Molecular Population Genetics of Drosophila Immune System Genes

Andrew G. Clark and Lei Wang

Institute of Molecular Evolutionary Genetics, Department of Biology, Pennsylvania State University, University Park, Pennsylvania 16802

Manuscript received July 9, 1996
Accepted for publication July 2, 1997

ABSTRACT

A striking aspect of many vertebrate immune system genes is the exceptionally high level of polymorphism they harbor. A convincing case can be made that this polymorphism is driven by the diversity of pathogens that face selective pressures to evade attack by the host immune system. Different organisms accomplish a defense against diverse pathogens through mechanisms that differ widely in their requirements for specific recognition. It has recently been shown that innate defense mechanisms, which use proteins with broad-spectrum bactericidal properties, are common to both primitive and advanced organisms. In this study we characterize DNA sequence variation in six pathogen defense genes of *Drosophila melanogaster* and *D. mauritiana*, including *Andropin*; cecropin genes *CecA1*, *CecA2*, *CecB*, and *CecC*; and *Diptericin*. The necessity for protection against diverse pathogens, which themselves may evolve resistance to insect defenses, motivates a population-level analysis. Estimates of variation levels show that the genes are not exceptionally polymorphic, but *Andropin* and *Diptericin* have patterns of variation that differ significantly from neutrality. Patterns of interpopulation and interspecific differentiation also reveal differences among the genes in evolutionary forces.

ONE of the most consistent findings in the field of molecular population genetics is that genes that serve a recognition function in immune systems are exceptionally variable (OTA and NEI 1994; PARHAM and OHTA 1996), and the cell-surface antigens of pathogens are also exceedingly diverse (TANABE *et al.* 1987; LI *et al.* 1990; SMITH *et al.* 1990; WEINBERG *et al.* 1990; HUGHES 1991). Analysis of the patterns of variation provides a convincing case that this diversity is generated by natural selection in favor of new antigenic determinants in the pathogens, followed by the improved success of hosts whose immune system can recognize the new diversity (HUGHES and NEI 1988). Recently much progress has been made in understanding how insects fight infections by pathogenic organisms, and the question naturally arises, how do insects deal with pathogenic diversity? Before getting into the population genetic aspects of this problem it is useful to review briefly the means by which insects defend themselves.

When bacteria are injected into insects, there are rapid responses at two distinct levels, cellular and humoral (reviewed by HULTMARK 1993). The cellular response removes bacteria through phagocytosis by a particular class of hemocytes called plasmatocytes (DUNN 1986). The molecular basis for the recognition of foreign substances in the hemolymph by plasmatocytes is not known, but several experiments demonstrate that it is not highly specific. Foreign objects too large to be phagocytized are encapsulated by protein from granular cells. Little is known about genetic regulation of recognition by plasmatocytes, but defects in the ability of Phormia flies to encapsulate eggs of a parasitoid wasp have been characterized (CARTON *et al.* 1992). Although the insect cellular response can effectively remove bacteria from the hemolymph, there is no clonal selection and future infections elicit no stronger response than the first exposure to a pathogen. In this sense the insect immune system appears to have no memory (FAYE and HULTMARK 1993).

In addition to the cellular response, insects respond to infection with a cell-free or humoral response. Injury to the cuticle results in activation of a phenoloxidase cascade, which allows hardening of the cuticle at the point of injury (and as a side effect produces melanin resulting in dark staining of wounded areas). If hemolymph is exposed to air, a serine protease cascade, not unlike that in mammals, is also activated and results in clotting of the hemolymph. In addition, stimulation by injection, ingestion, abrasion, or breaking the insect's cuticle with a sterile probe results in nearly immediate induction of transcription of a suite of bactericidal (antimicrobial) proteins (BOMAN and HULTMARK 1987). Although induction of antimicrobial proteins generally appears to be nonspecific, different classes of soluble peptidoglycan fragments result in two orders of magnitude difference in induction (IKETANI and MORISHIMA 1993). Alignments of molecular sequences have shown homology between antimicrobial proteins of insects and mammals, confirming an ancient ancestry (LAMBERT *et al.* 1989; LEE *et al.* 1989). The first of some 50 of these antimicrobial proteins to be isolated was

Corresponding author: Andrew Clark, Department of Biology, Pennsylvania State University, University Park, PA 16802.
E-mail: c92@psu.edu

cecropin, so called because its source was the cecropia moth, *Hyalophora cecropia* (HULTMARK *et al.* 1980).

Cecropins are small proteins, consisting of a signal peptide of 22 amino acids and a mature protein of 35–39 residues. Cecropins fold into two α helices and one Gly-Pro hinge, a structure that allows the cecropin molecules to mimic a voltage-dependent cation channel and integrate into the cell membrane of prokaryotes (DURELL *et al.* 1992). Integration into the membrane appears to cause sufficient leakage of cations to result in cell lysis. The action of cecropins is stoichiometric, not catalytic, consistent with the mechanism of cell killing. Micromolar concentrations are sufficient to kill most Gram-positive and Gram-negative bacteria, but there are exceptions. One micromolar will kill *Escherichia coli* and 14 μM will kill most Serratia, but to kill Pseudomonas and Bacillus require >100 μM, and these bacteria are often pathogenic to insects (BOMAN and HULTMARK 1987). Wounded Drosophila produce hemolymph concentrations of cecropins of up to 50 μM (SAMAKOVLIS *et al.* 1990). Eukaryotic cells are generally resistant to cecropins but, at 100 μM, Trypanosoma, Plasmodium, and Leischmania are also killed (JAYNES *et al.* 1988).

Diptericin, another antimicrobial protein produced by infected insects, was first isolated in the fly *Phormia terranovae* (DIMARCQ *et al.* 1988). Diptericin has a 19 amino acid signal peptide, which is cleaved to leave 82 amino acids in the mature protein. Diptericin kills only Gram-negative bacteria (LAMBERT *et al.* 1989; DIMARCQ *et al.* 1990). Postinfection induction of diptericin is nonspecific: it is transcribed whether Gram-positive or -negative bacteria infect the fly (DIMARCQ *et al.* 1990). As in the case of cecropins, the kinetics of synthesis after subsequent infections are no different, so there is no memory to the recognition system. In addition there are numerous other bactericidal proteins produced in insects and other organisms, including attacins, defensins, magainins, abiecin, hemolin, sarcotoxin, bactericidin, sapecin, and phormicin (ZASLOFF 1987; BOMAN 1991; DUCLOHIER 1994). An important aspect of these diverse proteins is that they function in different ways, so that their combined use makes it more difficult for bacteria to evolve resistance.

Drosophila is emerging as a model organism for studying the genetics and molecular biology of insect pathogen defense. The most potent of inducible bactericidal proteins in Drosophila are the cecropins. KYLSTEN *et al.* (1990) reported the sequence of a 4-kilobase (kb) genomic clone of a region of 99E that contains the genes *Andropin*; cecropins A1, A2, and B (*CecA1*, *CecA2*, and *CecB*); and two pseudogenes. Just 2.5 kb away lies the *CecC* gene (TRYSELIUS *et al.* 1992). Each cecropin gene has a single intron of 58–61 base pairs (Figure 1). Andropin (*Anp*) is male specific, it is induced by mating, and it is found in the ejaculatory duct (SAMAKOVLIS *et al.* 1991). *CecA1* and *CecA2* differ only in the signal peptide and synonymous sites, while *CecB* differs from *CecA1* at 10 amino acid positions (five in the signal peptide and five in the mature protein). The duplication of *CecA1* and *CecA2* is intriguing because they appear to be expressed identically. *CecB* and *CecC* are expressed during metamorphosis, when insects are especially vulnerable to microbial infection (TRYSELIUS *et al.* 1992). The TATA box, cap site, and polyadenylation signals of all cecropin genes are unexceptional. The only published report of polymorphism in this gene cluster is that of KYLSTEN *et al.* (1990), who observed variation among nine cDNA clones of the *CecA1* and seven cDNA clones of *CecA2*, all from the same *Canton-S* lab stock. Although there is a high level of interspecific sequence conservation of cecropins, and their role in bacterial defense does not guarantee that they face the same kind of diversifying selection that immunoglobulins face, one cannot help asking about the pattern of DNA sequence variation in bacterial defense molecules. Here we perform a molecular population genetic analysis of allelic sequences of six defense genes in order to infer evolutionary forces at play on these genes.

