DOI: 10.1534/genetics.104.034249

Elevated Polymorphism and Divergence in the Class C Scavenger Receptors of *Drosophila melanogaster* and *D. simulans*

Brian P. Lazzaro[1]

Department of Entomology, Cornell University, Ithaca, New York 14853

Manuscript received August 2, 2004
Accepted for publication January 5, 2005

ABSTRACT

Scavenger receptor proteins are involved in the cellular internalization of a broad variety of foreign material, including pathogenic bacteria during phagocytosis. I find here that nonsynonymous divergence in three class C scavenger receptors (*Sr-C*'s) between *Drosophila melanogaster* and *D. simulans* and between each of these species and *D. yakuba* is approximately four times the typical genome average. These genes also exhibit unusually high levels of segregating nonsynonymous polymorphism in *D. melanogaster* and *D. simulans* populations. A fourth *Sr-C* is comparatively conserved. McDonald-Kreitman tests reveal a significant excess of replacement fixations between *D. melanogaster* and *D. simulans* in the *Sr-C*'s, but tests of polymorphic site frequency spectra do not support models of directional selection. It is possible that the molecular functions of SR-C proteins are sufficiently robust to allow exceptionally high amino acid substitution rates without compromising organismal fitness. Alternatively, SR-Cs may evolve under diversifying selection, perhaps as a result of pressure from pathogens. Interestingly, *Sr-CIII* and *Sr-CIV* are polymorphic for premature stop codons. *Sr-CIV* is also polymorphic for an in-frame 101-codon deletion and for the absence of one intron.

SCAVENGER receptors (SRs) compose a protein family defined by gross structural similarities and participation in cellular internalization of foreign compounds. In contrast to the extreme specificity exhibited by most receptor-ligand systems, individual scavenger receptors have high affinity for a broad array of polyanionic ligands such as bacteria, apoptotic cell debris, and modified lipopolyprotein (reviewed in Krieger *et al.* 1993). *Drosophila melanogaster Sr-CI*, the one functionally characterized class C scavenger receptor (SR-C), shares with mammalian class A scavenger receptors (SR-As) the ability to bind compounds such as acetylated low-density lipopolyprotein, intact gram-negative and gram-positive bacteria, and bacterial molecular components (Abrams *et al.* 1992; Krieger and Herz 1994; Pearson *et al.* 1995; Krieger 1997; Gough and Gordon 2000). SR-CI and SR-As require interaction with other, unknown proteins at the membrane surface to internalize bound targets (Rämet *et al.* 2001; Underhill and Ozinsky 2002; Meister 2004). *Sr-A* mutant mice are susceptible to bacterial infection (Peiser *et al.* 2002) and *Sr-CI*-deficient *D. melanogaster* cells fail to phagocytose bacteria at wild-type levels (Rämet *et al.* 2001). Naturally occurring polymorphism in *D. melanogaster Sr-C*'s has been associated with phenotypic variation in the ability to suppress bacterial infection (Lazzaro *et al.* 2004). Due to the potential importance of *Sr-C*'s in pathogen recognition and host defense, understanding the evolutionary pressures experienced by these genes is of interest.

Three "predicted" gene homologs of *Sr-CI* (CG4099) can be recovered from the *Drosophila melanogaster* complete genome sequence (*Sr-CII*, CG8856; *Sr-CIII*, CG31962; *Sr-CIV*, CG3212). Because *Sr-CI* has been most thoroughly functionally characterized of the four, it will be considered the prototypical *Sr-C* throughout this article. Domain structures of the gene products from all four *Sr-C*'s are presented in Figure 1. The N terminus of the mature SR-CI protein is extracellular and is defined by two tandem complement-control protein (CCP) domains and one MAM domain. The CCP and MAM domains are sufficient for binding of bacteria (Rämet *et al.* 2001). A short spacer separates the MAM domain from a somatomedin B domain with unknown function and a long threonine-rich domain that extends the extracellular portion of the protein away from the membrane surface. The C terminus of SR-CI is cytoplasmic and may be heavily phosphorylated (Pearson *et al.* 1995). SR-CII is identical in domain structure to SR-CI, but SR-CIII and SR-CIV are putatively extracellularly secreted and lack Thr-rich, transmembrane and cytoplasmic domains. SR-CIII additionally lacks the somatomedin B domain (Figure 1). The precise functions of SR-Cs II, III, and IV have not been elucidated, but given their homology to

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. AY865019–AY865135.

[1]*Address for correspondence:* Department of Entomology, 4138 Comstock Hall, Cornell University, Ithaca, NY 14853.
E-mail: BL89@cornell.edu

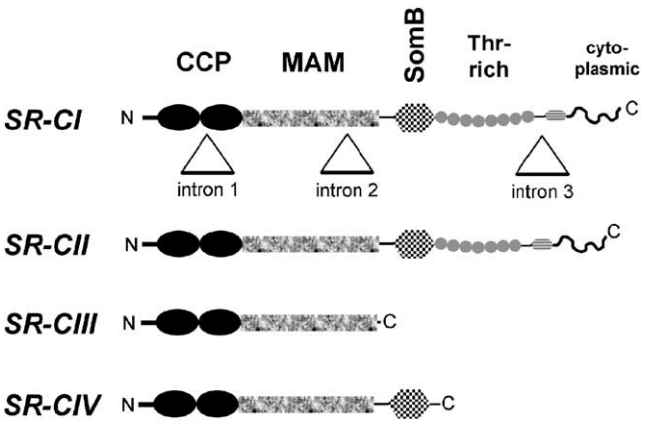

FIGURE 1.—Schematic of the four Drosophila *Sr-C* proteins. Protein functional domains are demarcated after PEARSON *et al.* (1995). Intron positions in the corresponding genes, indicated only for *Sr-CI*, are conserved across the four genes.

SR-CI and apparent contribution to antibacterial immunocompetence (LAZZARO *et al.* 2004), I consider that these proteins, like SR-CI, may be involved in the recognition of pathogens or pathogen-derived molecules.

Patterns of molecular variation have been used to infer evolutionary forces acting on many classes of proteins involved in the recognition of pathogens and foreign compounds. These data do not, however, paint a uniformly consistent picture. The archetypal recognition structure has been the vertebrate major histocompatibility complex (MHC), involved in the presentation of antigens to the immune system. In particular, the antigen-binding cleft of the MHC is extremely polymorphic. This variability expands the range of efficiency of antigen presentation by the MHC and is maintained by overdominant, diversifying natural selection (HUGHES and NEI 1988, 1999). Insects lack antibody-mediated immune responses, however, and are not known to have MHC-equivalent molecules. The best-characterized recognition proteins in invertebrates are pattern recognition receptors (PRRs) such as peptidoglycan recognition proteins (PGRPs) and gram-negative binding proteins (GNBPs). PRRs (also present in vertebrates) recognize microbial molecules such as peptidoglycan and β-glucan using functional domains that are highly conserved across distantly related species (*e.g.*, DZIARSKI 2004). PRR genes harbor very little intraspecific polymorphism (JIGGINS and HURST 2003; LITTLE *et al.* 2004; B. P. LAZZARO and A. G. CLARK, unpublished observations). The prevailing hypothesis is that the molecules bound by PRRs cannot tolerate extensive structural alteration, preventing microbes from evolving evasion and placing little pressure for adaptation on host PRRs. The PRR-based recognition system is aided by the fact that the target epitopes are unambiguously microbial, obviating the need for the host to make fine distinctions between self and non-self.

The expectation with respect to SR-C diversity levels is unclear. On one hand, SR-C diversity may be unnecessary if, like GNBPs and PGRPs, SR-Cs are responsive to conserved, easy-to-recognize epitopes such that potential pathogens cannot evolve evasion. Preliminary surveys, however, have suggested that patterns of *Sr-C* molecular diversity in wild North American populations of *D. melanogaster* (B. P. LAZZARO and A. G. CLARK, unpublished observations) and *D. simulans* (SCHLENKE and BEGUN 2003) are more complex than those of *PGRP*s and *GNBP*s. To more thoroughly characterize variability in Drosophila *Sr-C*'s, I have sampled alleles of all four *Sr-C* genes from both North American and African populations of *D. melanogaster* and *D. simulans*. Diversity levels are measured across gene regions encoding putative functional domains, and statistical tests are employed to detect evidence of potential adaptive evolution. The collected data are interpreted in the context of the current understanding of variability in invertebrate immune response genes specifically and the *D. simulans* and *D. melanogaster* genomes in general.

