. A. Carareto¹ & Cristina Vieira²✉

Interspecific hybridization is a stressful condition that can lead to sterility and/or inviability through improper gene regulation in *Drosophila* species with a high divergence time. However, the extent of these abnormalities in hybrids of recently diverging species is not well known. Some studies have shown that in *Drosophila*, the mechanisms of postzygotic isolation may evolve more rapidly in males than in females and that the degree of viability and sterility is associated with the genetic distance between species. Here, we used transcriptomic comparisons between two *Drosophila mojavensis* subspecies and *D. arizonae* (*repleta* group, *Drosophila*) and identified greater differential gene expression in testes than in ovaries. We tested the hypothesis that the severity of the interspecies hybrid phenotype is associated with the degree of gene misregulation. We showed limited gene misregulation in fertile females and an increase in the amount of misregulation in males with more severe sterile phenotypes (motile vs. amotile sperm). In addition, for these hybrids, we identified candidate genes that were mostly associated with spermatogenesis dysfunction.

Speciation is a complex process resulting from the divergence of two populations from an ancestral lineage by reproductive barriers capable of preventing gene flow^{1,2}. Among these barriers, postzygotic isolation mechanisms contribute to hybrid incompatibility, and their consequences can be observed by the presence of two main traits, hybrid sterility and/or hybrid inviability, which can evolve at different rates. Overall, hybrid sterility evolves faster than hybrid inviability, mainly when considering the different sexes, since these barriers evolve faster in males than in females³. Indeed, several studies considering intraspecific and interspecific hybridization have shown that the germline is primarily affected, and sterility is often detected^{4–10}. Although the process of speciation has been widely studied, the causes of postzygotic incompatibility in hybrids are not fully understood. According to the model of Dobzhansky–Muller^{11,12}, fixed mutations in genetically isolated populations can result in deleterious epistatic interactions, disrupting regulatory networks and leading to serious consequences in hybrids^{2,13–15}. In addition, according to Haldane's rule¹⁶, sterility in hybrids is more likely to affect the heterogametic sex, and the degree and extent of these genetic incompatibilities are related to the time of divergence and are likely the result of divergent regulatory sequences^{13,15,17,18}.

Most of the hybrid incompatibility genes identified so far seem to have species-specific effects leading to hybrid incompatibility, indicating that this is a complex polygenic trait that has different epistatic effects,

¹Department of Biology, UNESP - São Paulo State University, São José do Rio Preto, São Paulo State (SP), Brazil. ²Laboratoire de Biométrie et Biologie Evolutive, CNRS, UMR 5558, Université Claude Bernard Lyon 1, University of Lyon, 69622 Villeurbanne, France. ³Laboratory of Molecular Immunology, The Rockefeller University, New York, NY, USA. ✉email: cristina.vieira@univ-lyon1.fr

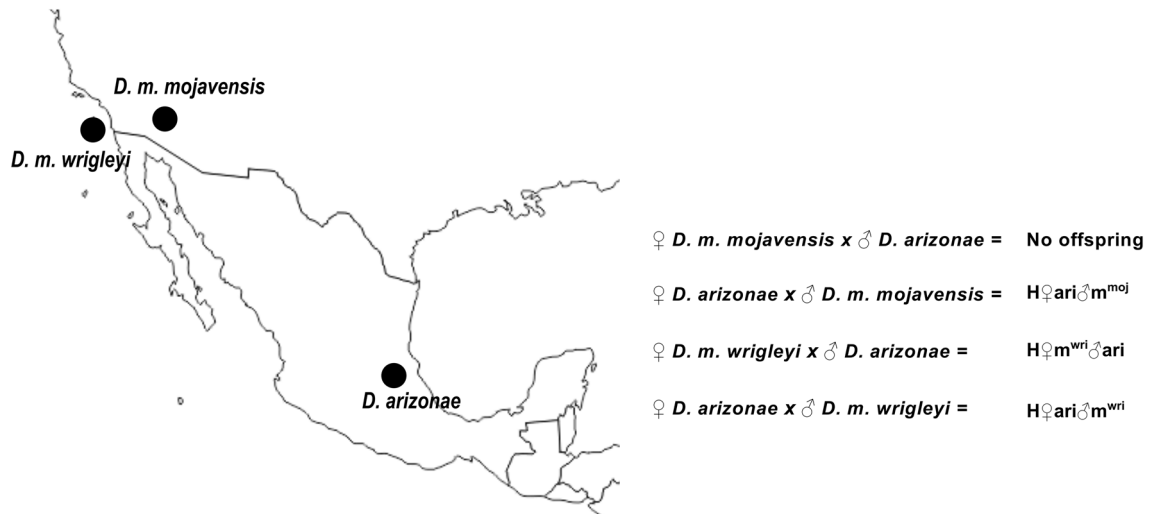


Figure 1. Crosses performed between *D. arizonae* and *D. mojavensis* subspecies and differential gene expression between parental lines and hybrids. Geographic distribution of *D. m. wrigleyi*, *D. m. mojavensis* and *D. arizonae* lines used in this study and their respective cross directions with offspring identifications (http://www.d-maps.com/carte.php?num_car=1404&lang=en). H♀m^{wri}♂ari and H♀ari♂m^{wri} are offspring from crosses between *D. arizonae* and *D. m. wrigleyi*, and H♀ari♂m^{moj} are hybrids from crosses between *D. arizonae* and *D. m. mojavensis*.

depending on the pair of species, which is likely related to the rapid accumulation of genetic changes over time. Studies of recently diverged species presenting different phenotypes regarding postzygotic isolation mechanisms can help to clarify early disruptions in gene regulation and expression that may influence the speciation process. *Drosophila mojavensis* and *D. arizonae* are appropriate for such a study. They are sibling species with divergence time estimates from 0.66–0.99 mya¹⁹ to 4.2 ± 0.99 mya²⁰, depending on the study. The most recent data, based on more than 5000 genes, indicate ~ 1.5 mya²¹ of divergence between the two species. This pair of species is widely used in speciation studies due to their ability to produce hybrids in the laboratory; however, introgression has not yet been found in nature despite the favourable ecological conditions for hybridization, mainly between sympatric populations²². *D. mojavensis* is endemic to the southwestern United States and northwestern Mexico. It is composed of four ecologically distinct subspecies, which are distributed in four different geographic regions, and each subspecies uses a specific host cactus as a feeding and breeding site, with no evidence of recent gene flow, constituting well-structured populations^{19,23–27}. In crosses between *D. mojavensis* subspecies and *D. arizonae*, hybrids can be produced in both directions. Nevertheless, they present incomplete and asymmetric postzygotic isolation, since in crosses with *D. mojavensis* females, the sterility of the male hybrids is variable, whereas in crosses with *D. arizonae* females, the male hybrids are always sterile because they do not have motile sperm^{10,28,29}.

With this in mind, we sought to assess the magnitude of differential gene expression (DEG) in male and female offspring from *D. arizonae* × *D. mojavensis* crosses and its association with hybrid phenotypes. We showed that in fertile hybrid females, very few DEGs were identified. However, in male hybrids, the degree of deregulation of gene expression was related to the severity of the sterile phenotype because males without motile spermatozoa had several DEGs, with a bias for underexpression, and the functions of these genes were related to spermatogenesis. By contrast, in sterile hybrids with motile spermatozoa, the degree of deregulation was lower and had a bias toward overexpression, and the gene functions were not directly related to spermatogenesis.

Results

The transcriptomes of the ovaries and testes of three parental allopatric strains (*D. m. mojavensis*, *D. m. wrigleyi* and *D. arizonae*) and their reciprocal hybrids (Fig. 1) were sequenced. We were able to recover 11,654 coding genes from *D. mojavensis* r1.04 transcripts. Of those, 9321 (80%) genes were expressed in *D. arizonae*, *D. m. mojavensis* and H♀ari♂m^{moj} ovaries, and 9700 (83.3%) were expressed in *D. arizonae*, *D. m. wrigleyi*, H♀m^{wri}♂ari and H♀ari♂m^{wri} female gonads. In testes, we found 11,146 (95.6%) genes expressed in *D. arizonae*, *D. m. mojavensis* and H♀ari♂m^{moj} and 11,223 (96.3%) expressed in *D. arizonae*, *D. m. wrigleyi*, H♀m^{wri}♂ari and H♀ari♂m^{wri}. The read alignment rate ranged from 81.7 to 86% in ovaries and from 76.4 to 80.8% in testes (Supplementary Table S1). A similar alignment rate for *D. arizonae*, *D. mojavensis* subspecies and their reciprocal hybrids indicated that the genome of the *D. mojavensis* r1.04 reference can be used in analyses involving *D. arizonae* and hybrids. Additionally, according to Lopez-Maestre et al.³⁰, the average nucleotide divergence between *D. arizonae* and *D. mojavensis* is very low, less than 2%; thus, the sequence alignments of parental species and their hybrids with the r1.04 genome of *D. mojavensis* should not affect mapping efficiency. Principal component analysis (PCA) was performed to verify the variance of the biological replicates. Within ovaries and testes, the replicates were grouped together (see Supplementary Fig. S1), indicating low variance between replicates.

Differential gene expression in *D. arizonae* and *D. mojavensis* gonads. We compared the transcriptomes of *D. arizonae* and *D. m. mojavensis* and those of *D. arizonae* and *D. m. wrigleyi* and identified 501