MATERIALS AND METHODS

Drosophila stocks: Twenty second and third chromosome extracted lines from Beltsville, MD were kindly provided by BRIAN CHARLESWORTH (CLARK *et al.* 1995). Five lines from Zimbabwe used in this study were a gift from DAVID BEGUN. To obtain Zimbabwe lines containing a single allele of each cecropin gene, stocks were crossed to a line bearing the deficiency *Df(3R)R133*, which covers cytological positions 99E4 to 100F5 and spans the cecropin gene cluster. Heterozygotes for this deficiency were hemizygous for the cecropin genes (including andropin) and were used for DNA extraction. The gene encoding Diptericin (*Dpt*) is located on 2R at 56A (WICKER *et al.* 1990), and four generations of inbreeding was sufficient to obtain homozygotes. Strain 31 of *Drosophila mauritiana* was obtained from the National Drosophila Species Resource Center.

PCR and sequencing: From the sequence of KYLSTEN *et al.* (1990), we designed oligonucleotide primers for PCR amplification of *Anp*, *CecA1*, *CecA2*, and *CecB*. The *CecC* sequence of TRYSELIUS *et al.* (1992) and the *Dpt* sequence of WICKER *et al.* (1990) were similarly used to design primers. A display of the cecropin gene cluster showing the location of primers and PCR products appears in Figure 1. Despite the duplicated nature of this gene cluster, it was possible to design primer pairs that uniquely amplified each gene. Genomic DNAs were extracted by a standard phenol-chloroform protocol, and the PCR was performed with 1 min denaturation at 94°, 1.5 min annealing at 55°, and 2 min synthesis at 72° for 30 cycles. All primer pairs yielded clean single-banded PCR products. Cycle sequencing and sequence determination were done with the ABI cycle sequencing chemistry following manufacturer's directions, and sequences were read with an ABI 373 automated DNA sequencer. Both strands of each gene-allele combination were sequenced with at least twofold redundancy, and sequences were assembled with the SEQMAN program of DNAStar.

Single-strand conformation polymorphism: To expand the sample size, we quantified variation in banding patterns by analysis of single-stranded conformation polymorphism (SSCP), as described by AGUADÉ *et al.* (1994). In this method, PCR products of up to ~300 nucleotides (nt) are denatured

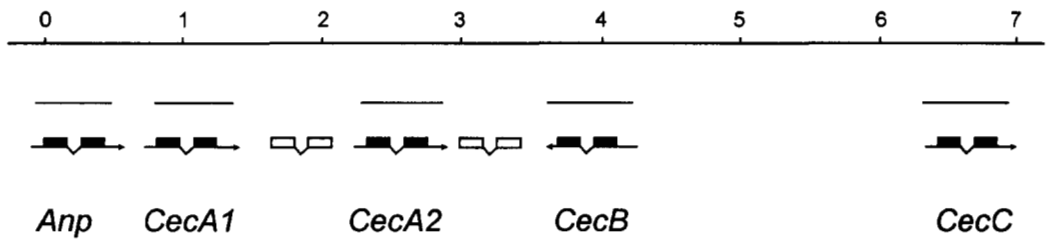

FIGURE 1.—The cecropin gene cluster on chromosome *3* (99E) of *D. melanogaster* (TRYSELIUS *et al.* 1992). Scale is in kilobase pairs. Sequenced regions are indicated by the horizontal lines. Open boxes are pseudogenes. The oligonucleotide primers, listed in 5′–3′ order, used to amplify each gene were *Anp* (GGTTGATTGCCTATAAAGCCACTTG and GCAGTGCAGAATTACTTC GG), *CecA1* (AGCTCTCGCCTTTTGTATC and TACAGGGAGCAACAGAAT), *CecA2* (AGCTCTCGGCGTTCCTGGTGC and CAT TCACAACTATTCGCTGA), *CecB* (AAGAGAAATGAGCGGGTCGAGG and GCGGTTCCGAACTGAGTCCATC), *CecC* (CCTGTG CCAGAAGTCCAGTCA and GCGTTATCCTGGTAGAGTCC), and *Dip* (GCCTATAAAAGAGCATCGAA and GCTAGACTCGGA TACCAATCGAG).

by boiling, cooled rapidly in ice, and separated electrophoretically on a nondenaturing gel (we used the MDE acrylamide mix of Hydrolink). We departed from the AGUADÉ *et al.* (1994) protocol in using 0.75 mm × 18 cm × 24 cm vertical acrylamide gels instead of larger sequencing gels and directly silver stained the gels. If SSCP patterns matched, they were treated as identical alleles, and repeatable mismatches were considered as distinct alleles.

RESULTS

Structure of the sequence variation: Our sample of cecropin and *Dpt* genes had 114 nt sites segregating in *D. melanogaster* and an addition 51 sites that had diverged from *D. mauritiana* (Table 1). Several differences among the patterns of variation are immediately apparent, including the lack of intermediate-frequency alleles in *Anp* compared to *CecC*. Formal analysis of these patterns begins by quantifying the variation in terms of the numbers of segregating sites per gene (S), the estimate of $\theta = 4N\mu$ (WATTERSON 1975), and the nucleotide diversity π (NEI 1987), defined as the probability a random pair of alleles will mismatch at a site. Values of these statistics are reported for each gene in Table 2. θ estimates for all sites in each gene range from 0.007 for *Anp* to 0.026 for *CecC*, tending to be somewhat higher than average for *D. melanogaster*, but not significantly so. There is a tendency for introns to be more variable and coding regions less so, but this is not universal. Our initial hypothesis that these pathogen defense genes might be hypervariable can be rejected.

One test of fit of the data to the neutral mutation model is based on the equivalence of estimates of θ based on numbers of segregating sites and on pairwise site heterozygosity (TAJIMA 1989). Tajima's test statistic D is less than zero for four of the six genes, but in no case was the value significantly different from zero. The power of this test is low for such small genes (SIMONSEN *et al.* 1995), and SSCP patterns were used to assemble the data into super-alleles consisting of the entire gene cluster. This assembled dataset had 19 Beltsville alleles and five Zimbabwe alleles. Within Beltsville, there were 14 distinct alleles, yielding 82 segregating sites, and nucleotide diversity $\pi = 0.025 \pm 0.013$ [Tajima $D = 1.311$, not significant (NS)]. Within Zimbabwe, there were five alleles, 101 segregating sites, $\pi = 0.0213 \pm 0.013$, and Tajima $D = 1.029$, NS). Pooling the data from the two populations, there were 114 sites, 19 alleles, and Tajima's $D = 2.53$. This value is significantly different from zero but, as we shall see, the populations exhibit significant genetic differentiation, so the test most likely rejects the null hypothesis due to population subdivision rather than departure from neutrality. Because there is some chance that matching SSCP patterns may differ in nucleotide sequence, all analyses that follow were based on sequence data only.

Neutral models allow calculation of θ from both the observed number of singleton sites and from the number of segregating sites. Correspondence between these two estimates provides the basis for the neutrality test of FU and LI (1993). In the nine alleles of *Anp* that were sampled, there were six segregating sites and all were singletons, a result that significantly departs from neutrality by this test ($D = -2.52$, $P < 0.01$). *Dpt* also had an excess of singleton sites, with 11 external sites out of a total of 16 segregating sites ($D = -2.22$, $P < 0.025$). The other four genes did not depart from neutrality by the Fu and Li test, and four of the six genes had $D < 0$, suggesting a tendency toward purifying selection (or recovery from a sweep).

Introns in the cecropin cluster are very short, with 62, 60, 58, 58, and 70 nt for *Anp*, *CecA1*, *CecA2*, *CecB*, and *CecC*, respectively. It is striking that all intron lengths are conserved between *D. melanogaster* and *D. mauritiana*. Some indel variation was found in the flanks, but these are not reported in Table 1 and are not used in subsequent analyses. Note that *Dpt* has no intron. Another interesting feature of the data is a nine-copy TA repeat upstream from *CecC*. This microsatellite repeat was found to be monomorphic in our sample, and *D. mauritiana* shared the nine-copy repeat.

Silent and replacement variation: A striking feature of DNA sequence variation in the major histocompatibility complex (MHC) alleles is the excess of nonsynonymous variation, consistent with selection favoring ele-

TABLE 1

Segregating nucleotide sites in Drosophila antibacterial genes

	5′			Exon 1					Intron 1									Exon 2					
	2	2	2	3	3	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
	3	7	8	1	3	6	6	8	0	1	1	1	2	2	2	3	4	4	5	6	7	8	9
	7	5	7	3	8	0	3	2	6	0	1	4	0	4	5	7	5	7	1	4	5	0	4
Andropin																							
Con	t	c	t	g	g	a	t	g	t	a	g	t	g	c	t	a	a	g	a	t	t	a	g
GB	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B009	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B141	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B205	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	g	a	g	.	.	.	.
B316	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.
Z10	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Z18	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Z22	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C
Z24	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
maur	c	a	a	.	C	g	c	A	g	t	t	a	a	t	a	g	.	.	.	A	C	.	.