MATERIALS AND METHODS

Origin of the Drosophila lines studied: Alleles were sampled from African and North American *D. melanogaster* and *D. simulans* populations. The North American *D. melanogaster* (n = 12) are chromosome 2 extracted lines derived from flies collected in Pennsylvania in 1998. These lines have previously been used to survey variation in genes encoding secreted antibacterial peptides (LAZZARO and CLARK 2001, 2003). African *D. melanogaster* (n = 10) were collected in 1992 from the Sangwa wildlife refuge in Zimbabwe and have since been maintained as isofemale lines (BEGUN and AQUADRO 1993). These lines were provided by C. F. Aquadro. North American *D. simulans* (n = 8) were collected in northern California in 1995 and have previously been used to survey variation at a large number of immunity-related and immunity-independent loci (BEGUN and WHITLEY 2000a; SCHLENKE and BEGUN 2003). African *D. simulans* (n = 9) were collected in Harare, Zimbabwe, in 1994 by associates of C. F. Aquadro and have since been maintained as isofemale lines (C. F. AQUADRO, personal communication). The *Drosophila yakuba* strain sequenced was obtained from the Drosophila Species Stock Center in Tuscon, Arizona.

Generation and analysis of sequence data: *Sr-CIII* and *Sr-CI* are tandemly arranged head-to-tail at *D. melanogaster* cytological position 24D. A total of 3794 bp of this locus were surveyed, including the entire coding sequence of each gene (*Sr-CIII*, 960 bp; *Sr-CI*, 1908 bp), 161 bp 5′ of the SR-CIII start codon, 95 bp 3′ of the SR-CI stop codon, an ~300-bp intergenic spacer between the two genes, the two introns of *Sr-CIII* (118 bp), and the three introns of *Sr-CI* (240 bp). The *Sr-CII* survey region (cytological position 48F) begins at the start codon, includes the entire coding sequence (1804 bp) and three introns (350 bp), and terminates 6 bp 3′ of the stop codon. The *Sr-CIV* survey region (cytological position 23F) begins 208 bp 5′ of the start codon, includes the 1221-bp coding region and both introns (137 bp), and terminates 65 bp 3′ of the stop codon.

DNA was extracted from pools of ~30 flies from each line by a standard phenol:chloroform extraction followed by ethanol precipitation. Primers to amplify and sequence each of the four *Sr-C* gene regions from *D. melanogaster* and *D. simulans* were designed using the *D. melanogaster* complete genome sequence (ADAMS *et al.* 2000). Primers for amplification and

sequencing of *D. yakuba* were designed using data deposited by the *D. simulans*/*D. yakuba* genome sequencing consortium into the NCBI trace archive (http://www.ncbi.nlm.nih.gov/Traces). Primer sequences and PCR amplification conditions are available upon request. PCR products were directly sequenced using Beckman Coulter (Fullerton, CA) CEQ or ABI BigDye (Applied Biosystems, Foster City, CA) technologies under slight modification of the manufacturers' suggested protocols. All alleles were sequenced on both strands.

Initial sequencing of the African *D. melanogaster* lines revealed them to be highly heterozygous. For these lines only, DNA was re-extracted from a single fly representing each line and each locus was sequenced again. The presumption was that the individual flies were likely to be homozygous for one of the alleles segregating in each line and that choosing a random homozygous individual from a known heterozygous strain is analogous to randomly choosing a single allele from the population. The individual fly representing one line was heterozygous at *Sr-CII*, and the individuals from two other lines were heterozygous at the *Sr-CIII,I* locus. These lines were dropped from the analysis of the relevant loci (thus, $n = 9$ in *Sr-CII* and $n = 8$ in *Sr-CIII,I*). In *Sr-CIV*, one-half of the individual African *D. melanogaster* were heterozygous at *Sr-CIV*. For this locus, amplification products were cloned using a TOPO TA kit (Invitrogen, Carlsbad, CA) such that single alleles could be sequenced. At least two clones of each allele were sequenced, and then a single allele from each line was randomly chosen for inclusion in population genetic analyses. The same approach of cloning and sequencing was adopted for the *D. yakuba* strain, which also turned out to be heterozygous at *Sr-C* loci.

Partial coding regions of all four SR-C genes from the North American *D. simulans* lines considered here have been previously and independently sequenced by Schlenke and Begun (2003) and submitted to GenBank (accession nos. AY349846–AY349861, AY349878–AY349885, and AY349713–AY349720). The two sets of sequences are in nearly perfect agreement, although several ambiguous bases in the Schlenke and Begun sequences (scored as "N") were resolved here. At the infrequent positions where there are disagreements between bases obtained here and those reported by Schlenke and Begun, the base sequence obtained here was used in analysis. *D. yakuba* sequences obtained by Schlenke and Begun were not included in the analyses in this article.

Data analysis: Unless otherwise noted, molecular population genetic test statistics were calculated using DnaSP 3.51 and DnaSP 4.00 (Rozas and Rozas 1999; Rozas *et al.* 2003). Alleles with very large deletions were excluded from calculations of polymorphism and divergence to avoid eliminating too many informative sites. Calculation of statistics including all alleles and eliminating all sites with alignment gaps yielded comparable results (not shown). McDonald-Kreitman tests (McDonald and Kreitman 1991) were calculated using the 2×2 test of independence in DnaSP. McDonald-Kreitman tests were also run under a Poisson random field framework (Bustamante *et al.* 2002) on servers maintained by the Cornell Computational Biology Service Unit (Cornell CBSU) using default input parameters. K^*_{ST} (Hudson *et al.* 1992) and H (Fay and Wu 2000) were calculated using scripts written in C++. For calculation of H and for a subset of the McDonald-Kreitman tests, it is necessary to assign mutation events to either the *D. melanogaster* or the *D. simulans* lineages. These assignments were made assuming mutational parsimony and disallowing back mutation. Mutations that could not be parsimoniously placed on a single lineage (for instance, where *D. yakuba* and *D. melanogaster* are fixed for alternate states of a *D. simulans* polymorphism) were discarded from the relevant tests. Critical values of H were determined by simulation of 10,000 neutral genealogies after Hudson (1990) using Hudson's "mksamples" coalescence simulator (Hudson 2002), conditioning on the observed number of segregating sites and assuming no recombination. Critical values of K^*_{ST} were determined by randomly permuting subpopulation assignments of alleles and recalculating the test statistic 10,000 times.

RESULTS

Polymorphism and divergence: Nonsynonymous divergence is substantially elevated in *Sr-C*'s *I*, *III*, and *IV* between *D. melanogaster* and *D. simulans* and between *D. melanogaster* or *D. simulans* and *D. yakuba*. Synonymous nucleotide divergence is normal or slightly elevated across all species pairs in all genes (Table 1).

The CCP and MAM domains of SR-CI have previously been shown to be sufficient for binding of bacteria (Rämet *et al.* 2001). Nonsynonymous divergence in the sequence encoding these domains is approximately four times (CCP) and twice (MAM) the average level among the genomes of these three Drosophila species (Table 1; Takano 1998). The gene regions encoding the transmembrane domain and cytoplasmic tail of SR-CI show 2- to 3-fold excess nonsynonymous divergence. The Thr-rich domain, which is likely to be more mutable due to the repetitive nature of the sequence and less constrained by precise primary sequence than by tertiary structure, is encoded by sequence three to seven times more divergent at replacement positions than the genome averages. Most strikingly, the sequence of the functionally uncharacterized somatomedin B domain, which is perfectly conserved in length, has a nonsynonymous replacement rate equivalent to typical synonymous divergence rates in these species. The somatomedin B domain of *Sr-CI* shows a Jukes-Cantor corrected divergence of 8.5% in nonsynonymous positions between *D. melanogaster* and *D. simulans* and of 27.7% (26.4%) between *D. melanogaster* (*D. simulans*) and *D. yakuba*. Synonymous divergence between *D. melanogaster* (*D. simulans*) and *D. yakuba* is 33.3% (31.0%) in *Sr-CI*. By way of contrast, replacement divergence between *D. melanogaster* and *D. simulans* is typically on the order of 1.2% and between *D. melanogaster* (*D. simulans*) and *D. yakuba* is 2.5% (2.2%) (Takano 1998; see also Begun and Whitley 2000a, Begun 2002). Thus, nonsynonymous divergence in the somatomedin B domain of *Sr-CI* is increased ~10-fold relative to genome norms, with $K_a \approx K_s$, a condition attributable to either a complete relaxation of purifying selection or strong and consistent diversifying selection (Yang and Bielawski 2000).