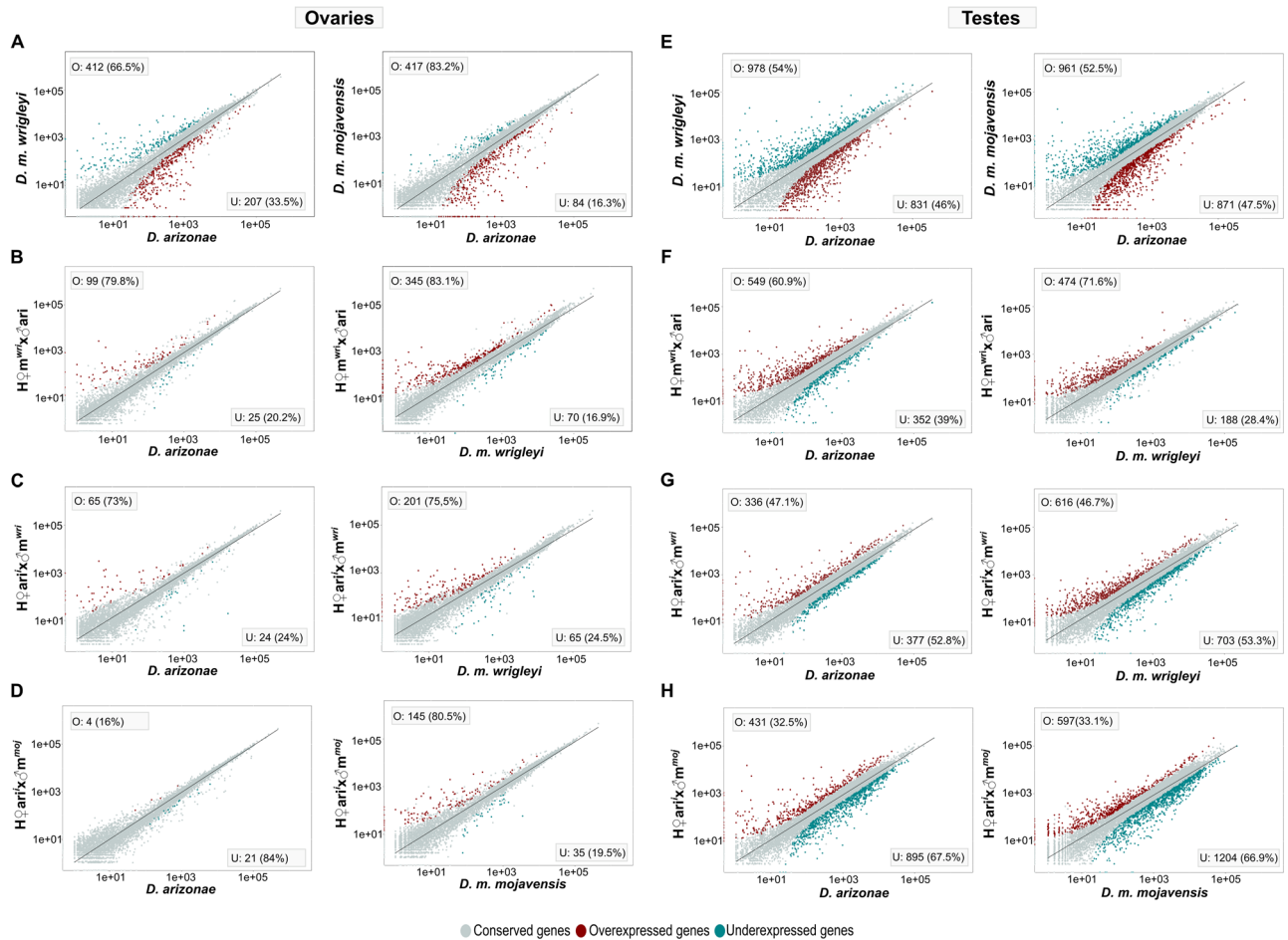


Figure 2. Differential gene expression between *D. arizonae* and *D. mojavensis* subspecies and between hybrids and their respective parental lines. Scatter plots representing differential gene expression in ovaries (left) and testes (right) between parental lines, as well as between hybrids and their respective parental lines. Genes were considered differentially expressed when they presented twofold differences and FDR-corrected p -values < 0.01 . Red dots = overexpressed genes; blue dots = underexpressed genes; grey dots = not differentially expressed. (A) Differential expression in ovaries of *D. m. mojavensis* \times *D. arizonae* and *D. m. wrigleyi* \times *D. arizonae* hybrids. (B) Differential expression in ovaries of $H^{\square}m^{wri} \times \text{ari} \times D. m. wrigleyi$ hybrids and *D. arizonae*. (C) Differential expression in ovaries of $H^{\square} \text{ari} \times m^{wri} \times D. m. wrigleyi$ hybrids and *D. arizonae*. (D) Differential expression in ovaries of $H^{\square} \text{ari} \times m^{moj} \times D. m. mojavensis$ hybrids and *D. arizonae*. (E) Differential expression in testes of *D. m. mojavensis* \times *D. arizonae* and *D. m. wrigleyi* \times *D. arizonae* hybrids. (F) Differential expression in testes of $H^{\square} m^{wri} \times \text{ari} \times D. m. wrigleyi$ hybrids and *D. arizonae*. (G) Differential expression in testes of $H^{\square} \text{ari} \times m^{wri} \times D. m. wrigleyi$ hybrids and *D. arizonae*. (H) Differential expression in testes of $H^{\square} \text{ari} \times m^{moj} \times D. m. mojavensis$ hybrids and *D. arizonae*. The scatter plot was generated using the ggplot2 package (version 3.3.3) in R⁷⁹. O: overexpressed genes; U: underexpressed genes.

(5.3%) and 619 (6.3%) DEGs in ovaries, respectively, and the majority of the genes were overexpressed in *D. arizonae* ($X^2 = 220$, $p = 2.2e-16$; $X^2 = 67.231$, $p = 2.415e-16$, respectively) (Fig. 2A, Supplementary Table S6). In testes, we observed that 16% of genes were differentially expressed (Fig. 2E) when comparing *D. arizonae* with both *D. mojavensis* subspecies, and there was a higher proportion of overexpressed genes in both comparisons ($X^2 = 4.34$, $p = 0.03759$; $X^2 = 11.783$, $p = 0.0005976$) (Fig. 2E, see Supplementary Table S6). Gene enrichment analyses between parental lines showed that in ovaries, the DEGs were mainly enriched for nervous system, metabolism, chemotaxis, and extracellular matrix organization, among others. In testes, DEGs were also enriched for metabolic function, sensory perception, nervous system, gene expression, and behaviour, among other functions (Fig. 3, Supplementary Tables S2 and S3 and Supplementary Fig. S2 for cellular component and molecular function enrichment). Among the DEGs with functions in metabolic processes, those involved in processes of oxidation–reduction and metabolism of carbohydrates and lipids can be highlighted (see Supplementary Tables S2, S3, S4 and S5).

Gene expression in hybrid female gonads is very similar to that in the parental lines. We compared the patterns of gene expression between hybrids and their parental lines. In hybrids between *D. m. wrigleyi* and *D. arizonae*, the percentage of DEGs between the hybrids and the parental lines ranged from 1.27 to 4.2%

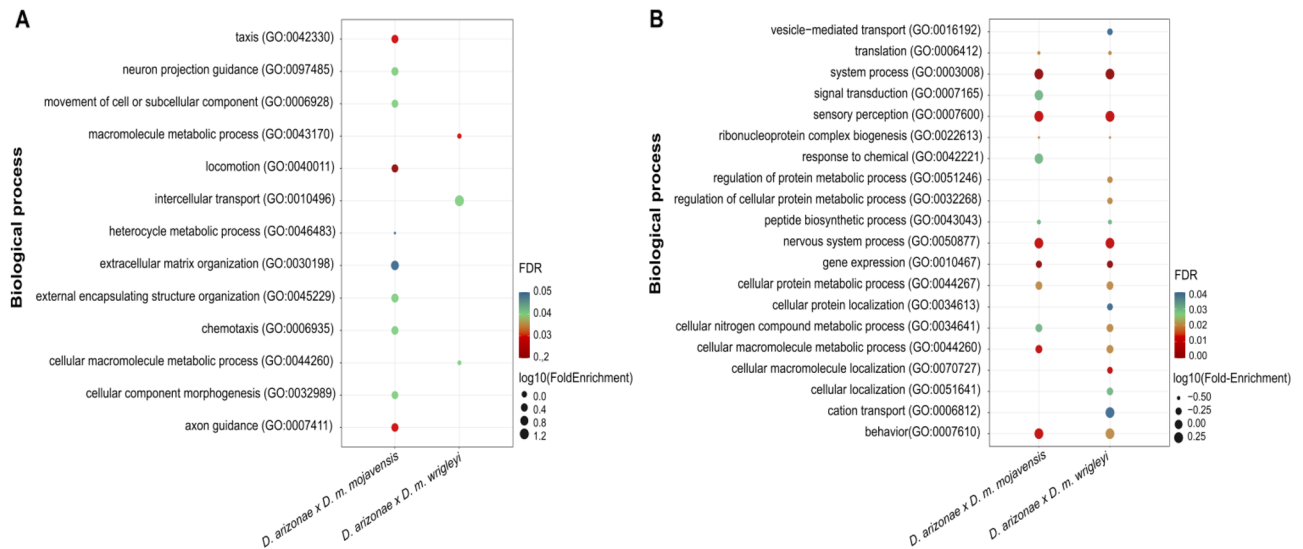


Figure 3. GO term enrichment analysis of DEGs between *D. arizonae* and *D. mojavensis* subspecies. Dot plot representing GO enrichment for the biological process category for DEGs in (A) ovaries and (B) testes between *D. arizonae*-*D. m. mojavensis* and *D. arizonae*-*D. m. wrightleyi* (FDR < 0.05). The dot plot was generated using the ggplot2 package (version 3.3.3) in R⁷⁹.

(Supplementary Table S7). For all the comparisons, the proportion of overexpressed genes was significantly higher than that of underexpressed genes and was always higher in female hybrids from *D. arizonae*-*D. m. wrightleyi* crosses (Fig. 2A–C, see Supplementary Table S7). In hybrids between *D. m. mojavensis* and *D. arizonae*, the percentage of DEGs between the hybrids and the parental lines was 0.26 and 1.93% when compared with *D. arizonae* and *D. m. mojavensis*, respectively (Supplementary Table S7). The proportion of overexpressed versus underexpressed genes was significantly different and dependent on the cross (Supplementary Table S7). When comparing $H\text{♀}ari\text{♂}m^{moj}$ with *D. arizonae*, there was a higher proportion of underexpressed genes than overexpressed genes ($X^2 = 10.24$, $p = 0.001374$), contrary to the comparison between $H\text{♀}ari\text{♂}m^{moj}$ and *D. m. mojavensis*, for which we identified a higher proportion of overexpressed genes in the hybrids ($X^2 = 66.006$, $p = 4.497e-16$) (Fig. 2D).

By filtering the genes that were over- or underexpressed relative to both parental lines for all crosses, we greatly reduced our set of DEGs. From the DEGs in $H\text{♀}m^{wri}\text{♂}ari$ ovaries, 16 genes were overexpressed in relation to both parental lines (Supplementary Table S8); however, no significant GO enrichment was found. Moreover, no underexpressed genes were found. $H\text{♀}ari\text{♂}m^{wri}$ ovaries showed 13 overexpressed genes (Supplementary Table S9) and only one underexpressed gene (FBgn0135298, with unknown function) when compared with both parental lines. Additionally, $H\text{♀}m^{wri}\text{♂}ari$ and $H\text{♀}ari\text{♂}m^{wri}$ ovaries showed only one shared overexpressed gene, FBgn0145754 (unknown function) (Fig. 4A). In $H\text{♀}ari\text{♂}m^{moj}$ ovaries, no overexpressed genes were found, while only three genes were underexpressed. These genes included FBgn014602, which corresponds to *alpha-Est5* and has carboxylic ester hydrolase activity; FBgn028050, corresponding to the *Maverick* gene, which has a role in signalling pathways, the regulation of neuromuscular junctions, dendrite development and imaginal disc-derived wing size; and FBgn0146651, orthologous to CG11854 in *D. melanogaster*, which is involved in reproduction, more specifically in courtship behaviour^{31,32}.

Gene expression in the hybrid male germline is transgressive when compared to the parental lines.

In comparison to ovaries, hybrid testes always presented a larger number of DEGs in relation to parental lines (Fig. 2E, Supplementary Table S7). In $H\text{♀}m^{wri}\text{♂}ari$ testes, we found that 8% and 5.8% of genes were differentially expressed compared with *D. arizonae* and *D. m. wrightleyi*, respectively, which displayed a significant bias for overexpression ($X^2 = 42.637$, $p = 6.59e-11$; $X^2 = 122.7$, $p = 2.2e-16$, Fig. 2F, Supplementary Table S7). In hybrid testes from the reciprocal cross ($H\text{♀}ari\text{♂}m^{wri}$), we found that 6.3% and 11.75% of genes were differentially expressed compared with *D. arizonae* and *D. m. wrightleyi* (Fig. 2G, Supplementary Table S7). Unlike $H\text{♀}m^{wri}\text{♂}ari$ testes, no bias towards over- or underexpression was observed in the comparison with *D. arizonae* ($X^2 = 2.244$, $p = 0.1341$) (see Supplementary Table S7). In testes from $H\text{♀}ari\text{♂}m^{moj}$, we observed that 11.9% and 16% of genes were differentially expressed compared with *D. arizonae* and *D. m. mojavensis*, respectively, with a significant proportion of underexpressed genes, 67.5% and 66.9%, respectively ($X^2 = 162.24$, $p = 2.2e-16$, $X^2 = 203.91$, $p = 2.2e-16$) (Fig. 2H, Supplementary Table S7).

