Rapid Evolution of Genomic Acp Complement in the melanogaster Subgroup of Drosophila

David J. Begun and Heather A. Lindfors
Section of Evolution and Ecology, University of California, Davis

Unusual properties of molecular evolution in reproduction-related Drosophila genes, including atypically rapid rates of protein evolution, support the idea that natural selection plays an important role in divergence of reproductive function in Drosophila. We used subtractive hybridization to investigate another potential side of evolution of the male reproductive transcriptome. We carried out a screen for genes with much greater transcript abundance in Drosophila simulans reproductive tracts than in Drosophila melanogaster reproductive tracts. Such genes could be present in both species but diverged dramatically in transcript abundance or could be present in D. simulans but absent from D. melanogaster. Here we report data from melanogaster subgroup species for three previously unknown accessory gland protein genes (Acps) identified in this screen. We found multiple Acps that were present in some lineages yet absent from other closely related melanogaster subgroup lineages, representing several losses of genes. An Acp that may have been lost in D. melanogaster and Drosophila erecta is segregating a null allele in Drosophila yakuba, yet shows evidence of adaptive protein evolution in contrasts of polymorphism and divergence within and between D. yakuba and its close relative, Drosophila teissieri. These data suggest that turnover of Acps occurs rapidly in Drosophila, consistent with rapid evolution of seminal fluid function.

Introduction

Proteins associated with male reproduction evolve unusually quickly in many taxa (reviewed in Swanson and Vacquier 2002), including Drosophila (Coulthart and Singh 1988; Civetta and Singh 1998; Begun et al. 2000; Swanson et al. 2001; Zhang, Hambuch, and Parsch 2004; Richards et al. 2005). One interpretation of such data is that functions associated with male reproduction experience greater directional selection compared to many other functional categories. In Drosophila, rapid evolution of several male reproduction-related phenotypes (e.g., Patterson and Stone 1952; Palopoli and Wu 1994; Wu et al. 1995; Liu et al. 1996; Pitnick 1996; Pitnick et al. 1997; Price 1997; Knowles and Markow 2001; Meiklejohn et al. 2003; Ranz et al. 2003) is consistent with this proposition.

Accessory gland proteins (ACPs) are a major component of Drosophila seminal fluid, which is transferred with sperm during mating. Drosophila melanogaster seminal fluid components have several effects on sperm utilization and female physiology (reviewed in Chen 1996; Woliner 1997; Chapman and Davies 2004). ACPs have been implicated in induction of oviposition, in rendering females recalcitrant to remating, and in mediating sperm displacement and sperm storage in females (Herndon and Woliner 1995; Neubaum and Woliner 1999; Tram and Woliner 1999; Chapman et al. 2003; Liu and Kubli 2003). In addition to its effect on female reproduction and sperm use, seminal fluid appears to have toxic effects on D. melanogaster females (Chapman et al. 1995; Rice 1996; Holland and Rice 1999). For example, Rice (1996) showed that sexual selection on males in laboratory populations of D. melanogaster caused evolution of elevated male toxicity toward females, which may be mediated by ACP70A (Wigby and Chapman 2005) or other seminal fluid proteins. Other evidence that properties of seminal fluid may evolve rapidly comes from investigation of the insemination reaction, a postcopulatory, inflammatory response of the female Drosophila reproductive tract (Patterson and Stone 1952; Knowles and Markow 2001). This reaction is less severe in matings between flies from closely related species or populations than in matings between more distantly related species or populations. From a molecular evolutionary genetics perspective, our knowledge of Drosophila Acp variation is minimal, though tantalizing. Acp evolve quickly compared to other genes (e.g., Begun et al. 2000; Swanson et al. 2001), at least in some cases under directional selection (Tsaur and Wu 1997; Aguadé 1999; Begun et al. 2000; Holloway and Begun 2004; Kern, Jones, and Begun 2004). Natural Acp variation in D. melanogaster populations may affect performance of males in sperm competition (Fiunera, Bumont, and Clark 2005).

The testis-expressed component of the Drosophila genome may also have unusual evolutionary properties. Testis-expressed genes evolve quickly in flies (Zhang, Hambuch, and Parsch 2004; Richards et al. 2005). Moreover, though data are scarce, anecdotal evidence suggests that testis expression will be common in lineage-restricted genes. For example, Sdrc codes for a sperm-specific axonemal dynein protein in D. melanogaster. The gene is a chimera which originated through a complex set of rearrangements including a gene-fusion event between the cell-adhesion protein annexin X and a cytoplasmic dynein intermediate chain (Nurminsky et al. 1998). Interestingly, the gene is absent in the sister species, Drosophila simulans, suggesting that it originated in the very recent past. Jingwei is a novel chimeric gene, which originated by the insertion of an Adh retrosequence into a duplicated locus in the melanogaster subgroup of Drosophila (Long and Langley 1993). It is present in Drosophila yakuba (and Drosophila teissieri) but absent from D. melanogaster and D. simulans. Jingwei is expressed in a male-specific manner in D. teissieri but not in its close relative, D. yakuba (Long and Langley 1993). Recent analyses show that Drosophila retrogenes tend to migrate from X chromosomes to autosomes and to
be enriched for expression in male reproductive tissues (Betran, Thornton, and Long 2002; Betran and Long 2003).

