J Mol Evol (2006) 62:793–802
DOI: 10.1007/s00239-005-0239-4

JOURNAL OF MOLECULAR EVOLUTION

Positive Selection on Nucleotide Substitutions and Indels in Accessory Gland Proteins of the *Drosophila pseudoobscura* Subgroup

Sheri Dixon Schully,* Michael E. Hellberg

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

Received: 6 October 2005 / Accepted: 20 February 2006 [*Reviewing Editor:* Dr. Willie Swanson]

Abstract. Genes encoding reproductive proteins often diverge rapidly due to positive selection on nucleotide substitutions. While this general pattern is well established, the extent to which specific reproductive genes experience similar selection in different clades has been little explored, nor have possible targets of positive selection other than nucleotide substitutions, such as indels, received much attention. Here, we inspect for the signature of positive selection in the genes encoding five accessory gland proteins (Acps) (*Acp26Aa*, *Acp32CD*, *Acp53Ea*, *Acp62F*, and *Acp70A*) originally described from *Drosophila melanogaster* but with recognizable orthologues in the *D. pseudoobscura* subgroup. We compare patterns of selection within the *D. psuedoobscura* subgroup to those in the *D. melanogaster* subgroup. Similar patterns of positive selection were found in *Acp26Aa* and *Acp62F* in the two subgroups, while *Acp53Ea* and *Acp70A* experienced purifying selection in both subgroups. These proteins have thus remained targets for similar types of selection over long (>21-MY) periods of time. We also found several indel substitutions and polymorphisms in *Acp26Aa* and *Acp32CD*. These indels occur in the same regions as positively selected nucleotide substitutions for *Acp26Aa* in the *D. pseudoobscura* subgroup but not in the *D. melanogaster* subgroup. Rates of indel substitution within *Acp26Aa* in the *D. pseudoobscura* subgroup were up to several times those in noncoding regions of the *Drosophila* genome. This suggests that indel substitutions may be under positive selection and may play a key role in the divergence of some Acps.

Key words: Positive selection — Insertion — Deletion — Repeats — Reproductive protein — *Drosophila pseudoobscura*

Introduction

Genes encoding reproductive proteins are often more divergent than genes encoding nonreproductive proteins (e.g., Civetta and Singh 1998; Singh and Kulathinal 2000). This divergence commonly stems from selection for nucleotide substitutions that result in amino acid changes (Swanson and Vacquier 2002). Such positive selection can be identified by comparing relative rates of nonsynonymous and synonymous changes at orthologous loci (McDonald and Kreitman 1991; Hughes and Nei 1988). More recent refinements to these methods allow for the identification of specific residues targeted by positive selection (Yang et al. 2000; Yang and Swanson 2002; Palomino et al. 2002). Such site-specific models have been used to detect positive selection in a variety of reproductive genes (Swanson and Vacquier 2002).

Protein divergence, however, is not brought about solely by nucleotide substitutions. Partial gene duplications contribute to the divergence of some reproductive proteins by producing variation in the number of internal repeats (e.g., bindin, a sperm-

**Current address:* National Institutes of Health, National Cancer Institute, 6116 Executive Plaza, Suite 502, Bethesda, MD 20892, USA

Correspondence to: Michael E. Hellberg 202 Life Sciences Building, Baton Rouge, LA 70803, USA; email: mhellbe@lsu.edu

Table 1. Functions of accessory gland proteins used in this study

Protein	Function(s) in *D. melanogaster*
Acp26Aa	Hormonal activity; increases egg-laying (Herndon and Wolfner 1995; Heifetz et al. 2000; Chapman et al. 2001; Heifetz et al. 2001); involved in sperm competition (Clark et al. 1995)
Acp32CD	Function unknown (M. Wolfner, personal communication, June 2005)
Acp53Ea	Potential hormonal activity; correlated with sperm competitive ability (Clark et al. 1995)
Acp62F	Protects sperm from proteolysis (Lung et al. 2002); decreases female's life span (Chapman et al. 1995; Lung et al. 2002)
Acp70A (sex peptide)	Hormonal activity; increases egg-laying (Chen et al. 1988; Aigaki et al. 1991; Soller et al. 1997, 1999); decreases female receptivity (Chen et al. 1988; Aigaki et al. 1991)

borne adhesion protein from sea urchins [Metz and Palumbi 1996; McCartney and Lessios 2004; Zigler and Lessios 2004]) or by altering posttranslational modifications to introduce new coding regions into the mature protein (e.g., TMAP, an acrosomal protein from marine snails [Hellberg et al. 2000]). Homogenization of internal repeats by concerted evolution may also contribute to the rapid divergence of reproductive proteins (e.g., VERL, the egg-borne vitelline envelope receptor for lysin from abalone [Swanson and Vacquier 1998]). Insertions and deletions (indels) are another source of variation upon which positive selection may act. The rate of spontaneous indel mutations may be as high as that for nucleotide substitutions (Britten et al. 2003; Denver et al. 2004), and recent studies have shown positive selection acting on indels in sperm-specific proteins in mammals (Podlaha and Zhang 2003; Podlaha et al. 2005).

Here we evaluate positive selection in some well-characterized examples of rapid divergence in reproductive proteins: the accessory gland proteins (Acps) of *Drosophila*. During mating, *D. melanogaster* males transfer 70–106 Acps to females in the seminal fluid that accompanies sperm (Mueller et al. 2005). These Acps elicit many behavioral and physiological changes in the mated female (Wolfner 2002): they increase egg-laying rate (Herndon and Wolfner 1995; Chapman et al. 2001; Heifetz et al. 2001), promote sperm storage (Neubaum and Wolfner 1999; Xue and Noll 2000), reduce female willingness to remate (Chen et al. 1988; Aigaki et al. 1991), reduce female life span (Chapman et al. 1995; Lung et al. 2002), and mediate sperm competition (Harshman and Prout 1994; Clark et al. 1995). The genes underlying these reproductive functions often diverge relatively rapidly: Acps in the *D. melanogaster* subgroup are on average twice as divergent between species as nonreproductive proteins (Civetta and Singh 1995; Singh and Kulathinal 2000; Swanson et al. 2001).