Regarding the DEGs in $H\text{♀}m^{wri}\text{♂}ari$ testes, 32 were overexpressed and 14 were underexpressed when compared with both parental species, while in $H\text{♀}ari\text{♂}m^{wri}$ testes, 56 and 128 genes were over- and underexpressed, respectively, when compared with *D. arizonae* and *D. m. wrightleyi*. Similar to those of $H\text{♀}ari\text{♂}m^{wri}$, testes of $H\text{♀}ari\text{♂}m^{moj}$ displayed more underexpressed genes (519) than overexpressed genes (57) when compared to both parental species.

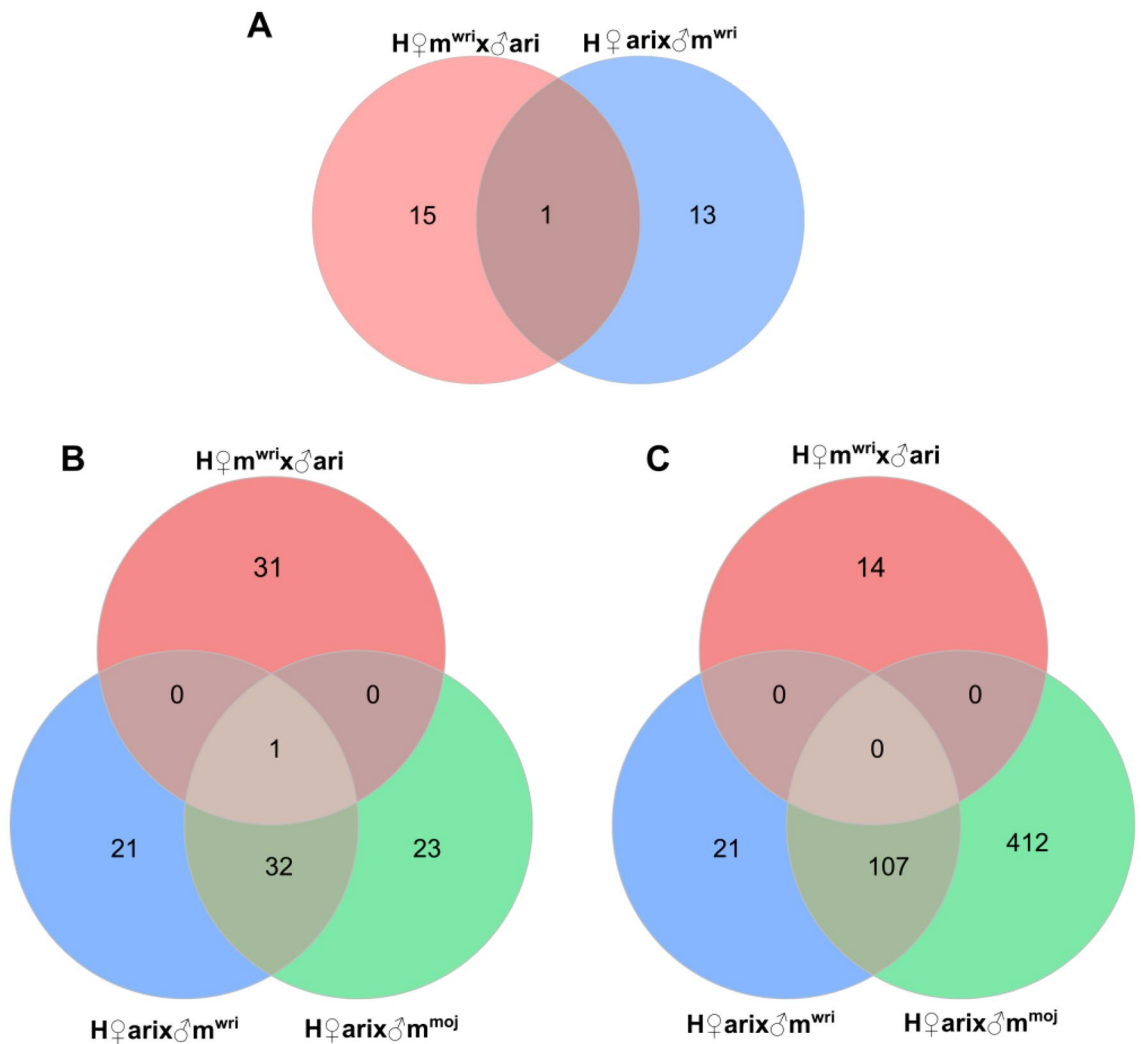


Figure 4. Shared DEGs between hybrids. (A) Number of shared overexpressed genes between $H_{\text{♀}}m^{\text{wri}} \times \text{♂ari}$ and $H_{\text{♀}}\text{ari} \times \text{♂}m^{\text{wri}}$ females. (B) Number of shared overexpressed genes between $H_{\text{♀}}m^{\text{wri}} \times \text{♂ari}$, $H_{\text{♀}}\text{ari} \times \text{♂}m^{\text{wri}}$ and $H_{\text{♀}}\text{ari} \times \text{♂}m^{\text{moj}}$ males. (C) Number of shared underexpressed genes between $H_{\text{♀}}m^{\text{wri}} \times \text{♂ari}$, $H_{\text{♀}}\text{ari} \times \text{♂}m^{\text{wri}}$ and $H_{\text{♀}}\text{ari} \times \text{♂}m^{\text{moj}}$ males. A Venn diagram was generated using the VennDiagram package (version 1.6.20) in R⁷⁹.

Spermatogenesis-related functions of DEGs in testes. In the male gonads, the functions of differentially expressed genes differed depending on the direction of the cross. In $H_{\text{♀}}m^{\text{wri}} \times \text{♂ari}$ testes, of the 46 DEGs found relative to both parental species, 39 had an orthologue in *D. melanogaster*. Most of the DEGs presented functional annotations associated with several functions, but no significant enrichment for GO terms was observed (see Supplementary Tables S10 and S11). On the other hand, in $H_{\text{♀}}\text{ari} \times \text{♂}m^{\text{wri}}$ testes, 183 genes were differentially expressed in both species, and 131 *D. melanogaster* orthologues were recovered. Functional annotation of these genes showed that several of them were related to processes such as reproduction and cellular division, including spermatogenesis, cilium movement, microtubule-based movement and nucleation, sperm motility, chromosome segregation and mitotic spindle elongation, as well as other functions related to several metabolic processes (*Cyp6a9/Cyp6a21*, *Ugt50B3*, *Gpo2*, CG7140, *P5CDh2*, *inaE*), transcription (*zen2*, *hb*, *Hr3*, *vnd*, *ap*) and sensory perception (*Obp8a*, *Or98a*) (see Supplementary Tables S12 and S13). In $H_{\text{♀}}\text{ari} \times \text{♂}m^{\text{moj}}$ testes, 576 genes were identified as differentially expressed, and 440 had a *D. melanogaster* orthologue. In these hybrids, most of the underexpressed genes were functionally annotated as having similar functions as those found in $H_{\text{♀}}\text{ari} \times \text{♂}m^{\text{wri}}$, but it is important to emphasize some other functions, such as spermatid differentiation and development (Supplementary Tables S14 and S15).

We searched for shared DEGs between hybrids and found that $H_{\text{♀}}m^{\text{wri}} \times \text{♂ari}$ and $H_{\text{♀}}\text{ari} \times \text{♂}m^{\text{wri}}$ testes shared only one overexpressed gene (Fig. 4B). However, when comparing male gonad expression from crosses of $H_{\text{♀}}\text{ari} \times \text{♂}m^{\text{moj}}$ and $H_{\text{♀}}\text{ari} \times \text{♂}m^{\text{wri}}$, which have the same *D. arizonae* mother but two different *D. mojavensis* subspecies fathers, we identified 33 shared overexpressed genes (Fig. 4B, Supplementary Table S16) and 107 shared underexpressed genes (Fig. 4C, Supplementary Table S17). Interestingly, some of these genes exhibited GO term enrichment for spermatogenesis-related functions, such as regulation of cytokinesis and cell division, protein localization to the microtubule plus-end, microtubule-based process, male germline cyst formation and cilium

movement, in addition to various metabolic processes (Fig. 5). Based solely on these reproductive genes, it is remarkable that some of them might have a direct role in reproduction, since they participate in male meiosis, male courtship, and mating behaviour, and play a role in gene silencing and pre-miRNA processing. Moreover, for many of these shared DEGs, the mutant phenotype has already been described for other *Drosophila* and corresponds to male sterility (Table 1)^{33–49}.

To better understand the results obtained for hybrid males, we examined some life-history traits, such as viability, sperm motility and fertility. Our analyses revealed that the viability of the hybrid was significantly reduced when compared to that of the parental lines (see Supplementary Table S18), since the average viability ranged from 3.85 to 14.2% in H♀ari♂m^{maj}, H♀m^{wri}♂ari and H♀ari♂m^{wri}, while in the parental lines, it ranged from 29.58 to 86.8% (Fig. 6, Supplementary Table S18). Additionally, we found that all hybrid females were fertile. Regarding F₁ males from the two types of crosses, we observed two phenotypes. H♀m^{wri}♂ari males produced motile sperm but were sterile, since no offspring were produced in backcrosses and F₁ × F₁ crosses. In contrast, H♀ari♂m^{wri} and H♀ari♂m^{maj} males produced amotile sperm and were also sterile (Fig. 6, Supplementary Table S18). These results suggest a link between the severity of the sterile phenotype and the number and functions of DEGs in male hybrids from crosses with *D. arizonae* mothers.

Inheritance of gene expression. We compared the level of gene expression in H♀ari♂m^{maj}, H♀m^{wri}♂ari and H♀ari♂m^{wri} with that in each parental line, following the six inheritance categories of McManus et al.¹⁵. Most of the genes in the ovaries and testes of the interspecific hybrids showed conserved expression (Fig. 7, Supplementary Tables S19 and S20). H♀ari♂m^{maj} showed conserved expression for 97.8% of the genes in ovaries and 78.5% of those in testes (Fig. 7, Supplementary Table S19). Few genes were classified as having additive expression in the female and male gonads (0.05 and 1.45%, respectively). In the dominant category, H♀ari♂m^{maj} exhibited an overrepresentation of genes with *D. arizonae*-like expression (1.8%, in ovaries and 9.5% in testes), and few genes were *D. mojavensis*-dominant (Fig. 7, Supplementary Table S19). In the overdominant and underdominant categories, almost no genes were found for ovaries (0.03% of DEGs). However, the testes showed several genes in these categories, reaching 0.5% for overdominant and 4.6% for underdominant genes (Fig. 7, Supplementary Table S19).

Similar to H♀ari♂m^{maj}, most of the expressed genes in H♀m^{wri}♂ari and H♀ari♂m^{wri} female and male gonads had conserved expression, but in testes, the level of conserved inheritance was lower than that in ovaries (Fig. 7, Supplementary Table S20). Additionally, very few genes in the ovaries (0.3 and 0.22%) and testes (1.1 and 1.7%) of H♀m^{wri}♂ari and H♀ari♂m^{wri} displayed additive expression (Fig. 7, Supplementary Table S20), as was observed for H♀ari♂m^{maj}. An interesting finding is that most of the genes classified in dominant categories displayed a *D. arizonae*-like pattern of expression in ovaries, but in testes, this pattern was related to the maternal line (Fig. 7, Supplementary Table S20). Considering the overdominant and underdominant categories, few genes were found for these categories in ovaries (0.16% in H♀m^{wri}♂ari and 0.13% in H♀ari♂m^{wri}) (Fig. 7, Supplementary Table S20). Similarly, in hybrid testes, very few genes were classified as over- (0.28% in H♀m^{wri}♂ari and 0.5% in H♀ari♂m^{wri}) or underdominant (0.12% in H♀m^{wri}♂ari and 1.1% in H♀ari♂m^{wri}) (Fig. 7, Supplementary Table S20).