Sr-CIII, which consists of only the CCP and MAM domains, displays a rate of nonsynonymous divergence similar to that of the homologous regions of *Sr-CI*. *Sr-IV*, which has CCP, MAM, and somatomedin B domains, also has an overall divergence rate similar to that of *Sr-CI*, although fewer of the substitutions are in the somatomedin B sequence and more are in the MAM (Table 1).

TABLE 1

Per-site nonsynonymous (synonymous) Jukes-Cantor corrected nucleotide divergence across SR-C domains

	CCP	MAM	Somatomedin B	Thr rich	Transmembrane/ cytoplasmic	Whole CDS	Intron, flank
No. of codons *Sr-CI*	107	185	34	133	93	624	(641 bp)
$K_{mel\text{-}sim}$	0.042 (0.124)	0.014 (0.102)	0.085 (0.176)	0.044 (0.140)	0.061 (0.161)	0.040 (0.125)	(0.153)
$K_{mel\text{-}yak}$	0.120 (0.265)	0.057 (0.383)	0.277 (0.256)	0.182 (0.374)	0.081 (0.340)	0.111 (0.333)	(0.350)
$K_{sim\text{-}yak}$	0.135 (0.287)	0.055 (0.330)	0.264 (0.316)	0.164 (0.345)	0.053 (0.323)	0.105 (0.310)	(0.321)
No. of codons *Sr-CII*	107	185	49	97	94	579	(238 bp)
$K_{mel\text{-}sim}$	0.004 (0.120)	0.004 (0.119)	0.009 (0.154)	0.041 (0.148)	0.012 (0.117)	0.017 (0.128)	(0.147)
$K_{mel\text{-}yak}$	0.015 (0.564)	0.023 (0.321)	0.044 (0.304)	0.138 (0.326)	0.048 (0.484)	0.057 (0.396)	(0.445)
$K_{sim\text{-}yak}$	0.018 (0.607)	0.023 (0.339)	0.035 (0.411)	0.127 (0.351)	0.043 (0.423)	0.054 (0.433)	(0.429)
No. of codons *Sr-CIII*	121	155				307	(641 bp)
$K_{mel\text{-}sim}$	0.058 (0.098)	0.017 (0.151)				0.045 (0.132)	(0.153)
$K_{mel\text{-}yak}$	0.137 (0.404)	0.064 (0.400)				0.115 (0.418)	(0.350)
$K_{sim\text{-}yak}$	0.105 (0.417)	0.069 (0.448)				0.105 (0.461)	(0.321)
No. of codons *Sr-CIV*	107	181	31			407	(383 bp)
$K_{mel\text{-}sim}$	0.025 (0.151)	0.048 (0.120)	0.053 (0.036)			0.049 (0.103)	(0.081)
$K_{mel\text{-}yak}$	0.062 (0.381)	0.123 (0.460)	0.166 (0.469)			0.129 (0.405)	(0.290)
$K_{sim\text{-}yak}$	0.066 (0.307)	0.120 (0.420)	0.135 (0.474)			0.126 (0.364)	(0.312)
Genome average							
$K_{mel\text{-}sim}$						0.013 (0.108)	
$K_{mel\text{-}yak}$						0.025 (0.233)	
$K_{sim\text{-}yak}$						0.022 (0.213)	

Sr-CII is far less divergent between species than are the other *Sr-C*'s and is generally in line with, although on the high end of, divergences observed in independent genes across these three species (Table 1; Takano 1998). The sequence encoding the Thr-rich domain of SR-CII displays nonsynonymous divergence equivalent to that of the corresponding sequence of *Sr-CI*, but this may reflect a relaxation of constraint on primary amino acid sequence of the Thr-rich domain (provided overall polarity is retained) or an increase in mutation rate due to the repetitive nature of the nucleotide sequence. In stark contrast to *Sr-CI*, the *Sr-CII* gene regions encoding CCP, MAM, and somatomedin B domains are highly conserved across species, actually diverging less than the genome average at nonsynonymous positions between *D. melanogaster* and *D. simulans* or between either species and *D. yakuba* (Table 1).

Paralleling the interspecific divergence data, intraspecific nonsynonymous polymorphism is also elevated in the *Sr-C* genes (Table 2). Replacement substitutions typically constitute 20–30% of the total number of polymorphic sites in *D. melanogaster* (Moriyama and Powell 1996; Andolfatto 2001; Fay *et al.* 2002; Mousset and Derome 2004) and 8–15% of polymorphic sites in *D. simulans* (Moriyama and Powell 1996; Begun and Whitley 2000a; Andolfatto 2001; Mousset and Derome 2004). The proportion of polymorphic sites in *Sr-C*'s predicted to change amino acid sequence ranges from 37% (*Sr-CII*) to 63% (*Sr-CIV*) in *D. melanogaster* and 20% (*Sr-CII*) to 53% (*Sr-CIV*) in *D. simulans* (Table 3). The probability that a certain proportion of polymorphic sites are nonsynonymous can be treated as a binomial sampling process with Prob(success) equal to the mean proportion of polymorphisms that are nonsynonymous genome wide. Using this test, the excess of nonsynonymous relative to synonymous polymorphisms between individual *Sr-C*'s and the relevant genome average is highly significant at *Sr-C*'s *I, III,* and *IV* ($P < 10^{-4}$), but not at *Sr-CII* ($P = 0.06$ in *D. melanogaster*; $P = 0.04$ in *D. simulans*).

The observed increases in both nonsynonymous polymorphism and divergence could conceivably result either from positive selection acting to diversify Drosophila *Sr-C*'s or from a relaxation of purifying selection on these genes relative to genome norms. An elevated mutation rate cannot explain the patterns observed in

TABLE 2

Summary statistics describing polymorphism Drosophila *Sr-C* genes

	n	bp	S_{non}	S_{syn}	π_{non}	π_{syn}	π_{noncod} (bp)
Sr-CI							
D. melanogaster, Africa	8	1875	24	23	0.0062	0.0201	0.0183 (859)
D. melanogaster, North America	12	1890	21	33	0.0057	0.0256	0.0111 (918)
D. simulans, Africa	9	1851	28	44	0.0050	0.0157	0.0251 (962)
D. simulans, North America	8	1809	13	13	0.0054	0.0318	0.0174 (971)
Sr-CIII							
D. melanogaster, Africa	8	927	23	15	0.0103	0.0292	See above
D. melanogaster, North America	12	954	17	20	0.0057	0.0355	
D. simulans, Africa	9	960	14	16	0.0084	0.0257	
D. simulans, North America	8	960	7	21	0.0050	0.0553	
Sr-CII							
D. melanogaster, Africa	9	1790	17	30	0.0036	0.0288	0.0287 (273)
D. melanogaster, North America	12	1800	10	11	0.0036	0.0128	0.0108 (285)
D. simulans, Africa	9	1746	15	56	0.0041	0.0564	0.0626 (290)
D. simulans, North America	8	1797	9	41	0.0019	0.0340	0.0398 (319)
Sr-CIV							
D. melanogaster, Africa	10	1220	20	19	0.0100	0.0287	0.0124 (356)
D. melanogaster, North America	12	1199	52	29	0.0185	0.0350	0.0168 (302)
D. simulans, Africa	9	1221	36	30	0.0151	0.0464	0.0142 (411)
D. simulans, North America	8	1221	15	18	0.0064	0.0257	0.0274 (407)

n, sample size at each locus; bp, size of the gene in base pairs excluding alignment gaps; S_{non}, the number of nonsynonymous segregating sites; S_{syn}, the number of synonymous segregating sites; π_{non}, estimate of nonsynonymous population heterozygosity; and π_{syn}, estimate of synonymous population heterozygosity. π_{noncod} is an estimate of heterozygosity in introns and 5′ and 3′ flanks, with the length in base pairs considered in parentheses. Noncoding base pairs are pooled across the tandem genes *Sr-CI* and *Sr-CIII*.