While the functions of Acps in the *D. melanogaster* subgroup have been widely studied, along with the role of positive selection on nucleotide substitutions in effecting their divergence, little is known about Acps in other drosophilid lineages. The recent publication of the *D. pseudoobscura* genome (Richards et al. 2005) permits the comparison of Acp evolution between two lineages (the *D. melanogaster* and *D. pseudoobscura* subgroups) that have been independent for 21–46 MYA (Beckenbach et al. 1993). Wagstaff and Begun (2005) used a combination of computational and molecular approaches to identify five orthologous Acp loci from the *D. melanogaster* subgroup in *D. pseudoobscura* (Table 1). Stevison et al. (2004) compared the divergence of four X-linked putative Acps in two species from the *D. melanogaster* subgroup and two species from the *pseudoobscura* subgroup. One of these putative Acp genes (*CG16707*) exhibted positive selection in both subgroups and contributed to an overall correlation between $d\mathrm{N}/d\mathrm{S}$ (the ratio of nonsynonymous substitutions per nonsynonymous site to synonymous substitutions per synonymous site [Hughes and Nei 1988]) values for 12 orthologues compared between the two clades. Genomic comparisons by Muller et al. (2005) found that the $d\mathrm{N}/d\mathrm{S}$ ratios of Acp genes with recognizable orthologues between *D. melanogaster* and *D. pseudoobscura* were lower within the *D. melanogaster* subgroup than were $d\mathrm{N}/d\mathrm{S}$ ratios for Acp genes with no identifiable orthologue in *D. pseudoobscura*. Mueller et al. (2005) also defined Acps more narrowly than previously, thereby excluding *CG16707*. Thus the degree to which selection on orthologous Acps is correlated in distant clades remains unknown.

Here, we test for positive selection on the five Acps in the *D. pseudoobscura* subgroup identified by Wagstaff and Begin (2005). In addition to nucleotide substitution rates, we evaluate the role that indels, a source of variation heretofore ignored in studies of Acps in *Drosophila*, play in the divergence of these proteins. We compare patterns of selected change within the *D. pseudoobscura* subgroup to those seen for the same Acps in the *D. melanogaster* subgroup. If the patterns of molecular evolution in these Acps are similar between these two clades, then the conserved functions of these proteins could remain a constant target for similar types of selection over large time scales. Alternatively, different patterns of selection on orthologous reproductive proteins in the two lineages

would suggest that selection opportunistically targets different loci among different clades.

Materials and Methods

Fly Stocks

Flies used in this study were obtained from Dr. Mohamed Noor, Dr. Carlos Machado, and the Tucson Stock Center (http://stockcenter.arl.arizona.edu) and largely overlap with those used by Machado et al. (2002). We used 20 lines of *D. pseudoobscura*: 4 lines from Mather, California (Mather17, Mather32, Mather52, and Mather1959); 4 lines from Mt. St. Helena, California (MSH9, MSH21, MSH24, and MSH32); 1 line from James Reserve, California; 4 lines from American Fort Canyon, Utah (AF2, AFC3, AFC7, and AFC12); 4 lines from Flagstaff, Arizona (Flagstaff5, Flagstaff14, Flagstaff16, and Flagstaff18); 1 line from Tucson, Arizona; 1 line from Baja, California (Baja 1); and 1 line from Sonora, Mexico (Sonora 3). We also used 11 lines of *D. p. bogotana* from near the city of Bogotá in Cundinamarca, Colombia (Bogotá 1960, Bogotá 1976, Potosí2, Potosí3, Susa2, Susa6, Sutatausa3, Sutatausa5, Toro1, Toro6, and Toro7); 7 lines of *D. persimilis*—3 lines from Mather, California (Mather37, Mather40, MatherG) and 4 lines from Mt. St. Helena, CA (MSH1, MSH3, MSH7, and MSH42); and 3 lines of *D. miranda* (MSH22, MSH38, and Mather 1993).

DNA Isolation, PCR Amplification, and Sequencing

DNA was extracted from whole male flies using the single fly squish protocol of Gloor and Engels (1992). PCR primers were designed from *D. pseudoobscura* Acp sequences from Wagstaff and Begun (2005) using Primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3). The PCR was performed on a PTC-200 (MJ Research, Watertown, MA) using the following conditions: 94°C for 2 min 30 sec, 50°C for 2 min, then 72°C for 2 min, followed by 38 cycles of 94°C for 45 sec, 50°C for 1 min, then 72°C for 1 min 15 sec. Resulting amplicons were purified using either a Strataprep PCR Purification Kit (Stratagene, La Jolla, CA) or a QuickStep2 96-Well PCR Purification Kit (Edge BioSystems, Gaithersburg, MD), then sequenced using both amplification primers on an ABI 377 automated sequencer, using Big Dye Terminators (V3.1; Applied Biosystems, Foster City, CA). Sequences are available from GenBank (DQ368868–DQ369012).

Sequence Analyses

Nucleotide sequences for each Acp were initially assembled and edited with Sequencher 3.0 (Gene Codes, Ann Arbor, MI). Inferred amino acid sequences were then aligned with ClustalW (http://www2.ebi.ac.uk/clustalw/) under default settings. Further alignment modifications were made by hand. Resulting amino acid alignments were then used to align nucleotides.

Measures of Acp polymorphism and divergence, as well as McDonald-Kreitman's (1991) test for nonneutrality, were calculated using DnaSP 4.0 (Rozas et al. 2003). These measures can reveal departures from neutrality that act across all sites of a protein by comparing the number of silent versus replacement polymorphisms. We tested for recombination with DnaSP using the algorithm described by Hudson and Kaplan (1985). No significant recombination was detected. DnaSP 4.0 was also used to calculate Tajima's *D*, Fu and Li's *D*, and Fay and Wu's *H*, with confidence levels for these estimated by the coalescent with 1000 replications. We obtained parsimony and neighbor-joining (Saitou and Nei 1987) trees for alleles using Kimura two-parameter distances in PAUP* v4.0b10 (Swofford 2001). Branch support was estimated by bootstrapping using 1000 replicates.