	5′		e 1			Intron 1										e 2		3′
	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	2	2	2	2	3	3	3	3	3	3	3	3	4	4	4	4	4	5
	4	5	8	8	2	6	6	7	7	7	8	8	8	9	0	0	1	3
	7	4	4	5	6	0	1	0	3	5	0	3	5	0	0	1	0	2
Cecropin A1																		
Con	a	c	c	g	g	t	c	g	a	c	a	a	g	c	a	a	c	t
GB	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B115	.	.	.	.	.	c	t	.	t	t	g	c	.	.	.	.	.	.
B137	.	.	.	.	.	c	t	.	t	t	g	c	.	.	.	.	.	.
B205	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B222	.	.	.	.	.	.	.	.	t	t	g	c	.	.	.	.	.	.
B226	.	.	g	C	c	c	t	.	t	t	g	c	.	.	.	c	.	.
B316	.	.	.	.	c	.	.	.	.	.	.	.	.	.	.	.	.	a
Z5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	a
Z10	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Z18	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Z22	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Z24	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.
maur	c	a	.	.	c	.	.	c	.	a	.	g	a	t	t	t	.	.

	5′								Exon 1								i 1			e 2			3′						
	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
	4	4	4	4	4	4	5	5	5	5	5	5	5	5	5	6	6	6	6	7	7	7	8	8	8	8	8	8	8
	6	6	7	7	8	8	1	2	4	4	5	6	6	7	7	7	2	7	8	8	0	1	2	3	4	4	6	6	8	8
	1	2	6	8	2	4	6	3	6	8	7	3	5	2	3	5	0	1	7	9	5	4	6	0	1	4	6	9	1	4
Cecropin A2																														
Con	a	c	t	c	g	a	c	t	a	c	t	c	t	c	g	c	a	a	c	c	t	g	c	t	t	t	g	a	t	t
GB	.	.	.	.	.	.	.	.	.	.	c	.	.	.	.	t	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B115	.	.	.	.	.	.	.	.	.	.	c	.	.	.	A	t	.	g	t	.	c	a	t	g	.	g	.	.	.	.
B141	.	.	.	.	.	.	.	c	.	.	.	t	.	t	.	.	.	.	.	.	.	.	.	g	.	g	.	.	.	.
B208	.	.	.	.	.	.	.	c	.	.	.	t	.	t	.	.	g	g	.	.	.	.	.	.	.	.	.	g	.	.
B225	.	.	.	.	.	.	.	c	.	.	.	t	.	t	.	.	.	.	.	.	.	.	.	.	a	.	.	.	.	.
B226	.	.	.	.	.	.	.	c	.	.	.	.	.	t	.	.	g	.	.	t	.	.	t	.	.	.	.	g	g	g
Z5	.	a	.	.	c	g	.	c	G	.	.	t	G	.	.	.	.	.	.	.	.	.	.	g	.	g	c	.	.	g
Z10	.	.	.	.	.	c	.	.	.	.	c	.	.	.	.	t	g	.	.	t	.	.	t	g	.	.	.	.	.	.
Z22	.	.	.	.	.	c	.	.	.	.	c	.	.	.	.	.	.	.	.	t	.	.	.	.	.	.	.	g	.	g
Z24	g	a	c	g	.	.	a	.	.	G	c	.	.	.	.	.	.	.	.	t	.	.	.	.	.	.	.	g	g	g

Con is the consensus and GB is the initially reported sequence in the GenBank database (*Anp, CecA1, CecA2,* and *CecB* are in GenBank accession X16972, *CecC* is accession Z11167, and diptericin is accession X16872). The numbers above each site refer to the position in the GenBank reference sequence. Note that the GenBank entry for *CecB* is in reverse order. The B lines are *D. melanogaster* from Beltsville, MD, and the Z lines are from Zimbabwe. *maur* is the strain 31 of *D. mauritiana.* Dots indicate matches with the consensus, upper case letters indicate nonsynonymous changes. 5′ refers to the 5′ flank, e 1 or exon 1 refers to sequences in the first exon, i1 or intron 1 refers to the intron sequences, and 3′ refers to the 3′ flanking sequences. GenBank accession numbers for the new sequences reported here are AF018964–AF019035.

TABLE 1

Continued

	5′									Exon 1				i 1			Exon 2				
	4120	4102	4092	4089	4088	4064	4051	4044	4042	4002	4001	3995	3936	3904	3879	3878	3863	3843	3815	3812	3795
Cecropin B																					
Con	c	g	g	a	a	t	g	a	c	t	g	c	a	a	c	c	t	c	a	g	c
GB	.	a	.	c	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B115	.	.	.	.	.	.	t	.	.	.	.	.	.	.	.	.	c	.	g	a	.
B137	.	.	.	.	.	.	t	.	.	.	.	.	.	.	.	.	c	.	.	a	.
B141	.	.	c	.	.	.	.	.	.	c	T	.	.	.	.	.	c	.	.	a	.
B202	.	.	.	.	.	.	.	.	.	c	.	.	.	.	.	.	c	.	.	a	.
B208	.	.	.	.	.	.	.	.	.	.	.	A	.	g	.	.	c	.	.	.	.
B222	.	.	.	.	.	.	.	.	.	c	.	.	g	g	.	.	.	.	.	.	G
B225	t	.	.	.	g	c	.	.	.	c	.	.	.	.	t	a	.	.	.	.	.
B226	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	c	.	.	a	G
Z10	.	a	.	c	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G
Z18	.	a	.	c	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G
Z22	.	a	.	c	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G
Z24	.	.	.	c	.	.	.	.	.	.	.	.	.	.	.	.	.	t	.	.	G
maur	.	.	.	.	g	c	.	g	t	c	.	.	.	.	.	.	.	.	.	.	.

	5′							Exon 1											Intron 1																	Exon 2					3′	
	249	274	275	299	301	315	324	338	349	352	385	388	394	398	401	402	405	412	428	432	433	434	447	448	452	456	460	463	464	478	479	482	484	487	489	524	542	553	555	571	592	594
Cecropin C																																										
Con	t	g	c	t	a	g	c	a	c	c	a	g	c	t	c	t	a	t	a	c	a	a	t	t	t	t	a	t	g	a	a	t	t	t	g	c	a	a	g	c	t	a
GB	.	.	.	.	.	.	.	.	t	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	t	.	.	.	.	.	a	.	.	.	.	.	.
B009	.	.	.	.	g	.	.	.	t	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	t	.	.	.	.	.	a	.	.	.	.	.	.
B101	c	.	.	.	g	.	.	.	.	.	.	.	.	.	.	.	.	.	t	t	.	.	.	g	.	.	c	a	t	.	.	.	.	.	.	.	.	.	.	.	.	.
B115	.	.	g	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	t	.	.	.	g	.	.	c	.	.	t	.	.	.	a	.	.	.	.	.	.	.	.
B202	.	.	.	.	g	.	.	.	.	.	.	.	t	.	.	.	.	.	t	.	.	.	.	g	.	.	c	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
B205	.	.	.	.	.	c	.	C	.	.	.	.	.	.	t	.	.	.	.	.	.	.	.	.	.	.	.	.	.	t	.	.	.	.	.	.	.	.	.	.	.	.
B208	c	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	g	.	.	c	.	.	.	t	.	.	.	.	.	.	.	.	.	.	.
B316	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	g	.	.	c	.	.	.	t	.	.	.	.	.	.	G	.	.	.	.
Z5	.	.	.	.	.	.	.	.	t	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	t	.	.	.	.	.	a	.	.	.	.	.	.
Z10	.	.	.	.	.	.	.	.	.	.	T	.	.	g	t	G	G	.	.	.	.	.	g	.	.	.	.	.	.	.	.	.	.	.	.	a	.	.	.	.	.	.
Z18	.	.	.	.	.	.	.	.	t	.	T	.	.	.	t	G	G	.	.	.	.	.	g	.	a	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Z22	.	.	.	.	.	.	.	.	t	.	T	.	.	g	t	G	.	.	g	.	g	.	g	.	a	g	.	.	.	t	.	.	.	.	.	.	.	.	.	.	.	.
Z24	.	.	.	.	.	.	.	.	t	.	T	.	.	.	.	G	.	.	g	.	g	.	g	a	a	g	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
maur	.	t	.	c	.	.	a	.	.	t	g	t	.	.	.	.	.	g	.	.	.	g	c	g	.	.	c	.	.	.	.	a	c	.	c	.	g	.	.	G	c	g