Lineage-restricted genes, which may be the result of gains of novel genes or loss of genes that were present in an ancestor, are interesting because they may make significant contributions to the unique biological properties of individual lineages. Though it is difficult to imagine that novel genes fix under some process other than directional selection, the question of gene loss is more problematic. Models explaining gene loss might invoke loss of redundant or partially redundant genes under genetic drift, changes (genomic or ecological) that render previously necessary genes dispensable, or fixation of gene deletions by natural selection. Addressing these alternatives requires functional and evolutionary analysis of recent evolutionary changes.

Here we report the results of a screen for male reproductive tract transcripts that are more abundant in *D. simulans* than in *D. melanogaster*. Such a screen could identify genes that are present in both species but have diverged dramatically in transcript abundance or could identify *D. simulans*-specific genes, which of course, would not be annotated in the *D. melanogaster* reference sequence. We use phylogenetic data to address the issue of the direction of evolution underlying gene presence/absence differences. Finally, we use molecular population genetic data to infer the evolutionary mechanisms associated with gene presence/absence differences or expression differences.

**Materials and Methods**

**Subtraction**

We performed polymerase chain reaction (PCR)–subtractive hybridization (Clontech, Mountain View, Calif.) on RNA isolated from male reproductive tracts dissected from flies derived from large outbred population samples of *D. melanogaster* (Napa, California) and *D. simulans* (Austin, Texas) to enrich for transcripts that were more abundant in *D. simulans*. Approximately 1,000 clones from the subtraction experiment were subjected to colony PCR, purified, placed in microtiter plates, and spotted (1 μl) onto replicated nylon filters. These replicated dot blots were then hybridized with 32P-labeled cDNA derived from either *D. melanogaster* male reproductive tracts or *D. simulans* male reproductive tracts to identify clones that hybridized much more strongly to *D. simulans* cDNA. Expression of genes associated with these candidate clones was then further characterized by hybridization with 32P-labeled cDNA from accessory glands, testes, male carcasses, and whole females. Clones that were highly enriched or specific for expression in *D. simulans* male reproductive tracts were identified for DNA sequencing and further characterization. Three of these clones form the basis of the work presented here.

DNA sequences of the three subtraction clones were used to design rapid amplification of cDNA ends (RACE) primers. The three corresponding genes were then characterized by RACE carried out on *D. simulans* male reproductive tract cDNA. The resulting products were cloned, sequenced, and, when possible, compared to corresponding *D. melanogaster* and *D. simulans* genomic sequence to infer intron/exon structure.

**Estimation of Transcript Abundance by Serial Dilution Dot Blots**

Ambion’s MicroPoly(A) kit (Austin, Tex.) was used to extract mRNA from tissues (accessory gland, testes, and carcass) dissected from 1- to 3-day-old males from *D. melanogaster*, *D. simulans*, and *D. yakuba*. cDNA was made using Clontech’s SMART RACE cDNA Amplification Kit. Concentrations of the resulting cDNAs were normalized using a spectrophotometer. These cDNAs were then subjected to twofold serial dilutions, which were subsequently spotted onto dot blots. For each species, the resulting blot was then hybridized with 32P-labeled PCR product from an Acp, exposed to X-ray film, stripped, and then rehybridized with a control gene generated from each species. For example, the *D. simulans* dot blot contained *D. simulans* cDNA, was probed with a labeled *D. simulans* Acp PCR product, stripped, and reprobed by labeled *D. simulans* Gapdh. A comparison of the relative signal intensity for Acp versus control gene across different tissues and species allowed us to infer the direction of expression evolution.

**Molecular Population Genetic Analyses**

*Drosophila simulans* sequence data were from inbred lines from Wolfskill, Calif. (Begun and Whitley 2000). *Drosophila melanogaster* isofemale lines from Malawi were from B. Ballard (University of Iowa). *Drosophila melanogaster* isogenic lines from North America, most of which were derived from Napa, were isolated in our laboratory. *Drosophila yakuba* isofemale lines were from P. Andolfatto (University of California–San Diego). *Drosophila santomea* and *D. teissieri* isofemale lines were from M. Long (University of Chicago).