Acp sequences from the *D. melanogaster* subgroup were downloaded from GenBank. These were chosen based on sequence length (>75% of the protein's open reading frame had to be available) and uniqueness (identical sequences were not included). These sequences were given initially by Tsaur et al. (2001; AF302208–AF302229), Begun et al. (2000; AY010527–AY010711), Panhuis et al. (2003; AY344246–AY344364), Holloway and Begun (2004; AY635196–AY635290), and Kern et al. (2004; AY505178–AY505293). For these analyses, *D. melanogaster* (Zimbabwe) and *D. p. bogotana* were considered as taxa separate from their nominal conspecifics.

We used the codeml program in PAML 3.14 (Yang 2004) to test for positive selection and to infer amino acid sites under positive selection under the maximum likelihood methods of Nielson and Yang (1998) and Yang et al. (2000). A Bayes Empirical Bayes (BEB; Deely and Lindley 1981) approach was subsequently used to calculate the posterior probabilities that each particular site fell into the different $d\mathrm{N}/d\mathrm{S}$ (or ω) classes (Yang et al. 2005).

We performed three tests for positive selection on nucleotide substitutions. First, we used a model (M0) that assumed a single ω value for all sites to estimate the level of positive selection averaged over all codons (Nielsen and Yang 1998). Second, a more robust test for adaptive evolution was performed by comparing the nested models M7 and M8 (Yang et al. 2000). The neutral model M7 allowed ω to take on beta-distributed values between 0 and 1 at each codon (i.e., no positive selection). This was compared with selection model M8, which used the same beta-distributed values for neutral codons but added another parameter that allows a proportion of codons to take on ω values greater than one. Finally, we compared selection model M8 to model M8A, which allows a proportion of sites to equal but not exceed a ω value of 1 (Swanson et al. 2003). Positive selection was inferred if $\omega > 1.0$. Significance was determined by comparing twice the difference between the likelihood values of M7 vs. M8 or M8 vs. M8A to a chi-square table of critical values with one degree of freedom. The default starting value of ω in PAML is 0.3 for all models. Because convergence is a concern for MCMC analyses, we varied initial ω values (set to 0, 0.5, and 1.0). Results were consistent regardless of these starting values (data not shown); we report values from the default priors here.

Codon-based maximum likelihood approaches have had success in identifying residues under selection, as evidenced by their ability to identify residues already functionally implicated as being under positive selection (e.g., Yang and Swanson 2002; Palomino et al. 2002). We used the BEB approach to identify positively selected residues instead of alternative parsimony-based approaches (Suzuki and Nei 2004; Zhang 2004) because (1) while the parsimony methods have a low rate of false positives, they also have little power for detecting positive selection or identifying positively selected sites (Wong et al. 2004), and (2) while the older Naive Empirical Bayesian approach (NEB) can have high false-positive rates, the BEB approach corrects for past problems and reduces the false-positive rate considerably (Yang et al. 2005). Through the BEB approach, sites under positive selection can be identified, even if the average $d\mathrm{N}/d\mathrm{S}$ over all sites is <1. Sites with a high probability of belonging to the class with $\omega > 1$ are likely to be under positive selection.

Determining whether positive selection promotes indels is not as straightforward as for nucleotide substitutions because there are no natural within-gene comparisons analogous to synonymous substitutions. To test for positive selection on indels, Podlaha and Zhang (2003) compared the rates of indel substitutions in the reproductive protein of interest to those in neutral (noncoding) sequences with the simple ratio (number of nucleotide indels)/(total number of base pairs)/(divergence time). Ideally, such comparisons

Table 2. Acp polymorphism in the *D. pseudoobscura* subgroup

Locus	Species	n[a]	L[b]	S[c]	Syn[d]	Non[d]	θ[e]	π[f]	D[g]	Div[h]
Acp26Aa	*D. pseudoobscura*	20	558	56	24	32	0.0308	0.0273	−0.091	0.0938
	D. p. bogotana	11	618	16	9	7	0.0125	0.0145	−0.067	0.0831
	D. persimilis	7	654	17	6	11	0.0117	0.0120	−0.110	0.0862
	D. miranda	3	579	2	0	2	0.0023	0.0017	—	
Acp32CD	*D. pseudoobscura*	20	875	36	25	11	0.0113	0.0083	−0.061	0.0306
	D. p. bogotana	11	900	9	3	6	0.0054	0.0083	0.024	0.0345
	D. persimilis	7	879	17	13	4	0.0093	0.0096	−0.066	0.0278
	D. miranda	3	885	0	0	0	0.0000	0.0000	—	
Acp53Ea	*D. pseudoobscura*	20	330	7	4	3	0.0064	0.0036	−0.106	0.0120
	D. p. bogotana	11	330	2	1	1	0.0023	0.0024	−0.020	0.0116
	D. persimilis	7	330	1	1	0	0.0016	0.0015	−0.010	0.0109
	D. miranda	3	330	5	3	2	0.0101	0.0101	—	
Acp62F	*D. pseudoobscura*	20	408	27	14	13	0.0192	0.0140	−0.118	0.0334
	D. p. bogotana	11	408	6	3	3	0.0057	0.0066	−0.083	0.0338
	D. persimilis	7	408	9	2	7	0.0106	0.0098	0.010	0.0265
	D. miranda	3	408	7	2	5	0.0114	0.0116	—	
Acp70A	*D. pseudoobscura*	20	165	1	1	0	0.0019	0.0016	−0.076	0.0433
	D. p. bogotana	11	165	0	0	0	0.0000	0.0000	—	0.0545
	D. persimilis	7	165	1	1	0	0.0033	0.0040	−0.014	0.0454
	D. miranda	3	165	2	2	0	0.0121	0.0121	—	

[a] Number of lines sequenced.
[b] Average length (bp) of the sequences from each species.
[c] Number of polymorphic sites.
[d] Syn, number of synonymous polymorphisms in the coding regions; non, number of nonsynonymous polymorphisms in the coding regions.
[e] Estimate of 4Nu per base pair using the number of polymorphic sites (Watterson 1975).
[f] Estimate of 4Nu using the average number of nucleotide differences per site (Nei 1987).
[g] Tajima's (1989b) statistic. No values were significantly different from zero.
[h] Average divergence per base pair between alleles from each taxon and alleles of *D. miranda*.

would be made involving the same genomes, but such estimates are not always available. For their comparisons among primates, Podlaha and Zhang (2003) used estimates of neutral indel rates between humans and chimpanzees, even though this pair showed no indels for the *Catsper1* gene of interest. We estimated indel substitution rates in *Drosophila* from differences between intronic, 5′-intergenic, and 3′-intergenic regions in *D. simulans* and *D. sechellia* (Halligan et al. 2004). Note that this method of comparison is conservative, because indels occurring in exonic sequences must occur in multiples of 3 bp so as not to disrupt open reading frames, a constraint not present for noncoding regions. Indels were counted without regard to their size.