	5′				Exon																							
	2232	2240	2241	2245	2267	2281	2293	2294	2305	2311	2338	2386	2391	2392	2412	2413	2422	2442	2449	2465	2470	2485	2491	2516	2521	2541	2565	2566
Diptericin																												
Con	a	a	c	t	g	c	t	t	t	g	c	c	g	c	c	a	t	g	t	g	g	a	c	g	g	c	t	c
GB	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B009	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B101	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.
B115	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B137	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B141	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
B202	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B208	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B222	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B225	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B226	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B306	c	.	.	.	.	t	.	.	.	.	a	t	.	A	G	.	.	.	.	.	a	c	t	.	.	.	.	.
B314	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Z18	.	.	.	.	.	.	.	.	.	.	.	.	.	A	G	.	.	.	.	.	.	.	.	.	.	.	.	.
Z22	c	t	.	.	.	.	.	.	.	.	.	.	.	A	G	c	.	c	.	.	.	.	t	.	.	.	.	.
Z24	.	t	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	C	.	.	.	.
maur	c	t	t	c	T	.	c	G	a	c	.	.	.	A	G	.	a	.	c	.	a	c	.	.	t	.	a	a

TABLE 2

Counts of segregating sites and estimates of nucleotide diversity

Defense genes[a]	Whole region			Coding			Intron			Flank		
	No. of sites	S	θ-π-D	No. of sites	S	θ-π-D	No. of sites	S	θ-π-D	No. of sites	S	θ-π-D
Andropin (n = 9)	326	6	0.71 ± 0.44	174	5	1.11 ± 0.72	62	1	0.62 ± 0.69	90	0	0.00 ± 0.00
			0.46 ± 0.35			0.59 ± 0.50			0.53 ± 0.67			0.00 ± 0.00
			−1.63			−1.58			−1.02			0.00
Cecropin A1 (n = 12)	442	12	0.93 ± 0.48	192	5	0.89 ± 0.55	61	6	3.36 ± 1.98	189	1	0.18 ± 0.20
			0.59 ± 0.38			0.58 ± 0.46			1.41 ± 1.21			1.21 ± 0.81
			0.22			−1.45			1.76			−0.09
Cecropin A2 (n = 10)	457	30	2.42 ± 1.18	192	12	2.30 ± 1.24	58	3	1.90 ± 1.40	207	15	2.67 ± 1.39
			1.33 ± 0.79			1.41 ± 0.94			1.46 ± 1.29			1.56 ± 1.01
			0.04			0.00			0.29			0.06
Cecropin B (n = 13)	411	19	1.53 ± 0.72	192	9	1.55 ± 0.82	58	3	1.71 ± 1.22	161	7	1.44 ± 0.81
			1.53 ± 0.88			1.53 ± 0.97			1.49 ± 1.27			1.46 ± 0.96
			−0.58			−0.03			−1.14			−0.75
Cecropin C (n = 13)	394	28	2.61 ± 1.17	192	11	1.90 ± 0.97	70	16	7.57 ± 3.66	132	4	1.00 ± 0.65
			1.57 ± 0.90			1.60 ± 1.01			2.12 ± 1.55			2.04 ± 1.31
			0.41			0.22			0.90			−0.72
Diptericin (n = 16)	391	16	1.26 ± 0.58	322	14	1.34 ± 0.63	—	—	—	70	2	0.88 ± 0.71
			1.92 ± 1.06			1.82 ± 1.03						1.73 ± 1.31
			−1.59			−1.68						−0.48

S, the number of segregating sites per gene; $\theta = 4N\mu$; π, nucleotide diversity; D, Tajima's test statistic. θ-π-D are entries top, middle and bottom, respectively.

[a] n is the number of sequences.

vated amino acid sequence diversity in these proteins (HUGHES and NEI 1988). Table 3 reports the estimates of the proportion of synonymous sites with synonymous variation and the proportion of nonsynonymous sites with nonsynonymous variation. In no case was there an excess of nonsynonymous variation, and the ratio of p_s:p_n is similar to that of other conserved Drosophila proteins. We conclude that, despite their role in pathogen defense, there is no evidence that allelic diversity of cecropins is favored by natural selection.

The status of silent and replacement variation within *D. melanogaster* can also be compared to the divergence with *D. mauritiana* through a series of tests of independence with contingency tables (TEMPLETON 1987; MCDONALD and KREITMAN 1991). Sites may be partitioned in a 2 × 2 table according to whether they are replacement *vs.* silent differences and according to whether they are polymorphic within *D. melanogaster vs.* fixed in *D. melanogaster* for a nucleotide different from that in *D. mauritiana*. The tail probabilities of these chi-squares indicate that none can reject the null hypothesis (Table 3). Even when the data are pooled across the genes, yielding 24 silent polymorphisms, 13 silent fixed differences, 25 replacement polymorphisms, and seven replacement fixed, the chi-square was just 1.47 (P = 0.226). Rejection of the null hypothesis would imply that there is a disparity between the factors that maintain polymorphism and those that determine whether interspecific differences accumulate. Failure to reject the null hypothesis implies that the forces that impact interspecific and intraspecific variation are indistinguishable, a result that is consistent with a highly conserved role for these genes.

Tests based on polymorphism and divergence: The Hudson-Krietman-Aguadé (HKA) test considers the goodness of fit to a neutral model, which predicts that the amount of nucleotide divergence and the number of segregating sites should covary (HUDSON *et al.* 1987). The two estimates of θ and the divergence between *D. melanogaster* and *D. mauritiana* for each gene can be compared graphically (Figure 2). Table 4 shows that, of all the tests that could be performed, only the *Anp-CecB* rejected the null hypothesis. In this case, *CecB* had many more polymorphic sites compared to *Anp*, while *Anp* had a high ratio of divergence to polymorphism. The HKA tests were also performed on the average pairwise difference (KREITMAN and HUDSON 1991), and none of these tests rejected the null hypothesis. This may not be surprising, because the variance of estimates of θ based on average pairwise distance are generally larger than those based on numbers of segregating sites, making this test less powerful. In conclusion, despite the low power of the test, there is a suggestion of departure from neutrality in the difference between *Anp* and *CecB*.

The HKA test compares levels of polymorphism and divergence, so if the same forces perturbing polymorphism and divergence are acting on the tested genes,

TABLE 3

Synonymous and nonsynonymous polymorphism and divergence

Gene	p_s[a]	p_n	P[b]
Andropin	—	0.004 ± 0.0002	0.417
Cecropin A1	0.013 ± 0.006	0.002 ± 0.0009	0.667
Cecropin A2	0.082 ± 0.001	0.005 ± 0.0017	—
Cecropin B	0.044 ± 0.006	0.006 ± 0.0020	—
Cecropin C	0.034 ± 0.006	0.001 ± 0.0046	0.100
Diptericin	0.008 ± 0.003	0.004 ± 0.0011	0.260

[a] p_s and p_n are the proportion of synonymous sites with synonymous substitutions and the proportion of nonsynonymous sites with nonsynonymous substitutions.

[b] Fisher exact test tail probability for the McDONALD-KREITMAN test. No figure is given for *CecA2* because the *D. mauritiana* allele was not obtained and *CecB* had no fixed differences.

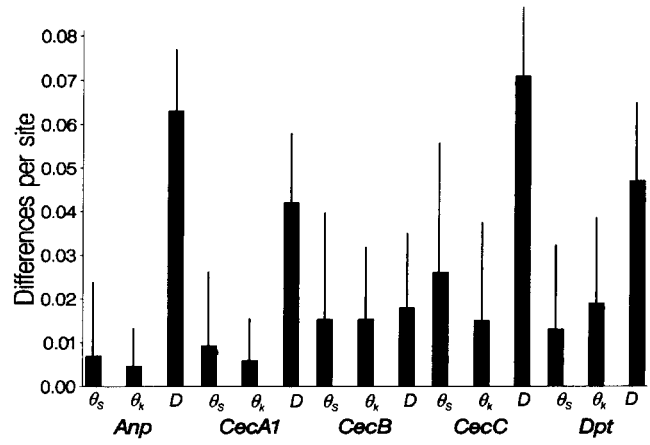

FIGURE 2.—Polymorphism and divergence in bactericidal genes. θ_S is the estimate of $4N\mu$ from the number of segregating sites in each gene, θ_k is the estimate from per site heterozygosity, and D is the nucleotide divergence (number of substitutions per site) between *D. melanogaster* and *D. mauritiana*. The error bars indicate the upper 95% confidence bound on θ as defined by KREITMAN and HUDSON (1991). For D, the upper bound is based on the variance of the divergence estimate.

the null hypothesis may not be rejected even if there are strong departures from neutrality. In an attempt to get around this problem, we compared polymorphism and divergence between the cecropin genes, the *Adh* 5′ flanking region (HUDSON *et al.* 1987) and the *vermilion* flanking region (BEGUN and AQUADRO 1995). Each of the five genes and the pooled cecropin cluster listed in Table 4 were tested against *Adh*-5′ flank, *vermilion* flanking region in Africa, and the *vermilion* flanking region in non-African populations. Of these 18 tests, only the *CecB vs. Adh*-flank was significant ($\chi^2 = 8.12$, $P < 0.005$), but the significance vanished after performing a conservative correction for multiple comparisons.

