arms of the different species had already been noticed in the 1940s (*54*). Most of the interspecies rearrangements can be attributed to the occurrence of paracentric inversions (pericentric inversions degrade the integrity of the chromosomes). Additional processes such as simple or Robertsonian translocations (although occurring much less frequently than inversions in *Drosophila*) presumably would most easily explain major exchanges between chromosomal arms, which our analysis indicated. Finally, transposon-mediated rearrangements involving large chromosomal segments (*60*, *61*) could also have led to the extensive recombinations observed in our interspecies comparisons. The sequencing of additional insect genomes in the future will certainly help elucidate some of these evolutionary consequences.

References

1. M. W. Gaunt, M. A. Miles, *Mol. Biol. Evol.* **19**, 748 (2002).
2. D. K. Yeates, B. M. Wiegmann, *Annu. Rev. Entomol.* **44**, 397 (1999).
3. N. J. Besansky, J. R. Powell, *J. Med. Entomol.* **29**, 125 (1992).
4. M. Ashburner, in *Drosophila: A Laboratory Handbook* (Cold Spring Harbor Laboratory Press, Plainview, NY, 1989) p. 74.
5. R. A. Holt *et al.*, *Science* **298**, 129 (2002).
6. W. M. Fitch, *Syst. Zool.* **19**, 99 (1970).
7. R. L. Tatusov, E. V. Koonin, D. J. Lipman, *Science* **278**, 631 (1997).
8. P. Bork, *Genome Res.* **10**, 398 (2000).
9. E. S. Lander *et al.*, *Nature* **409**, 860 (2001).
10. S. Aparicio *et al.*, *Science* **297**, 1301 (2002).
11. M. Ashburner *et al.*, *Nature Genet.* **25**, 25 (2000).
12. G. K. Christophides *et al.*, *Science* **298**, 159 (2002).
13. A. T. Monnerat *et al.*, *Mem. Inst. Oswaldo Cruz* **97**, 589 (2002).
14. M. Affolter, T. Marty, M. A. Vigano, A. Jazwinska, *EMBO J.* **20**, 3298 (2001).
15. E. M. Zdobnov, R. Apweiler, *Bioinformatics* **17**, 847 (2001).
16. R. Apweiler *et al.*, *Bioinformatics* **16**, 1145 (2000).
17. A. Bateman *et al.*, *Nucleic Acids Res.* **30**, 276 (2002).
18. I. Letunic *et al.*, *Nucleic Acids Res.* **30**, 242 (2002).
19. S. R. Schmid, P. Linder, *Mol. Microbiol.* **6**, 283 (1992).
20. G. Dimopoulos *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 6619 (2000).
21. G. Dimopoulos *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 8814 (2002).
22. C. M. Adema, L. A. Hertel, R. D. Miller, E. S. Loker, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 8691 (1997).
23. S. Gokudan *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 10086 (1999).
24. N. Kairies *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 13519 (2001).
25. H. Ranson *et al.*, *Science* **298**, 179 (2002).
26. S. M. Kanzok *et al.*, *Science* **291**, 643 (2001).
27. E. M. Zdobnov *et al.*, data not shown.
28. J. M. Ribeiro, J. G. Valenzuela, *J. Exp. Biol.* **202**, 809 (1999).
29. C. Barillas-Mury, unpublished results.
30. K. J. Schmid, D. Tautz, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 9746 (1997).
31. P. Green *et al.*, *Science* **259**, 1711 (1993).
32. H. D. Youn, J. O. Liu, *Immunity* **13**, 85 (2000).
33. H. D. Youn, L. Sun, R. Prywes, J. O. Liu, *Science* **286**, 790 (1999).
34. A. J. Clark, K. Bloch, *J. Biol. Chem.* **234**, 2578 (1959).
35. R. H. Dadd, in *Comprehensive Insect Physiology, Biochemistry and Pharmacology*, L. I. Gilbert, Ed. (Pergamon, Oxford, 1985), vol. 4, pp. 313–390.
36. F. Lyko, *Trends Genet.* **17**, 169 (2001).
37. D. A. Petrov, *Trends Genet.* **17**, 23 (2001).
38. W. H. Li, T. Gojobori, M. Nei, *Nature* **292**, 237 (1981).
39. D. Torrents *et al.*, data not shown.
40. D. A. Petrov, E. R. Lozovskaya, D. L. Hartl, *Nature* **384**, 346 (1996).
41. A. Stoltzfus, J. M. Logsdon Jr., J. D. Palmer, W. F. Doolittle, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 10739 (1997).
42. W. Gilbert, S. J. de Souza, M. Long, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 7698 (1997).
43. I. B. Rogozin, J. Lyons-Weiler, E. V. Koonin, *Trends Genet.* **16**, 430 (2000).
44. D. Schmucker *et al.*, *Cell* **101**, 671 (2000).
45. I. Letunic, R. R. Copley, P. Bork, *Hum. Mol. Genet.* **11**, 1561 (2002).
46. E. L. George, M. B. Ober, C. P. Emerson Jr., *Mol. Cell. Biol.* **9**, 2957 (1989).
47. W. Dietmaier, S. Fabry, *Curr. Genet.* **26**, 497 (1994).
48. S. Aparicio *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 1684 (1995).
49. C. Hardison, *Trends Genet.* **16**, 369 (2000).
50. D. Thomasova *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 8179 (2002).
51. I. Bancroft, *Trends Genet.* **17**, 89 (2001).
52. S. Wong, G. Butler, K. H. Wolfe, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 9272 (2002).
53. K. H. Wolfe, D. C. Shields, *Nature* **387**, 708 (1997).
54. H. J. Muller, in *The New Systematics*, J. H. Huxley, Ed. (Clarendon, Oxford, 1940), pp. 185–268.
55. V. N. Bolshakov *et al.*, *Genome Res.* **12**, 57 (2002).
56. R. L. Kelley *et al.*, *Cell* **98**, 513 (1999).
57. V. H. Meller, *Trends Cell Biol.* **10**, 54 (2000).
58. A. Pannuti, J. C. Lucchesi, *Curr. Opin. Genet. Dev.* **10**, 644 (2000).
59. C. Schutt, R. Nothiger, *Development* **127**, 667 (2000).
60. R. Paro, M. L. Goldberg, W. J. Gehring, *EMBO J.* **2**, 853 (1983).
61. P. Hatzopoulos, M. Monastirioti, G. Yannopoulos, C. Louis, *EMBO J.* **6**, 3091 (1987).
62. T. F. Smith, M. S. Waterman, *J. Mol. Biol.* **147**, 195 (1981).
63. J. E. Blair, K. Ikeo, T. Gojobori, S. B. Hedges, *Biomed. Central Evol. Biol.* **2**, 7 (2002).
64. A. M. Aguinaldo *et al.*, *Nature* **387**, 489 (1997).
65. J. O. Korbel, B. Snel, M. A. Huynen, P. Bork, *Trends Genet.* **18**, 158 (2002).
66. R. R. Sokal, F. J. Rohlf, *Biometry: The Principles and Practice of Statistics in Biological Research* (Freeman, New York, 1995).
67. A. N. Clements, *The Biology of Mosquitoes* (Chapman & Hall, London, 1992), vol. 1.
68. G. Wegener, *Experientia* **52**, 404 (1996).