Results

Intraspecific Variation

Intraspecific sequence variation at the five Acp loci examined (Table 2) was comparable to the range of values reported for neutral sequence regions by Machado et al. (2002) for these same taxa. For both Θ_w and Nei's π, *D. pseudoobscura* had the highest levels of nucleotide variation at *Acp26Aa*, *Acp32CD*, and *Acp62F*, while *D. miranda* had the highest nucleotide variation for *Acp53Ea* and *Acp70A*.

For all phylogenetic analyses of Acps, alleles from *D. miranda* fell basal to the other taxa (not shown). Individuals from the same taxon generally grouped together, and with the same topology as generally accepted for this subgroup, although support was weak. *Acp26Aa* (Fig. 1) provided the strongest exception to this pattern, with many *D. persimilis* alleles grouping with *D. p. pseudoobscura* alleles, to the exclusion of a basal group of *D. p. pseudoobscura* and *D. p. bogotana* alleles.

Tests for Neutrality

Tajima's (1989) *D* was not significantly different from zero in any taxon within the *D. pseudoobscura* subgroup for any of the Acp loci (Table 2). Fu and Li's *D* and Fay and Wu's *H* were also not significantly different from zero. Therefore, we cannot reject the hypothesis that these loci are evolving neutrally using these tests. McDonald-Kreitman tests also failed to reveal any departure from neutral behavior at *Acp53Ea*, *Acp62F*, or *Acp70A* (Supplementary Table 1). The only comparisons that showed deviation from neutrality were between *D. p. bogotana* and *D. miranda* for *Acp26Aa* ($p = 0.005$) and between *D. persimilis* and *D. miranda* for *Acp32CD* ($p = 0.0079$). All other interspecific comparisons for *Acp26Aa* and *Acp32CD* did not deviate from neutrality under this test (Supplementary Table 1). These results remained significant after applying the Williams correction for independence (Sokal and Rohlf 1995).

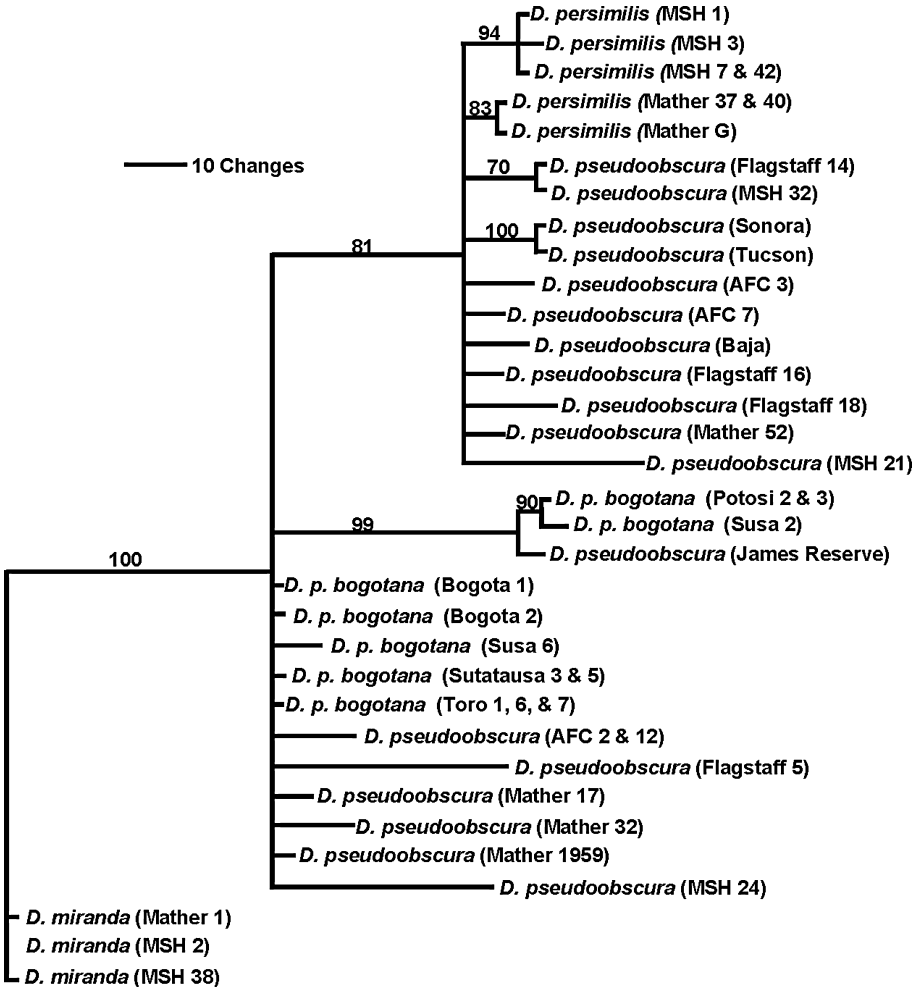

Fig. 1. Neighbor-joining tree (using Kimura two-parameter distances) for alleles of *Acp26Aa* of the *Drosophila pseudoobscura* subgroup. Numbers above branches indicate bootstrap support values. Geographic origin and line numbers are shown in parentheses.

Tests for Positive Selection on Nucleotide Substitutions

Raw sequence comparisons suggest the possibility of positive selection, with more replacement than silent polymorphisms in at least some comparisons for *Acp26Aa*, *Acp32CD*, and *Acp62F* (Table 2). $d\mathrm{N}/d\mathrm{S}$ ratios (ω) averaged across lineages and sites (M0) were <1 for all Acps in both subgroups, with the exception of *Acp32CD* in the *D. melanogaster* subgroup (Table 3). Under the positive selection model (M8) positive selection was detected in *Acp26Aa*, *Acp32CD*, and *Acp62F* in both the *D. pseudoobscura* subgroup and the *D. melanogaster* subgroup, with *Acp53Ea* under selection in the *D. melanogaster* group as well (Table 3).