We did not collect polymorphism data from *D. mauritiana*, so least squares estimates of f, the ratio of effective sizes of *D. melanogaster* and *D. mauritiana* could not be obtained. Nevertheless, it was possible to determine the effects of varying f on the significance of the HKA tests. The *Anp-CecB* tests yielded a $\chi^2 = 5.56$, when it was assumed that $f = 1$ (equal effective sizes), and a $\chi^2 = 6.52$, when $f = 0.05$ (population size of *D. mauritiana* assumed to be 5% that of *D. melanogaster*). The *Adh*-flank *vs. CecB* tests had a $\chi^2 = 8.12$ with $f = 1$, and this increased to 8.97 with $f = 0.05$. The only test that went from nonsignificant to significant as f declined was the *vermilion* flank *vs. CecB*, whose $\chi^2 = 3.24$ for $f = 1$ and 4.10 for $f = 0.05$. Intermediate values of f gave intermediate chi-squares. We are confident that the results of Table 4 would not be altered greatly by better estimates of the relative effective sizes of *D. melanogaster* and *D. mauritiana* populations.

Population subdivision: Following HUDSON *et al.*

TABLE 4

Tests of homogeneity of polymorphism and divergence

		Within species				Between species							
Gene 1	Gene 2		S_1		S_2		D_1		D_2	θ_1	θ_2	T	χ^2
Anp	*CecA1*	6	9.63	12	8.37	17	13.37	8	11.63	0.0084	0.0054	3.93	2.14
	CecB	6	13.07	19	11.93	17	9.93	2	9.07	0.0113	0.0082	1.70	5.56
	CecC	6	12.03	28	21.97	17	10.97	14	20.03	0.0104	0.0158	2.23	2.26
	Dpt	6	9.92	16	12.08	17	13.08	12	15.92	0.0086	0.0087	3.68	1.75
CecA1	*CecB*	12	15.12	19	15.88	8	4.88	2	5.12	0.0097	0.0109	0.14	0.81
	CecC	12	12.90	28	27.10	8	7.10	14	14.90	0.0082	0.0194	0.95	0.04
	Dpt	12	11.67	16	16.33	8	8.33	12	11.67	0.0075	0.0118	1.53	0.01
CecB	*CecC*	19	15.67	28	31.33	2	5.33	14	10.67	0.0108	0.0225	0.21	0.43
	Dpt	19	15.00	16	20.00	2	6.00	12	8.00	0.0103	0.0145	0.42	1.00
CecC	*Dpt*	28	26.40	16	17.60	14	15.60	12	10.40	0.0189	0.0127	1.10	0.14
Dpt	*pool*	16	16.93	65	64.07	12	11.07	41	41.93	0.0122	0.0092	1.32	0.02

S_1 and S_2, the observed and expected numbers of segregating sites within *D. melanogaster*, respectively; D_1 and D_2, the observed and expected numbers of sites that differ between *D. melanogaster* and *D. mauritiana*, respectively. θ_1, θ_2, and T are estimated from the equations in HUDSON *et al.* (1987) and the chi-square (1 d.f.) is a test of goodness-of-fit.

TABLE 5

Divergence between Beltsville and Zimbabwe populations

Gene	H_w	H_b	F_{st}	K^*	P^a
Andropin	0.005	0.005	0.000	0.740	0.279
Cecropin A1	0.008	0.011	0.312	1.578	0.031
Cecropin A2	0.021	0.024	0.131	2.445	0.067
Cecropin B	0.011	0.016	0.285	1.977	0.005
Cecropin C	0.025	0.039	0.358	2.590	0.002
Diptericin	0.005	0.012	0.541	1.629	0.015

H_w, H_b, F_{st} and K^* are estimated following HUDSON *et al.* (1992b), and P is the tail probability from the permutation test of HUDSON *et al.* (1992a).

(1992b), H_w is the average number of differences per site for pairs of alleles drawn from within a population, H_b is the average number of differences per site for pairs of alleles drawn from the two different populations, and an estimate of F_{st} is $1 - H_b/H_w$. Table 5 reports the divergence between Beltsville and Zimbabwe populations. *Anp* is exceptional in exhibiting no differentiation between populations, and the other genes show F_{st} values comparable to other comparisons with Zimbabwe, which appear to be more divergent than non-African populations (BEGUN and AQUADRO 1995). The significance of the population subdivision was assessed by a permutation test, calculating a divergence statistic for each of 1000 datasets randomized with respect to geographic origin of each allele and comparing the observed statistic to this distribution. Haplotype tests were not adequate for these small samples and, following the recommendations of HUDSON *et al.* (1992a), we used the statistic $K^* = \Sigma(\log(d_{ij})/n$, where n is the number of allele pairs and d_{ij} is the number of sites that differ between i and j. All the genes except *Anp* and *CecA2* were significantly heterogeneous.

Multigene family aspects: The four cecropin genes are readily aligned with one another and, while the alignment with *Anp* is not as strong, it is possible to add *Anp* to the cecropin alignment, indicating that they share a common ancestor. To infer the history of duplication events, it is necessary to determine pairwise divergences among the genes and to compare these to distances of homologous genes between the two species. Considering first the interspecific divergence, it is evident that there is significant heterogeneity among these genes in their divergence between *D. melanogaster* and *D. mauritiana*, ranging from 0.018 substitutions per site in *CecB* to 0.071 substitutions per site for *CecC* (Table 6). There was also significant heterogeneity among functional regions (coding *vs.* intron *vs.* flanks), consistent with variation in selective pressures at play on the genes.

Pairwise distances obtained from this alignment show that homologous genes in the two species (such as *Anp* in *D. melanogaster* and *D. mauritiana*) are more similar than are nonhomologous pairs (Table 7). This pattern is consistent with the radiation of the multigene family predating the *D. melanogaster-D. mauritiana* split. An expanded distance matrix was used to construct a neighbor-joining tree to illustrate the relationships graphically (Figure 3). In the case of *CecA1* and *CecC*, this tree places *D. mauritiana* as a outgroup with high bootstrap probability. There is a tendency for the Zimbabwe alleles to be more divergent than Beltsville alleles. As indicated previously, *CecB* exhibits less *D. melanogaster-D. mauritiana* divergence, and in fact the *D. mauritiana* allele appears clustered with the *D. melanogaster* alleles. There is no significant bootstrap confidence to the joining of the A and B clades, but rooting a tree with the cecropia moth amino acid sequence of cecropin A produces a tree with toplogy (A,(B,C)) (not shown).

An important phenomenon in multigene families is the exchange of sequences across genes in the array. Such exchange events can lead to homogenization of sequences across genes in the array in a process known as concerted evolution. Figure 3 shows that, if such exchanges are occurring, they are not sufficient to homogenize the array within species, since the closest ancestor of each gene is not another gene within the same species, but is the homologue in the other species. Nevertheless, we can ask whether there is any evidence for exchange of sequence among cecropin genes by applying statistical methods designed to detect gene conversion events. The algorithm of SAWYER (1989) was applied to the alignment of all the cecropin genes in *D. melanogaster*. The null hypothesis is that the lengths of runs of identity are no longer than one would get from randomly permuted sequences. Runs are tallied for all pairs of sequences and are scored by two statistics, the sum of squared run lengths and the maximum run length. Significance is determined by constructing the

TABLE 6

Nucleotide sequence divergence between *D. melanogaster* and *D. mauritiana*

Gene	Whole region	Coding	Intron	Flank
Andropin	0.063 ± 0.017	0.044 ± 0.020	0.127 ± 0.052	0.041 ± 0.027
Cecropin A1	0.042 ± 0.014	0.005 ± 0.005	0.052 ± 0.021	0.051 ± 0.037
Cecropin B	0.018 ± 0.007	0.011 ± 0.008	0.034 ± 0.017	0.006 ± 0.005
Cecropin C	0.071 ± 0.016	0.044 ± 0.016	0.149 ± 0.053	0.067 ± 0.035
Diptericin	0.047 ± 0.012	0.041 ± 0.012	—	0.118 ± 0.073