Comparing all inheritance categories of H♀m^{wri}♂ari and H♀ari♂m^{wri}, significant differences in the expression profiles of reciprocal hybrids were observed in ovaries ($X^2 = 146.67$, $p < 0.001$), which were mainly influenced by paternal effects, since 201 H♀m^{wri}♂ari genes and 23 H♀ari♂m^{wri} genes showed paternally dominant inheritance. In testes, significant differences in the expression profiles were also found between reciprocal hybrids ($X^2 = 823.2$, $p < 0.001$); however, unlike in the ovaries, these differences were mainly influenced by maternal and underdominant inheritance. More specifically, 140 and 695 genes were found to have exclusive maternal inheritance and 14 and 128 genes had underdominant expression in H♀m^{wri}♂ari and H♀ari♂m^{wri}, respectively.

Discussion

In our study, the target species *D. m. mojavensis*, *D. m. wrighti* and *D. arizonae* showed approximately 6% of genes being differentially expressed in the ovaries and approximately 16% of genes being differentially expressed in the testes, agreeing with previous findings from our research group for hybrid female transcriptomes³⁰ and with results obtained for hybrid male gonads⁷. This proportion of DEGs between species was intermediate when compared to other pairs of species presenting higher or lower divergence times. For example, *D. yakuba* and *D. melanogaster* (~6.1 mya divergence time²⁰) have a proportion of DEGs varying from 29.59 to 42.58%³⁰. In more recently diverged species, fewer DEGs are generally observed, such as in males of *D. yakuba* and *D. santomea* (~1.18 mya⁵¹, DEG: 19%⁷) and those of *D. p. pseudoobscura* and *D. p. bogotana* (~0.25 mya⁵², DEG: 14.6%⁵³).

The differences in the proportion of DEGs in ovaries and testes could be related to the faster evolution of male-biased genes, as has been previously reported in other *Drosophila* species. According to previous studies, during the divergence process, male-biased genes display higher evolutionary rates, driven by positive selection, and most of them are preferentially expressed in gonad tissues^{54–57}. This is the case in *D. arizonae* and *D. mojavensis* and can explain the high proportion of DEGs in testes when compared with the parental lines⁵⁸.

Hybrid female gonads exhibited few DEGs when compared with both parental lines, and most of these genes were overexpressed. In agreement with our results, previous studies in hybrids of *D. arizonae*-*D. m. mojavensis*³⁰ and *D. melanogaster*-*D. simulans*¹³ have reported a bias towards overexpression in female offspring. However, this bias was not observed in *D. buzzatii* and *D. koepferae* (~4.49 mya) female hybrids¹⁴. The absence of massive gene deregulation in ovaries could be related to the higher stability of gene expression in females, which is mainly influenced by the slower evolutionary rates of female-biased genes and the presence of two X-chromosomes, as was reported in other *Drosophila* species^{17,57}, and this phenomenon could be important in the fertile phenotype observed in these hybrids. Thus, we suggest that in hybrid ovaries from recently diverged species, DEGs tend to

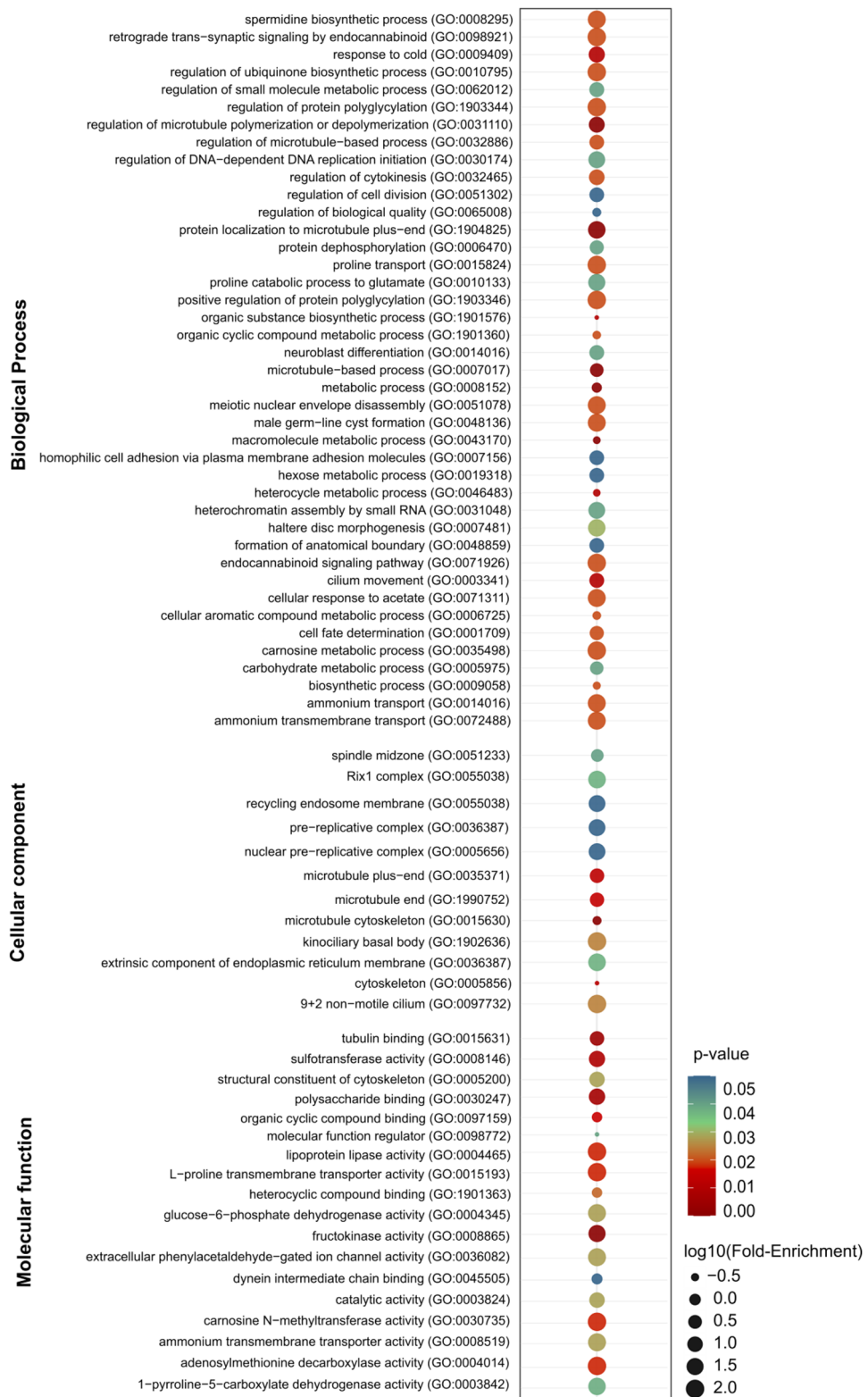


Figure 5. Functions of shared over- and underexpressed genes between $H_{\text{ari}}^{\text{wri}}$ and $H_{\text{ari}}^{\text{moj}}$ males. Dot plot representing GO enrichment for the biological process, cellular component and molecular function categories. The plot was generated using the ggplot2 package (version 3.3.3) in R⁷⁹.

| <i>D. mojavensis</i> ID | <i>D. melanogaster</i> ID | Gene name | Biological process | Mutant phenotype | References | Expression | | |
|-------------------------|-----------------------------|-------------------------|--|---|------------|-------------------------|------------------------|------------------------|
| | | | | | | H♀m ^{wri} ♂ari | H♀ari♂m ^{wri} | H♀ari♂m ^{maj} |
| FBgn0140112 | FBgn0028858 | <i>CG10839</i> | GO:0007018—microtubule-based movement | – | – | Not DF | Under | Under |
| FBgn0145039 | FBgn0039812 | <i>CG15548</i> | GO:000022—mitotic spindle elongation | – | – | Not DF | Under | Under |
| FBgn0135509 | FBgn0035581 | <i>Dnah3</i> | GO:0007018—microtubule-based movement | Amotile sperm | 41 | Not DF | Under | Under |
| FBgn0140104 | FBgn0028901 | <i>CG18109</i> | GO:0007020—microtubule nucleation | – | – | Not DF | Under | Under |
| FBgn0147429 | FBgn0063261 | <i>CG31275</i> | GO:0007018—microtubule-based movement | – | – | Not DF | Under | Under |
| FBgn0142040 | FBgn0027066 | <i>Eb1</i> | GO:0000226: microtubule cytoskeleton organization | Defects in spindle elongation and orientation; reduction in astral microtubules | 44 | Not DF | Under | Under |
| FBgn0143126 | FBgn0032225 | <i>CG5022</i> | GO:0031032—actomyosin structure organization | – | – | Not DF | Not DF | Under |
| FBgn0137431 | FBgn0283476 | <i>Dhc16F</i> | GO:0001539—cilium or flagellum-dependent cell motility | Male sterility | 49 | Not DF | Not DF | Under |
| FBgn0136866 | FBgn0039925 | <i>Kif3C</i> | GO:0007018—microtubule-based movement | Roles in flagellar/ciliary motilities | 38,39 | Not DF | Not DF | Under |
| FBgn0140824 | FBgn0031952 | <i>cdc14</i> | GO:0071850—mitotic cell cycle arrest | – | – | Not DF | Under | Under |
| FBgn0140999 | FBgn0262123 | <i>l(2)41Ab</i> | GO:0070286—axone-mal dynein complex assembly | – | – | Not DF | Under | Under |
| FBgn0146436 | FBgn0038565 | <i>CG7794</i> | GO:0007017—microtubule-based process | – | – | Not DF | Under | Under |
| FBgn0280294 | FBgn0023090 | <i>dtr</i> | GO:0060271—cilium morphogenesis | Male sterility | 49 | Not DF | Not DF | Under |
| FBgn0140392 | FBgn0002673 | <i>twe</i> | GO:0007140—male meiosis, GO:0007283—spermatogenesis | Male sterility. Absence of meiotic divisions in male germline and no motile sperm | 34,43 | Not DF | Under | Under |
| FBgn0144612 | FBgn0267326 | <i>Nil</i> | GO:0030317—sperm motility | Male sterility. Amotile sperm and fail to be transferred to the seminal vesicle | 33 | Not DF | Under | Under |
| FBgn0281134 | FBgn0001313 | <i>kl2</i> | GO:0003341: cilium movement | Male sterility. Defects in sperm individualization | 48 | Not DF | Under | Not DF |
| FBgn0142949 | FBgn0265512 | <i>mlt</i> | GO:0007291—sperm individualization | Male sterility. Defects in sperm individualization | 37 | Not DF | Not DF | Under |
| FBgn0142705 | FBgn0002865/ FBgn0004171 | <i>Mst98Ca/ Mst98Cb</i> | GO:0007286—spermatid development | Structural proteins in the sperm tail | 45 | Not DF | Not DF | Under |
| FBgn0146037 | FBgn0260942 | <i>bond</i> | GO:0007112—male meiosis cytokinesis | Failure of cytokinesis in dividing spermatocytes | 46 | Not DF | Not DF | Under |
| FBgn0138561 | FBgn0052529 | <i>Hers</i> | GO:0006342: chromatin silencing | Affect the regulation process of cell proliferation/differentiation | 40 | Not DF | Over | Over |
| FBgn0141892 | FBgn0010052 | <i>Jhe</i> | GO:0060179—male mating behavior | Reduced courtship mutant males | 36,47 | Not DF | Not DF | Over |
| FBgn0140780 | FBgn0086681/ FBgn0261349 | <i>mst36F/ mst36Fb</i> | GO:0007018: microtubule-based movement | Reduction of male fertility | 35 | Not DF | Not DF | Under |

Table 1. DEG in male hybrids gonads with reproductive functions. Deregulated genes in male hybrid gonads presenting spermatogenesis-related functions and their mutant phenotype in other *Drosophila* species associated to them. DF, Differentially expressed; Under, Underexpressed; Over, Overexpressed.