To identify the particular residues underlying this positive selection, we used the BEB approach of Yang et al. (2005). Many residues were subject to positive selection in three of the Acps examined: *Acp26Aa*, *Acp32CD*, and *Acp62F* (Table 3). For *Acp26Aa*, a higher proportion of sites underwent positive selection in the *D. pseudoobscura* subgroup than in the *D. melanogaster* subgroup. A similar number of sites underwent positive selection between the groups for *Acp62F*. *Acp53Ea* also had a $\omega > 1$ in the *D. melanogaster* subgroup, but this was not significant (Table 3). The extensive divergence between orthologous loci prevented us from determining whether the same residues were under selection in the two clades.

Acp26Aa had the highest $d\mathrm{N}/d\mathrm{S}$ ratio in the *D. pseudoobscura* subgroup and under strong selection in the *D. melanogaster* subgroup. *Acp62F* was also undergoing significant positive selection in both groups, but at fewer sites and with lower ω values. No significant positive selection was detected in *Acp53Ea* or *Acp70A* for either subgroup. *Acp32CD* was undergoing significant positive selection in the *D. melanogaster* subgroup, but not in the *D. pseudoobscura* subgroup, although positive selection was suggested at more sites in this Acp than in either *Acp53Ea* or *Acp70A*.

Indel Substitutions

Nucleotide substitutions were not the only source of variation in *Acp26Aa*. Amino acid alignments of *Acp26Aa* revealed several indels in both the

Table 3. LRTs of positive selection for Acps in the *D. pseudoobscura* and *D. melanogaster* subgroups

Gene	Subgroup	No. alleles/ taxon	L (codons)	S	*d*N/ *d*S	2Δ1, M8vs. M7	2Δ1, M8vs. M8a	Parameter estimate under M8 (ß & ω)	Positively selected sites
Acp26Aa	*pseudoobscura*	32/4	244	4.867	0.710	146.21**	101.21**	p_0 = 0.857 p_1 = 0.143 ω = 4.834	23, 30, 31, 33, 35, 39, 42, 43, 45, 47, 49, 55, 57, 59, 61, 62, 63, 64, 74, 75, 76, 77, 89, 97, 98, 100, 101, 102, 237
	melanogaster	12/5	256	0.849	0.870	13.8**	13.34**	p_0 = 0.938 p_1 = 0.062 ω = 6.753	**23**, **24**, 25, **30**, 101, 178, 200
Acp32CD	*pseudoobscura*	31/4	303	0.394	0.263	2.24	1.40	p_0 = 0.855 p_1 = 0.145 ω = 1.641	145, 270, **272**
	melanogaster	8/2	260	0.086	1.580	13.2**	22.38**	p_0 = 0.892 p_1 = 0.108 ω = 17.589	29, 40, 145, 185, 187
Acp53Ea	*pseudoobscura*	31/4	110	0.130	0.182	0.001	0.02	NA	None detected
	melanogaster	19/5	110	0.463	0.266	6.68	6.08*	p_0 = 0.767 p_1 = 0.233 ω = 1.194	None detected
Acp62F	*pseudoobscura*	34/4	135	0.808	0.371	7.92*	8.63**	p_0 = 0.943 p_1 = 0.057 ω = 4.112	**13**, **16**, **86**, 89, 130, 131, 132, **134**
	melanogaster	17/5	92	1.133	0.450	6.26*	6.25*	p_0 = 0.721 p_1 = 0.279 ω = 1.802	7, **9**, 10, 13, 22, 23, 24, 30, 34, **38**, 44, 58, 70, 85, **91**
Acp70A	*pseudoobscura*	21/4	55	0.237	0.266	0.006	0.02	NA	None detected
	melanogaster	7/4	55	0.330	0.365	0.002	0.03	p_0 = 1.000 p_1 = 0.000 ω = 1.000	None detected

Note. *S* is the tree length, measured as the number of nucleotide substitutions per codon. *d*N/*d*S is the average ratio over sites and branches, both calculated under model M0. The proportion of sites under positive selection (p_1) or under selective constraint (p_0) is given under model M8. Positively selected sites with posterior probability >0.9 are underlined and those with posterior probability = 0.8–0.9 are in boldface. Positively selected sites are identified under Bayes Empirical Bayes (BEB) analysis and are subgroup specific because comparisons of positively selected sites could not be made between subgroups. *D. melanogaster* subgroup sites are numbered as the sequences appear in GenBank and *D. pseudoobscura* subgroup sites are numbered at the start codon with the exception of *Acp26Aa*, in which residues analyzed began immediately after the intron. NA, not applicable. *Significant at 5% level. **Significant at 1% level based on comparison to a chi-square table of critical values with 1 df.

Fig. 2. Amino acid alignment of insertions and deletions in part of *Acp26Aa* from (**a**) the *Drosophila pseudoobscura* subgroup and (**b**) the *D. melanogaster* subgroup. Positively selected sites with posterior probabilities >0.8 are highlighted in gray. *D. pseudoobscura* subgroup sites are numbered starting immediately after the sole intron. *D. melanogaster* subgroup sites are numbered as the sequences appear in GenBank.

Table 4. Estimated indel substitution rates for *Acp26Aa*, intronic, and gene flanking regions

Region	Taxon 1	Taxon 2	Divergence	Indels	Total bp	Indel substitution rate (subs/billion years)
Acp26Aa	*D. p. bogotana*	*D. miranda*	2.1 MY[a]	5	732	1.63
Acp26Aa	*D. pseudoobscura*	*D. persimilis*	0.5 MY[a]	7	732	9.56
Acp26Aa	*D. pseudoobscura*	*D. p. bogotana*	0.15 MY[a]	6	732	27.32
Intronic[b]	*D. simulans*	*D. sechellia*	0.9 MY[c]	44	6302	3.88
5′ intergenic[b]	*D. simulans*	*D. sechellia*	0.9 MY[c]	9	3094	1.62
3′ intergenic[b]	*D. simulans*	*D. sechellia*	0.9 MY[c]	18	3159	3.17