Supporting Online Material

www.sciencemag.org/cgi/content/full/298/5591/149/DC1
Materials and Methods
Figs. S1 to S10
Tables S1 to S5
References

6 August 2002; accepted 6 September 2002

Immunity-Related Genes and Gene Families in *Anopheles gambiae*

George K. Christophides,[1]* Evgeny Zdobnov,[1]* Carolina Barillas-Mury,[2] Ewan Birney,[3] Stephanie Blandin,[1] Claudia Blass,[1] Paul T. Brey,[4] Frank H. Collins,[5] Alberto Danielli,[1] George Dimopoulos,[6] Charles Hetru,[7] Ngo T. Hoa,[8] Jules A. Hoffmann,[7] Stefan M. Kanzok,[8] Ivica Letunic,[1] Elena A. Levashina,[1] Thanasis G. Loukeris,[9] Gareth Lycett,[1] Stephan Meister,[1] Kristin Michel,[1] Luis F. Moita,[1] Hans-Michael Müller,[1] Mike A. Osta,[1] Susan M. Paskewitz,[10] Jean-Marc Reichhart,[7] Andrey Rzhetsky,[11] Laurent Troxler,[7] Kenneth D. Vernick,[12] Dina Vlachou,[1] Jennifer Volz,[1] Christian von Mering,[1] Jiannong Xu,[12] Liangbiao Zheng,[8] Peer Bork,[1] Fotis C. Kafatos[1]†

We have identified 242 *Anopheles gambiae* genes from 18 gene families implicated in innate immunity and have detected marked diversification relative to *Drosophila melanogaster*. Immune-related gene families involved in recognition, signal modulation, and effector systems show a marked deficit of orthologs and excessive gene expansions, possibly reflecting selection pressures from different pathogens encountered in these insects' very different life-styles. In contrast, the multifunctional Toll signal transduction pathway is substantially conserved, presumably because of counterselection for developmental stability. Representative expression profiles confirm that sequence diversification is accompanied by specific responses to different immune challenges. Alternative RNA splicing may also contribute to expansion of the immune repertoire.

Malaria transmission requires survival and development of the *Plasmodium* parasite in two invaded organisms: the human host and the mosquito vector. Interactions between the immune system of either organism with the parasite can hinder or even abort its development. The mosquito is known to mount robust immune reactions (*1*), accounting in part for the major parasite losses that occur within the vector. For example, melanotic encapsulation in a refractory strain of *A. gambiae*, the major vector of

human malaria, completely blocks parasite transmission (*2*).

The goal of this article is to describe potential molecular components and thus facilitate future in-depth analysis of the mosquito immune system's impact on the malaria parasite. This goal is best served by a comparative genomic analysis of the *Anopheles* and *Drosophila* immune systems. *Drosophila* is the best model system for the study of invertebrate immunity (*3*); it is a dipteran insect like the mosquito, and it also has a fully sequenced and extensively annotated genome, which has been compared with the *Anopheles* genome (*4*).

Insect immune reactions do not belong to adaptive immunity (which occurs only in chordates) but to the ancient defense system of innate immunity, which is relied upon by the vast majority of metazoans for dealing with invasive organisms, including pathogens and parasites. This system uses a wide range of gene families, some of which also have other physiological or developmental functions. It consists of both cellular and humoral responses, occurring first at the barrier epithelia (essentially the epidermis, gut, and tracheal respiratory organs of insects). Responses then become systemic, using the hemolymph-filled hemocoel, the open circulatory system of insects. Epithelial immunity is less well studied at present and occurs by direct interaction between epithelial cells and microorganisms. For malaria transmission, the key interaction is between the ookinete parasite stage and the midgut epithelial cells that it invades (*5*). In the systemic phase, key actors are the fat body (the insect's functional analog of liver and the main source of circulating immune-related components) and the hemocytes. The latter cells also engage in the cellular defenses of phagocytosis or encapsulation of larger invaders.

A Comparison of Immunity Gene Content in *Anopheles* and *Drosophila*

In this study, we have analyzed 18 mosquito gene families and a number of individual genes, for which comparative evidence from *Drosophila* and other organisms strongly suggested involvement in immune responses. We have used comparative bioinformatic analysis and manual annotation to characterize 242 *Anopheles* genes and relate them to 185 homologs from *Drosophila* (table S1). To facilitate future work, we named the mosquito genes systematically, using nomenclature rules that we propose for *Anopheles*, which are based largely on the HUGO nomenclature for the human genome.