these genes because silent divergence (Table 1) and polymorphism (Table 2) are not substantially increased. The McDonald-Kreitman (MK) test can reveal heterogeneity between the ratio of synonymous to replacement polymorphisms within species compared to the ratio of synonymous to replacement fixations between species (McDonald and Kreitman 1991). Application of the MK test reveals a marginally significant excess of nonsynonymous fixations between *D. melanogaster* and *D. simulans* in *Sr-CI* ($G = 4.446$, $P = 0.035$) and nearly significant excesses of nonsynonymous fixations at *Sr-CII* and *Sr-CIII* ($G = 3.198$, $P = 0.074$ and $G = 3.519$, $P = 0.061$, respectively; Table 3). Exclusion of the sequence encoding the hypothetically unconstrained (and highly polymorphic) Thr-rich domains from analyses of *Sr-CI* and *Sr-CII* reveals a more substantial excess of nonsynonymous fixations in the remaining domains of these genes (in *Sr-CI*, $G = 8.62$ and $P = 0.003$; in *Sr-CII*, $G = 7.82$ and $P = 0.005$). The MK test statistic is not significant in *Sr-CIV* due to the extreme number of replacement polymorphisms (the 406 codon gene harbors 81 nonsynonymous polymorphisms across both *D. melanogaster* and *D. simulans*). When data from the four genes are pooled, the MK test indicates a highly significant excess of nonsynonymous fixations between species ($G = 9.364$, $P = 0.002$), consistent with positive selection driving the diversification of these genes.

Interpretation is slightly clouded, however, when sequence from *D. yakuba* is used to localize mutation events to either the *D. melanogaster* or the *D. simulans* lineage. When mutations are polarized in this way (Table 3), *D. simulans* shows a significant excess of nonsynonymous fixations at *Sr-CI* ($G = 5.751$, $P = 0.016$), a nearly significant excess in *Sr-CIV* ($G = 3.163$, $P = 0.057$), and a significant excess across all four genes combined ($G = 5.825$, $P = 0.016$). The effect is not seen in *Sr-CII* and *Sr-CIII* ($G = 0.236$, $P = 0.627$ and $G = 0.259$, $P = 0.611$, respectively). Again, exclusion of the sequence encoding the Thr-rich domain results in a more profound pattern in *Sr-CI* ($G = 11.27$, $P = 0.0008$). No significant MK tests are observed on the *D. melanogaster* lineage due to the much higher proportional level of nonsynonymous polymorphism in that species (Table 3), although *Sr-CIII* showed a tendency toward excess nonsynonymous fixation ($G = 2.983$, $P = 0.084$). When substitutions along the *D. melanogaster* lineage in all four genes are considered jointly, $G = 1.928$ ($P = 0.165$).

A similar finding of significant MK tests between *D. melanogaster* and *D. simulans* and along the *D. simulans*

TABLE 3

McDonald-Kreitman comparisons of polymorphism and divergence

	Polymorphic		Fixed			
	Synonymous	Nonsynonymous	Synonymous	Nonsynonymous	*G*	*P*-value
			Between *D. melanogaster* and *D. simulans*			
Sr-CI	83	62	27	38	4.446	0.035[a]
Sr-CII	95	36	16	13	3.198	0.074[a]
Sr-CIII	51	50	9	20	3.519	0.061
Sr-CIV	51	74	8	14	1.235	0.267
All genes	280	222	60	85	9.364	0.002
			Polarized to the *D. melanogaster* lineage			
Sr-CI	31	32	16	20	0.209	0.648
Sr-CII	22	20	9	11	0.286	0.597
Sr-CIII	17	21	3	12	2.983	0.084
Sr-CIV	32	44	7	14	0.535	0.464
All genes	102	117	35	57	1.928	0.165
			Polarized to the *D. simulans* lineage			
Sr-CI	38	26	5	13	5.751	0.016[a]
Sr-CII	50	16	9	4	0.236	0.627
Sr-CIII	24	19	4	2	0.259	0.611
Sr-CIV	26	36	2	11	3.163	0.057
All genes	138	97	20	30	5.825	0.016

[a] Exclusion of the gene sequence encoding the Thr-rich domain results in highly significant MK test statistics. See text for details.

lineage, but not along the *D. melanogaster* lineage, was previously observed in the immune-inducive transcription factor *Relish* (Begun and Whitley 2000b). In *Relish*, failure to reject homogeneity resulted from a reduction in synonymous polymorphism on the *D. melanogaster* lineage, such that levels of synonymous and replacement heterozygosity were comparable. Because the number of fixed replacements was elevated on both lineages and similar between species, Begun and Whitley (2000b) concluded that positive selection probably acts on *Relish* in both species. As in *Relish*, replacement fixations in *Sr-C*'s are elevated on both the *D. melanogaster* and the *D. simulans* lineages relative to genome norms.

To test whether amino acid replacements in SR-Cs may be adaptive, the average estimate of $\gamma = 4N_e s$ was determined for nonsynonymous substitutions in each of the four *Sr-C* genes under a Poisson random field framework (Bustamante *et al.* 2002). γ was estimated at 1.28 (SE = 0.65) in *Sr-CI* and 0.73 (SE = 0.48) in *Sr-CIII*, indicating that the average replacement fixation is moderately selectively favored. These values are typical for Drosophila genes (Bustamante *et al.* 2002). In *Sr-CIV*, γ was estimated to be -0.07 (SE = 0.33) suggesting that replacement substitutions in this gene are effectively neutral, an observation unusual for functional Drosophila genes (Bustamante *et al.* 2002). Nonsynonymous fixations in *Sr-CII* seem to have been more strongly favored, with γ estimated at 3.27 (SE = 0.99). Only 4 of 34 Drosophila genes examined by Bustamante *et al.* (2002) have estimated γ as large as or larger than that estimated from *Sr-CII*.

Analyses of site frequency spectra, population structure, and linkage disequilibrium: Patterns in the distribution of allele frequencies at polymorphic sites are frequently used to infer departure from null models of selective neutrality. Two such tests are Tajima's *D* (Tajima 1989) and Fay and Wu's *H* (Fay and Wu 2000). Negative values of *D* result from an excessive proportion of polymorphic sites at which the rarer allele is at very low frequency in the population and may be attributed to mutational recovery after a strong directional selective event. Positive values of *D* occur when a high proportion of the polymorphic sites have both states at intermediate frequencies, as under some scenarios of balancing selection. *H* is analogous to *D*, except that it tests for an excess of sites at which the derived state is common, which can also result from positive directional selection. Application of the *H* test to the SR-C data yielded no significant results in any of the four genes in either species (data not shown). Similarly, no significantly negative values of *D* were observed (Table 4). There is thus no indication from the site frequency data that these

TABLE 4

Summary statistics describing Drosophila *Sr-C* gene regions

	Tajima's *D*	4*Nc* (per base)
Sr-CIII,I		
D. melanogaster, Africa	−0.256	0.0184
D. melanogaster, North America	+0.052	0.0051
D. simulans, Africa	−0.490	0.0264
D. simulans, North America	+2.080**	0.0000
Sr-CII		
D. melanogaster, Africa	+0.280	0.0297
D. melanogaster, North America	+1.869*	0.0046
D. simulans, Africa	+0.216	0.0265
D. simulans, North America	−0.894	0.0000
Sr-CIV		
D. melanogaster, Africa[a]	+1.040/0.373	0.0091/0.0131
D. melanogaster, North America	+0.003	0.0378
D. simulans, Africa	+0.188	0.0556
D. simulans, North America	+0.349	0.0000

**$P = 0.003$, Bonferroni $P < 0.05$; *$P = 0.017$.

[a] Statistics in African *D. melanogaster Sr-CIV* are calculated first excluding alleles with large deletions and then including all alleles but excluding sites with alignment gaps.

genes have experienced recent directional selection. In two cases *D* is significantly positive (Table 4). In North American *D. melanogaster SR-CII, D* = +1.87 ($P = 0.017$), although this value is not statistically significant after correction for multiple tests. North American *D. simulans* have $D = +2.08$ ($P = 0.003$) in the *Sr-CIII,I* cluster, a value that is significantly positive at $\alpha < 0.05$ even after Bonferroni correction.

