

Letter to the Editor

Association of Misexpression with Sterility in Hybrids of *Drosophila simulans* and *D. mauritiana*

Pawel Michalak,* Mohamed A.F. Noor

Department of Biological Sciences, Life Sciences Building, Louisiana State University, Baton Rouge, LA 70803, USA

Received: 16 January 2004 / Accepted: 8 March 2004 [Reviewing Editor: Martin Kreitman]

Abstract. Recent studies have identified genes associated with hybrid sterility and other hybrid dysfunctions, but the consequences of introgressions of these "speciation genes" are often poorly understood. Previously, we identified a panel of genes that are underexpressed in sterile male hybrids of Drosophila simulans and D. mauritiana relative to pure species. Here, we build on this reverse-genetics approach to demonstrate that the underexpression of at least five of these genes in hybrids is associated with hybrid sterility and that these five genes are coordinately regulated. We map one upstream regulator of these genes to a region previously shown to harbor one or more factors causing hybrid sterility. Finally, we show that the genes underexpressed in hybrids are often highly conserved, as might be predicted for downstream targets of the genetic changes that cause hybrid sterility. This approach integrates forward genetics with reverse genetics to show a proximate consequence of the introgression of particular hybrid sterility-conferring regions between species: underexpression of genes necessary for normal spermatogenesis.

Key words: Hybrid sterility — Gene expression — Speciation

Introduction

The last decade has been an exciting time for genetic studies of attributes that prevent gene flow between species, such as hybrid sterility, hybrid inviability, and mating discrimination. Advances in this area are highly significant, as they give insights into the genetic changes that cause the formation of new species (speciation). Recently, genetic studies have moved from coarse recombinational mapping of genomic regions conferring these traits to fine-mapping and identifying genes that appear to be directly or indirectly involved (Noor 2003). Largely through this "forward-genetics" approach, four loci that contribute to hybrid dysfunctions such as sterility or inviability have been identified (Ting et al. 1998; Walter and Kazianis 2001; Barbash et al. 2003; Presgraves et al. 2003). Three of these four loci (Ting et al. 1998; Walter and Kazianis 2001; Barbash et al. 2003) appear to regulate transcription of other genes, based on sequence motifs identified within them. This consistency in function suggests that a reverse-genetics approach may be a fruitful complement to genetic studies of speciation.

The Drosophila melanogaster species group has been the focus of many of the most recent advances. For instance, two of its close relatives, *D. simulans* and *D. mauritiana*, produce sterile hybrid males but fertile hybrid females, allowing for genetic crosses. The architecture of this hybrid male sterility has been studied extensively using forward genetics, and this led to the identification of *Odysseus*, one of many genes that contribute to this hybrid dysfunction. The

^{*}Present address: Department of Biology, University of Texas at Arlington, Arlington, TX 76019-0498, USA

Correspondence to: Mohamed A.F. Noor; email: mnoor@lsu.edu

putative *Odysseus* gene produces a regulatory homeodomain protein (*OdsH*) (Ting et al. 1998).

Recently, we began a reverse genetics approach to examine hybrid dysfunctions between D. simulans and D. mauritiana by surveying global patterns of gene expression in these species and their hybrids using microarrays (Michalak and Noor 2003). We identified a panel of genes that were under- or overexpressed in hybrids relative to pure species, and many of these were confirmed via real-time PCR. We found that genes expressed primarily or exclusively in males, including several involved in fertility, were disproportionately underexpressed in hybrids. This observation is consistent with the more rapid expression evolution of male-specific genes within Drosophila (Meiklejohn et al. 2003; Ranz et al. 2003, 2004). Because germline transcription occurs early in Drosophila spermatogenesis (premeiotic) (Fuller 1998) while sterility arises at a later time when translation occurs (postmeiotically) (Wu et al. 1992), this underexpression may be directly associated with subsequent sterility. Alternatively, hybrid sterility and underexpression of various transcripts may be indirectly associated as separate consequences of the introgression of particular genes between species that cause developmental disruptions. However, our initial study provided no evidence for either kind of association between misexpression and sterility.