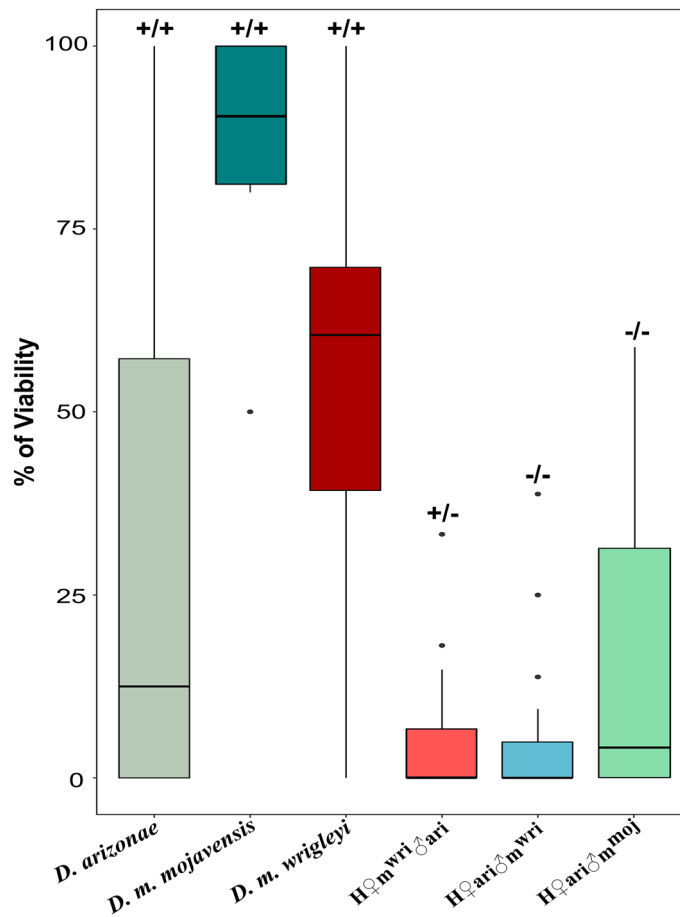


Figure 6. Average viability and fertility of intraspecific and interspecific offspring. Average viability (%) of *D. arizonae*, *D. m. wrigleyi* and *D. m. mojavensis* and their respective hybrids, $H_{\text{♀}m}^{\text{wri}} \text{♂}^{\text{ari}}$, $H_{\text{♀}ari}^{\text{wri}} \text{♂}^{\text{m}}$ and $H_{\text{♀}ari}^{\text{m}} \text{♂}^{\text{moj}}$. Sperm motility and fertility are represented by— and + signals. +/+ : motile sperm/fertile, +/- : motile sperm/sterile, -/- : amotile sperm/sterile. The boxplot was generated using the ggplot2 package (version 3.3.3) in R⁷⁹.

be overexpressed, but over time, global deregulation will increase due to the accumulation of genetic changes, and the number of over- and underexpressed genes will become symmetric.

Male hybrids from crosses performed in different directions exhibited distinct proportions of genes with transgressive expression (over- or underexpressed relative to the parental species). In hybrids in which the mother was *D. m. wrigleyi*, we observed an overexpression of genes when compared to the parental lines. In contrast, if the mother was *D. arizonae*, there was either no difference between the proportion of over- or underexpressed genes or we observed a bias towards underexpression in relation to the parental lines. The differences observed between the reciprocal crosses could be due to the origin of the sex chromosomes. It has been reported that some specific epistatic factors among sex chromosomes and autosomes play a role in the differential expression profiles of interspecific hybrids⁵⁹, which could explain the differences we observed.

Moreover, differences in sex-autosome interactions between reciprocal hybrids could also affect the severity of the sterility phenotype. There is evidence that in hybrids between *D. arizonae* females and *D. m. mojavensis* males, amotile sperm can result from interactions between the *D. m. mojavensis* Y chromosome and the 3rd autosome and/or the X chromosome from *D. arizonae*⁶⁰. However, in the reciprocal cross, male hybrid sterility was associated with interactions between the Y chromosome of *D. arizonae* and the 4th autosome of *D. mojavensis*⁶¹. Our analyses of sperm motility and fertility showed that hybrid males from crosses between *D. arizonae* females and *D. m. mojavensis* or *D. m. wrigleyi* males had amotile sperm and were sterile, as expected. The sperm of hybrid males from crosses between *D. mojavensis* females and *D. arizonae* males was motile, exhibiting normal tails, but the individuals were sterile, indicating that the disruption of spermatogenesis also occurred in these males. These findings are in agreement with Hardy et al.⁶², who observed that *D. arizonae*-*D. m. mojavensis* hybrids display abnormal spermatid development and disruption of the spermatid tails.

The differential gene expression pattern in hybrid testes that we observed for crosses with *D. arizonae* mothers, with a tendency for an excess of underexpressed genes in hybrids when compared to the parental lines, is thus associated with a more severe male sterile phenotype. The functions associated with these underexpressed genes are related to spermatogenesis and could indicate a breakdown in gene regulation during spermatogenesis^{55–57,63–65}. Moreover, we also identified genes related to metabolism, which supports the reduction

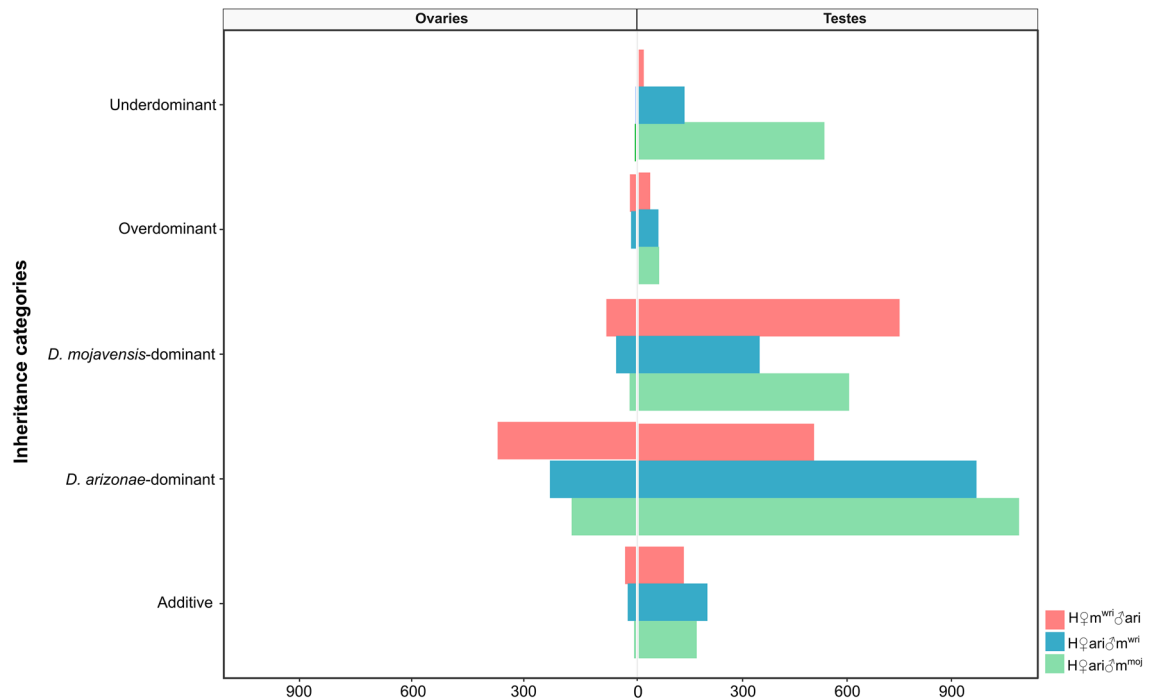


Figure 7. Inheritance of gene expression patterns in ovaries and testes of $H_{\text{♀}}m^{\text{wri}}\delta_{\text{ari}}$, $H_{\text{♀}}\text{ari}\delta_{\text{m}^{\text{wri}}}$ and $H_{\text{♀}}\text{ari}\delta_{\text{m}^{\text{moj}}}$. Gene expression patterns were classified into six categories of inheritance depending on the significance of the differential expression measured by performing pairwise comparisons in all conditions, according to McManus et al.¹³. These categories are as follows: Conserved, when the level of gene expression in hybrids is similar to that of both parental lines (not show in the plot). Additive, when the expression levels are different between the two parental lines, but the hybrid expression level is intermediate. *D. arizonae*-dominant or *D. mojavensis*-dominant, when the hybrid expression level is similar to that of only one parental line. Overdominant, when the expression level in the hybrids is significantly higher than that in both parental lines. Underdominant, when the hybrid expression level is significantly lower than that in both parental lines. The bar plot was generated using the ggplot2 package (version 3.3.3) in R⁷⁹.

of hybrid viability. These findings are in agreement with several studies that reported that most of the DEGs in male hybrids are underexpressed, and their functions are associated with reproduction, such as meiotic arrest, spermatogenesis^{4,6,8} and spermiogenesis, which are often observed in sterile hybrids^{66–69}.

Analyses of the inheritance of gene expression patterns showed that most of the genes had conserved expression in hybrid ovaries and testes. This result is similar to previous findings for *D. arizonae*-*D. m. mojavensis* female hybrids³⁰ but is quite different from those found in *D. sechellia*-*D. melanogaster* and *D. simulans*-*D. melanogaster* hybrids, since only 6% and 11% of their genes had conserved expression, respectively^{9,15}. Furthermore, given that *D. simulans*-*D. melanogaster* and *D. arizonae*-*D. mojavensis* may have similar divergence times^{21,70}, the number of genes classified as conserved is quite different (11% vs. ~95%), indicating that the interactions between different genetic changes accumulated in distinct species might have stronger or weaker effects on hybrids, causing disturbances in gene expression to variable degrees.

Among the DEGs with dominant inheritance in the ovaries and testes of all hybrids, most showed *D. arizonae*-like expression, regardless of the crossing direction. These findings could indicate that in the gonads of hybrids, the effect of the *D. arizonae* genome is more important than the maternal effects. Therefore, we speculate that for hybrid ovaries, the *D. arizonae* genome has a stronger effect on hybrid gene expression, likely due to regulatory sequences. This result corroborates our previous findings³⁰ that among the genes classified in the dominant category, for both crossing directions, the *D. arizonae*-dominant inheritance pattern was stronger. Likewise, the expression profile in the ovaries of hybrids from crosses between *D. melanogaster* females and *D. sechellia* males showed that 49% of the genes were classified as dominant, and among these, 84% showed *D. sechellia*-like expression¹⁵, indicating that the maternal species does not influence the expression profile of the hybrids. In the target species of the current study, the hybrid female gonads showed very few genes classified as over- or underdominant, agreeing with previously reported results³⁰ and demonstrating the greater stability of their gene expression. In testes, on the other hand, the number of differentially expressed genes classified in these categories was higher, with a higher number of underdominant genes for one specific cross direction. In hybrid male gonads, the underexpression of genes, mainly related to reproduction, has often been observed^{15,56,71,72}, indicating that this can be related to disruptions in regulatory networks driven by the rapid divergence of male-biased genes, leading to sterility^{7,56}.