In none of the loci is there a significant difference between the frequency spectrum of synonymous and nonsynonymous polymorphisms. Values of *D* are slightly smaller for nonsynonymous substitutions than for synonymous polymorphisms in 3 of the 16 comparisons across the four genes and populations (the exceptions being North American *D. melanogaster Sr-CI* and *Sr-CII* and African *D. melanogaster Sr-CIV*). This pattern is significantly unexpected ($P = 0.011$) if comparisons are considered Bernoulli trials where either synonymous or nonsynonymous polymorphisms are equally likely to exhibit smaller *D*. The significant tendency for nonsynonymous polymorphisms to segregate at lower population frequencies than synonymous polymorphisms is consistent with a general action of purifying selection on amino acid mutation.

The North American and African populations of both species are highly significantly differentiated at all loci as measured by K^*_{ST} (Table 5; Hudson *et al.* 1992). In genes encoding membrane-bound SR-Cs, variability in North American populations is a subset of the variation present in African populations, with most mutations shared across populations or private to African samples. Genome-wide genetic differentiation of this sort has previously been observed between African and North American Drosophila populations in antibacterial peptide genes (Clark and Wang 1997) and several immunity-independent loci (*e.g.*, Andolfatto 2001; Caracristi and Schlötterer 2003). The pattern is generally attributed to Africa being the ancestral home of the *D. melanogaster* and *D. simulans* species, with subsequent founding events leading to the colonization of North America (David and Capy 1988; Hamblin and Veuille 1999; Andolfatto 2001; Wall *et al.* 2002). The genes encoding secreted SR-Cs, *Sr-CIII* and *Sr-CIV*, are unusual in that up to 30% of the segregating polymorphisms are private to the North American samples (not shown). This is unexpected under a simple founding model and may be consistent with adaptation to a new environment.

Both *D. melanogaster* and *D. simulans* also exhibit increased linkage disequilibrium in North American populations relative to African. In *D. melanogaster*, per-site estimates of 4*Nc* (Hudson 1987) are approximately four- and sixfold reduced in North America relative to Africa in the *Sr–CIII,I* cluster and *Sr-CII*, respectively

TABLE 5

Genetic differentiation between African and North American subpopulations

	D. melanogaster K^*_{ST} (*P*-value)	*D. simulans* K^*_{ST} (*P*-value)
Sr-CIII,I	0.059 (0.002)	0.132 (0.001)
Sr-CII	0.012 (<0.001)	0.019 (0.002)
Sr-CIV[a]	0.105 (0.003)/0.050 (0.010)	0.057 (0.008)

[a] Statistics in African *D. melanogaster Sr-CIV* are calculated first excluding alleles with large deletions and then including all alleles but excluding sites with alignment gaps.

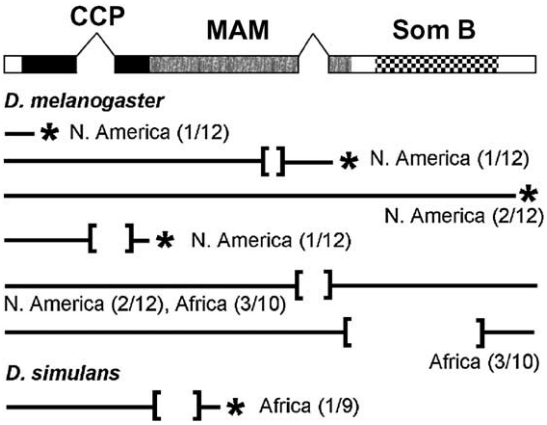

FIGURE 2.—Positions and population frequencies of polymorphic large deletions and premature stop codons in *Sr-CIV*. Asterisks indicate premature stop codons, open brackets mark the boundaries of deletions. Although the intron 2 absence deletion is present in only 2 of the 12 North American *D. melanogaster* alleles sequenced, a genotyping survey of 101 lines suggested the deletion has a population frequency of 33% (LAZZARO *et al.* 2004).

(Table 4). Increased linkage disequilibrium on this order has previously been observed in non-African relative to African populations in other, functionally unrelated, loci around the genome (*e.g.*, ANDOLFATTO and PRZEWORSKI 2000; ANDOLFATTO and WALL 2003). There is no apparent reduction in $4Nc$ at *D. melanogaster Sr-CIV*. In North American *D. simulans*, however, the situation is extreme and highly unusual. All of the segregating sites in all four genes (88 in the *Sr-CIII,I* cluster, 89 in *Sr-CII*, and 46 in *Sr-CI*) are blocked into only two or three major haplotypes at each locus. With such strong linkage disequilibrium, estimated $4Nc = 0$. Such a complete arrangement of so-called "yin-yang" haplotypes involving a large number of sites is unheard of in either *D. melanogaster* or *D. simulans* (BEGUN and WHITLEY 2000a; ZHANG *et al.* 2003; but see ROZAS *et al.* 2001; QUESADA *et al.* 2003). This observation is explored in more detail in the accompanying article by SCHLENKE and BEGUN (2005, this issue).

Large deletions, premature termination, and evolutionary modification of start and stop codons: Perhaps the most striking aspect of the polymorphism observed in the Drosophila *Sr-C* genes is the presence of large deletions segregating in *Sr-CIV* (Figure 2). One of these is an in-frame deletion eliminating 101 codons, including those encoding 50 amino acids of the MAM domain and 26 amino acids of the somatomedin B domain. This deletion is present in 3 of the 10 African *D. melanogaster* alleles considered here, but is absent from the North American *D. melanogaster* alleles sampled, and was not detected in a previous screen of 101 North American *D. melanogaster* alleles of *Sr-CIV* (LAZZARO *et al.* 2004). A second in-frame *D. melanogaster* deletion is an imprecise excision of the second *Sr-CIV* intron. This deletion begins seven bases into the intron and extends one base into the third exon, resulting in a net insertion of 2 amino acids into the MAM domain. Three of the 10 African *D. melanogaster* lines and 2 of the 12 North American *D. melanogaster* alleles carry the deletion, although genotyping data from the larger North American sample places its frequency at 32.7% (LAZZARO *et al.* 2004).

Three frameshift deletions and two point mutations result in premature stop codons in *Sr-CIV* (Figure 2). Two of the deletions are found in the North American *D. melanogaster* sample, and one is found in the African *D. simulans*; all are singletons. The *D. simulans* deletion eliminates 169 bases of exon 2, resulting in a stop codon after 90 codons. One of the *D. melanogaster* deletions eliminates most of the first intron and 16 bases of the second exon, resulting in a premature termination after 85 codons. The other *D. melanogaster* stop is 4 bp in length and terminates the predicted protein after 252 amino acids. Three additional North American *D. melanogaster* alleles carry premature stops caused by point mutations, one harboring a unique stop 30 codons into the protein and the other two sharing a termination codon after 385 codons. The predicted wild-type protein length is 406 amino acids, so the latter pair of alleles could potentially retain functionality. On the other hand, the preponderance of large deletions and segregating premature stop codons raises the possibility that *Sr-CIV* may not be a functional gene at all, but instead may be a young pseudogene. This possibility is further examined in DISCUSSION.