Here, we extend our previous study to address a suite of questions regarding hybrid under- or overexpression of six specific transcripts and hybrid male sterility in D. simulans/D. mauritiana. First, we use backcross analyses (first generation and fifth generation) to determine whether under- or overexpression of these transcripts may be directly or indirectly associated with sterility. Second, we examine whether several of the misexpressed genes may be coordinately regulated, suggesting changes in one or more trans-acting regulators in hybrids. Third, we explore whether such trans-acting regulators can be genetically mapped. Finally, we sequenced several of the genes underexpressed in hybrids to determine if they have experienced high rates of nonsynonymous substitution in the divergence of these species, as has been documented for various other reproductionrelated genes (Swanson and Vacquier 2002a, b).

Materials and Methods

Fertility Assay in First-Generation Backcrosses

 F_1 hybrid females (*Drosophila simulans* Florida City females \times *D. mauritiana* SYN males) were backcrossed to *D. simulans* FC males, and the male progeny were collected. Seven-day-posteclosion backcross males were individually mated with virgin *D. simulans* females. One hour after copulation, the females' reproductive tracts were removed in Insect Ringer's solution. Based on the sperm in the female tracts, males were assigned to one of three

categories: "sterile" (S: no full-length sperm observed in reproductive tract), "fertile" (F: full-length, motile sperm present), and "semifertile" (SF: full-length, immotile sperm present). Fertile males and semi-fertile males likely represent the same category, as confirmed by an independent experiment in which we scored purespecies *D. simulans* males as both F and SF (but never S) when mated to *D. simulans* females.

Quantitative Real-Time PCR

One hour after the fertility assay, RNA was extracted from individual backcross males using Bertucci and Noor's (2001) single-fly RNA isolation protocol and quantified by spectrophotometry. Samples from the three fertility categories were randomized in each step of their processing from RNA extraction to loading the PCR plates. Two hundred fifty nanograms of total RNA was reverse transcribed in a reaction of 5 mM MgCl₂, 50 mM KCl, 10 mM Tris-HCl, 1 mM dNTPs, 20 U RNasin, 20 U reverse transcriptase, a 2.5 µM concentration of target primer, and a 50 nM concentration of 18S rRNA reverse primer. QRT-PCR primers and fluorescent probe sequences are available upon request. Only DNA sequences that matched the sequences of both species were used for this purpose. Target probes were FAM-labeled and the control probe (18S rRNA) was VIC-labeled. Predeveloped ABI TaqMan assay reagents and ABI standard protocols were used to prepare the QRT-PCR reaction mixture. PCR conditions consisted of an initial 2 min at 50°C and 10 min at 95°C, followed by 40 cycles of 95°C for 15 s and 60°C for 1 min. ABI Prism 7000 SDS software was used to visualize and quantify the amplification products. Given the large difference in variance observed between categories, a nonparametric Mann-Whitney U-test was used to test for difference in threshold cycle number (C_T) between fertile and sterile males. The results were robust to various cycle thresholds tested and to whether normalization for 18S rRNA was applied.

Fifth-Generation Backcross Expression Assays

While a good "first-pass" assay, associations between fertility and expression in first-generation backcross progeny may be caused by linkage between genes affecting these two traits, particularly in a small genome such as that of Drosophila. As such, we produced fifth-generation backcross progeny (BC5) to rigorously test whether infertility and underexpression were consequences of the same genetic changes. Assuming normal recombination, BC5 flies possess on average only 3% of a haploid D. mauritiana genome. Single first-generation backcross females were mated to D. simulans males to produce second-generation backcross progeny. This process was repeated until fifth-generation backcross (BC5) males were produced. The lineage with the largest number of both fertile (eight) and sterile (nine) BC5 males was selected for analysis. The BC5 males were mated to D. simulans females as above and scored as S or F (none were SF) based on sperm morphology and motility. One hour after mating, the males were decapitated, RNA was extracted from their bodies, and QRT-PCR was performed as above. RNA extracted from these single flies was enough to survey expression of only five transcripts via QRT-PCR, so Mst98Cb was not surveyed.