TABLE 7

Nucleotide sequence divergence among genes in the cecropin cluster

	Anp	Anp_{mau}	*A1*	$A1_{mau}$	*A2*	*B*	B_{mau}	*C*	C_{mau}
Anp		0.055	1.035	1.048	1.090	1.090	1.090	1.149	1.197
Anp_{mau}	0.013		0.996	1.022	1.035	1.134	1.134	1.181	1.213
A1	0.113	0.108		0.027	0.142	0.421	0.409	0.307	0.349
$A1_{mau}$	0.115	0.111	0.009		0.162	0.432	0.421	0.320	0.365
A2	0.122	0.113	0.023	0.025		0.415	0.398	0.349	0.409
B	0.122	0.128	0.047	0.048	0.046		0.020	0.450	0.444
B_{mau}	0.122	0.128	0.046	0.047	0.045	0.008		0.457	0.463
C	0.131	0.136	0.037	0.039	0.041	0.049	0.050		0.065
C_{mau}	0.139	0.142	0.041	0.042	0.046	0.049	0.050	0.015	

Jukes-Cantor distances among the genes within and between species in the cecropin gene cluster. Distance (numbers of substitutions per site) are above the diagonal and standard errors below. *mau, D. mauritiana.*

null distribution by randomly permuting the sequences. Performing this test on sets of alleles of a single gene rejects the null hypothesis for every gene in the cluster, indicating that exchanges are occurring among homologues. On the other hand, the null hypothesis was not rejected in any test of exchange across nonhomologous genes, consistent with the tree in Figure 3, which suggests little nonhomologous exchange.

The expansion of this small gene family allows a test of relative rates of evolution in each gene since the time of the *D. melanogaster-D. mauritiana* separation. The approach of TAJIMA (1993) was applied by considering triplets of alleles: a *D. melanogaster* and *D. mauritiana* allele of the same gene (call them *A* and *B*) and an allele of a different gene from *D. melanogaster*, which can be considered as the outgroup (call it *C*). Counts of sites that fall into two categories are made: category 1 consists of sites in which *A* and *C* match but differ from *B*, and category 2 sites have *B* and *C* matching, but differ from *A*. The counts of these two classes of sites are expected to be the same, and TAJIMA (1993) showed that a simple chi-square test is appropriate. We applied this test to 12 combinations of alleles of *Anp, CecA1, CecB*, and *CecC* in *D. melanogaster* and *D. mauritiana*. In no case could the null hypothesis be rejected, consistent with the conclusion that the *D. melanogaster* and *D. mauritiana* genes have been evolving at approximately the same rate.

DISCUSSION

Patterns of cecropin variation: The clearest result from this study is that cecropins and *Dpt* do not exhibit the enormous excess of variation that is typical of vertebrate immune system genes. At the same time, several attributes of the patterns of variation depart from a strictly neutral model. The HKA test reveals a significant difference in levels of polymorphism and divergence between *Anp* and *CecB*. The Fu and Li test rejects neutrality for both *Anp* and *Dpt*. In Figures 2 and 3, *CecB* is evidently much less divergent between *D. melanogaster* and *D. mauritiana* than would be expected from its level of polymorphism. It is remarkable that, while *Anp* exhibits departures from neutrality, it shows no significant heterogeneity between the Beltsville and Zimbabwe populations, a pattern at odds with the other genes and findings of other studies (BEGUN and AQUADRO 1995). The *D. melanogaster-D. mauritiana* sequence divergence for *Anp* is comparable to that of the other genes, despite the lack of population heterogeneity. The pattern of interpopulation and interspecific divergence is reversed in *CecB*, which shows significant population differentiation but little *D. melanogaster-D. mauritiana* difference.

Complex demographic history can render tests of neutrality ambiguous, because factors such as population subdivision can produce positive Tajima's *D* even for a neutral gene. Tests that make comparisons across sites that share the same demographic history are less sensitive to these problems. In particular, the tests that compare patterns of synonymous *vs.* nonsynonymous sites are robust to demography, while tests that compare polymorphism and divergence may not be robust. Although the only rejections of neutrality were in tests that may be sensitive to demographic history, the results cannot be entirely explained by neutral demographic processes because the six different genes examined gave markedly different patterns of variation.

A word on statistical power: The small size of the cecropin genes may cause one to wonder whether the power of tests of neutrality is so low that rejection becomes unlikely even with very strong selection. A thorough study of the power of tests of TAJIMA (1989) and FU and LI (1993) showed that in the trade-off between increasing the number of alleles in a sample *vs.* increasing the number of nucleotides sequenced, power is best improved by increasing the number of alleles surveyed (SIMONSEN *et al.* 1995). This study also found that the Tajima *D* statistic was generally more powerful than either tests of Fu and Li, but it should be noted that only the no-outgroup test was examined. Our sample gave no rejections of neutrality by the Tajima test, but *Anp* and *Dpt* both rejected neutrality by the Fu and Li test.

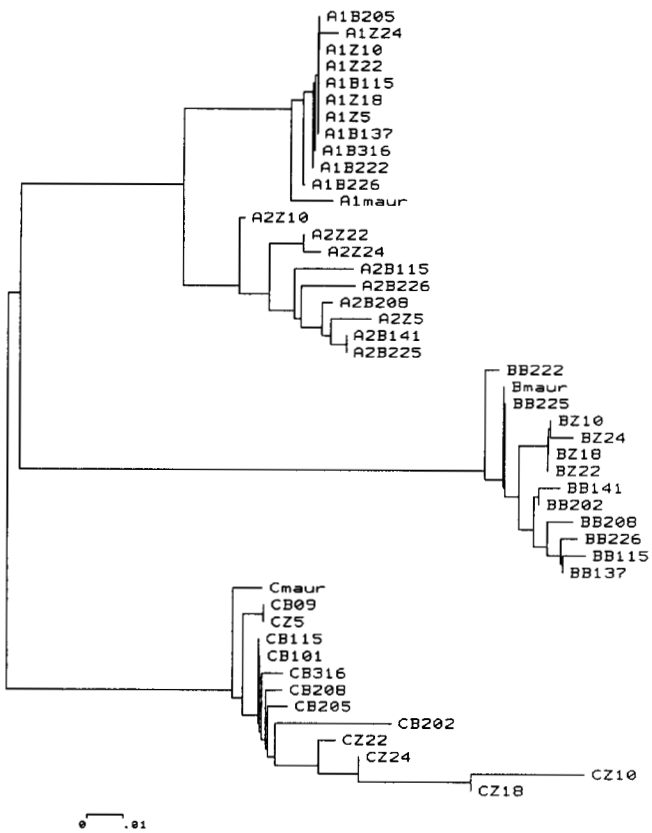

FIGURE 3.—Neighbor-joining tree constructed from Kimura two-parameter distances of the aligned nucleotide sequences of the cecropin genes.

This would appear to be at odds with the expected result from SIMONSEN *et al.* (1995), except that we used the outgroup form of the Fu and Li test. Use of an outgroup would have to improve the power of the Fu and Li test, because it allows unambiguous assignment of external branches. The cecropin genes and *Dpt* gave negative Tajima's *D*, indicating excess low-frequency alleles. This is consistent with a past selective sweep or purifying selection and is the opposite of the departure we might have expected *a priori*, given the function of these genes. Regardless of the considerations of power, our initial hypothesis was that cecropins may exhibit levels of polymorphism comparable to immune system genes in other organisms, and the samples examined clearly allow rejection of this hypothesis. It would seem safe to conclude that cecropins do not experience the same magnitude of selection for increased diversity as do MHC or immunoglobulin genes in vertebrates.

Regulation of expression of cecropins and *Dpt* in Drosophila: A functional immune response requires not only the ability to synthesize molecules that kill bacteria but also the ability to recognize invasion by nonself tissue. Despite the observation that different bacterial components, such as lipopolysaccharides, produce different levels of response when injected into insects (KAPPLER *et al.* 1993), we remain ignorant of the primary recognition molecules. Understanding the primary recognition would be useful for interpreting how bacteria can evade recognition. Other aspects of regulation of insect immunity are becoming much clearer. Cecropins are expressed mostly in the fat body, although, when the cuticle is abraded, there is some localized cecropin expression in underlying epithelial cells (BREY *et al.* 1993). Cultured Schneider cells synthesize cecropins when the cell culture is infected with bacteria or when bacterial cell wall fragments are added. These observations have afforded tests of which components elicit the strongest response (SAMAKOVLIS *et al.* 1992).