[a] Divergence times are based on the *amylase* gene and are from Aquadro et al. (1991).
[b] Estimates from Halligan et al. (2004).
[c] Divergence times from Hey and Kliman (1993).

pseudoobscura and the *melanogaster* subgroups, including polymorphisms within species (Fig. 2a). In contrast to these exonic indels, there were no indels present in an immediately adjacent 68-bp intron of *Acp26Aa* in any of the seven individuals from the *pseudoobscura* subgroup (obtained from GenBank; our sequencing started immediately after the intron). In addition, 22 of the 29 positively selected sites (with posterior probabilities >0.8) fell within the indel regions of *Acp26Aa* in the *D. pseudoobscura* subgroup (Fig. 2a), even though these regions constituted only 39% of the total aligned protein-coding region. In contrast, only four of seven positively selected sites (with posterior probabilities >0.8) fell within indel regions in the *D. melanogaster* subgroup (Fig. 2b). The indels sometimes prevented unambiguous alignment of sequences (especially a 12-residue repeat shared by some *D. p. pseudoobscura* and *D. p. bogatana* sequences (Fig. 2a), however, analysis of several alternative alignments produced very similar results in terms of the number of residues under selection and overall values of *Dn*/*Ds* (not shown).

Comparisons of indel substitution rates in *Acp26Aa* to those in noncoding regions of *Drosophila* genomes suggest that indels may be under positive selection. The indel substitution rates in *Acp26Aa* are higher than, or of the same order of magnitude as, those in noncoding regions of *Drosophila* genomes (Table 4).

Acp32CD also contained several indels, including a single indel polymorphism within *Acp32CD* of *D. pseudoobscura*. Alignments of *Acp32CD* revealed one 6-bp insertion/deletion between *D. melanogaster* (United States and Zimbabwe) and *D. simulans*. This indel did not fall in a positively selected region of *Acp32CD*. No indels were present in *Acp53Ea*, *Acp62F*, or *Acp70A* in either of these groups.

Discussion

We have shown that the accessory gland proteins *Acp26Aa* and *Acp62F* have sites that are undergoing positive selection in the *D. pseudoobscura* subgroup. Similar proportions of positively selected sites are

found in these same two Acps in the *D. melanogaster* subgroup, and in *Acp32CD* as well. Two additional Acps, *Acp53Ea* and *Acp70A*, were not subject to positive selection in either of these subgroups. In addition to positive selection acting on nucleotide substitutions, we also found several indel replacements and polymorphisms in *Acp26Aa* and *Acp32CD*. The regions where these indels occur are the same places that harbor positively selected nucleotide substitutions for *Acp26Aa* in the *D. pseudoobscura* subgroup, but not in the *D. melanogaster* subgroup. The deep divergence in Acps from the two subgroups prevented us from determining whether the same residues are subject to positive selection in both subgroups, as Acps from the different subgroups could not be aligned. *Acp26Aa* has already been demonstrated to undergo positive selection in the *D. melanogaster* subgroup (Tsaur and Wu 1997; Tsaur et al. 1998; Begun et al. 2000) and in the *D. pseudoobscura* subgroup (Wagstaff and Begun 2005). However, this is the first study to identify positive selection at particular sites for *Acp26Aa* or any other drosophilid Acp or to note extensive indel variation or high rates of indel substitution within any Acp.

Mueller et al. (2005) suggested that, because most Acps from *D. melanogaster* could not be detected in *D. pseudoobscura*, Acps might be undergoing different evolutionary paths in these divergent lineages. Stevison et al. (2004), however, found that $d\mathrm{N}/d\mathrm{S}$ values were correlated for 12 orthologous genes in the *melanogaster* and *pseudoobscura* subgroups, 4 of which were putative Acps (although only 2 of these would qualify as Acps under the definitions of Mueller et al. [2005]). For the subset of five Acps where orthologues in the two subgroups had been recognized by Wagstaff and Begun (2005), we found that the relative strength of positive selection on nucleotide substitutions is similar. This suggests that the presumably conserved functions of these proteins have remained targets for the same type of selection, diversifying or stabilizing, over long periods of time.

The functions of the two Acps shown here to be under positive selection suggest a potential role in some observed reproductive incompatibilities within the two subgroups. *Acp62F* protects sperm from proteolysis (Lung et al. 2002), which could potentially protect the sperm in the female's reproductive tract. The protease inhibitor class to which *Acp62F* belongs was noted by Mueller et al. (2005) as being especially lacking in orthologues between the *melanogaster* and the *pseudoobscura* subgroups. However, whether the action of *Acp62F* is species specific in the *D. melanogaster* subgroup remains unknown.

Acp26Aa (ovulin) increases egg-laying (Herndon and Wolfner 1995; Heifetz et al. 2001). In addition, Clark et al. (1995) showed that *Acp26Aa* genotypes correlate with sperm displacement ability within *D. melanogaster*. If these observed intraspecific effects carried over to interactions between subspecies, the allelic variation at *Acp26Aa* might play a role in the conspecific sperm precedence observed between subspecies of *D. pseudoobscura* (Dixon et al. 2003). Here, we found that *Acp26Aa* alleles from the same *D .p. pseudoobscura* populations used by Dixon et al. (2003) fell into two different (modestly supported) phylogenetic groups: one basal and including all alleles from *D. p. bogotana*, the other derived and containing all *D. persimilis* alleles but none from *D. p. bogotana* (Fig. 1). Studies that simultaneously genotyped *Acp26Aa* alleles and evaluated mating success (as Clarke et al. 1995) may reveal whether some conspecific sperm precedence seen between *D. p. pseudoobscura* and *D. p. bogotana* (Dixon et al. 2003) owes to divergence at this locus (possibly from introgressed *D. persimilis* alleles).