Fifth-Generation Backcross Genotype Assays

DNA was extracted from the heads using a DNA squish protocol (Gloor and Engels 1992) and used for amplification of a panel of 11 microsatellites differentiating the two lines. The microsatellites were selected primarily based on their cytological positions. Most markers used are described by Colson et al. (1999), and their



Fig. 1. Threshold cycle (Ct) for real-time RT-PCR amplification of various transcripts from individual fertile (f) and sterile (s) fifthgeneration backcross hybrid males.

cytological positions in *D. melanogaster* are presented parenthetically: AF047180 (01), DS09021 (08), AC005732 (24), AC005555 (29), DM1433 (46), DS00361 (54), AC004658 (63), and DM22F11T (73). We also included one microsatellite within an intron of the *OdysseusH* gene (16D: primers: 5'-TGACTTTGCATGTCAGTCG TGTG-3', 5'-AGTTACGTTCGCCATTCGAATGAC-3'), one 2 megabases from the *OdysseusH* gene (13E: primers: 5'-CGAGCT GGTGCCGAGAATCTATATAA-3', 5'-ATTCCAAGGCAACT GTGTCGC-3'), and one approximately 16 kilobases from *Mst84Dc* (84D: 5'-AAAAAACTGCATTTGGCAGCCG-3', 5'-GAGAGCAGAAATCGAGAATCAGGC-3').

Sequence Analyses

Two lines of Drosophila simulans and one line of D. mauritiana were sequenced for Acylphosphatase (Acyp; 363 bp; cytological position 34E2), Mst98Cb (545 bp; cytological position 98C3), CG5762 (603 bp; cytological position 95F15), Mst84Dc (168 bp; cytological position 84D5), and CG14718 (605 bp; cytological position 86E16). These five genes are underexpressed in hybrids relative to pure species (Michalak and Noor 2003). Our analyses are not intended as tests for the action of natural selection but, rather, address whether extensive amino acid sequence divergence occurred between D. simulans and D. mauritiana. The published D. melanogaster sequence (Adams et al. 2000) was used as an outgroup where needed. Primers used for amplification and sequencing are available upon request. PCR was performed on genomic DNA from inbred lines directly and without cloning. Sequencing reactions were performed in both directions using ABI BigDye Terminator v. 3 and visualized on an ABI 3100. Sequences were analyzed using the computer programs SITES (Hey and Wakeley 1997) and PAML (Yang 1997). The sequences reported in this paper have been deposited in the GenBank database (accession Nos. AY549946-AY549961).

Results

Expression in First-Generation Backcross Males

We assayed expression of six genes putatively misexpressed in hybrid males relative to pure species in firstgeneration backcross hybrid males: *Acylphosphatase* (*Acyp*), *CG5762*, *CG11266*, *CG14718*, *Mst84Dc*, and *Mst98Cb*. Sterile males (S) consistently exhibited a more variable pattern of expression than non-sterile (F and SF) males, and this pattern was consistent with or without normalization using 18S rRNA. Normalization had no substantial effect on any of our analyses, so for clarity, we present only our analyses on threshold cycle number (C_T) values directly. High C_T values correspond with low expression of a transcript, as more PCR cycles are required to amplify the transcript. We detected no difference between fertile (F) and semifertile (SF) backcross hybrid males for expression of any transcript, and these categories were subsequently pooled for comparison to sterile males. The mean expression among fertile and semifertile males was significantly higher than that for sterile males for all transcripts except *CG11266* (p < 0.01 for all but *CG14718*; p = 0.019 for *CG14718*). Withinsample replicate effects were negligible.

Expression in Fifth-Generation Backcross Males

Expression of Acyp, CG5762, CG11266, CG14718, and Mst84Dc was assayed in the fifth-generation backcross hybrid males. The same patterns as before were noted, but the results were even more striking: fertile individuals exhibited extremely little variance in $C_{\rm T}$, while sterile individuals exhibited much higher variance and a significantly higher $C_{\rm T}$ for all genes except CG11266 (Fig. 1). Unlike the first-generation backcross males, we observed very little variation in CG11266 among both fertile and sterile individuals. If disruption of the other four transcripts is associated with sterility, then the variation observed among fertile males could represent incomplete developmental constraint, while that among sterile males could result from more major coordinated disruptions.