DNA sequences from inbred or isogenic lines were obtained by direct sequencing of PCR products. Data from isofemale lines were obtained by PCR amplification using high-fidelity polymerase, followed by cloning and sequencing. Basic summary statistics and tests of neutrality were calculated in DnaSP v4.0 (Rozas et al. 2003). Hudson-Kreitman-Aguade tests (Hudson, Kreitman, and Aguade 1987) were carried out by comparison of Acp data to previously collected data from the *vermilion* locus (Begun and Aquadro 1995). Sequence data for this paper have been submitted to GenBank under accession numbers DQ096191–DQ096280 and DQ097147–DQ097159.

**Results**

**Identification and Characterization of Novel Acps**

Three novel transcripts showing much greater expression in *D. simulans* than in *D. melanogaster* accessory glands (more than 50-fold) were identified by the use of PCR-based subtractive hybridization. None correspond to annotated *D. melanogaster* genes or previously identified *D. simulans* accessory gland–derived expressed sequence tags (Swanson et al. 2001). The three predicted *D. simulans* proteins have strongly predicted signal peptides (Bendtsen et al. 2004; www.cbs.dtu.dk/services/SignalP), which supports the
notion that they are ACPs (Swanson et al. 2001). Two of the three *D. simulans* *Acps* have homologous open reading frames (ORFs) in the *D. melanogaster* reference sequence. They are referred to as Acp24A4 (79 codons) and Acp54A1 (46 codons) based on their cytological locations in *D. melanogaster*.

Acp24A4 organization is unusual in two respects. First, in *D. simulans* (but not in *D. melanogaster*) its intron has a rare noncanonical GC 5′-splice junction (Misra et al. 2002). Second, 5′ RACE analysis of accessory gland miRNA from this gene revealed a rare (about 5%) longer transcript which contains a Kunitz domain–containing ORF distinct from Acp24A4, which also has an inferred Kunitz domain. The putative dicistronic message (Misra et al. 2002), likely containing the transcript of another previously uncharacterized Acp, was subsequently confirmed using reverse transcriptase (RT)–PCR in *D. simulans*. We have no data on whether *D. melanogaster* produces a dicistronic message at this locus. Analyses presented here are restricted to the 3′ region of the dicistronic message containing the Acp24A4 protein-coding region.

One of the three *D. simulans* genes identified in the subtraction screen, which codes for a predicted protein of 150 residues, returned no significant BlastN hits to the *D. melanogaster* genome. Lack of a significant Blast hit between a *D. simulans* gene and the *D. melanogaster* reference sequence is highly unusual given the low level of sequence divergence typically observed between these species. Southern blot analysis (not shown) of restriction-digested *D. melanogaster* and *D. simulans* genomic DNAs suggested that the gene is single copy in *D. simulans* and showed that its apparent absence from the *D. melanogaster* reference sequence is real, rather than a result of an error in the *D. melanogaster* assembly.

Given that the gene was not homologous to *D. melanogaster* DNA, additional *D. simulans* sequence from flanking regions was needed to address the issue of homology over a larger chromosomal region. We isolated a *D. simulans* genomic phage clone containing the gene of interest and sequenced a fragment containing the gene plus 5′- and 3′-flanking DNA. These flanking sequence data, which were easily alignable to the *D. melanogaster* genome, revealed that the *D. simulans* gene is located in the region corresponding to cytological position 23D4 of *D. melanogaster*. The alignment also revealed that the *D. melanogaster* reference sequence is deleted for the entire *D. simulans* transcription unit, as well as roughly 100 bp of unique 5′ sequence and 230 bp of unique 3′ sequence. PCR assays from multiple *D. simulans* lines (*n* = 8) and *D. melanogaster* lines (*n* = 16) suggest that presence/absence of this gene is a fixed difference between species. We refer to this gene as Acp23D4a, as two other tightly linked *Acp*s were later discovered in *D. yakuba* (see below). The predicted *D. simulans* Acp23D4a protein appears to contain a C-type lectin–binding domain based on National Center for Biotechnology Information (NCBI) Conserved Domain analysis (Marchler-Bauer et al. 2003) and weak BlastP hits to the *D. melanogaster* and *D. simulans* ACP29B (Mueller et al. 2004).