The characterization and comparison of genes and families is summarized in Table 1. A basic conceptual framework of this analysis is that 1:1 orthologs correspond to well-conserved functions; orthologous groups (OG) represent functions that have begun to diversify; specific expansions (SE) represent major diversifications toward species-specific functions; and other genes (OT) represent genes that may have become highly specialized, or lost from the other species. A comparison of these global genome data against the immune genes is shown in Fig. 1A. In both species (and to a greater extent in *Anopheles*) we note that, relative to the genome as a whole (*4*), the immunity system has a deficit of 1:1 orthologs, contrasting

Table 1. Summary of potential immune components. Columns show gene numbers in orthologous pairs (1:1), total genes (TO), orthologous groups (OG), specific gene expansions (SE), and other homologs (OT). SCRB12 is not included in the analysis; CTL groups are not defined in *Drosophila*.

Family		*A. gambiae*				1:1	*D. melanogaster*			
		OT	SE	OG	TO		TO	OG	SE	OT
					Recognition					
PGRP	S	1	2	–	3	–	7	–	5	2
	L	–	–	1	4	3	6	2	–	1
TEP		2	10	2	15	1	6	1	4	–
GNBP	A	1	–	–	2	1	3	–	2	–
	B	–	4	–	4	–	–	–	–	–
SCR	A	1	–	–	5	4	5	–	–	1
	B	5	3	–	16	8	12	–	–	4
	C	–	–	–	1	1	4	–	2	1
CTL	MA	–	5	–	6	1	4	–	1	2
	GA	1	–	–	4	3	5	–	2	–
	SE	–	–	–	2	2	2	–	–	–
	O	5	–	–	10	5	24	–	16	3
GALE		–	5	–	8	3	5	–	–	2
FBN		3	52	–	57	2	13	–	11	–
					Modulation					
CLIP	A	2	6	–	10	2	11	–	7	2
	B	4	9	1	17	3	10	2	2	3
	C	3	2	2	7	–	5	1	2	2
	D	–	–	4	7	3	9	4	–	2
SRPN		1	–	8	10	1	17	6	5	5
IAP		1	2	–	6	3	4	–	–	1
					Signal transduction					
TOLL		–	2	4	10	4	9	2	2	1
MyD88		–	–	–	1	1	1	–	–	–
Tube		–	–	–	1	1	1	–	–	–
Pelle		–	–	–	1	1	1	–	–	–
Cactus		–	–	–	1	1	1	–	–	–
REL		–	–	–	2	2	3	–	–	1
Imd		–	–	–	1	1	1	–	–	–
STAT		–	–	2	2	–	1	1	–	–
					Effector molecules					
PPO		–	8	–	9	1	3	–	–	2
DEF		3	–	–	4	1	1	–	–	–
CEC		–	4	–	4	–	4	–	4	–
CASP	L	–	–	–	2	2	2	–	–	–
	S	2	–	8	10	–	5	3	–	2
SUM		35	114	32	242	61	185	22	65	37

[1]European Molecular Biology Laboratory, Meyerhofstrasse 1, D-69117 Heidelberg, Germany. [2]Department of Microbiology, Immunology and Pathology (MIP), Colorado State University, Fort Collins, CO 80523–1682, USA. [3]European Bioinformatics Institute–European Molecular Biology Laboratory, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK. [4]Unité de Biochimie et Biologie Moléculaire des Insectes, Institut Pasteur, 25 rue du Dr. Roux, 75724 Paris Cedex 15, France. [5]Center for Tropical Disease Research and Training, University of Notre Dame, Post Office Box 369, Notre Dame, IN 46556–0369, USA. [6]Department of Biological Sciences, Centre for Molecular Microbiology and Infection, Imperial College of Science, Technology and Medicine, London SW7 2AZ, UK. [7]Institut de Biologie Moléculaire et Cellulaire, Unité Propre de Recherche, 9022 du Centre National de la Recherche Scientifique, 15 rue Descartes, F67084 Strasbourg Cedex, France. [8]Yale University School of Medicine, Epidemiology and Public Health, 60 College Street, New Haven, CT 06520, USA. [9]Institute of Molecular Biology and Biotechnology–Foundation of Research and Technology Hellas, Vassilika Vouton, Post Office Box 1527, GR-711 10 Heraklion, Crete, Greece. [10]Department of Entomology, 237 Russell Lab, 1630 Linden Drive, University of Wisconsin, Madison, WI 53706, USA. [11]Columbia Genome Center and Department of Medical Informatics, Columbia University, Russ Berrie Medical Science Pavilion, 1150 St. Nicholas Avenue, New York, NY 10032, USA. [12]Department of Medical and Molecular Parasitology, New York University School of Medicine, 341 East 25th Street, Room 613, New York, NY 10010, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: dg-office@embl-heidelberg.de

with an overabundance of specific gene expansions (Table 1). The same features are evident in the large immunity-related fibrinogen-domain (FBN) family, which we discuss in a companion comparative genomics paper as a prime example of gene family diversification (*4*). It would appear that in many immune families, orthologs are under pressure to diversify, or are lost, whereas certain immune genes reduplicate and then diversify. Our working hypothesis is that these prominent features reflect strong selective pressures to adjust and expand the innate immune repertoire in response to new challenges related to new ecological and physiological conditions; in the case of *Anopheles* the challenges include blood-borne infectious agents such as *Plasmodium*. When the immune genes are divided into four major categories (Fig. 1B) corresponding to the four major steps of the immune response, the ortholog deficit is greatest in the recognition, modulation, and effector categories; in contrast, the signal transduction category shows abundant 1:1 pairs and groups of orthologs, but minimal specific gene expansion.

Recognition of Infectious Nonself

In the terminology proposed by C. Janeway (*6*), innate immune responses begin when specialized, soluble or cell-bound pattern-recognition receptors (PRRs) recognize (and bind to) pathogen-associated molecular patterns (PAMPs) that are common in microorganisms but rare or absent in the responding species. PRRs can serve as opsonins facilitating phagocytosis; as receptors for signal transduction pathways that lead to synthesis of anti-pathogen effectors; and as initiators of clotting, melanization, or other protein modification cascades that are implicated in different steps of immunity. We have analyzed potential PRRs belonging to six gene families, two of which we will discuss in detail here and four primarily in the supplementary material.