Polymorphic premature stop codons are not unique to *Sr-CIV* among the Drosophila *Sr-C* genes. Two *D. melanogaster* alleles of *Sr-CIII*, encoding the other putatively secreted SR-C, also harbor singleton point mutations resulting in truncation of the predicted protein sequence. One of these is a North American allele and ends the open reading frame after 10 codons. The other is an African allele that terminates after 22 codons. Interestingly, the *Sr-CIII* start codon varies among *D. melanogaster*, *D. simulans*, and *D. yakuba*. The predicted *D. melanogaster* protein sequence begins with the three amino acids Met-Ala-Met. It is not experimentally known which of the two methionine codons is actually the primary start in *D. melanogaster*, but the first methionine is absent from *D. yakuba* and the second methionine is mutated to leucine in *D. simulans*. Without sequence from an outgroup species, it is not possible to determine which *Sr-CIII* start codon is ancestral.

One African *D. melanogaster* allele of *Sr-CII* carries a four-base deletion in the last two codons of the gene. A stop codon in the new reading frame terminates translation, altering the C terminus of the protein from DL* to ERG*. There are no premature stop codons in *Sr-CII*, and there is no variability in start or stop in *Sr-CI*.

DISCUSSION

The class C scavenger receptors compose a four-member gene family in *D. melanogaster*, *D. simulans*, and *D.*

yakuba (Figure 1). *D. melanogaster* SR-CI directly binds bacteria to facilitate phagocytosis (Rämet *et al.* 2001) and is involved in endocytosis of lipopolyprotein (Abrams *et al.* 1992; Pearson *et al.* 1995). Detailed functional characterization of the other three genes is lacking, although naturally occurring polymorphism in all four *D. melanogaster Sr-C* genes has been implicated as contributing to phenotypic variation in the suppression of bacterial infection (Lazzaro *et al.* 2004). It seems reasonable to hypothesize that these genes may be important for the recognition of bacteria during an immune response, raising the question of what evolutionary pressures they face. Population genetic analysis of this *Sr-C* gene family in two populations each of *D. melanogaster* and *D. simulans* reveals them to be on different evolutionary trajectories.

SR-CI and SR-CIII: *Sr-CI* shows rapid nonsynonymous divergence in the gene regions encoding the CCP, MAM, somatomedin B, and transmembrane/cytoplasmic domains (Table 1), all of which are conserved in length and unambiguously alignable across species. In particular, the rate of nonsynonymous divergence in the sequence encoding the *Sr-CI* somatomedin B domain is approximately equivalent to the rate of synonymous divergence ($K_a \approx K_s$), a condition attributable to either positively selected diversification or a complete relaxation of functional constraint. The gene encoding the putatively secreted SR-CIII, which consists of only CCP and MAM domains, exhibits divergence rates similar to those of the homologous regions in *Sr-CI*.

McDonald-Kreitman tests reveal a significant excess of nonsynonymous fixations in *Sr-CI* between *D. melanogaster* and *D. simulans* ($G = 4.446$, $P = 0.035$; $G = 8.62$, $P = 0.003$ excluding the Thr-rich domain) and along the *D. simulans* lineage ($G = 5.751$, $P = 0.016$; $G = 11.27$, $P = 0.0008$ excluding the Thr-rich domain). The excess of replacement fixations in *Sr-CIII* is nearly significant between species ($G = 2.983$, $P = 0.084$). When considered jointly, these two genes show a highly significant excess of nonsynonymous fixations between *D. melanogaster* and *D. simulans* ($G = 7.167$, $P = 0.007$) consistent with adaptively driven divergence in these two genes, particularly in *Sr-CI*. Nonsynonymous fixations are, on average, estimated to be selectively favored in *Sr-CI* and *Sr-CIII* ($\gamma = 1.27$ in *Sr-CI* and 0.73 in *Sr-CIII*), although these values are unexceptional for Drosophila (Bustamante *et al.* 2002).

If the elevated nonsynonymous divergence observed in these genes were driven by pathogen diversity, intuition would suggest that substitutions would be concentrated in regions of the protein involved in binding. Contrary to that intuition, high amino acid divergence is also suggested in the SR-CI somatomedin B and transmembrane domains, neither of which is known to have direct interactions with bacteria. Interestingly, a similar substitution pattern is observed in mouse SR-A, where a small number of strain-specific nucleotide differences result in a high proportion of amino acid polymorphisms localized in regions of the protein flanking the interior and exterior of the membrane surface (Daugherty *et al.* 2000; Fortin *et al.* 2000). Although SR-CI and SR-A share ligand affinity and may have similar three-dimensional structures, they are unrelated in primary amino acid sequence and their functional similarities have arisen through evolutionary convergence. Amino-carboxy orientation is reversed in SR-A relative to SR-CI and the domain structure is completely distinct (Peiser *et al.* 2002). Given that SR-As and SR-Cs are structural analogs, not homologs, the convergence in substitution patterns may indicate similarity in evolutionary pressures experienced by those genes.

Why might the highest rates of substitution in both SR-CI and SR-A be observed in regions of the protein presumably involved not in pathogen binding but in protein-protein interactions? It is well known that pathogenic bacteria can actively inhibit host immune responses through a variety of mechanisms (*e.g.*, Lindmark *et al.* 2001; Fauvarque *et al.* 2002). One could speculate that pathogens might also seek to evade engulfment by disrupting interactions between scavenger receptors and other proteins required for phagocytotic internalization, and then SR domains outside those responsible for direct bacterial binding may experience pressure driving evolutionary diversification. Begun and Whitley (2000b) put forward a similar hypothesis to explain the rapid evolution of the immune-related transcription factor *Relish*. The sequence encoding the Relish domain that is cleaved to activate the transcription factor shows extraordinarily high levels of nonsynonymous substitution, which Begun and Whitley (2000b) proposed is driven by the secretion into host cells of pathogen repressor molecules designed to prevent Relish activation. This hypothesis was later bolstered by the observation that Dredd, a caspase involved in the activation of Relish, shows correlated rapid evolution (Schlenke and Begun 2003).

SR-CII: In contrast to *Sr-C*'s *I* and *III*, *Sr-CII* shows a high degree of conservation among *D. melanogaster*, *D. simulans*, and *D. yakuba* (Table 1). The gene regions encoding the MAM and CCP domains show higher than expected conservation in comparisons of *D. melanogaster/D. simulans* to *D. yakuba*, although divergence is slightly elevated in sequences encoding the somatomedin B and transmembrane domains (Table 1). The gene region encoding the threonine-rich domain exhibits high levels of replacement polymorphism and divergence in *Sr-CII*, likely due to the repetitive nature and presumed low functional constraint on the primary sequence of this domain, which is also variable in length between *D. yakuba* and *D. melanogaster/D. simulans*. It is conservative to imagine that many amino acid substitutions in this domain are functionally neutral. A McDonald-Kreitman test reveals a nearly significant excess of replacement fixations between *D. melanogaster* and *D. simulans* ($G = 3.198$, $P = 0.074$; $G = 7.82$, $P = 0.005$ when the Thr-rich domain is excluded), consistent with

the action of positive selection on this gene, although it is clear that *Sr-CII* does not evolve under strong diversifying selection. The mean $\gamma = N_e s$ estimated for nonsynonymous substitutions in *Sr-CII* is 3.27, indicating adaptive favorability of the nonsynonymous fixations observed in this locus. This value is in the top 10% of estimates collected from a panel of unrelated Drosophila genes (Bustamante *et al.* 2002).

Published data suggest that *Sr-CII* is expressed only in early embryos (Rämet *et al.* 2001), although low-level tissue-specific expression remains a possibility. This observation, coupled with the molecular evolutionary data, suggests that SR-CII may be functionally distinct from the other SR-Cs. As of June 2004, a BLAST search of the unassembled *D. pseudoobscura* genome sequence database at the Baylor College of Medicine (http://www.hgsc.bcm.tmc.edu/blast/?organism=Dpseudoobscura) yielded a high-quality match only to the comparatively conserved *Sr-CII*, but no clear matches to *Sr-CI*, *Sr-CIII*, or *Sr-CIV* (B. P. Lazzaro, unpublished observations).