**Phylogenetic Distribution of Acps**

To determine whether the *D. melanogaster/D. simulans* data from Acp23D4a represent loss in *D. melanogaster* versus gain in *D. simulans*, we used PCR to isolate and sequence this genomic region for multiple *D. yakuba* isofemale lines. We obtained data for 12 chromosomes, all of which harbored a copy of Acp23D4a, suggesting that absence of Acp23D4a in *D. melanogaster* is a result of gene loss in this lineage. However, the PCR product from *D. yakuba* was significantly longer than that expected if organization of the region was similar in *D. yakuba* and *D. simulans*. Sequence data from the amplified *D. yakuba* region showed that *D. yakuba* contains a tandem, highly diverged duplicate (53% amino acid identity) of Acp23D4a. This gene, which was shown by RT-PCR to have accessory gland–specific expression, is referred to as Acp23D4b. It was present in 11 of 12 *D. yakuba* chromosomes (fig. 1), with 1 of the 12 chromosomes being deleted for the entire Acp23D4b-coding region. Thus, Acp23D4b is polymorphic for presence/absence in

![Organizational Structure of Acp23D4a, Acp23D4b, and Acp23D4c](https://academic.oup.com/mbe/article-abstract/22/10/2010/113445/113445)

**Fig. 1.**—Organization of Acp23D4a gene region in five species of the *melanogaster* subgroup. Filled boxes represent protein-coding regions; solid lines indicate homologous noncoding sequences; dashed lines represent deleted regions. The “yakuba common allele” was present in 11 of 12 clones as indicated by a 2.1-kb PCR amplicon. However, DNA sequence data from the 3′ end of the region was determined only in one of the clones. The “yakuba deletion allele,” as indicated by a 1.1-kb amplicon, was present in one line (which was sequenced). Eight of nine *Drosophila santomea* common allele, and one contained the “santomea deletion allele.” Question marks indicate regions not sequenced. All *Drosophila simulans, Drosophila melanogaster,* and *Drosophila erecta* surveyed contained the structure indicated in the figure.
**Fig. 2.**—Alignment of non–protein-coding regions of Acp23D4a gene region in four species of the melanogaster subgroup. Completely conserved positions are marked by asterisks. For *Drosophila yakuba*, the region from the initiation codon of Acp23D4a to the termination codon of Acp23D4c (fig. 1) was omitted from the alignment; this deleted DNA corresponds to an insertion between bases 695 and 696 of the *D. yakuba* DNA in the figure. For *Drosophila simulans*, the region from the initiation to the stop codon of Acp23D4a (fig. 1) was omitted from the alignment; this deleted DNA corresponds to an insertion between bases 648 and 649 of the *D. simulans* DNA in the figure.
D. yakuba. The predicted D. yakuba Acp23D4b protein shows evidence for a C-type lectin–binding domain in an NCBI Conserved Domain search (Marchler-Bauer et al. 2003), similar to the result observed for Acp23D4a.

We used PCR and sequencing to examine this genomic region from 10 chromosomes of D. santomea, a sister species of D. yakuba (Lachaise et al. 2000). All 10 chromosomes had a complete copy of Acp23D4a. However, one chromosome had a roughly 280-bp deletion that included about 100 bp of the 5’ end of the protein-coding region of Acp23D4b. Thus, both D. santomea and D. yakuba are segregating for Acp23D4b copy number, though the molecular lesion at Acp23D4b in the two species are not obviously identical by descent (fig. 1). We also observed three D. santomea point mutations causing a premature termination codon in Acp23D4a and one D. santomea point mutation causing a premature termination codon in Acp23D4b.

Further inspection of D. yakuba sequence data from the amplified region revealed a third potential ORF (fig. 1). Based on amino acid similarity, the predicted protein is not obviously homologous to that coded for by Acp23D4a or Acp23D4b. However, the predicted protein has a strongly predicted signal peptide and a lectin domain (as evidenced by BlastP vs. D. melanogaster lectin-21Cb [CG13686] of E = 10⁻¹⁹), suggesting that this ORF might also code for an Acp. Indeed, RT-PCR results on D. yakuba cDNA showed accessory gland–specific expression for this gene, which we refer to as Acp23D4c. The phylogenetic distribution of Acp23D4c (absence in D. melanogaster/D. simulans and presence in D. yakuba) is consistent with either gain in the D. yakuba lineage or loss in the common ancestor of D. melanogaster/D. simulans.

We were able to amplify and sequence intact copies of each of the three putative Acps in this genomic region from D. teissieri, a sister species of D. yakuba/D. santomea. However, our sample sizes were small, such that low-frequency polymorphic loss-of-function alleles might well have gone undetected. Our overall conclusion is that D. yakuba and D. teissieri may have up to three effectively single-copy Acp genes in this genomic region, while D. simulans has one (Acp23D4a) and D. melanogaster has none (fig. 1). Our D. santomea data show that many chromosomes harbor Acp23D4a and Acp23D4b. Though we have no data from the downstream region of D. santomea, it seems highly likely that such data would reveal that this species also harbors Acp23D4c. At least one of the Acps from this region, Acp23D4b, is polymorphic for presumptive loss-of-function alleles in both D. yakuba and D. santomea.