Peptidoglycan Recognition Proteins (*PGRPs*). This family, distinguished by the PGRP domain (IPR002502), plays central and diverse roles in activating insect immune reactions. These include the melanization cascade, phagocytosis, and signal transduction pathways for production of anti–Gram-positive ($Gram^+$) and anti–Gram-negative ($Gram^-$) effectors (see supplementary material). We have identified seven distinct genes of this family in the *Anopheles* genome, of which three belong to the short (S) subfamily that encodes secreted proteins (*PGRPS1*, *S2*, and *S3*), while four belong to the long (L) subfamily (*PGRPLA*, *LB*, *LC*, and *LD*) encoding transmembrane or intracellular products. By comparison, *Drosophila* has 13 *PGRP* genes, six in the L subfamily (including the orthologs of the *Anopheles* L genes) and seven in the S subfamily (*7*).

Of special interest is the *Anopheles* chromosomal locus 21F (2L) encompassing two adjacent *PGRPLA* and *PGRPLC* genes within ~21 kb (Fig. 2A). The corresponding ~14-kb-long *Drosophila* locus at 67A8 (2R) includes an additional gene, *PGRP-LF*, an apparent product of species-specific tandem duplication. Except for *Drosophila PGRP-LF*, these genes have two or more PGRP domains, each domain encoded by two exons separated by introns at conserved positions (Fig. 2B). This gene architecture is compatible with alternative splicing, leading to proteins with alternative PGRP domains. Using a polymerase chain reaction–based approach on an adult cDNA library, we detected three main RNA isoforms (1, 2, and 3) from the *Anopheles PGRPLC* gene (see below); they carry alternative PGRP domains linked to a common backbone, which encodes a putative signal peptide and transmembrane domain. In *Drosophila*, isoforms of *PGRP-LC* are involved in the Imd signaling pathway and phagocytosis (*8–10*). The domains of this gene are more similar within a species than across species, indicating either that in the common ancestor this gene had one domain, which subsequently triplicated independently, or that a multidomain ancestral gene has followed concerted evolution after speciation (Fig. 2C). Similarly, the *PGRPLA* gene is represented in both species; however, the mosquito gene contains duplicated PGRP domains that are differentially spliced, leading to two distinct detected isoforms, *PGRPLA1* and *2* (Fig. 2A).

Microarray analysis (Fig. 2D) confirmed that some isoforms, which differ between species, are differentially regulated and functionally equivalent to gene expansions in other immunity gene families. After immune and oxidative challenges, the *Anopheles* isoform *PGRPLC2* is up-regulated by all four treatments tested, *PGRPLC1* by none, and *PGRPLC3* only by bacteria. Similarly, both *PGRPLA* isoforms respond to *Escherichia coli*, but additionally *PGRPLA1* responds to peptidoglycan (PGN) whereas *PGRPLA2* responds to *Staphylococcus aureus*. Taken together, these results suggest involvement of immunity signals in splice selection on transcripts of the *PGRPLA/C* gene cluster. Finally, the expression analysis revealed that *PGRPS1* is the only short *PGRP* to be induced by bacteria. *PGRPS2* is not up-regulated by these treatments, and *PGRPS3* is actually down-regulated by *S. aureus* and PGN.

Thioester-containing proteins (*TEPs*). This family is represented in many metazoa, from *Caenorhabditis elegans* to humans. It encodes proteins that play an important role in immune responses as part of the complement system and as the universal protease inhibitors, α2-macroglobulins. Recently, complement-like opsonin function for $Gram^-$ bacteria has been demonstrated (*11*) for the first member of this family studied in the mosquito, aTEP-I (now renamed TEP1). Another member of the family, TEP4, was shown to be up-regulated in *Plasmodium*-infected mosquitoes (*12*). A hallmark of the family is the conserved thioester (TE) motif.

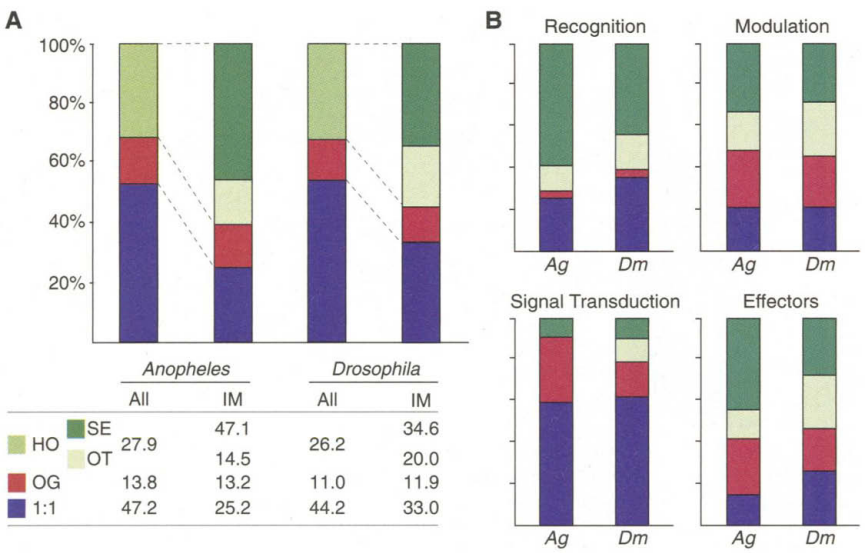

Fig. 1. Comparative analysis of immunity proteins in *Anopheles* and *Drosophila*, and comparison with the respective total proteomes (*4*). Proteins are divided into categories with their sizes shown as percentages. Category 1:1, orthologous pairs; OG, orthologous groups; HO, homologous proteins. The HO category is subdivided for the immunity studies as species-specific expansion (SE) and other homologs (OT). (**A**) Comparison of protein categories between whole genomes and immunity proteins. (**B**) Comparisons of protein categories in gene sets corresponding to the steps of recognition, modulation, signal transduction, and effectors.

After proteolytic activation, TEPs use TE for binding covalently to a nearby target, which is then cleared by phagocytic cells or destroyed by the membrane attack complex (MAC).