Previous studies evaluating positive selection on Acps have only examined nucleotide substitutions. Two recent studies, however, have shown positive selection acting on indels in a sperm-specific protein (Catsper1) in both primates (Podlaha and Zhang 2003) and rodents (Podlaha et al. 2005). *Catsper1* encodes a voltage-gated calcium ion channel that is necessary for proper sperm motility (Ren et al. 2001) and may help mediate sperm competition. Positive selection on nucleotides also occurs in indel-rich regions of the gamete recognition protein bindin from sea urchins (Metz and Palumbi 1996; McCartney and Lessios 2004, Zigler and Lessios 2004). Previous studies evaluating the molecular evolution of Acps in *Drosophila*, however, have either implicitly or explicitly excluded indels from their analyses (e.g., Tsaur and Wu 1997; Begun et al. 2000), although Mueller et al. (2005) noted that two Acp loci (*CG14560* and *CG9074*) contained repetitive regions. Our results suggest that indel substitutions play a significant role in the divergence of some Acps. Indels appear to be concentrated in the same part of *Acp26Aa* of the *D. pseudoobscura* subgroup as where most residues under positive selection occur. This correlation we found between positively selected residues and indel sites in the *D. pseudoobscura* subgroup should not arise as an artifact of the PAML analysis (and indeed is not present in the *D. melanogaster* subgroup) because gaps are treated as ambiguities and dropped from the analysis in pairwise fashion. Further, the high rates of indel substitution in *Acp26Aa* (Table 4) suggest that positive selection may act on the indels themselves.

Positive selection often drives the rapid evolution of reproductive proteins (Swanson and Vacquier 2002). We have demonstrated that the strength of positive selection on nucleotide substitutions acting on five orthologous Acps is similar in two drosophilid lineages that split 21–46 MYA (Beckenbach et al.

1993). In addition, indels also contribute to the divergence of some Acps and may even be promoted by positive selection.

Acknowledgments. We thank Carlos Machado and Mohamed Noor for providing fly stocks and Kevin Schully and Pat Arbour-Reily for technical help. We thank Daniel Ortiz-Barrientos, Mike Taylor, Mariana Wolfner, Willie Swanson, and two anonymous reviewers for critical reviews and helpful discussions. Sequencing expenses were supported in part by an NSF Multiuser Equipment Grant (DBI-0400797) to the LSU Museum of Natural Sciences.

References

Aigaki T, Fleischmann I, Chen PS, Kubli E (1991) Ectopic expression of sex peptide alters reproductive behavior of female *Drosophila melanogaster*. Neuron 7:557–563

Beckenbach AT, Wei YW, Liu H (1993) Relationships in the *Drosophila obscura* species group, inferred from mitochondrial *cytochrome oxidase II* sequences. Mol Biol Evol 10:619–634

Begun DJ, Whitley P, Todd BL, Waldrip-Dail HM, Clark AG (2000) Molecular population genetics of male accessory gland proteins in *Drosophila*. Genetics 156:1879–1888

Britten RJ, Rowen L, Williams J, Cameron RA (2003) Majority of divergence between closely related DNA samples is due to indels. Proc Natl Acad Sci USA 100:4661–4665

Chapman T, Liddle LF, JKalb JM, Wolfner MF, Partridge L (1995) Cost of mating in *Drosophila melanogaster* females is mediated by male accessory gland products. Nature 373:241–244

Chapman T, Herndon LA, Heifetz Y, Partridge L, Wolfner MF (2001) The *Acp26Aa* seminal fluid protein is a modulator of early egg-hatchability in *Drosophila melanogaster*. Proc Roy Soc Lond B 268:1647–1654

Chen PS, Stumm-Zollinger E, Aigaki T, Balmer J, Bienz M, Bohen P (1988) A male accessory gland peptide that regulates reproductive behaviour of female *D. melanogaster*. Cell 54:291–298

Civetta A, Singh RS (1995) High divergence of reproductive tract proteins and their association with postzygotic reproductive isolation in *Drosophila melanogaster* and *Drosophila virilis* group species. J Mol Evol 41:1085–1095

Civetta A, Singh RS (1998) Sex-related genes, directional sexual selection, and speciation. Mol Biol Evol 15:901–909

Clark AG, Aguadé M, Prout T, Harshman LG, Langley CH (1995) Variation in sperm displacement and its association with accessory gland protein loci in *Drosophila melanogaster*. Genetics 139:189–201

Deely JJ, Lindley DV (1981) Bayes empirical Bayes. J Am Stat Assoc 76:833–841

Denver DR, Morris K, Lynch M, Thomas WK (2004) High mutation rate and predominance of insertions in the *Caenorhabditis elegans* nuclear genome. Nature 430:679–682

Dixon SM, Coyne JA, Noor MAF (2003) The evolution of conspecific sperm precedence in *Drosophila*. Mol Ecol 12:1179–1184

Gloor GB, Engels WR (1992) Single-fly DNA preps for PCR. Droso Inform Serv 71:148–149

Halligan DL, Eyre-Walker A, Andolfatto P, Keightley PD (2004) Patterns of evolutionary constraints in intronic and intergenic DNA of *Drosophila*. Genet Res 14:273–279

Harshman LG, Prout T (1994) Sperm displacement without sperm transfer in *Drosophila melanogaster*. Evolution 48:758–766

Heifitz Y, Tram U, Wolfner MF (2001) Male contributions to egg production:the role of accessory gland products and sperm in *Drosophila melanogaster*. Proc Roy Soc Lond B 268:175–180

Hellberg ME, Moy GM, Vacquier VD (2000) Positive selection and propeptide repeats promote rapid interspecific divergence of a gastropod sperm protein. Mol Biol Evol 17:458–466

Herndon LA, Wolfner MF (1995) A *Drosophila* seminal fluid protein, *Acp26aa*, stimulates egg laying in females for 1 day after mating. Proc Natl Acad Sci USA 92:10114–10118

Holloway AK, Begun DJ (2004) Molecular evolution and population genetics of duplicated accessory gland protein genes in *Drosophila*. Mol Biol Evol 21:1625–1628

Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics 111:147–164

Hughes AL, Nei M (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature 335:167–170

Kern AD, Jones CD, Begun DJ (2004) Molecular population genetics of male accessory gland proteins in the *Drosophila simulans* complex. Genetics 167:725–735

Lung O, Tram U, Finnerty C, Eipper-Mains M, Kalb J, Wolfner MF (2002) The *Drosophila melanogaster* seminal fluid protein *Acp62F* is a protease inhibitor that is toxic upon ectopic expression. Genetics 160:211–224

Machado CA, Kliman RM, Markert JA, Hey J (2002) Inferring the history of speciation from multilocus sequence data: the case of *Drosophila pseudoobscura* and its close relatives. Mol Biol Evol 19:472–488