The intriguing patterns of gene presence/absence variation in this genomic region motivated us to investigate the homologous region of Drosophila erecta, a sister taxon to the D. yakuba/D. teissieri/D. santomea clade (Ko, David, and Akashi 2003). PCR and sequence analysis of the region revealed that the D. erecta genome (at least the strain used in our experiment) does not contain Acp23D4a, Acp23D4b, or Acp23D4c. This result is consistent with our Southern blot analysis (not shown) of restriction-digested D. erecta genomic DNA using PCR-amplified D. simulans Acp23D4a as a probe, which showed no convincing evidence of hybridization in D. erecta. An alignment of this genomic region for four species (minus the gene-containing regions in D. yakuba and D. simulans) is shown in figure 2. This alignment supports the idea that the homologous region has been identified in all four species but that genes present in D. simulans and D. yakuba/D. teissieri are absent from D. melanogaster and D. erecta. The phylogeny of these
four species (D. melanogaster and D. simulans as sister taxa and D. yakuba/D. teissieri and D. erecta as sister taxa; Ko, David, and Akashi 2003; Parsch 2003) suggests that multiple events are required to explain the distribution of genes. For example, under the assumption that gene loss is more likely than independent gains of the same gene, the more parsimonious explanation is that Acp23D4a, Acp23D4b, and Acp23D4c were independently lost in D. melanogaster and D. erecta. Figure 3 shows a summary of our hypothesis regarding Acp presence/absence variation for the region.

**Acp54AI**

Turnover of Acps in the Acp23D4a region could be peculiar to this genomic region. However, data from the Acp54AI region suggest that this is not the case. Though PCR and sequence analysis of the Acp54AI region of D. yakuba revealed that the 5'- and 3'-flanking regions of D. melanogaster and D. yakuba were highly similar, there was no evidence of Acp54AI in D. yakuba. Moreover, there was no evidence for an ORF coding for a protein possessing a signal peptide in the region of ambiguous alignment between the species (fig. 4). We conclude that Acp54AI was either gained in the common ancestor of D. melanogaster and D. simulans or lost in the D. yakuba lineage. Preliminary analyses of recent whole-genome shotgun data from D. erecta suggest that the gene is in this species, which would imply loss in D. yakuba as the more parsimonious explanation.

**Acp24A4**

Acp24A4 is present in D. melanogaster, D. simulans, and D. yakuba. We were unable to amplify regions bearing on its status in D. erecta. However, analysis of D. erecta whole-genome shotgun data suggests that the gene is also present in this species.

**Expression Evolution of Acp24A4**

The phylogenetic distribution of genes coding for the three transcripts isolated from our subtraction library allowed investigation of the direction of expression evolution in D. melanogaster versus D. simulans for only one, Acp24A4. To distinguish evolution of elevated transcript abundance in D. simulans versus evolution of reduced transcript abundance in D. melanogaster, we used dot blots to measure transcript abundance for Acp24A4 in D. yakuba (fig. 5). Those data strongly suggest that, like D. simulans, D. yakuba males express high levels of Acp24A4 transcript, supporting the evolution of decreased abundance in D. melanogaster.

**Molecular Population Genetics of Acps**

The presence of polymorphic and fixed null Acp alleles and the evolution of dramatically reduced Acp24A4 expression in D. melanogaster suggest that hypomorphic or amorphic Acp mutations can rise to appreciable frequencies in Drosophila populations. We used molecular population genetics data from D. melanogaster and D. simulans to try to gain some insight into the process underlying this phenomenon. For example, if a deletion of Acp23D4a had recently been driven through D. melanogaster populations by strong directional selection, we might expect to observe reduced sequence variation in that genomic region as a result of the associated hitchhiking effect (Maynard Smith and Haigh 1974; Kaplan, Hudson, and Langley 1989). Alternatively, molecular population genetic data allow us to investigate whether Acps that are absent from some species (e.g., Acp23D4a) show evidence of adaptive protein divergence in the taxa in which they are present.