The *Drosophila* genome contains six *TEP* genes (*dTep*) (*13*). In strong contrast, after excluding putative haplotypes (designated *TEP16-19*), we have identified 15 *TEP* genes in the *Anopheles* genome (Fig. 3A and Table 1). Only a single 1:1 ortholog and one OG are shared. The majority of *TEP*s (4 in *Drosophila* and 10 in *Anopheles*) represent species-specific expansions, possibly permitting finely tuned responses to multiple pathogenic environments distinct in the two species. In addition, two dTeps and nine *Anopheles* TEPs lack the TE motif; as in vertebrate C5, the TE motif may not always be essential for the functions of insect TEPs.

A notable feature of the *Anopheles TEP* genes is arrangement in multiple chromosomal clusters (Fig. 3B). Genes that are either extensively diverged or resemble *Drosophila* most closely are all located at 29A-30E (3R). The two most similar genes (*TEP2*, *15*) are very close together, whereas the others (*TEP12*, *13*, *14*) are farther apart. Members of the major *Anopheles*-specific expansion are all located at 39C-40B (3L) in three clusters separated by 0.1 and 0.5 Mb. Close resemblance is evident between some genes in different clusters (e.g., *TEP5*, *7*, and *11*), as is a two-step specific expansion (*TEP8*, *9*, and *10*). The structural analysis of the *TEP* family is consistent with a model of sequential gene reduplications, potentially enabling diversified pathogen recognition.

Other recognition factors. We have analyzed four additional families associated with immune recognition in other species (see supplementary material). The Gram-Negative Binding Protein (GNBP) family includes members that are known to bind to Gram$^-$ bacteria, lipopolysaccharide (LPS), and β-1,3-glucan; to be involved in innate immune signaling in response to LPS (*14*); and to be up-regulated by immune challenge (*15*). In *Anopheles*, this family includes only one 1:1 ortholog and five other genes, four of which belong to a mosquito-specific subfamily derived from gene expansion (fig. S1). The multidomain scavenger receptor (SCR) family shows three disparate subfamilies (fig. S2) and is involved in immunity and development, recognizing multiple ligands and helping dispose of bacteria and apoptotic cells. Members of the large B subfamily are associated with uptake of multiple ligands, apoptotic corpses, and *Plasmodium*-infected erythrocytes; the fruit fly *croquemort* (*crq*) (*16*) is represented in the mosquito by a specific gene expansion. Two distinct carbohydrate-binding (lectin) families were also studied. C-type lectins (CTL), which bind to various sugars and LPS or are involved in cell adhesion, show prominent gene expansions (fig. S3). The Galectins (GALE) are associated with multiple functions, including apoptosis and innate immunity; in *Anopheles* several members (fig. S4) are induced by both bacteria and *Plasmodium* (*17*). Taken together (Fig. 1B), these six recognition families show great diversification by species-specific expansions and a deficit of 1:1 orthologs (less so in the case of SCR and CTLs).

Signal Modulation and Amplification

After recognition of infectious nonself, extracellular cascades of activating serine proteases and countervailing serine protease inhibitors (serpins) process the signal by either amplifying a strong "danger signal" or dampening false alarms (see also supplementary material). These modulatory families have a clear 1:1 ortholog deficit, but show increased numbers of OGs and only modest specific gene expansions.

The clip domain serine proteases (CLIPs) are characterized by the homonymous domain, a compact disulfide-bridged structure thought to regulate and localize the activity of the catalytic protease domain. One CLIP, Persephone (CG6367), acts to activate the Toll signaling cascade (*18*), whereas others