SR-CIV: The data from *Sr-CIV* are perplexing. Five of 12 North American *D. melanogaster* alleles sampled carry premature stop codons, and 3 of 10 African *D. melanogaster* alleles carry an in-frame 101-codon deletion. One of the 18 *D. simulans Sr-CIV* alleles sampled carries a premature stop codon. (Premature stops are not exclusive to *Sr-CIV*, as two alleles of *Sr-CIII*, the other secreted SR-C, are also predicted to terminate early in the protein.) The prevalence of stop codons and an apparently disruptive deletion in *Sr-CIV* raises the possibility that *Sr-CIV* may be a young pseudogene, at least in *D. melanogaster*, resulting in an absolute relaxation of constraint. If *Sr-CIV* is a pseudogene, though, it is an extremely young one, and may even be polymorphic for activity in *D. melanogaster*. Divergence among species is not substantially higher in *Sr-CIV* than in the other *Sr-C* genes, and there is no evidence for an acceleration of nonsynonymous divergence on the *D. melanogaster* lineage. Synonymous polymorphism and divergence far exceed nonsynonymous polymorphism and divergence in *Sr-CIV*, suggesting at least an historical functional constraint on mutations. McDonald-Kreitman tests yield no evidence of adaptive diversification in *Sr-CIV* (Table 3) and the estimate of $\gamma\ (=N_e s) = -0.07$ on nonsynonymous fixations is completely consistent with neutrality. K_a/K_s is 0.5 between *D. melanogaster* and *D. simulans* and ~0.4 between either of these species and *D. yakuba*. On balance, the data suggest that functional constraint on *Sr-CIV* is greatly relaxed, although probably recently so. The additional presence of null alleles in *Sr-CIII* suggests that secreted SR-Cs may be dispensable with minimal effect on organismal fitness. Very young immune-related pseudogenes have previously been observed in *Cecropin* gene family in the *D. melanogaster* species subgroup, where two apparent pseudogenes in *D. melanogaster* continue to be transcribed despite exhibiting polymorphism for nonsense mutations (Ramos-Onsins and Aguadé 1998). A naturally occurring null allele has also been recovered in the gene encoding the antibacterial peptide *Attacin A* (Lazzaro and Clark 2001).

Sr-CIV is also exceptional in being polymorphic for the presence/absence of an intron, with the intron-absent state segregating at ~40% frequency in North American and African *D. melanogaster*. To date, only one other intermediate-frequency intron presence-absence polymorphism has been described from any eukaryote, in the *jingwei* gene of *D. teissieri* (Llopart *et al.* 2002). As in *jingwei*, the *Sr-CIV* polymorphism derives from an imprecise genomic deletion that eliminates the intron but retains reading frame.

Linkage disequilibrium in North American *D. simulans*: The North American *D. simulans* lines are clear outliers with respect to the other lines in terms of linkage disequilibrium. The 88 polymorphic sites in the *Sr-CIII,I* locus segregate in only two major haplotypes (with two additional haplotypes separated by three mutations), the 89 sites in *Sr-CII* form two haplotypes, and the 46 sites in *Sr-CIV* form three major haplotypes (with a fourth distinguished by two mutations). The failure to find such strong haplotype structure outside the *Sr-C*'s and two other immunity-related genes (Schlenke and Begun 2005, this issue) even though 56 additional genes have been surveyed in these same lines (Begun and Whitley 2000a; Schlenke and Begun 2003) largely precludes any hypotheses based on demography or admixture. Chromosomal inversions are an unlikely and unparsimonious explanation, since the same phenomenon is observed in *Sr-C*'s on both arms of chromosome 2 and disequilibrium is not absolute between the physically proximal *Sr-CIII,I* locus at (*D. melanogaster*) cytological position 24D and *Sr-CIV* at 23F. *Sr-CII* has probably been affected by a strong selective sweep at a genetically linked locus conferring insecticide resistance (Schlenke and Begun 2004), but such strong linked sweeps are expected to be rare. Furthermore, Schlenke and Begun report that disequilibrium decays in regions flanking the *Sr-C* loci and that the haplotype structure observed in the *D. simulans* lines described here is not present in other populations sampled from North America (Schlenke and Begun 2005, this issue). Significant, although less severe, haplotype structure previously observed in European *D. simulans* has been attributed to the actions of positive selection (Rozas *et al.* 2001; Quesada *et al.* 2003).

Conclusions: Drosophila class C scavenger receptors are a multigene family exhibiting different evolutionary trajectories. *Sr-C*'s *I*, *III*, and *IV* are rapidly evolving. In the case of *Sr-CI* and, possibly, in *Sr-CIII*, rapid evolution seems to be driven by positive selection in both *D. melanogaster* and *D. simulans*. *D. melanogaster Sr-CIV*, however, seems to be evolving under a lack of functional constraint and may be a young (or even polymorphic) pseudogene. In contrast to the other *Sr-C*'s, *Sr-CII* is evolutionarily conserved, although it too displays indications

of adaptive evolution. It will be interesting to couple future functional studies of these genes with the molecular evolutionary observations. It will be of particular interest to determine the degree to which the extensive nonsynonymous polymorphism exhibited by Drosophila *Sr-C*'s influences protein function and whether there is any detriment to animals carrying apparent null mutations in the secreted *Sr-C*'s. Notably, *Sr-CI* may be the first Drosophila pathogen recognition gene characterized as evolving under adaptive diversification.

I extend special thanks to Todd Schlenke for open discussion of unpublished results and sharing of *D. simulans* lines. I also thank Chip Aquadro for flies and Brian Bettencourt for assistance in searching the unassembled *D. pseudoobscura* genome sequence for SR-C homologs. Early stages of this work were supported by National Institutes of Health grant AI46402 to A. G. Clark. This research was facilitated by prompt public release of data by the *D. simulans* and *D. yakuba* genome sequencing consortium. Finally, I acknowledge the Institute for *Drosophila* Immunomics in Ithaca, New York.

LITERATURE CITED

Abrams, J. M., A. Lux, H. Steller and M. Krieger, 1992 Macrophages in *Drosophila* embryos and L2 cells exhibit scavenger receptor-mediated endocytosis. Proc. Natl. Acad. Sci. USA **89:** 10375–10379.

Adams, M. D., S. E. Celniker, R. A. Holt, C. A. Evans, J. D. Gocayne *et al.*, 2000 The genome sequence of *Drosophila melanogaster*. Science **287:** 2185–2195.

Andolfatto, P., 2001 Contrasting patterns of X-linked and autosomal nucleotide variation in *Drosophila melanogaster* and *Drosophila simulans*. Mol. Biol. Evol. **18:** 279–290.

Andolfatto, P., and M. Przeworski, 2000 A genome-wide departure from the standard neutral model in natural populations of Drosophila. Genetics **165:** 257–268.

Andolfatto, P., and J. D. Wall, 2003 Linkage disequilibrium patterns across a recombination gradient in African *Drosophila melanogaster*. Genetics **165:** 1289–1305.

Begun, D. J., 2002 Protein variation in *Drosophila simulans*, and comparison of genes from centromeric versus noncentromeric regions of chromosome 3. Mol. Biol. Evol. **19:** 201–203.

Begun, D. J., and C. F. Aquadro, 1993 African and North American populations of *Drosophila melanogaster* are very different at the DNA level. Nature **365:** 548–550.

Begun, D. J., and P. Whitley, 2000a Reduced X-linked nucleotide polymorphism in *Drosophila simulans*. Proc. Natl. Acad. Sci. USA **97:** 5960–5965.

Begun, D. J., and P. Whitley, 2000b Adaptive evolution of relish, a Drosophila NF-kappaB/IkappaB protein. Genetics **154:** 1231–1238.

Bustamante, C. D., R. Nielson, S. A. Sawyer, K. M. Olsen, M. D. Purugganan *et al.*, 2002 The cost of inbreeding in *Arabidopsis*. Nature **416:** 531–534.

Caracristi, G., and C. Schlötterer, 2003 Genetic differentiation between American and European *D. melanogaster* populations could be attributed to admixture of African alleles. Mol. Biol. Evol. **20:** 792–799.

Clark, A. G., and L. Wang, 1997 Molecular population genetics of Drosophila immune system genes. Genetics **147:** 713–724.

Daugherty, A., S. C. Whitman, A. E. Block and D. L. Rateri, 2000 Polymorphism of class A scavenger receptors in C57BL/6 mice J. Lipid Res. **41:** 1568–1577.