Statistical associations of expression level between transcripts within individuals may reflect linkage disequilibrium of their regulators or involvement in the same molecular pathways. The former explanation is improbable after several generations of backcrossing. Within BC5 males, we found strikingly high correlations of expression among genes—r > 0.90and highly statistically significant (p < 0.0001) for any pair of *Acyp*, *CG5762*, *CG14718*, and *Mst84Dc*—but



Fig. 2. Regression of threshold cycle (Ct) for real-time RT-PCR amplification of two pairs of transcripts in fifth-generation back-cross hybrids.

low (r < 0.4) and nonsignificant correlations (p > 0.10) for any comparison to *CG11266* (Fig. 2). Because *Mst98Cb* and *Mst84Dc* are known to be corregulated (White-Cooper et al. 1998), we predict that expression of *Mst98Cb* would also have been correlated with that of the other four genes had we been able to test it in BC5 males. Moreover, we did observe that expression of *Mst98Cb* was correlated with that of *Mst98Cb* was correlated with that of *Mst98Cb* males (r = 0.87, p < 0.0001), confirming this prediction.

Localizing the Trans-Acting Factor(s)

The strong correlations in expression among Acyp, CG5762, CG14718, and Mst84Dc suggest that they are coregulated and that their disruption in hybrids is caused in part by a change/disruption in one or more trans-acting factors. To evaluate whether such transacting factors can be genetically mapped, we genotyped the BC5 males for 11 microsatellites. Except for the microsatellite within OdsH and the one 2 megabases from this gene, all loci were homozygous or hemizygous for the D. simulans allele. Hence, little D. mauritiana genome remains after the five generations of backcrossing. The microsatellite immediately upstream of *Mst84Dc* was also homozygous for the D. simulans allele, further supporting the role of divergence in trans-acting factors causing the hybrid expression disruptions. We identified a perfect correspondence between fertility and genotype at the microsatellite within the *OdysseusH* gene (16/16): all males bearing the *D. simulans* allele were fertile, while all males bearing the *D. mauritiana* allele were sterile. We observed a weaker association between fertility and genotype at the microsatellite 2 megabases upstream of *OdysseusH* (14/16), consistent with some recombination between this locus and the gene causing sterility.

DNA Sequence Analyses

Our sequence analyses of five genes underexpressed in sterile male hybrids identified very low differentiation between D. simulans and D. mauritiana (Table 1). PAML's (Yang 1997) maximum likelihood procedure estimated the ratio of nonsynonymous-to-synonymous substitution (dN/dS) for all pairwise comparisons between species for all four genes to be below 0.20, consistent with strong purifying selection. We observed only a single replacement polymorphism between the two D. simulans strains (in CG14718), so we focus our subsequent analyses on one of these two lines. For Mst84Dc, the entire gene was sequenced in two D. simulans and two D. mauritiana lines, and one line of D. simulans bore exactly the same sequence as one line of D. mauritiana. The other two sequences differed by 9-bp insertions/deletions in the repetitive CGP amino acid motif characteristic of this gene family. Similarly, in all 363 coding bases of Acyp and in the 545 bp of *Mst98Cb* examined, no amino acid replacement differences separated the two species. We did observe an apparent recent acceleration in the rate of nonsynonymous substitution in the lineage leading to D. simulans in the CG14718 transcript, as the number of nonsynonymous differences between D. simulans and D. mauritiana is greater than between the outgroup species (D. melanogaster) and D. simulans. However, PAML estimated the rate of nonsynonymous-to-synonymous substitution (dN/dS) to be only 0.31 in the branch from the common ancestor of D. simulans and D. mauritiana to present-day D. simulans, still consistent with purifying selection and not clearly indicative of an accelerated amino acid substitution rate.

Discussion

Previously, we identified a panel of transcripts underexpressed in hybrids relative to pure species (Michalak and Noor 2003). Here we have shown that underexpression of four of these transcripts, *Acyp*, *CG5762*, *CG14718*, and *Mst84Dc*, and likely a fifth, *Mst98Cb*, is associated with infertility of hybrid males, even after five generations of recombination. This observation suggests that these genes are

	Mst84Dc			Mst98Cb			CG5762			CG14718			Acyp		
	Mel	Sim	Mau	Mel	Sim	Mau	Mel	Sim	Mau	Mel	Sim	Mau	Mel	Sim	Mau
Mel		3	3	_	13	16	_	26	24	_	13	16		10	9
Sim	0		0	0		12	13		5	9		6	8		3
Mau	0	0		0	0	—	13	3	_	6	7	_	8	0	

Table 1. DNA sequence differences between *D. melanogaster* (mel), *D. simulans* (sim), and *D. mauritiana* (mau) at five genes: Above diagonal, synonymous nucleotide differences; below diagonal, nonsynonymous nucleotide differences

downstream targets of some of the same genetic difference(s) between species that cause hybrid sterility. Consistent with this observation, two of these genes, *Mst84Dc* and *Mst98Cb*, are known to function in spermiogenesis at approximately the same stage (Schafer et al. 1993). We also found a strikingly high correlation in expression of these transcripts.