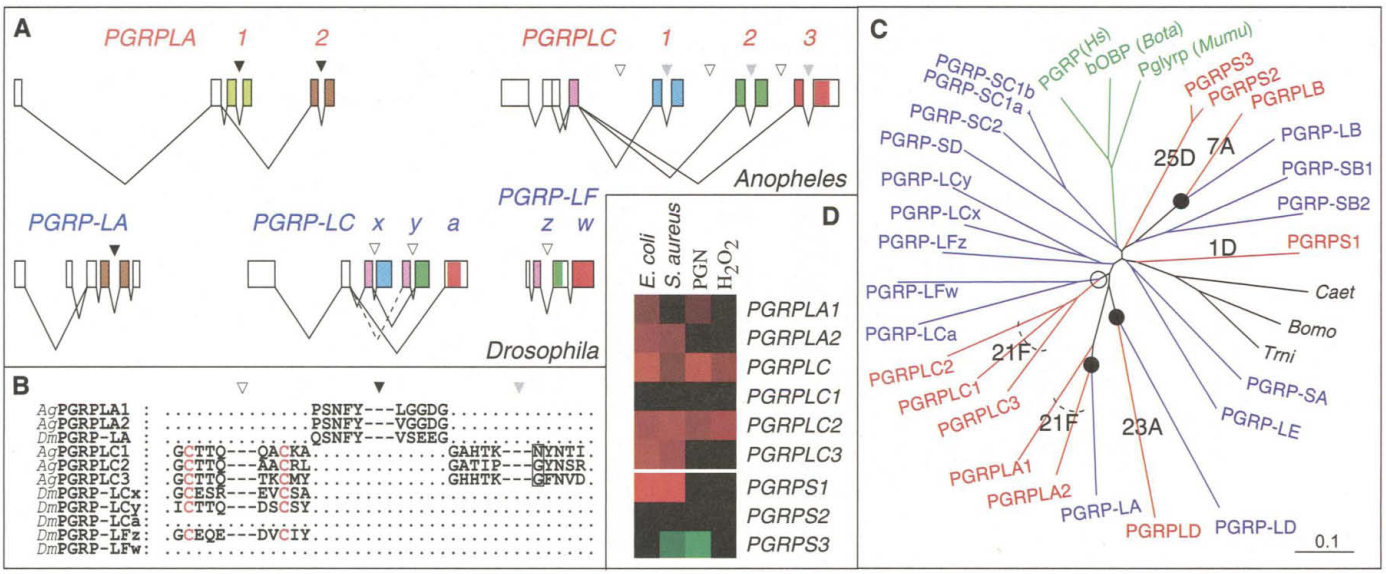

Fig. 2. Gene organization, transcriptional activity, and phylogenetic analysis of the *PGRP* gene family. (**A**) Exon/intron organization of the *Anopheles* 21F and *Drosophila* 67A8 PGRP loci. Exons coding for PGRP domains are colored. Arrowheads indicate introns as positioned in Fig. 2C. Numbers and letters designate PGRP domains included in alternative isoforms. (**B**) Intra- and interspecies conservation of introns in PGRP domains shown in 2A. Genes have maintained identical intron/exon boundaries, except that the *Drosophila* PGRP-LCa and -LFw domains lack introns possibly lost secondarily. *Anopheles* PGRP-LC has an additional exon in each of the PGRP domains. Amino acids encoded by codons spanning intron boundaries are boxed. A cysteine pairing conserved in almost all known PGRPs is highlighted in red; one is changed to a Tyr in the *Drosophila* PGRPSA of *semmelweis* mutants (*43*). (**C**) Phylogenetic analysis of the PGRP domains. In this and subsequent dendrograms, *Anopheles* genes/proteins are indicated as red branches, and *Drosophila* (blue), vertebrates (green), and invertebrates and common gene stems (black) are colored as shown; dots on nodes indicate orthologous pairs, and circles indicate orthologous groups. Numbers accompanying or grouping branches indicate chromosomal locations. (**D**) Expression profiles of PGRP isoforms in cultured cells challenged with *E. coli*, *S. aureus*, peptidoglycan (PGN), and H_2O_2. Color intensities indicate fold regulation relative to reference (naïve) cells (see Methods in supplemental material). Regulation values below 1.5-fold are masked.

are associated with immune effector cascades (e.g., the phenoloxidase cascade in Lepidoptera and hemolymph clotting in the horseshoe crab) or serve in development (e.g., Snake, Easter, and Stubble in *Drosophila*) (*19*). The *Anopheles* and *Drosophila* genomes encode 41 and 35 CLIPs, respectively, in four subfamilies (Fig. 4A). This apparent numerical conservation is deceptive, as only eight orthologous pairs and five OGs exist; numerical conservation appears to be the net effect of counterbalancing species-specific expansions. The developmental genes are conserved, unlike the single well-characterized *Drosophila* immune CLIP *Persephone*, which is not conserved in the mosquito.

Most serpins (SRPNs) are irreversible inhibitory substrates for proteases, often but not exclusively of the serine class. Noninhibitory serpins are less well characterized; some were shown to function in hormone transport or blood-pressure regulation. In mammals, serpins account for 10% of the plasma proteins and affect blood coagulation, fibrinolysis, phagocytosis, inflammation, microbial infection, and complement activation. The mosquito genome encodes 14 serpins, 10 of which are inhibitory. Again, gene expansions/losses result in species-specific diversification; only one orthologous pair and four OGs are evident (Fig. 4B). The *Drosophila* serpin encoded by the *nec* locus, which is a partner of the Persephone Clip-domain protease in the Toll-mediated antifungal response (*20*), also has no ortholog in the mosquito. The functions corresponding to Persephone and Nec must be served by independently evolved *Anopheles* CLIPs and SRPNs. The *Drosophila* serpin-27A (CG 11331), involved in control of melanization, forms an OG with three mosquito serpins, which constitute interesting potential modulators of prophenoloxidases (PPOs) (see below). In a separate study (*21*), we have determined that the mosquito SRPN10 (lacking a 1:1 ortholog, and initially named spi21F) is intracellular and has isoforms with distinct biochemical inhibitory specificities; thus, as in the *PGRPL* subfamily, alternative splicing augments SRPN diversification. Notably, one of these isoforms is greatly up-regulated in midgut cells during *Plasmodium* invasion.

Signal Transduction Pathways

Signal transduction pathways link recognition and amplification of the "danger" signal with transcriptional activation. In *Drosophila*, antimicrobial responses use two major signal transduction pathways, Toll and Imd (*3*), and at least in the mosquito a third pathway, STAT, is also involved (*22*). Here we will consider the well-characterized Toll pathway, which has both developmental and immune functions and engages many genes and families (Table 1 and supplementary material). Most steps in the pathway are served by well-conserved individual genes, presumably reflecting conservation of balanced functions. The signaling receptor family shows a modest level of diversification.

Anopheles has 11 *TOLL* genes (Fig. 4C), of which four (*TOLL 6*, *7*, *8*, and *9*) are unambiguous orthologs of *Drosophila* counterparts (*23*). However, orthologs of *Toll-2*, *-3*, and *-4* have not been detected in *Anopheles*, which shows instead a species-specific expansion (*TOLL 10* and *11*). Gene reduplication has also generated four mosquito genes—*TOLL 1A*, *1B*, *5A*, and *5B*—that together with the fruit fly *Toll-1* and *-5* genes form an interesting OG. The most parsimonious hypothesis is that single type 1 and 5 genes were ancestrally linked and that this pair reduplicated and translocated in the mosquito, forming the 1A/5A pair at chromosomal site 6C and the 1B/5B pair at 39C; in the fruit fly the ancestral pair separated to different locations. It remains to be determined whether the immune function of the *Drosophila* type 1 gene is ancestral and retained by both *Anopheles* 1A and 1B, and whether any of the type 5 genes have immune functions.

In *Drosophila*, Toll signal transduction is initiated by binding of a cleaved peptide ligand, Spaetzle, on the extracellular domain of Toll, the intracellular domain of which interacts with MyD88, Tube, and Pelle, probably forming a multimeric inactive protein kinase complex (*24*, *25*). Upon Spaetzle binding, Pelle phosphorylates (directly or indirectly) Toll, itself, and Cactus; Cactus phosphorylation causes release of the Rel transcription factors Dorsal and DIF, which translocate into the nucleus and activate numerous genes, including those encoding antifungal peptides (*26*). The intracellular pathway is intact in *A. gambiae*: We have identified single genes encoding orthologs of *MyD88*, *Tube*, *Pelle*, and *Cactus* (Table 1). Another Pelle-like domain is found in the COOH end of a predicted, unusually large, protein sequence whose NH_2-terminal part is homologous to Tube. The mosquito ortholog of *Dorsal*, *Gambif-1*, was identified previously (*27*), but surprisingly, no ortholog of *DIF* was found.