Therefore, here, we showed that in hybrids from recently diverged species, such as *D. arizonae* and *D. mojavensis*, the degree of misregulated gene expression is related to the severity of the sterile phenotype. In female hybrids, which are fertile, very few DEGs were identified when compared with both parental lines. In

contrast, in male hybrids, the degree of misregulated gene expression was dependent on the subspecies of *D. mojavensis* and, most importantly, the cross direction. Hybrid males from *D. mojavensis* females, which are sterile but had motile sperm, displayed a smaller number of misregulated genes, with a bias towards overexpression; these genes are not directly related to the sterile phenotype, since few genes acting on reproduction were found. Nevertheless, in male hybrids carrying the X chromosome of *D. arizonae*, which are sterile with amotile sperm, the disruption of gene expression was higher and presented a bias towards underexpression. Surprisingly, most of these genes were directly related to spermatogenesis-related functions, as well as sperm movement. However, more analyses must be undertaken to clarify the regulatory differences between this pair of species. These analyses include investigating the divergence of male-biased gene sequences, regulatory studies (cis–trans regulatory changes and the impact of microRNAs on gene expression), and functional analyses of the genes identified in the current study. Thus, we can obtain a better understanding of their relationship with the sterile phenotype in the initial steps of hybrid incompatibility.

Methods

***Drosophila* strains and RNA sequencing.** Intraspecific and interspecific reciprocal crosses were performed between *D. arizonae* from Metztitlan, Hidalgo, México (stock number: 15081-1271.17) and two subspecies of *D. mojavensis*: *D. m. mojavensis* from the Anza Borrego Desert, California, USA (stock number: 15081-1352.01) and *D. m. wrightleyi* from Catalina Island, California, USA (stock number: 15081-1352.22). Crosses were performed with 3-day-old flies, ten males and ten females, in 2.3 × 9.5 cm vials containing standard *Drosophila* medium supplemented with yeast under the same temperature (23 °C) and humidity conditions. One-day-old virgin female and male offspring (control and F₁ hybrids) were collected after hatching and were isolated until they reached sexual maturity. The male and female reproductive tracts of 9- to 12-day-old flies were dissected in PBS (phosphate-buffered saline) and stored at –80 °C until use for RNA extraction.

To verify the hybrid status of the F₁ offspring of interspecific crosses, 10 individuals of each cross were randomly collected for DNA extraction, and PCR for the ribosomal ITS-1 (*internal transcribed spacer 1*) from the 18S gene region (NCBI Reference Sequence: EU306666.1)⁷³ was performed. The oligonucleotide primer for ITS-1 amplified 500 bp and 550 bp amplicons in *D. arizonae* and *D. mojavensis*, respectively. Therefore, in hybrids, two different fragments corresponding to *D. arizonae* and *D. mojavensis* alleles were expected.

After confirming the hybrid status of the offspring, 30 pairs of ovaries and 50 pairs of testes were used to perform total RNA extraction with two biological replicates using the RNeasy kit (Qiagen). The samples were treated with DNase (DNA-free Kit, Ambion) and stored at –80 °C. The samples were quantified by fluorescence in a Bioanalyzer 2100 (Agilent). Sequencing was performed using the GenomEast platform by a member of the France Génomique consortium (ANR-10-INBS-0009) with an Illumina HiSeq 4000. The samples were sequenced in 2 × 100 paired-end reads, and the average size of the inserts was 300 base pairs.

Twelve transcriptomes were sequenced with two biological replicates each: *D. arizonae*, *D. m. mojavensis*, and *D. m. wrightleyi* (controls, ovaries and testes) and hybrids from crosses between *D. arizonae* and both *D. mojavensis* subspecies (ovaries and testes). The hybrid transcriptomes from *D. m. mojavensis* female and *D. arizonae* male crosses were not sequenced because the hybrid incompatibility in this direction was very high, and it was not possible to obtain enough material to perform RNA extraction. The low number of replicates is due to the high index of prezygotic reproductive isolation between *D. arizonae* and *D. mojavensis* subspecies (ranging from 0.56 to ~0.70²⁵) and the low average viability of the hybrid offspring, which limited the number of hybrids obtained to perform RNA extraction.

Mapping and quantification of expression. The sequenced transcriptomes were trimmed using UrQ⁷⁴ to remove polyA tails and low-quality nucleotides. The sequence quality was then checked with FastQC software⁷⁵. The transcriptomes were aligned against all annotated coding sequences (CDSs) of the *D. mojavensis* r1.04 public genome⁷⁶ (available at <http://flybase.org/>). Overall, 21,915 Ref-Seq sequences were downloaded from <https://www.ncbi.nlm.nih.gov/refseq/>. From those sequences, 20,110 corresponding to mRNA were used as a reference to perform the alignments. This approach was used because the public genome of *D. mojavensis* presents the best quality of sequences and because *D. mojavensis* and *D. arizonae* are recently diverged species, a large divergence in their coding protein genes was not expected. Kallisto⁷⁷ was used to map the reads from parental and hybrid transcriptomes against the *D. mojavensis* r1.04 reference transcripts. Kallisto is able to perform rapid pseudoalignment to quickly determine the compatibility of the reads with their respective targets. The pseudoalignment of reads preserves the key information needed for quantification and is robust against errors, presenting a similar accuracy as other alignment tools⁷⁷. After the mapping procedure, BioMart⁷⁸, an R (3.6.1)⁷⁹ Bioconductor package, was used to recover the gene names corresponding to each transcript from the reference. This was possible because the BioMart database is maintained by Ensembl⁸⁰, providing direct access to a diverse set of data and enabling a wide range of powerful online queries, from gene annotation to database mining. Subsequently, due to several genes displaying different isoforms, the package tximport was used to summarize the transcript level estimation for the gene level analysis, allowing us to use these data for the differential expression of gene-level counts.

Differential expression analyses. Differential expression analyses were performed using DESeq2⁸¹, an R (3.6.1) package⁷⁹, using raw read counts to identify differentially expressed genes in the hybrids compared to the parental species (controls lines) for each gonad tissue. This package normalizes the counts using size factors that are estimated according to the median counts taken for all genes. Additionally, DESeq2 estimates the means and variances of raw read counts and tests for differential expression based on a model using a negative binomial distribution and uses Benjamini–Hochberg multiple test correction (FDR level of 0.01). The low number of rep-

licates can influence the statistical support of differential expression analyses. Therefore, to be conservative, we implemented stringent statistical thresholds. Genes were classified as significantly differentially expressed when the p -value, which was adjusted by FDR level, was below 0.01 and a higher than twofold change in expression was observed (corresponding to $\log_2(\text{FC}) > |1|$). Transcripts that presented fewer than ten mapped reads in all conditions tested were excluded from the analyses. The number of over- and underexpressed genes was analysed by a proportion test (`prop.test`) with R software⁷⁹.

Functional annotation and gene ontology enrichment analyses. Functional annotation was performed for all DEGs identified in the ovaries and testes of parents and hybrids. For this analysis, an orthologous gene table for *Drosophila* species was downloaded from <http://flybase.org/>. The *D. melanogaster* orthologs corresponding to DEGs in the hybrids were submitted to DAVID GO^{82,83}. In addition, gene enrichment was investigated using a list of specific DEGs. Thus, a target gene list was compared with a reference list, which contained all the genes in the PANTHER Classification System platform^{84–86} (available at <http://geneontology.org/>) for a selected organism, using Fisher's exact tests with FDR corrections or p -values. Then, we selected all significant GO terms from our target gene list and submitted them to the REVIGO web server⁸⁷. By using REVIGO, we were able to summarize and remove redundant GO terms.

Inheritance classifications. The R (3.6.1) package⁷⁹ was used to sort genes in terms of differences in their expression levels between each parental line and the reciprocal hybrids separately, according to McManus et al.¹⁵. The expression data were transformed into log percentages, and a threshold of twofold change and adjusted p -value < 0.01 were set to determine the significance of differentially expressed genes. Genes that were not differentially expressed were considered to have the same expression level as the parental lines, thus being considered conserved. Genes considered differentially expressed were classified as additive, dominant, underdominant or overdominant. Additive expression means that the expression level of a given gene is different between the two parental lines but intermediate in the hybrid. *D. arizonae*-dominant or *D. mojavensis*-dominant expression is when the hybrid expression is similar to only one parental line. Overdominant expression means that expression in the hybrids is significantly higher than that in both parental lines, while underdominant expression means that the hybrid expression of a given gene is significantly lower than that in both parental lines. Chi-square statistical tests were performed in the R (3.6.1) package⁷⁹.

Viability and sterility analyses. Virgin males and females of each strain were separated by sex 10 h after eclosion and stored separately in yeasted cactus-banana vials, with 10 flies per vial, until flies were sexually mature (9 days of age). Crosses were performed between *D. m. mojavensis*, *D. m. wrightleyi* and *D. arizonae*, as well as within the parental lines, as control crosses. Five replicates were performed per cross with 10 couples in each vial, which favours mating³⁰. Mating was performed for 72 h under similar temperature (23°) and light/dark (10/14 h) conditions. After 72 h, males from all crosses were discarded, and females (in pairs) were transferred to new fresh vials to lay eggs. This process was repeated five times every 48 h, so the females laid eggs for 10 days. Immediately after removing the females, the eggs that were laid were counted under a stereomicroscope, and after eclosion (~19 days after crossing), the number of imagoes was verified once a week for four weeks. The offspring viability ($\text{adults/eggs} \times 100$) was calculated based on the number of eggs and adults.

For sterility analysis, interspecific and control crosses were performed using three-day-old virgin flies to obtain as many hybrids as possible. In previous tests, we noticed an increased production of hybrids when the two species were kept together before they reached sexual maturity. All crosses were performed with five replicates under the same temperature and light/dark conditions for 12 days. Then, the parents were discarded, and the imagoes were separated by sex daily. The offspring were maintained in yeasted food vials until they reached 10 days of age (sexually mature). Sperm motility analyses were carried out for 20 F_1 male testes and seminal vesicles of each control and interspecific cross, according to Reed et al.⁵⁹. No statistical analyses were performed because for each cross, all males presented the same phenotype, motile or amotile sperm. Additionally, fertility analyses were carried out by backcrossing female and male hybrids with their respective parents, *D. arizonae*, *D. m. mojavensis* and *D. m. wrightleyi*. Crosses were performed with five couples per replicate with five replicates by cross. To ensure that the absence of offspring was not due to possible prezygotic, post-mating-prezygotic or postzygotic isolation mechanisms, we increased the crossing time and allowed the couples to mate for 15 days. After that, all parents were discarded, and fertility was evaluated based on the presence or absence of offspring, as reported by Carnelossi et al.²⁸. $F_1 \times F_1$ crosses were also performed using the offspring of each interspecific cross under the same conditions as the backcrosses. To check that tubes containing only eggs would not produce offspring, the tubes were maintained for 20 days after parent removal and then discarded. Statistical analyses were performed for average fecundity by and viability for each replicate of intraspecific and interspecific crosses by using R (3.6.1) package⁷⁹. Normality and variance tests (Shapiro–Wilk and Levene's test, respectively) were carried out, and when we obtained significant p -values (non-normal distribution), a nonparametric Kruskal–Wallis test was performed. Then, a post hoc Wilcoxon test was performed to determine significant differences between the treatments. For results with no significant p -values for normality and variance tests, one-way ANOVA was performed using Tukey's post hoc test.