The most exciting recent result regarding the regulation of cecropin expression is the finding of NF-κB binding sites in the promoter of cecropins and *Dpt* and the cloning of the gene for the factor that binds those sites (IP *et al.* 1993). The mammalian transcription factor NF-κB, first described by LENARDO and BALTIMORE (1989), regulates a number of genes involved in immune and acute phase responses. It does so by interacting with a nucleotide sequence element, the κB-motif (GGGR($\frac{Y}{A}$)TYYCC). NF-κB binding sites, with consensus sequence GGGRAYYYYY, were first noted in insect cecropin genes by FAYE's group (SUN *et al.* 1991) and have subsequently been found in several other antimicrobial proteins. We did not score variation in and around the NF-κB sites, which are located at −214 for *CecA2*, −178 for *CecB*, and −42, −142, and −794 in the case of *Dpt* (SUN *et al.* 1991; REICHHART *et al.* 1992). ENGSTRÖM *et al.* (1993) made artificial promoters with NF-κB sites and tested expression after transfection into a Drosophila blood cell line. Deletion of the NF-κB sequences destroys the ability to respond to induction. They found that a trimer of κB-like motif confers high levels of inducible expression from a reporter gene. As in the moth *Hyalophora cecropia*, stimulation with bacterial lipopolysaccharide induces a nuclear factor that specifically binds to the κB-like motif.

Recently IP *et al.* (1993) identified the relevant nuclear factor and discovered an intriguing parallel between the regulation of cell fate along the dorsoventral (DV) axis of Drosophila embryos and cecropin induction. Transcription is regulated in both functions by factors containing *rel* domains (*dorsal* and NF-κB) that are controlled at the level of nuclear transport. IP *et al.* (1993) characterized *Dif* (dorsal-related immunity factor), a novel *Rel*-containing gene that provides a potential link between these seemingly disparate processes. *Dif* does not appear to play any role in dorsoventral patterning, but instead mediates an immune response in Drosophila larvae. Dif protein is normally localized in the cytoplasm of the larval fat body, but quickly accumulates in the nucleus upon bacterial infection or injury. In the nucleus, Dif binds to κB-like sequence motifs in promoter regions of antimicrobial genes. These findings suggest that mammalian and in-

sect immunity share a common evolutionary origin, and, along with the homology to *dorsal*, they suggest a role in cell-cell adhesion and/or recognition.

Evolutionary considerations: The evolution of a defense against pathogens is tied intimately to the modes of transmission of the pathogen and to the population dynamics of the host and the pathogen. Simply put, on a population level, the rate of immune response must exceed the rate of division of the pathogen. Rapidly breeding insects can achieve some level of protection by simply staying ahead of pathogen transmission. In large organisms with a slow reproductive rate, a more sophisticated mechanism is necessary to fight infection. By having adaptive subpopulations of cells that produce specific defense molecules (antibodies), vertebrates have achieved an effective defense. However, bacterial pathogens can divide 50 times faster than mammalian B cells, so the vertebrate immune systems need to have memory. Antibacterial proteins can be made 150 times faster than immunoglobulin M and they diffuse faster than defensive cells, so they appear well suited to serve as a first line of defense for vertebrates and for rapidly reproducing organisms like insects (BOMAN 1991).

Because both vertebrates and invertebrates possess homologous antimicrobial molecules, the origin of innate defense must have predated that major evolutionary split (600 mya). To develop the ability to recognize novel invaders and retain a memory of their molecular signature, an immune system must also be able to recognize self and avoid eliciting attack against self-peptides. As a result, the evolution of an immune system with memory had to co-occur with the development of a complex system of defining self. In mammals, the major histocompatibility system provides part of this signal, which is used in a complex process of tolerization of T-cells (JANEWAY 1989). Membrane specificity of bactericidal proteins ensures that they lyse prokaryotic and not eukaryotic cells, but the existence of resistant bacteria shows that the recognition is not foolproof. The similarities between the regulation of dorsal and cecropin genes suggests that early evolution of an adaptive immune system involved recruitment of cellular adhesion molecules. The antibacterial protein hemolin, for example, is related to Drosophila neuroglian (FAYE and HULTMARK 1993). To effect killing of invaders, the vertebrate immune system uses the protein complement, which is analogous to the antimicrobial proteins of the innate immune system (although there is no apparent homology).

Coevolution of host and pathogen is expected to result in very high diversity of both antigen determinants and immune recognition molecules. The vertebrate immune system is able to make millions of different immunoglobulins, and it is able to adaptively expand that diversity. If a cecropin can kill both Gram-positive and -negative, one has to ask why make multiple cecropins? Members of the cecropin family in Drosophila exhibit different temporal and tissue specificity in their expression, and having multiple genes allows this targeting. The use of more than one slightly different defense molecule is analogous to use of pairs of antibiotics and greatly retards the evolution of resistance by pathogens. Heterogeneity between individuals may also be of critical importance to the invasion and spread of a pathogen. Pathogens that can evade defense systems often have an enormous advantage, so selection favors increased diversity of pathogens. While there is still much to learn at the level of the molecular developmental function of innate immunity, it is this arms race aspect of immunity that demands a population level approach as well (reviewed in TRAVIS 1993).

We have established that there is genetic heterogeneity in the molecules that provide pathogen defense to insects, but this heterogeneity is far lower than the heterogeneity seen in vertebrate immune system genes. Genes for insect bactericidal proteins also do not exhibit the excess nonsynonymous variation seen in vertebrate immune system genes. Cecropins and diptericin are effector molecules, and one might expect them to exhibit heterogeneity in order to be functional in pathogens that may evolve resistance. At the same time, there may be a premium on high levels of diversity in the molecules that recognize pathogens, an aspect that was not touched in this study. The relation between sequence variation in cecropins and functional variation in defense is the topic of an article in progress.

We thank RITA MUKHERJEE, GAIL LEVINE and DAVID MOZER for technical assistance. This work was supported by National Science Foundation grant DEB-9527592.

LITERATURE CITED

AGUADÉ, M., W. MEYERS, A. D. LONG and C. H. LANGLEY, 1994 Single-strand conformation polymorphism analysis coupled with stratified DNA sequencing reveals reduced sequence variation in the *su(s)* and *su(wa)* regions of the *Drosophila melanogaster* X chromosome. Proc. Natl. Acad. Sci. USA **91:** 4658–4662.

BEGUN, D. J., and C. F. AQUADRO, 1995 Molecular variation at the *vermilion* locus in geographically diverse populations of *D. melanogaster* and *D. simulans.* Genetics **140:** 1019–1032.

BOMAN, H.G., 1991 Antibacterial peptides: key components needed in immunity. Cell **65:** 205–207.

BOMAN , H. G., and D. HULTMARK, 1987 Cell-free immunity in insects. Annu. Rev. Microbiol. **41:** 103–126.

BREY, P. T., W.-J. LEE, M. AMICI, Y. CAESIUM, S. PARROT *et al.*, 1993 Role of the integument in insect immunity: epicuticular abrasion and induction of cecropin synthesis in cuticular epithelial cells. Proc. Natl. Acad. Sci. USA **90:** 6275–6279.

CARTON, Y., F. FREY and A. NAPPI, 1992 Genetic determinism of the cellular immune reaction in *Drosophila melanogaster.* Heredity **69:** 393–399.

CLARK, A. G., M. AGUADÉ, T. PROUT, L. G. HARSHMAN and C. H. LANGLEY, 1995 Variation in sperm displacement and its association with accessory gland protein loci in *Drosophila melanogaster.* Genetics **139:** 189–201.

DIMARCQ, J. L., E. KEPPI, B. DUNBAR, J. LAMBERT, J.-M. REICHHART *et al.*, 1988 Insect immunity. Purification and characterization of a family of novel inducible antibacterial proteins from immunized larvae of the dipteran *Phormia terranovae* and complete amino-acid sequence of the predominant member, diptericin A. Eur. J. Biochem. **171:** 17–29.

DIMARCQ, J. L., D. ZACHARY, J. A. HOFFMANN, D. HOFFMANN and

J.-M. REICHHART, 1990 Insect immunity—expression of the two major inducible antibacterial peptides, defensin and diptericin, in *Phormia terranovae.* EMBO J. **9:** 2507–2515.

DUCLOHIER, H., 1994 Anion pores from magainins and related defensive peptides. Toxicology **87:** 175–188.

DUNN, P. E., 1986 Biochemical aspects of insect immunology. Annu. Rev. Entomol. **31:** 321–339.

DURELL, S. R., G. RAGHUNATHAN and H. R. GUY, 1992 Modeling the ion channel structure of cecropin. Biophys. J. **63:** 1623–1631.

ENGSTRÖM, Y., L. KADALAYIL, S. C. SUN, C. SAMAKOVLIS, D. HULTMARK *et al.,* 1993 κB-like motifs regulate the induction of immune genes in *Drosophila.* J. Mol. Biol. **232:** 327–333.