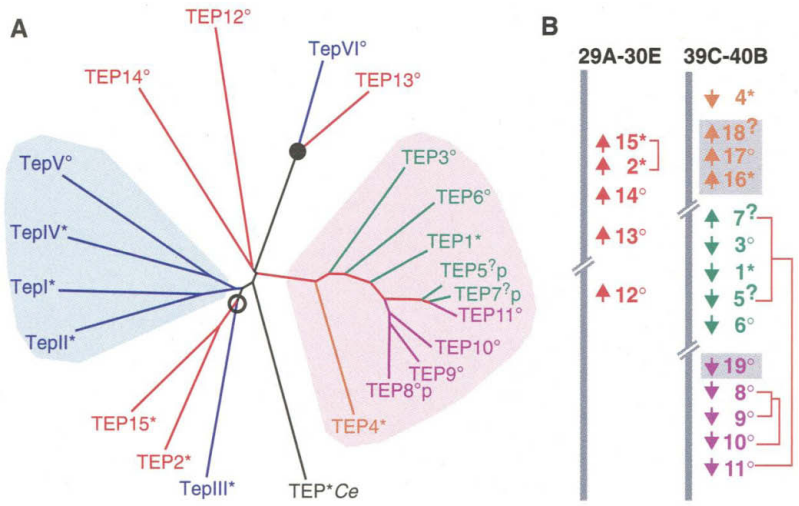

Fig. 3. Protein sequence comparison and chromosomal distribution of the *TEP* gene family. **(A)** Phylogenetic tree of complete sequence alignment. In this and subsequent figures, shading indicates gene expansions in *A. gambiae* (pink) and in *D. melanogaster* (blue). **(B)** *Anopheles* predicted *TEPs* (arrows) are physically located in four clusters and one isolated locus [identified by consistent colors in (A) and (B)]. Closest *Anopheles* paralogs are connected with brackets. Putative haplotypes are shaded in gray and are not discussed further. Superscript symbols after names indicate that the thioester motif is (*) present, (°) absent, or (?) unknown; p, partial sequence. Color scheme: blue, *D. melanogaster*; red, orange, green, and purple, *A. gambiae*; black, other invertebrates.

Effector Response Systems

After microbial recognition, signal modulation, and transduction, the transcriptional responses engage a large number of genes, including many with unknown function (*26*). However, three broad categories of effector systems are well recognized: antimicrobial peptides, the phenoloxidase-dependent melanization system, and the system of apoptosis-related genes. All three systems show a marked paucity of orthologs (Table 1).

Prophenoloxidases (*PPOs*): Melanization is an important immune response in insects and crustacea, and possibly in other arthropod classes. PPO proenzymes circulate through the hemolymph and, upon activation by clip domain proteases, catalyze key steps in the synthesis of melanin, thereby promoting cuticle sclerotization, wound healing, and melanotic encapsulation of pathogens (*28*, *29*); recently PPOs were also associated with hemolymph clotting (*30*). The genes show no signal peptide signature, suggesting that PPOs are released not by secretion but by rupture of hemocytes.

The *A. gambiae* genome encodes nine PPOs, threefold as many as the *Drosophila* genome. Six of the genes have been described (*31–33*) and numbered in order of discovery (*33*). Melanotic encapsulation of *Plasmodium* in refractory mosquitoes (*2*) suggests that antiparasitic defense may be one function of extra mosquito genes. Interestingly, the newly discovered *PPO9* is strongly induced in blood-fed *A. gambiae* (*34*) and may facilitate melanotic encapsulation. Other potential functions of extra genes may be to rapidly repair injuries endured by the swollen blood-fed mosquitoes, or by larvae living in swiftly running rainwater. The mosquito eggshell is also tanned after fertilization, and the adult mosquito cuticle and scales are more broadly melanized than those of *Drosophila*. Interestingly, most *Anopheles* PPO genes are part of a major expansion that may have occurred early in the mosquito lineage (Fig. 4D). Consistent with this hypothesis, all PPOs from other mosquitoes cluster with the *A. gambiae* genes. The sole exception is the *Anopheles PPO1* gene, which appears to be primitive; it clusters together with two members from *Drosophila* and one each from the fleshfly *Sarcophaga* and the beetle *Tenebrio*.

Other effector systems. Antimicrobial peptides (AMPs) are produced in the fat body, hemocytes, and epithelial tissues. Several hundred are now described, and their rapid evolution has been noted. The most important families are the widely distributed anti-Gram$^+$ insect defensins (DEF) and the predominantly anti-Gram$^-$ cecropins (CEC, in Diptera and Lepidoptera). Four *DEF* and four *CEC* genes exist in *Anopheles*, more numerous and more diverged than in *Drosophila*. Several other AMP families are specific to *Drosophila* but absent in *Anopheles*. Conversely, Gambicin (GAM1) (*35*) is mosquito specific. It appears that mosquitoes use few AMP families but may expand the spectrum of antibiotic activities, substantially diversifying both *DEF* and *CEC* sequences (see supplementary material).

The apoptotic machinery acts at three conceptually distinct levels: pro-apoptotic and anti-apoptotic regulators modulate the activity of initiator caspases (often by way of associated adaptor molecules), and these in turn activate effector caspases, the direct cell executioners. This system plays well-recognized roles in discarding unwanted cells during development but is also implicated in immunity. Intriguing evidence suggests that apoptosis and cell elimination is an important response of the *A. gambiae* midgut epithelium to *Plasmodium* invasion (*5, 21*). The initiator caspase DREDD, an important protease controlling morphogenic apoptosis, is also central to the *Drosophila* IMD (Immune Deficiency) pathway that resists Gram$^-$ bacterial infections (*36*). The number of caspase genes in *Anopheles* is somewhat higher than in humans (12 as compared to 11) and considerably higher than in *Drosophila* and *C. elegans* (7 and 3, respectively) (*37*). This overabundance is due to effector caspases, which have undergone a specific expansion, unlike initiator caspases (fig. S7). The negative regulators of caspases, IAPs (Inhibitor of Apoptosis Proteins), show both conservation (at least three orthologs with *Drosophila*) and specific expansion (two new IAPs) (fig. S8). The search for mosquito pro-apoptotic genes has been hampered by the extensive sequence diversification of the main players (*37*).