David, J. R., and P. Capy, 1988 Genetic variation of *Drosophila melanogaster* natural populations. Trends Genet. **4:** 106–111.

Dziarski, R., 2004 Peptidoglycan recognition proteins (PGRPS). Mol. Immunol. **40:** 877–886.

Fauvarque, M.-O., E. Bergeret, J. Chabert, D. Dacheux, M. Satre *et al.*, 2002 Role and activation of Type III secretion system genes in *Pseudomonas aeruginosa*-induced *Drosophila* killing. Microb. Pathog. **32:** 287–295.

Fay, J. C., and C.-I Wu, 2000 Hitchhiking under positive Darwinian selection. Genetics **155:** 1405–1413.

Fay, J. C., G. J. Wyckoff and C.-I Wu, 2002 Testing the neutral theory of molecular evolution with genomic data from *Drosophila*. Nature **415:** 1024–1026.

Fortin, A., M. Penman, M. M. Stevenson, M. Kregier and P. Gros, 2000 Identification of and characterization of naturally occurring variants of the macrophage scavenger receptor (SR-A). Mamm. Genome **11:** 779–785.

Gough, P. J., and S. Gordon, 2000 The role of scavenger receptors in the innate immune system. Microbes Infect. **2:** 305–311.

Hamblin, M. T., and M. Veuille, 1999 Population structure among African and derived populations of *Drosophila simulans*: evidence for ancient subdivision and recent admixture. Genetics **153:** 305–317.

Hudson, R. R., 1987 Estimating the recombination parameter of a finite population model without selection. Genet. Res. **50:** 45–50.

Hudson, R. R., 1990 Gene genealogies and the coalescent process, pp. 1–44 in *Oxford Surveys in Evolutionary Biology*, edited by D. Futuyma and J. Antonovic. Oxford University Press, Oxford.

Hudson, R. R., 2002 Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics **18:** 337–338.

Hudson, R. R., D. D. Boos and N. L. Kaplan, 1992 A statistical test for detecting geographic subdivision. Mol. Biol. Evol. **9:** 138–151.

Hughes, A. L., and M. Nei, 1988 Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature **335:** 167–170.

Hughes, A. L., and M. Nei, 1999 Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. Proc. Natl. Acad. Sci. USA **86:** 958–962.

Jiggins, F. M., and G. D. Hurst, 2003 The evolution of parasite recognition genes in the innate immune system: purifying selection on *Drosophila melanogaster* peptidoglycan recognition proteins. J. Mol. Evol. **57:** 598–605.

Krieger, M., 1997 The other side of scavenger receptors: pattern recognition for host defense. Curr. Opin. Lipidol. **8:** 275–280.

Krieger, M., and J. Herz, 1994 Structures and functions of multiligand lipoprotein receptors: macrophage scavenger receptors and LDL receptor-related protein (LRP). Annu. Rev. Biochem. **63:** 601–637.

Krieger, M., S. Acton, J. Ashkenas, A. Pearson, M. Menman *et al.*, 1993 Molecular flypaper, host defense and atherosclerosis—structure, binding-properties, and functions of macrophage scavenger receptors. J. Biol. Chem. **268:** 4569–4572.

Lazzaro, B. P., and A. G. Clark, 2001 Evidence for recurrent paralogous gene conversion and exceptional allelic divergence in the *Attacin* genes of *Drosophila melanogaster*. Genetics **159:** 659–671.

Lazzaro, B. P., and A. G. Clark, 2003 Molecular population genetics of inducible antibacterial peptide genes in *Drosophila melanogaster*. Mol. Biol. Evol. **20:** 914–923.

Lazzaro, B. P., B. K. Sceurman and A. G. Clark, 2004 Genetic basis of natural variation in D. melanogaster antibacterial immunity. Science **303:** 1873–1876.

Lindmark, H., K. C. Johansson, S. Stöven, D. Hultmark, Y. Engström *et al.*, 2001 Enteric bacteria counteract lipopolysaccharide induction of antimicrobial peptide genes. J. Immunol. **167:** 6920–6923.

Little, T. J., J. K. Colbourne and T. J. Crease, 2004 Molecular evolution of *Daphnia* immunity genes: polymorphism in a gram negative binding protein and an alpha-2-macroglobulin. J. Mol. Evol. **59:** 498–506.

Llopart, A., J. M. Comeron, F. G. Brunet, D. Lachaise and M. Long, 2002 Intron presence-absence polymorphism in *Drosophila* driven by positive Darwinian selection. Proc. Natl. Acad. Sci. USA **99:** 8121–8126.

McDonald, J. H., and M. Kreitman, 1991 Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature **351:** 652–654.

Meister, M., 2004 Blood cells of *Drosophila*: cell lineages and role in host defense. Curr. Opin. Immunol. **16:** 10–15.

Moriyama, E. N., and J. R. Powell, 1996 Intraspecific nuclear DNA variation in *Drosophila*. Mol. Biol. Evol. **13:** 261–277.

Mousset, S., and N. Derome, 2004 Molecular polymorphism in *Drosophila melanogaster* and *D. simulans*: What have we learned from recent studies? Genetica **120:** 79–86.

Pearson, A., A. Lux and M. Krieger, 1995 Expression cloning of *dSr-CI*, a class C macrophage-specific scavenger receptor from *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **92:** 4056–4060.

Peiser, L., S. Mukhopadhyay and S. Gordon, 2002 Scavenger receptors in innate immunity. Curr. Opin. Immunol. **14:** 123–128.

Quesada, H., U. E. M. Ramírez, J. Rozas and M. Aguadé, 2003 Large-scale adaptive hitchhiking upon high recombination in *Drosophila simulans*. Genetics **165:** 895–900.

Rämet, M., A. Pearson, P. Manfruelli, X. Li, H. Koziel *et al.*, 2001 *Drosophila* scavenger receptor CI is a pattern recognition receptor for bacteria. Immunity **15:** 1027–1038.

Ramos-Onsins, S., and M. Aguadé, 1998 Molecular evolution of the *Cecropin* multigene family in Drosophila: functional genes *vs.* pseudogenes. Genetics **150:** 157–171.

Rozas, J., and R. Rozas, 1999 DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics **15:** 174–175.

Rozas, J., M. Gullaud, G. Blandin and M. Aguadé, 2001 DNA variation at the *RP49* gene region of *Drosophila simulans*: evolutionary inferences from an unusual haplotype structure. Genetics **158:** 1147–1155.

Rozas, J., J. C. Sánchez-DelBarrio, X. Messeguer and R. Rozas, 2003 DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics **19:** 2496–2497.

Schlenke, T. A., and D. J. Begun, 2003 Natural selection drives Drosophila immune system evolution. Genetics **164:** 1471–1480.

Schlenke, T. A., and D. J. Begun, 2004 Strong selective sweep associated with a transposon insertion in *Drosophila simulans*. Proc. Natl. Acad. Sci. USA **101:** 1626–1631.

Schlenke, T. A., and D. J. Begun, 2005 Linkage disequilibrium and recent selection at three immunity receptor loci in *Drosophila simulans*. Genetics **169:** 2013–2022.

Tajima, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **132:** 585–595.

Takano, T. S., 1998 Rate variation of DNA sequence evolution in the Drosophila lineages. Genetics **149:** 959–970.

Underhill, D. M, and A. Ozinsky, 2002 Phagocytosis of microbes: complexity in action. Annu. Rev. Immunol. **20:** 825–852.

Wall, J. D., P. Andolfatto and M. Przeworski, 2002 Testing models of selection and demography in *Drosophila simulans*. Genetics **162:** 203–216.

Yang, Z., and J. P. Bielawski, 2000 Statistical methods for detecting molecular adaptation. Trends Evol. Ecol. **15:** 496–503.

Zhang, J., W. L. Rowe, A. G. Clark and K. H. Buetow, 2003 Genomewide distribution of high-frequency, completely mismatching SNP haplotype pairs observed to be common across human populations. Am. J. Hum. Genet. **73:** 1073–1081.

Communicating editor: M. Aguadé