McCartney MA, Lessios HA (2004) Adaptive evolution of sperm bindin tracks egg incompatibility in neotropical sea urchins of the genus *Echinometra*. Mol Biol Evol 21:732–745

McDonald JH, Kreitman M (1991) Adaptive protein evolution at the Adh locus in *Drosophila*. Nature 351:652–654

Metz EC, Palumbi SR (1996) Positive selection and sequence rearrangements generate extensive polymorphism in the gamete recognition protein bindin. Mol Biol Evol 13:397–406

Mueller JL, Ravi Ram K, McGraw LA, Bloch Qazi MC, Siggia ED, Clark AG, Aquadro CF, Wolfner MF (2005) Cross-species comparison of *Drosophila* male accessory gland protein genes. Genetics 171:131–143

Neilsen R, Yang Z (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 *envelope* gene. Genetics 148:929–936

Neubaum DM, Wolfner MF (1999) Mated *Drosophila melanogaster* females require a seminal fluid protein, Acp36DE, to store sperm efficiently. Genetics 153:845–857

Palomino MM, Meyers BC, Michelmore RW, Gaut BS (2002) Patterns of positive selection in the complete NBS-LRR gene family of *Arabidopsis thaliana*. Genet Res 12:1305–1315

Panhuis TM, Swanson WJ, Nunney L (2003) Population genetics of accessory gland proteins and sexual behaviour in *Drosophila melanogaster* populations from Evolution Canyon. Evolution 57:2785–2791

Podlaha O, Zhang J (2003) Positive selection on protein-length in the evolution of a primate sperm ion channel. Proc Natl Acad Sci USA 100:12241–12246

Podlaha O, Webb DM, Tucker PK, Zhang J (2005) Positive selection for indel substitutions in the rodent sperm protein Catsper1. Mol Biol Evol 22:1845–1852

Ren DJ, Navarro B, Perez G, Jackson AC, Hsu SF, Shi Q, Tilly JL, Clapham DE (2001) A sperm ion channel required for sperm motility and male fertility. Nature 413:603–609

Richards S, Liu Y, Bettencourt BR, Hradecky P, Letovsky S, Nielsen R, Thornton K, Hubisz MJ, Chen R, Meisel RP, Couronne O, Hua S, Smith MA, Zhang P, Liu J, Bussemaker HJ, van Batenburg MF , Howells SL, Scherer SE, Sodergren E, Matthews BB, Crosby MA, Schroeder AJ, Ortiz-Barrientos D, Rives CM, Metzker ML, Muzny DM, Scott G, Steffen D, Wheeler DA, Worley KC, Havlak P, Durbin KJ, Egan A, Gill R, Hume J, Morgan MB, Miner G, Hamilton C, Huang

Y, Waldron L, Verduzco D, Clerc-Blankenburg KP, Dubchak I, Noor MAF, Anderson W, White KP, Clark AG, Schaeffer SW, Gelbart W, Weinstock GM, Gibbs RA (2005) Comparative genome sequencing of *Drosophila pseudoobscura*: chromosomal, gene, and cis-element evolution. Genet Res 15:1–18

Rozas J, Sánchez-DelBarrio JC, Messegyer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496–2497

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Singh RS, Kulathinal RJ (2000) Sex gene pool evolution and speciation: a new paradigm. Genes Genet Syst 75:119–130

Sokal RR, Rohlf FJ (1995) Biometry. W.H. Freeman, New York

Stevison LS, Counterman BA, Noor MAF (2004) Molecular evolution of X-linked accessory gland proteins in *Drosophila pseudoobscura*. J Hered 95:114–118

Suzuki Y, Nei M (2004) False positive selection identified by ML-based methods: examples from the *sig1* gene of the diatom *Thalassiosira weissflogii* and the *tax* gene of a human T-cell lymphotropic virus. Mol Biol Evol 21:914–921

Swanson WJ, Vacquier VD (1998) Concerted evolution in an egg receptor for a rapidly evolving abalone sperm protein. Science 281:710–712

Swanson WJ, Vacquier VD (2002) Reproductive protein evolution. Annu Rev Ecol Syst 33:161–179

Swanson WJ, Clark AG, Waldrip-Dail HM, Wolfner MF, Aquadro CF (2001) Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. Proc Natl Acad Sci USA 98:7375–7379

Swanson WJ, Nielsen R, Yang Q (2003) Pervasive adaptive evolution in mammalian fertilization proteins. Mol Biol Evol 20:18–20

Swofford DL (2001) PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, MA

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595

Tsaur SC, Wu C-I (1997) Positive selection and the molecular evolution of a gene of male reproduction, *Acp26Aa* of *Drosophila*. Mol Biol Evol 14:544–549

Tsaur SC, Ting CT, Wu C-I (2001) Sex in *Drosophila mauritiana*: extremely high level of replacement polymorphism in a male reproductive gene. Mol Biol Evol 18:22–26

Wagstaff BJ, Begun DJ (2005) Comparative genomics of accessory gland protein genes in *Drosophila melanogaster* and *D. pseudoobscura*. Mol Biol Evol 22:818–832

Wolfner MF (2002) The gifts that keep on giving: physiological functions and evolutionary dynamics of male seminal proteins in *Drosophila*. Heredity 88:85–93

Wong WSW, Yang Z, Goldman N, Nielsen R (2004) Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. Genetics 168:1041–1051

Xue L, Noll M (2000) Drosophila female sexual behavior induced by sterile males showing copulation complementation. Proc Natl Acad Sci USA 97:3272–3275

Yang Z (2004) A probabilist's account of modern molecular population genetics. Heredity 92:474

Yang Z, Swanson WJ (2002) Codon substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. Mol Biol Evol 19:49–57

Yang Z, Nielsen N, Goldman N, Pedersen A-M (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155:431–449

Yang Z, Wong WSW, Nielsen R (2005) Bayes empirical Bayes inference of amino acid sites under positive selection. Mol Biol Evol 22:1107–1118

Zhang J (2004) On the evolution of codon volatility. Genetics 169:495–501

Zigler KS, Lessios HA (2004) Speciation on the coasts of the New World: phylogeography and the evolution of binding in the sea urchin genus *Lytechinus*. Evolution 58:1225–1241