Data availability

The datasets used and/or analysed during the current study are at <https://www.ncbi.nlm.nih.gov/sra>, Submission PRJNA691040.

Received: 24 November 2020; Accepted: 19 April 2021

Published online: 10 May 2021

References

- Coyne, J. A. & Orr, H. A. Patterns of speciation in *Drosophila* revisited. *Evolution* **51**, 295–303. <https://doi.org/10.1111/j.1558-5646.1997.tb02412.x> (1997).
- Orr, H. A. & Coyne, J. A. The genetics of postzygotic isolation in the *Drosophila virilis* group. *Genetics* **121**, 527–537 (1989).
- Turissini, D. A., McGirr, J. A., Patel, S. S., David, J. R. & Matute, D. R. The rate of evolution of postmating-prezygotic reproductive isolation in *Drosophila*. *Mol. Biol. Evol.* **35**, 312–334. <https://doi.org/10.1093/molbev/msx271> (2018).
- Brill, E., Kang, L., Michalak, K., Michalak, P. & Price, D. K. Hybrid sterility and evolution in Hawaiian *Drosophila*: differential gene and allele-specific expression analysis of backcross males. *Heredity (Edinb.)* **117**, 100–108. <https://doi.org/10.1038/hdy.2016.31> (2016).
- Hill, T., Schlotterer, C. & Betancourt, A. J. Hybrid dysgenesis in *Drosophila simulans* associated with a rapid invasion of the P-element. *PLoS Genet.* **12**, e1005920. <https://doi.org/10.1371/journal.pgen.1005920> (2016).
- Li, R. *et al.* Specific down-regulation of spermatogenesis genes targeted by 22G RNAs in hybrid sterile males associated with an X-chromosome introgression. *Genome Res.* **26**, 1219–1232. <https://doi.org/10.1101/gr.204479.116> (2016).
- Llopart, A. The rapid evolution of X-linked male-biased gene expression and the large-X effect in *Drosophila yakuba*, *D. santomea*, and their hybrids. *Mol. Biol. Evol.* **29**, 3873–3886. <https://doi.org/10.1093/molbev/mss190> (2012).
- Moehring, A. J., Teeter, K. C. & Noor, M. A. Genome-wide patterns of expression in *Drosophila* pure species and hybrid males. II. Examination of multiple-species hybridizations, platforms, and life cycle stages. *Mol. Biol. Evol.* **24**, 137–145. <https://doi.org/10.1093/molbev/msl142> (2007).
- Ranz, J. M., Namgyal, K., Gibson, G. & Hartl, D. L. Anomalies in the expression profile of interspecific hybrids of *Drosophila melanogaster* and *Drosophila simulans*. *Genome Res.* **14**, 373–379. <https://doi.org/10.1101/gr.2019804> (2004).
- Ruiz, A., Heed, W. B. & Wasserman, M. Evolution of the *mojavensis* cluster of cactophilic *Drosophila* with descriptions of two new species. *J. Hered.* **81**, 30–42. <https://doi.org/10.1093/oxfordjournals.jhered.a110922> (1990).
- Dobzhansky, T. Studies on hybrid sterility. II. Localization of sterility factors in *Drosophila Pseudoobscura* hybrids. *Genetics* **21**, 113–135 (1936).
- Muller, H. J. Isolating mechanisms evolution and temperature. *Biol. Symp.* **6**, 71–125 (1942). *Biol. Symp.* **6**, 44 (1942).
- Kelleher, E. S., Edelman, N. B. & Barbash, D. A. *Drosophila* interspecific hybrids phenocopy piRNA-pathway mutants. *PLoS Biol.* **10**, e1001428. <https://doi.org/10.1371/journal.pbio.1001428> (2012).
- Romero-Soriano, V. *et al.* Transposable element misregulation is linked to the divergence between parental piRNA pathways in *Drosophila* hybrids. *Genome Biol. Evol.* **9**, 1450–1470. <https://doi.org/10.1093/gbe/evx091> (2017).
- McManus, C. J. *et al.* Regulatory divergence in *Drosophila* revealed by mRNA-seq. *Genome Res.* **20**, 816–825. <https://doi.org/10.1101/gr.102491.109> (2010).
- Haldane, J. B. S. Sex ratio and unisexual sterility in animal hybrids. *J. Genet.* **12**, 8 (1922).
- Llopart, A., Brud, E., Pettie, N. & Comeron, J. M. Support for the dominance theory in *Drosophila* transcriptomes. *Genetics* **210**, 703–718. <https://doi.org/10.1534/genetics.118.301229> (2018).
- Ryazansky, S., Mikhaleva, E., Akulenko, N. & Olenkina, O. Testis-expressed cluster of microRNAs 959–964 controls spermatid differentiation in *Drosophila*. *013243*. <https://doi.org/10.1101/013243> (2015).
- Reed, L. K., Nyboer, M. & Markow, T. A. Evolutionary relationships of *Drosophila mojavensis* geographic host races and their sister species *Drosophila arizonae*. *Mol. Ecol.* **16**, 1007–1022. <https://doi.org/10.1111/j.1365-294X.2006.02941.x> (2007).
- Russo, C. A., Takezaki, N. & Nei, M. Molecular phylogeny and divergence times of drosophilid species. *Mol. Biol. Evol.* **12**, 391–404. <https://doi.org/10.1093/oxfordjournals.molbev.a040214> (1995).
- Sanchez-Flores, A. *et al.* Genome evolution in three species of cactophilic *Drosophila*. *G3 (Bethesda)* **6**, 3097–3105. <https://doi.org/10.1534/g3.116.033779> (2016).
- Jennings, J. H. & Etges, W. J. Species hybrids in the laboratory but not in nature: a reanalysis of premating isolation between *Drosophila arizonae* and *D. mojavensis*. *Evolution* **64**, 587–598. <https://doi.org/10.1111/j.1558-5646.2009.00834.x> (2010).
- Etges, W. J., De Oliveira, C. C., Noor, M. A. & Ritchie, M. G. Genetics of incipient speciation in *Drosophila mojavensis*. III. Life-history divergence in allopatry and reproductive isolation. *Evolution* **64**, 3549–3569. <https://doi.org/10.1111/j.1558-5646.2010.01096.x> (2010).
- Machado, C. A., Matzkin, L. M., Reed, L. K. & Markow, T. A. Multilocus nuclear sequences reveal intra- and interspecific relationships among chromosomally polymorphic species of cactophilic *Drosophila*. *Mol. Ecol.* **16**, 3009–3024. <https://doi.org/10.1111/j.1365-294X.2007.03325.x> (2007).
- Massie, K. R. & Markow, T. A. Sympatry, allopatry and sexual isolation between *Drosophila mojavensis* and *D. arizonae*. *Heredity* **142**, 51–55. <https://doi.org/10.1111/j.1601-5223.2005.01911.x> (2005).
- Matzkin, L. M. Population genetics and geographic variation of alcohol dehydrogenase (Adh) paralogs and glucose-6-phosphate dehydrogenase (G6pd) in *Drosophila mojavensis*. *Mol. Biol. Evol.* **21**, 276–285. <https://doi.org/10.1093/molbev/msh019> (2004).
- Matzkin, L. M. The molecular basis of host adaptation in cactophilic *Drosophila*: molecular evolution of a glutathione S-transferase gene (GstD1) in *Drosophila mojavensis*. *Genetics* **178**, 1073–1083. <https://doi.org/10.1534/genetics.107.083287> (2008).
- Carnelossi, E. A. *et al.* Specific activation of an I-like element in *Drosophila* interspecific hybrids. *Genome Biol. Evol.* **6**, 1806–1817. <https://doi.org/10.1093/gbe/evu141> (2014).
- Reed, L. K. & Markow, T. A. Early events in speciation: polymorphism for hybrid male sterility in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **101**, 9009–9012. <https://doi.org/10.1073/pnas.0403106101> (2004).
- Lopez-Maestre, H. *et al.* Identification of misexpressed genetic elements in hybrids between *Drosophila*-related species. *Sci. Rep.* **7**, 40618. <https://doi.org/10.1038/srep40618> (2017).
- Dauwalder, B., Tsujimoto, S., Moss, J. & Mattox, W. The *Drosophila* takeout gene is regulated by the somatic sex-determination pathway and affects male courtship behavior. *Genes Dev.* **16**, 2879–2892. <https://doi.org/10.1101/gad.1010302> (2002).
- Vanaphan, N., Dauwalder, B. & Zufall, R. A. Diversification of takeout, a male-biased gene family in *Drosophila*. *Gene* **491**, 142–148. <https://doi.org/10.1016/j.gene.2011.10.003> (2012).
- Chatterjee, N., Rollins, J., Mahowald, A. P. & Bazinet, C. Neurotransmitter Transporter-like: a male germline-specific SLC6 transporter required for *Drosophila* spermiogenesis. *PLoS ONE* **6**, e16275. <https://doi.org/10.1371/journal.pone.0016275> (2011).
- Courtot, C., Fankhauser, C., Simanis, V. & Lehner, C. F. The *Drosophila* cdc25 homolog twine is required for meiosis. *Development* **116**, 405–416 (1992).
- Di Cara, F., Cavaliere, D., Galliero, V., Polito, L. C. & Digilio, F. A. Expressional and functional analysis of the male-specific cluster mst36F during *Drosophila* spermatogenesis. *Insect. Mol. Biol.* **19**, 807–813. <https://doi.org/10.1111/j.1365-2583.2010.01033.x> (2010).
- Ellis, L. L. & Carney, G. E. Mating alters gene expression patterns in *Drosophila melanogaster* male heads. *BMC Genom.* **11**, 558. <https://doi.org/10.1186/1471-2164-11-558> (2010).
- Fabrizio, J. J. *et al.* Mulet (mlet) encodes a tubulin-binding cofactor E-like homolog required for spermatid individualization in *Drosophila melanogaster*. *Fly (Austin)* **6**, 261–272. <https://doi.org/10.4161/fly.21533> (2012).