Genetically mapping the disruptions of expression of these transcripts will identify the locations of one or more upstream trans-acting regulators. Such regulators are candidate genes that may confer hybrid sterility through disruptions in gene expression. We applied this approach to the fifth-generation backcross hybrid males for which we surveyed expression and found a perfect correspondence between genotype at a microsatellite within *OdysseusH* and fertility: all males bearing the D. mauritiana allele were sterile, while those bearing the D. simulans allele were fertile. This finding suggests that the D. mauritiana allele of OdysseusH and/ or gene(s) linked to it caused the disruptions in expression of Acyp, CG5762, CG14718, and Mst84Dc that are associated with sterility in backcross hybrids. This region contains several genes that cause sterility in male hybrids of D. simulans and D. mauritiana (Perez and Wu 1995; True et al. 1996), and OdsH itself is one candidate.

The genes identified thus far that contribute to hybrid sterility or inviability directly, and many other associated with male reproduction, often exhibit high rates of nonsynonymous substitution between species (Swanson and Vacquier 2002a, b; Noor 2003). We do not necessarily expect to observe this same pattern with the transcripts we studied, as they are genetically downstream of "speciation genes." Correspondingly, we observed little or no amino acid sequence differences between *D. simulans* and *D. mauritiana* in these transcripts.

Based on our results, we suggest that at least some of the genetic changes causing hybrid male sterility in *D. simulans* and *D. mauritiana* also cause transcriptional deregulation of multiple genes involved in spermatogenesis in these same hybrids. Sterile hybrids had much more variable expression of these genes than fertile hybrids and generally expressed these transcripts at lower levels. Because nearly all transcription in *Drosophila* spermatogenesis is premeiotic (Fuller 1998) while spermatogenesis breaks down from postmeiotic disruptions in hybrids (Wu et al. 1992), there was ample opportunity for transcription of these genes. It is reasonable to assume that at least some of this deregulation could contribute to sterility directly, although we cannot implicate one or more specific downstream targets when many are coordinately disrupted. The misexpression of these particular transcripts may be only indirectly associated with hybrid sterility as common downstream developmental consequences caused by introgression of a particular genomic region between species.

Our approach has identified possible functions in spermatogenesis of one previously undescribed gene, *CG5762*, and two incompletely described genes, *Acylphosphatase* and *CG14718*. Their regulation is coordinated with that of *Mst84Dc* and *Mst98Cb*, which have been studied in the genetic control of spermatogenesis (Schafer et al. 1993; White-Cooper et al. 1998). Further application of this approach may identify numerous other candidate genes that can be assayed for spermatogenetic function through direct manipulation.

We conclude that continued use of complementary forward-genetic and reverse-genetic approaches might shed light on the genetic control of fertility and its disruption in hybrids. Here, we have combined reverse genetics (gene expression assays) with forward genetics (genetic mapping) to identify both a possible proximate developmental cause of infertility (underexpression of transcripts involved in spermatogenesis) and the location of one or more genes that produce this proximate phenotype. Nonetheless, hybrid sterility in this species group and many others is caused by multiple independent incompatibilities (Wu et al. 1992). As such, repeating this process with more multigenerational backcross lineages will identify multiple factors causing infertility in hybrids as well as the developmental consequences of their introgression. This will simultaneously identify whether the same regulatory pathways are being disrupted by these different incompatibilities and, if so, if the pathways are disrupted at the same points or at different points. With approaches such as this, speciation genetic research will move from only identifying genes causing reproductive isolation to also embarking on detailed analyses of the functional consequences of the substitutions of these genes.