**Acp23D4a Region**

Table 1 shows that the ratio of polymorphism to divergence and the frequency spectrum in the region of the Acp23D4a deletion in D. melanogaster are consistent with neutral evolution. There is no evidence of hitchhiking effects reducing nucleotide variation in this region of the D. melanogaster genome. Table 2 shows estimates of silent and replacement divergence and the results of McDonald-Kreitman (MK) tests (McDonald and Kreitman 1991) for intact alleles of Acp23D4a, Acp23D4b, and Acp23D4c in D. yakuba versus D. teissieri. The null hypothesis of neutral evolution is not rejected for Acp23D4a or Acp23D4c. However, Acp23D4b, which is segregating a null allele in D. yakuba and has apparently been lost from D. melanogaster and D. erecta, shows strong evidence of adaptive protein evolution between D. yakuba and D. teissieri. It is also of interest to ask whether duplicates Acp23D4a and Acp23D4b show any evidence for adaptive protein divergence within the D. yakuba lineage. Although overall paralogous divergence between these genes is too great for an MK test, a sliding window analysis of paralogous silent divergence showed several regions having divergence less than 0.20. The results of MK tests on these regions (table 3) support the hypothesis of adaptive protein divergence between the D. yakuba paralogues, though we cannot determine whether one or both genes are evolving under directional selection. We have no population genetic data from the Acp23D4b deletion in D. yakuba and no data bearing on the potential fitness effects of low-frequency Acp nulls in D. yakuba/D. santomea.
FIG. 4.—Alignment of Acp54A1 gene region in Drosophila melanogaster (above) versus Drosophila yakuba (below). The region from start to stop codon of D. melanogaster Acp54A1 is in bold.
Wagstaff and Begun (2005) showed turnover of PCR product or with subgroup of Drosophila simulans were included in the analysis.

Gene Region of n Location Molecular Population Genetic Data from Table 1

Acp24A4 and Acp54A1

Population genetic data (table 4) from Acp24A4 suggest that protein divergence between species is a result of directional selection, though there is no evidence of reduced, linked variation associated with hitchhiking effects in D. melanogaster, the lineage in which reduced expression evolved. Acp54A1, which is likely a loss of a D. yakuba gene, shows no evidence of directional selection in D. melanogaster (table 5).

Discussion

In several cases, Acps revealed in our D. simulans/ D. melanogaster screen are present in only a subset of the melanogaster subgroup species surveyed. Assuming that the ancestral state of the Acp23D4a region is three genes and gene losses are independent, we would infer under parsimony a total of nine “fixed” Acp losses and two genes polymorphic for loss-of-function mutations across four species for the genes investigated here. This may well be an overestimate, but it seems clear that there have been multiple losses of Acps among this small sample of genes from the melanogaster subgroup. The one case in which we could investigate the direction of Acp expression evolution revealed a dramatic reduction of transcript abundance in D. melanogaster. Even given uncertainty regarding ancestral states, the data strongly suggest that hypomorphic or null Acp alleles can spread through Drosophila populations. Wagstaff and Begun (2005) showed turnover of Acps in a comparison of the distantly related species, D. melanogaster and D. pseudoobscura. Data from the relatively closely related species investigated here suggest that Acp turnover may occur quite rapidly.

One interpretation is that some Acps are maintained under weak selection (e.g., as a consequence of functional redundancy) such that at equilibrium, deleterious loss-of-function or partial loss-of-function alleles can reach appreciable frequencies or fix under drift. However, at least two aspects of the data militate against this as a general explanation. First, Acp23D4b shows strong evidence of adaptive protein evolution in D. yakuba/D. teissieri. Under the premise that genes evolving under directional selection are functionally important in at least some lineages, evidence for adaptive protein evolution would seem to render the equilibrium, mutation-selection-drift model highly unlikely as an explanation for Acp23D4b loss in D. melanogaster/D. simulans and D. erecta. Similarly, evidence of adaptive protein divergence between Acp23D4a and Acp23D4b in the D. yakuba lineage is difficult to reconcile with the notion that they are functionally redundant and generally susceptible to loss by drift in all melanogaster subgroup lineages. Second, high silent site divergence between Acp23D4a and Acp23D4b in D. yakuba is poor, the fact that we discovered a null in a small D. yakuba sample (as well as a “broken” Acp23D4b allele segregating in D. santomea) suggests that alleles are ancient. The predilection of the Drosophila genome for DNA loss (Petrov and Hartl 1998) suggests that genes susceptible to loss under mutation pressure should tend to be young.

Patterns of polymorphism and divergence at Acp23D4b in D. yakuba and D. teissieri are unusual. Although the MK test suggests a history of adaptive protein evolution, we also discovered polymorphic null alleles at this gene in D. yakuba and a premature stop codon in D. santomea. Though our estimate of the frequency of nulls is poor, the fact that we discovered a null in a small D. yakuba sample (as well as a “broken” Acp23D4b allele segregating in D. santomea) suggests that such alleles are not maintained at very low frequency under strong mutation-selection balance. A traditional view of gene function and evolution is that genes evolving under directional selection should have important functions affecting fitness. Such...
 genes might not be expected to segregate loss-of-function mutations at appreciable frequencies.