38. Gumy, L. F. *et al.* The kinesin-2 family member KIF3C regulates microtubule dynamics and is required for axon growth and regeneration. *J. Neurosci.* **33**, 11329–11345. <https://doi.org/10.1523/JNEUROSCI.5221-12.2013> (2013).
39. Henson, J. H. *et al.* The heterotrimeric motor protein kinesin-II localizes to the midpiece and flagellum of sea urchin and sand dollar sperm. *Cell Motil. Cytoskeleton* **38**, 29–37. [https://doi.org/10.1002/\(SICI\)1097-0169\(1997\)38:1%3c29::AID-CM4%3e3.0.CO;2-C](https://doi.org/10.1002/(SICI)1097-0169(1997)38:1%3c29::AID-CM4%3e3.0.CO;2-C) (1997).
40. Ito, S. *et al.* Epigenetic silencing of core histone genes by HERS in *Drosophila*. *Mol. Cell* **45**, 494–504. <https://doi.org/10.1016/j.molcel.2011.12.029> (2012).
41. Karak, S. *et al.* Diverse roles of axonemal dyneins in *Drosophila* auditory neuron function and mechanical amplification in hearing. *Sci. Rep.* **5**, 17085. <https://doi.org/10.1038/srep17085> (2015).
42. Lin, C. J. *et al.* Characterization of a TUTase/RNase complex required for *Drosophila* gametogenesis. *RNA* **23**, 284–296. <https://doi.org/10.1261/rna.059527.116> (2017).
43. Lin, T. Y. *et al.* Coordinate developmental control of the meiotic cell cycle and spermatid differentiation in *Drosophila* males. *Development* **122**, 1331–1341 (1996).
44. Rogers, S. L., Rogers, G. C., Sharp, D. J. & Vale, R. D. *Drosophila* EB1 is important for proper assembly, dynamics, and positioning of the mitotic spindle. *J. Cell Biol.* **158**, 873–884. <https://doi.org/10.1083/jcb.200202032> (2002).
45. Schafer, M., Borsch, D., Hulster, A. & Schafer, U. Expression of a gene duplication encoding conserved sperm tail proteins is translationally regulated in *Drosophila melanogaster*. *Mol. Cell Biol.* **13**, 1708–1718. <https://doi.org/10.1128/mcb.13.3.1708> (1993).
46. Szafer-Glusman, E. *et al.* A role for very-long-chain fatty acids in furrow ingression during cytokinesis in *Drosophila* spermatocytes. *Curr. Biol.* **18**, 1426–1431. <https://doi.org/10.1016/j.cub.2008.08.061> (2008).
47. Wijesekera, T. P., Saurabh, S. & Dauwalder, B. Juvenile hormone is required in adult males for *Drosophila* courtship. *PLoS ONE* **11**, e0151912. <https://doi.org/10.1371/journal.pone.0151912> (2016).
48. Zhang, J., Luo, J., Chen, J., Dai, J. & Montell, C. The role of Y chromosome genes in male fertility in *Drosophila melanogaster*. *Genetics* **215**, 623–633. <https://doi.org/10.1534/genetics.120.303324> (2020).
49. Zur Lage, P., Newton, F. G. & Jarman, A. P. Survey of the ciliary motility machinery of *Drosophila* sperm and ciliated mechanosensory neurons reveals unexpected cell-type specific variations: a model for motile ciliopathies. *Front. Genet.* **10**, 24. <https://doi.org/10.3389/fgene.2019.00024> (2019).
50. Rifkin, S. A., Kim, J. & White, K. P. Evolution of gene expression in the *Drosophila melanogaster* subgroup. *Nat. Genet.* **33**, 138–144. <https://doi.org/10.1038/ng1086> (2003).
51. Turissini, D. A. & Matute, D. R. Fine scale mapping of genomic introgressions within the *Drosophila yakuba* clade. *PLoS Genet.* **13**, e1006971. <https://doi.org/10.1371/journal.pgen.1006971> (2017).
52. Wang, R. L., Wakeley, J. & Hey, J. Gene flow and natural selection in the origin of *Drosophila pseudoobscura* and close relatives. *Genetics* **147**, 1091–1106 (1997).
53. Gomes, S. & Civetta, A. Hybrid male sterility and genome-wide misexpression of male reproductive proteases. *Sci. Rep.* **5**, 11976. <https://doi.org/10.1038/srep11976> (2015).
54. Artieri, C. G., Haerty, W. & Singh, R. S. Association between levels of coding sequence divergence and gene misregulation in *Drosophila* male hybrids. *J. Mol. Evol.* **65**, 697–704. <https://doi.org/10.1007/s00239-007-9048-2> (2007).
55. Civetta, A., Rajakumar, S. A., Brouwers, B. & Bacik, J. P. Rapid evolution and gene-specific patterns of selection for three genes of spermatogenesis in *Drosophila*. *Mol. Biol. Evol.* **23**, 655–662. <https://doi.org/10.1093/molbev/msj074> (2006).
56. Haerty, W. & Singh, R. S. Gene regulation divergence is a major contributor to the evolution of Dobzhansky–Muller incompatibilities between species of *Drosophila*. *Mol. Biol. Evol.* **23**, 1707–1714. <https://doi.org/10.1093/molbev/msl033> (2006).
57. Meiklejohn, C. D., Parsch, J., Ranz, J. M. & Hartl, D. L. Rapid evolution of male-biased gene expression in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **100**, 9894–9899. <https://doi.org/10.1073/pnas.1630690100> (2003).
58. Bono, J. M., Matzkin, L. M., Hoang, K. & Brandsmeier, L. Molecular evolution of candidate genes involved in post-mating-prezygotic reproductive isolation. *J. Evol. Biol.* **28**, 403–414. <https://doi.org/10.1111/jeb.12574> (2015).
59. Reed, L. K., LaFlamme, B. A. & Markow, T. A. Genetic architecture of hybrid male sterility in *Drosophila*: analysis of intraspecific variation for interspecific isolation. *PLoS ONE* **3**, e3076. <https://doi.org/10.1371/journal.pone.0003076> (2008).
60. Vigneault, G. & Zouros, E. The genetics of asymmetrical male sterility in *Drosophila mojavensis* and *Drosophila arizonensis* hybrids: interactions between the Y-chromosome and autosomes. *Evolution* **40**, 1160–1170. <https://doi.org/10.1111/j.1558-5646.1986.tb05741.x> (1986).
61. Pantazidis, A. C. & Zouros, E. Location of an autosomal factor causing sterility in *Drosophila mojavensis* males carrying the *Drosophila arizonensis* Y chromosome. *Heredity (Edinb.)* **60**(Pt 2), 299–304. <https://doi.org/10.1038/hdy.1988.46> (1988).
62. Hardy, R. W., Lougheed, A. & Markow, T. A. Reproductive tract and spermatid abnormalities of hybrid males from reciprocal crosses between *Drosophila mojavensis* and *D. arizonae*. *Fly (Austin)* **5**, 76–80. <https://doi.org/10.4161/fly.5.2.15571> (2011).
63. Sundararajan, V. & Civetta, A. Male sex interspecific divergence and down regulation of expression of spermatogenesis genes in *Drosophila* sterile hybrids. *J. Mol. Evol.* **72**, 80–89. <https://doi.org/10.1007/s00239-010-9404-5> (2011).
64. Wu, C. I. & Davis, A. W. Evolution of postmating reproductive isolation: the composite nature of Haldane's rule and its genetic bases. *Am. Nat.* **142**, 187–212. <https://doi.org/10.1086/285534> (1993).
65. Bauer DuMont, V. L., Flores, H. A., Wright, M. H. & Aquadro, C. F. Recurrent positive selection at *bgn*, a key determinant of germ line differentiation, does not appear to be driven by simple coevolution with its partner protein *bam*. *Mol. Biol. Evol.* **24**, 182–191. <https://doi.org/10.1093/molbev/msl141> (2007).
66. Jiang, J. & White-Cooper, H. Transcriptional activation in *Drosophila* spermatogenesis involves the mutually dependent function of *aly* and a novel meiotic arrest gene *cookie monster*. *Development* **130**, 563–573. <https://doi.org/10.1242/dev.00246> (2003).
67. Lindsley, D. L., Roote, J. & Kennison, J. A. Anent the genomics of spermatogenesis in *Drosophila melanogaster*. *PLoS ONE* **8**, e55915. <https://doi.org/10.1371/journal.pone.0055915> (2013).
68. Moehring, A. J., Llopart, A., Elwyn, S., Coyne, J. A. & Mackay, T. F. The genetic basis of postzygotic reproductive isolation between *Drosophila santomea* and *D. yakuba* due to hybrid male sterility. *Genetics* **173**, 225–233. <https://doi.org/10.1534/genetics.105.052985> (2006).
69. White-Cooper, H., Schafer, M. A., Alphey, L. S. & Fuller, M. T. Transcriptional and post-transcriptional control mechanisms coordinate the onset of spermatid differentiation with meiosis I in *Drosophila*. *Development* **125**, 125–134 (1998).
70. Cutter, A. D. Divergence times in *Caenorhabditis* and *Drosophila* inferred from direct estimates of the neutral mutation rate. *Mol. Biol. Evol.* **25**, 778–786. <https://doi.org/10.1093/molbev/msn024> (2008).
71. Gomes, S. & Civetta, A. Misregulation of spermatogenesis genes in *Drosophila* hybrids is lineage-specific and driven by the combined effects of sterility and fast male regulatory divergence. *J. Evol. Biol.* **27**, 1775–1783. <https://doi.org/10.1111/jeb.12428> (2014).
72. Ometto, L., Ross, K. G., Shoemaker, D. & Keller, L. Disruption of gene expression in hybrids of the fire ants *Solenopsis invicta* and *Solenopsis richteri*. *Mol. Ecol.* **21**, 2488–2501. <https://doi.org/10.1111/j.1365-294X.2012.05544.x> (2012).
73. Baffi, M. A. & Ceron, C. R. Molecular analysis of the rDNA ITS-1 intergenic spacer in *Drosophila mulleri*, *D. arizonae*, and their hybrids. *Biochem. Genet.* **40**, 411–421. <https://doi.org/10.1023/a:1020729612672> (2002).
74. Modolo, L. & Lerat, E. UrQC: an efficient software for the unsupervised quality trimming of NGS data. *BMC Bioinform.* **16**, 137. <https://doi.org/10.1186/s12859-015-0546-8> (2015).
75. S., A. *FastQC: a quality control tool for high throughput sequence data.* <http://www.bioinformatics.babraham.ac.uk/projects/fastqc> (2010).

76. Drosophila 12 Genomes, C. *et al.* Evolution of genes and genomes on the Drosophila phylogeny. *Nature* **450**, 203–218. <https://doi.org/10.1038/nature06341> (2007).
77. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527. <https://doi.org/10.1038/nbt.3519> (2016).
78. Durinck, S. *et al.* BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* **21**, 3439–3440. <https://doi.org/10.1093/bioinformatics/bti525> (2005).
79. Team, R. C. R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*. <https://www.R-project.org/>. (2018).
80. Yates, A. D. *et al.* Ensembl 2020. *Nucl. Acids Res.* **48**, D682–D688. <https://doi.org/10.1093/nar/gkz966> (2020).
81. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550. <https://doi.org/10.1186/s13059-014-0550-8> (2014).
82. da Huang, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57. <https://doi.org/10.1038/nprot.2008.211> (2009).
83. da Huang, W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucl. Acids Res.* **37**, 1–13. <https://doi.org/10.1093/nar/gkn923> (2009).
84. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29. <https://doi.org/10.1038/75556> (2000).
85. Mi, H., Muruganujan, A. & Thomas, P. D. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucl. Acids Res.* **41**, D377–386. <https://doi.org/10.1093/nar/gks1118> (2013).
86. Thomas, P. D. *et al.* PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res.* **13**, 2129–2141. <https://doi.org/10.1101/gr.772403> (2003).
87. Supek, F., Bosnjak, M., Skunca, N. & Smuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* **6**, e21800. <https://doi.org/10.1371/journal.pone.0021800> (2011).

Acknowledgements

We thank Nelly Burette, Mariana Ferrarini and Sonia Martinez for the technical support. This work was supported by the Agence Nationale de la Recherche (Grant Number 14-CE19-0016); by Brazilian agency FAPESP—Fundação de Amparo à Pesquisa do Estado de São Paulo (Grant Number 2016/19271-2); and CNPq—Conselho Nacional de Desenvolvimento Científico e Tecnológico (Grant Numbers 303455/2017-9, 141413/2016-6); by fellowships from the Eiffel program; and IDEX Lyon.

Author contributions

Conceptualization: C.A.B., C.M.A.C., C.V. Data curation: C.A.B. Formal analysis: C.A.B., V.M. Funding acquisition: C.M.A.C., C.V. Investigation: C.A.B., C.M.A.C., C.V. Methodology: C.A.B., V.M., T.Y.K.O. Writing – original draft: C.A.B. Writing – review and editing: C.A.B., C.M.A.C., C.V.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-89366-2>.

Correspondence and requests for materials should be addressed to C.V.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021