Acknowledgments. This research was supported by National Science Foundation Grants DEB-9980797, 0211007, and 0314552 and Louisiana Board of Regents Governor's Biotechnology Initiative Grant 005 to M.A.F.N. P.M. was also funded by a post-doctoral fellowship from the Lalor Foundation.

References

- Adams MD, Celnicker SE, Holt RA, Evans CA, Gocayne JD, et al. (2000) The genome sequence of *Drosophila melanogaster*. Science 287:2185–2195
- Barbash DA, Siino DF, Tarone AM, Roote J (2003) A rapidly evolving MYB-related protein causes species isolation in Drosophila. Proc Natl Acad Sci USA 100:5302–5307
- Bertucci LA, Noor MAF (2001) Single fly RNA preparations for RT-PCR. Drosoph Inf Serv 84:166–168
- Colson I, Macdonald SJ, Goldstein DB (1999) Microsatellite markers for interspecific mapping of *Drosophila simulans* and *D. sechellia*. Mol Ecol 8:1951–1955
- Fuller MT (1998) Genetic control of cell proliferation and differentiation in *Drosophila* spermatogenesis. Semin Cell Dev Biol 9:433–444
- Gloor GB, Engels WR (1992) Single-fly DNA preps for PCR. Drosoph Inf Serv 71:148–149
- Hey J, Wakeley J (1997) A coalescent estimator of the population recombination rate. Genetics 145:833–846
- Meiklejohn CD, Parsch J, Ranz JM, Hartl DL (2003) Rapid evolution of male-biased gene expression in Drosophila. Proc Natl Acad Sci USA 100:9894–9899
- Michalak P, Noor MAF (2003) Genome-wide patterns of expression in Drosophila pure-species and hybrid males. Mol Biol Evol 20:1070–1076
- Noor MAF (2003) Evolutionary biology: Genes to make new species. Nature 423:699–700
- Perez DE, Wu C-I (1995) Further characterization of the *Odysseus* locus of hybrid sterility in Drosophila: One gene is not enough. Genetics 140:201–206

- Presgraves DC, Balagopalan L, Abmayr SM, Orr HA (2003) Adaptive evolution drives divergence of a hybrid inviability gene between two species of Drosophila. Nature 423:715– 719
- Ranz JM, Castillo-Davis CI, Meiklejohn CD, Hartl DL (2003) Sexdependent gene expression and evolution of the Drosophila transcriptome. Science 300:1742–1745
- Ranz JM, Namgyal K, Gibson G, Hartl DL (2004) Anomalies in the expression profile of interspecific hybrids of *Drosophila melanogaster* and *D. simulans*. Genome Res 14:373–379
- Schafer M, Borsch D, Hulster A, Schäfer U (1993) Expression of a gene duplication encoding conserved sperm tail proteins is translationally regulated in *Drosophila melanogaster*. Mol Cell Biol 13:1708–1718
- Swanson WJ, Vacquier VD (2002a) The rapid evolution of reproductive proteins. Nat Rev Genet 3:137–144
- Swanson WJ, Vacquier VD (2002b) Reproductive protein evolution. Annu Rev Ecol Syst 33:161–179
- Ting C-T, Tsaur S-C, Wu M-L, Wu C-I (1998) A rapidly evolving homeobox at the site of a hybrid sterility gene. Science 282:1501–1504
- True JR, Weir BS, Laurie CC (1996) A genome-wide survey of hybrid incompatibility factors by the introgression of marked segments of *Drosophila mauritiana* chromosomes into *Drosophila simulans*. Genetics 142:819–837
- Walter RB, Kazianis S (2001) Xiphophoyus interspecies hybrids as genetic models of induced neoplasia. ILAR J 42:299– 321
- White-Cooper H, Schäfer MA, Alphey LS, Fuller MT (1998) Transcriptional and post-transcriptional control mechanisms coordinate the onset of spermatid differentiation with meiosis I in Drosophila. Development 125:125–134
- Wu C-I, Perez DE, Davis AW, Johnson NA, Cabot EL, Palopoli MF, Wu M-L (1992) Molecular genetic studies of postmating reproductive isolation in Drosophila. In: Takahata N, Clark AG (eds) Molecular paleopopulation biology. Springer-Verlag, Berlin, pp 191–212
- Yang Z (1997) PAML: A program package for phylogenetic analysis by maximum likelihood. CABIOS 13:555–556