One interpretation of our data, however, is that they represent a case of nonequilibrium function and evolution revealed by the different timescales of the polymorphism versus divergence data. In spite of its history of directional selection, the high nucleotide heterozygosity (table 2) and presence of a polymorphic null allele at \textit{D. yakuba} \textit{Acp23D4b} provide no evidence of recent directional selection. Moreover, null alleles in \textit{D. yakuba} and \textit{D. santomea} could be interpreted as evidence that standing variation at this locus has relatively small fitness effects. This suggests an episodic model of function and evolution alternating between periods of protein evolution dominated by strong directional selection with periods dominated by little directional selection and weak purifying selection. During periods of diminished selection, \textit{Acps} could segregate loss-of-function mutants or be lost from populations by drift as their functions are supplanted by other genes evolving of-function mutants or be lost from populations by drift periods of diminished selection, directional selection and weak purifying selection. During this scenario of alleles at mutation-selection balance subsequently fixing under directional selection after a change of the environment might be difficult to detect.

The biology of \textit{Acps} leads to some interesting, albeit speculative theories about why single-copy \textit{Acps} might experience rapid turnover under selection. Two examples of gene loss plausibly driven by natural selection are the CCR5 deletion in humans (Carrington et al. 1997; Stephens et al. 1998; Galvani and Slatkin 2003), which confers resistance to HIV and the \textit{R} genes in Arabidopsis (Stahl et al. 1999). The fact that these examples of gene loss are thought to involve resistance to pathogens suggests the possibility of a more general pattern of protein loss as a mechanism for inhibiting pathogen-infection pathways.

For example, \textit{ACP23D4a}, \textit{ACP23D4b}, and \textit{ACP23D4c} have similarity to C-type lectin galactose-binding domain proteins. C-type lectin proteins, which are thought to mediate cell-cell or cell-matrix interactions (Theopold et al. 1999; Dodd and Drickamer 2001), are found in several proteins including other \textit{Drosophila} ACPs.

| Table 2 | Polymorphism and Divergence Data and McDonald-Kreitman Tests from \textit{Acp23D4a-c} in \textit{Drosophila yakuba} and \textit{Drosophila teissieri} |
|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| Gene          | Species        | \(\pi_S\) | \(\pi_A\) | \(K_S\)  | \(K_A\)  | Fixed   | Polymorphic |
| \textit{Acp23D4a} | \textit{yakuba} (n = 12) | 0.026  | 0.012  | 0.14   | 0.11   | Silent  | 16   | 14   |
| \textit{Acp23D4b} | \textit{yakuba} (n = 10) | 0.052  | 0.005  | 0.16   | 0.22   | Silent  | 16   | 28   |
| \textit{Acp23D4c} | \textit{teissieri} (n = 5) | 0.036  | 0.017  | 0.19   | 0.08   | Silent  | 14   | 9    |

\(\pi_S\) and \(\pi_A\) are the average pairwise difference within species at silent and replacement sites, respectively. \(K_S\) and \(K_A\) are silent and replacement divergence estimated using the Jukes-Cantor correction.

Note: Sites surveyed include bases on the 5' and 3' flanking regions of the gene. Only bases clearly alignable to \textit{Drosophila simulans} were included in the analysis. McDonald-Kreitman test (= Value MK Test) is from the protein-coding region only, using all \textit{D. melanogaster} alleles and eight alleles isolated from inbred Wolf-skull, Calif. strains of \textit{D. simulans} (fixed silent = 5, fixed replacement = 16, polymorphic silent = 11, and polymorphic replacement = 4). The same MK result is obtained from separate analysis of U.S. and African \textit{D. melanogaster} alleles.

| Table 3 | Polymorphism and Divergence Data and McDonald-Kreitman Tests of \textit{Drosophila yakuba Acp23D4a} (n = 12) Versus \textit{Acp23D4b} (n = 10) in Segments Showing Silent Divergence Less Than 0.20 as Identified in a Sliding Window Analysis |
|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| Segment       | Nucleotide Positions | \(K_A\) | \(K_S\) | Fixed   | Polymorphic |
| 1              | 18–69              | 0.12   | 0.07   | Silent  | 1   | 0   |
| 2              | 187–264            | 0.24   | 0.11   | Silent  | 1   | 5   |
| 3              | 274–381            | 0.25   | 0.13   | Silent  | 1   | 6   |
| 4              | 420–481            | 0.41   | 0.18   | Silent  | 3   | 6   |
| Pooled         |                     |         |         | Silent  | 7   | 17  |

\(P = 0.0004\)