FAYE, I., and D. HULTMARK, 1993 The insect immune proteins and the regulation of their genes, pp. 25–53 in *Parasites and Pathogens of Insects, Vol. 2, Pathogens,* edited by N. E. Beckage, S. N. Thompson and B. A. Federici. Academic Press, San Diego.

FU, Y., and W.-H. LI, 1993 Statistical tests of neutrality of mutations. Genetics **133:** 693–709.

HUDSON, R. R., M. KREITMAN and M. AGUADÉ, 1987 A test of neutral molecular evolution based on nucleotide data. Genetics **116:** 153–159.

HUDSON, R. R., D. D. BOOS and N. L. KAPLAN, 1992a A statistical test for detecting geographic subdivision. Mol. Biol. Evol. **9:** 138–151.

HUDSON, R. R., M. SLATKIN and W. P. MADDISON, 1992b Estimation of levels of gene flow from DNA sequence data. Genetics **132:** 583–589.

HUGHES, A. L., 1991 Circumsporozoite protein genes of malaria parasites (*Plasmodium spp.*): evidence for positive selection on immunogenic regions. Genetics **127:** 345–353.

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.

HULTMARK, D., 1993 Immune reactions in Drosophila and other insects: a model for innate immunity. Trends Genet. **9:** 178–183.

HULTMARK, D., H. STEINER, T. RASMUSON and H. G. BOMAN, 1980 Insect immunity: purification and properties of three inducible bactericidal proteins from hemolymph of immunized pupae of *Hyalophora cecropia.* Eur. J. Biochem. **106:** 7–16.

IKETANI, M., and I. MORISHIMA, 1993 Induction of antibacterial protein synthesis by soluble peptidoglycan in isolated fat body from larvae of the silkworm, *Bombyx mori.* Insect Biochem. Mol. Biol. **23:** 913–917.

IP, Y. T., M. REACH, Y. ENGSTRÖM, L. KADALAYIL, H. CAI *et al.,* 1993 *Dif,* a dorsal-related gene that mediates an immune response in *Drosophila.* Cell **75:** 753–763.

JANEWAY, C. A. JR., 1989 Approaching the asymptote? Evolution and revolution in immunology. Cold Spring Harbor Symp. Quant. Biol. **54:** 1–13.

JAYNES, J. M., C. A. BURTON, S. B. BARR, G. W. JEFFERS, G. R. JULIAN *et al.,* 1988 In vitro cytocidal effect of novel lytic peptides on *Plasmodium falciparum* and *Trypanosoma cruzi.* FASEB J. **2:** 2878–2883.

KAPPLER, C., M. MEISTER, M. LAGUEUX, E. GATEFF, J. A. HOFFMANN *et al.,* 1993 Insect immunity. Two 17 bp repeats nesting a κB-related sequence confer inducibility to the Diptericin gene and bind a polypeptide in bacteria-challenged Drosophila. EMBO J. **12:** 1561–1568.

KREITMAN, M., and R. R. HUDSON, 1991 Inferring the evolutionary histories of the *Adh* and *Adh-dup* loci in *Drosophila melanogaster* from patterns of polymorphism and divergence. Genetics **127:** 565–582.

KYLSTEN, P., C. SAMAKOVLIS and D. HULTMARK, 1990 The *cecropin* locus in *Drosophila*; a compact gene cluster involved in the response to infection. EMBO J. **9:** 217–224.

LAMBERT, J., E. KEPPI, J. L. DIMARCQ, C. WICKER, J. M. REICHHART *et al.,* 1989 Insect immunity: isolation form immune blood of the dipteran *Phormia terranovae* of two insect antibacterial peptides with sequence homology to rabbit ling macrophage bactericidal peptides. Proc. Natl. Acad. Sci. USA **86:** 262–266.

LEE, J.-Y., A. BOMAN, S. CHUANXIN, M. ANDERSSON, H. JÖRNVALL *et al.,* 1989 Antibacterial peptides from pig intestine: isolation of a mammalian cecropin. Proc. Natl. Acad. Sci. USA **86:** 9159–9162.

LENARDO, M. J., and D. BALTIMORE, 1989 NF-κB: a pleiotropic mediator of inducible and tissue specific gene control. Cell **58:** 227–229.

LI, J., J. M. MUSSER, P. BELTRAN, M. W. KLINE, and R. K. SELANDER, 1990 Genotypic heterogeneity of strains of *Citrobacter diversus* expressing a 32-kilodalton outer membrane protein associated with neonatal meningitis. J. Clin. Microbiol. **28:** 1760–1772.

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

NEI, M., 1987 *Molecular Evolutionary Genetics.* Columbia University Press, New York.

OTA, T., and M. NEI, 1994 Divergent evolution and evolution by the birth-and-death process in the immunoglobulin VH gene family. Mol. Biol. Evol. **11:** 469–482.

PARHAM, P., and T. OHTA, 1996 Population biology of antigen presentation by MHC class I molecules. Science **272:** 67–74.

REICHHART, J.-M., M. MEISTER, J. L. DIMARCQ, D. ZACHARY, D. HOFFMANN *et al.,* 1992 Insect immunity: developmental and inducible activity of the *Drosophila* diptericin promoter. EMBO J. **11:** 1469–1477.

SAMAKOVLIS, C., D. A. KIMBRELL, P. KYLSTEN, Å. ENGSTRÖM and D. HULTMARK, 1990 The immune response in Drosophila: pattern of cecropin expression and biological activity. EMBO J. **9:** 2969–2976.

SAMAKOVLIS, C., P. KYLSTEN, D. A. KIMBRELL, Å. ENGSTRÖM and D. HULTMARK, 1991 The *Andropin* gene and its product, a male-specific antibacterial peptide in *Drosophila melanogaster.* EMBO J. **10:** 163–169.

SAMAKOVLIS, C., B. ÅSLING, H. G. BOMAN, E. GATEFF and D. HULTMARK, 1992 In vitro induction of cecropin genes—an immune response in a Drosophila blood cell line. Biochem. Biophys. Res. Commun. **188:** 1169–1175.

SAWYER, S., 1989 Statistical tests for detecting gene conversion. Mol. Biol. Evol. **6:** 526–538.

SIMONSEN, K. I., G. A. CHURCHILL and C. F. AQUADRO, 1995 Properties of statistical tests of neutrality for DNA polymorphism data. Genetics **141:** 413–429.

SMITH, N. H., P. BELTRAN and R. K. SELANDER, 1990 Recombination of Salmonella phase 1 flagellin genes generates new serovars. J. Bacteriol. **172:** 2209–2216.

SUN, S. C., B. ÅSLING and I. FAYE, 1991 Organization and expression of the immunoresponsive lysozyme gene in the giant silk moth, *Hyalophora cecropia.* J. Biol. Chem. **266:** 6644–6649.

TAJIMA, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123:** 585–595.

TAJIMA, F., 1993 Simple methods for testing the molecular evolutionary clock hypothesis. Genetics **135:** 599–607.

TANABE, K., M. MACKAY, M. GOMAN and J. G. SCAIFE, 1987 Allelic dimorphism in a surface antigen of malaria parasite *Plasmodium falciparum.* J. Mol. Biol. **195:** 273–287.

TEMPLETON, A. R., 1987 Genetic systems and evolutionary rates. pp. 218–234 in *Rates of Evolution,* K. S. W. CAMPBELL and M. F. DAY, editors. Allen and Unwin, London.

TRAVIS, J., 1993 Tracing the immune system's evolutionary history. Science **261:** 164–165.

TRYSELIUS, Y., C. SAMAKOVLIS, D. A. KIMBRELL and D. HULTMARK, 1992 *CecC,* a cecropin gene expressed during metamorphosis in Drosophila pupae. Eur. J. Biochem. **204:** 395–399.

WATTERSON, G. A., 1975 On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. **7:** 256–276.

WEINBERG, G. A., D. LEHMANN, T. E. TUPASI and D. M. GRANOFF, 1990 Diversity of outer membrane protein profiles of nontypable *Haemophilus influenzae* isolated from children from Papua New Guinea and the Philippines. Rev. Infect. Dis. **12s:** 1017–1020

WICKER, C., J.-M. REICHHART, D. HOFFMANN, D. HULTMARK, C. SAMAKOVLIS *et al.,* 1990 Insect immunity: characterization of a Drosophila cDNA encoding a novel member of the diptericin family of immune peptides. J. Biol. Chem. **265:** 22493–22948.

ZASLOFF, M., 1987 Magainin, a class of antimicrobial peptides from Xenopus skin: isolation characterization of two active forms and partial cDNA sequence of a precursor. Proc. Natl. Acad. Sci. USA **84:** 5449–5453.

Communicating editor: G. B. GOLDING