Note: Nucleotide positions refer to coordinates in aligned paralogous copies of \textit{Acp23D4a} and \textit{Acp23D4b} in \textit{D. yakuba}.
Table 5
Molecular Population Genetic Data from Acp54A1 Gene Region of Drosophila melanogaster

<table>
<thead>
<tr>
<th>Location</th>
<th>n</th>
<th>Sites</th>
<th>S</th>
<th>Tajima D</th>
<th>HKA P Value</th>
<th>MK Test P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>U.S.</td>
<td>9</td>
<td>1274</td>
<td>33</td>
<td>-1.125</td>
<td>0.51</td>
<td></td>
</tr>
<tr>
<td>Africa</td>
<td>10</td>
<td>1264</td>
<td>47</td>
<td>-0.997</td>
<td>0.86</td>
<td>0.46</td>
</tr>
</tbody>
</table>

NOTE.—Sites surveyed include bases on the 5'- and 3' flanking regions of the gene. Only bases clearly alignable to Drosophila simulans were included in the analysis. McDonald-Kreitman test (P value from Fisher's exact test) is from the protein-coding region only, using all D. melanogaster alleles and eight alleles isolated from inbred Wolfskill, Calif. strains of D. simulans.

(Swanson et al. 2001; Holloway and Begun 2004; Mueller et al. 2004). Though the function of lectin-binding proteins in Drosophila reproduction is unclear, their role in cell-cell interactions suggests that they could mediate sperm-female interactions (e.g., storage of sperm in seminal receptacle or spermathecae) or sperm-sperm interactions. An interesting, speculative possibility is that ACPs that facilitate or regulate physical interactions between sperm cells and the female reproductive tract (Bertram, Neubau, and Wolflner 1996; Neubau and Wölfner 1999; Chapman et al. 2003; Liu and Kubli 2003) could alter female reproductive tract cells in such a way as to increase the rates of pathogen entry. The presence of an Acp that facilitated such a sperm-female interaction could be deleterious to females in certain pathogen environments, thereby favoring Acp gene deletion. We also note that C-type lectin function in vertebrate innate immunity (Theopold et al. 1999; Dodd and Drickamer 2001; Fujita 2002), though there is no experimental evidence for a role of these proteins in fly immunity (De Gregorio et al. 2001). Nevertheless, it seems plausible that fly lectins transferred to females could also play a regulated role in fly-pathogen interactions, which is consistent with the fact that antimicrobial peptides are transferred to Drosophila females during mating (Samakovlis et al. 1991; Lung, Kuo, and Wölfner 2001).

Alternatively, in models of sexually antagonistic evolution (Rice 1996; Holland and Rice 1999; Wigby and Chapman 2004), selection favors sex-limited adaptations and counteradaptations in a chronic struggle for control over reproduction. Thus, one could imagine that a fixed Acp allele that favors males at the expense of females could become a target of a female counteradaptation. Subsequent to such female adaptation, males not transferring a particular Acp could have an advantage over other males in female-male (or female-male-male) interactions. Such a scenario predicts that female-expressed molecules in Drosophila that mediate antagonistic interactions with male proteins may also show hypomorphic or null allele frequencies similar to those documented here for Acps. Other evolutionary scenarios involving direct effects of male-female or male-male interactions on fitness of polymorphic Acp deletions are also plausible given the severe limitations of the data.

Regardless of our uncertainty about the population genetic mechanisms of the phenomena described here, including the forces controlling the spread of loss-of-function Acp alleles, our data raise new questions regarding genomic turnover of Acp genes. For example, the unambiguous cases of gene loss in our data leave open the question of the mechanisms underlying gain of Acps and the extent to which seminal fluid functions in different species are coded for by nonhomologous genes (Wagstaff and Begun 2005). Acps having phylogenetically restricted distributions in the melanogaster subgroup were identified in a screen for differential expression, raising the question of whether Acps that rapidly evolve in expression will be overrepresented among lineage-restricted genes (and whether this may be a general property of lineage-restricted genes). Insights into the population genetic mechanisms of Acp turnover will require additional functional and evolutionary data. For example, reverse genetic analysis of Acp23D4a in D. simulans or D. yakuba could rule out some alternative hypotheses. Further population genetic and functional investigation of polymorphic null or hypomorphic Acp alleles could also be interesting.

Acknowledgments

Penn Whitley played an important role in the early stages of the project. We thank C. Langley for interesting discussions and C. Langley, M. Wayne, and two anonymous reviewers for useful comments on the manuscript. This work was supported by National Institutes of Health GM55298 and National Science Foundation DEB 0327049.

Literature Cited


Marta Wayne, Associate Editor

Accepted June 6, 2005