Diversified Gene Expression and Beyond

Immune gene sequence diversification suggests diversified functions. As a first step toward functional analysis, we evaluated the developmental regulation in whole mosquitoes, and in greater depth responsiveness to sterile injury or infections with bacteria (Gram$^+$ or Gram$^-$) and *Plasmodium*, for 24 representative mosquito genes belonging to 12 immunity families (Fig. 5), including one (*FBN*) described in a companion article (*4*). In immune-challenged mosquitoes, the expression profiles were specific to the gene and the particular challenge. By comparing sterile injury and bacterial infections of the mosquitoes, we determined that *E. coli* but not *S. aureus* specifically induces *GNBPB1*. Both types of bacteria induced *SRPN10* and *4* sequentially, whereas *SRPN9* was only induced late in *S. aureus* infection. *E. coli* induced *GAM1* late and robustly, whereas *S. aureus* induced *GAM1* only early and transiently. *CLIPB14* and *15* were up-regulated by both bacteria in a sustained manner, but *CLIPA6* was only induced transiently and modestly. Amongst six members of the *FBN*

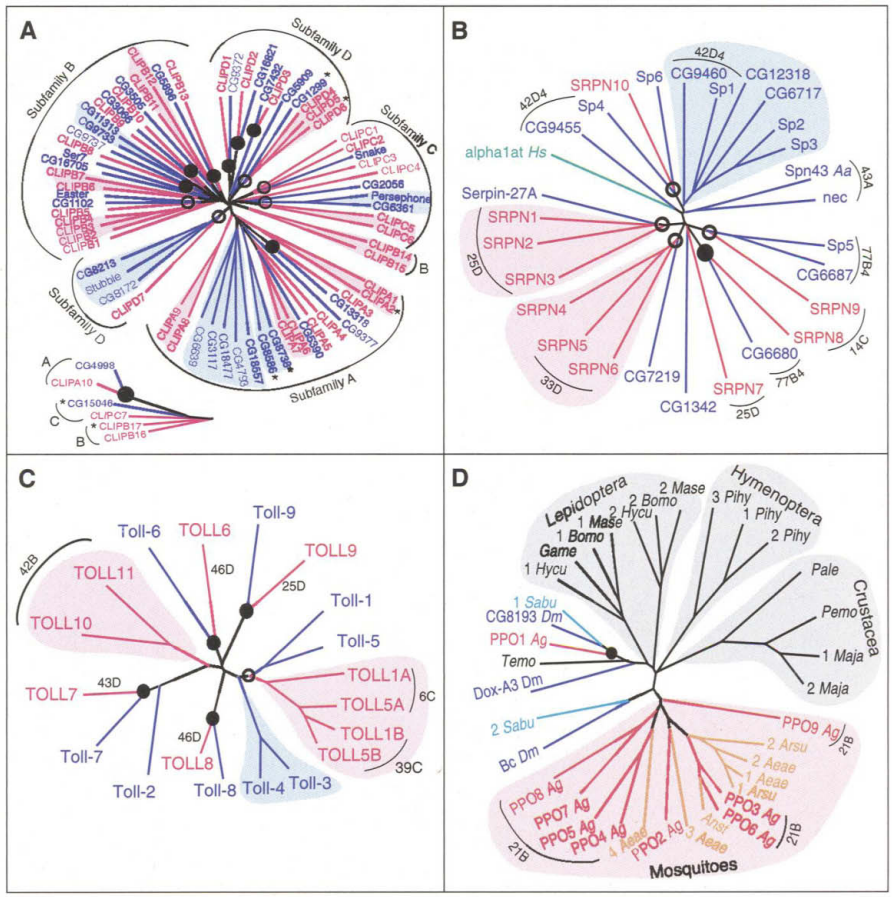

Fig. 4. Phylogenetic trees of CLIPs, SRPNs, TOLLs, and PPOs. (**A**) CLIP family. Tree was based on alignments that include the clip and serine protease domains. Proteins cluster into subfamilies A, B, C, and D and five hybrid sequences (shown at bottom of figure). Note: Long insertions within the clip domain were omitted for the orthologs CLIPA10 and CG4998 before sequence alignment; CG4914 was not aligned because of exceptionally arranged C residues in the clip domain. *, proteins containing two clip domains. (**B**) Inhibitory SRPNs. (**C**) TOLL family. (**D**) PPO family. Gray shading indicates groups of proteins from Hymenoptera, Lepidoptera, and Crustacea. Light blue branches indicate proteins from the dipteran, *Sarcophaga bullata*. For taxonomic abbreviations, see supplementary material.

family, only *FBN9* showed a sustained induction by bacteria. Finally, *TEP3* and *4* were strongly induced by both bacteria, possibly with different kinetics. Induction by sterile injury was rare, transient, and usually late (possibly in response to inadvertent infection of the wound, or cell damage; see late induction of *CASPL2* and *IAPB1* in Fig. 5); exceptionally, *PGRPLB* and *SCRBQ2* were induced similarly by sterile injury and bacteria.

During the life cycle of the parasite in the mosquito, six different genes (*FBN9*, *23*, and *CLIPB14*; *SRPN9*, *10*, and *4*) were activated primarily at 28 hours after infection of the mosquito, i.e., specifically when the midgut epithelium is invaded by ookinetes. In contrast, the parasite caused sustained induction of *PGRPLB*, *TEP4*, and *CLIP15* throughout its life cycle in the vector, suggesting an ongoing systemic rather than epithelial immune response. A delayed induction of *CEC1* and *GNBPA1* (which was not seen with bacteria) apparently represents reactions to the oocyst and sporozoite stages of the parasite.

The developmental profiles (for individual versus pooled stages) indicated stage-specific gene expression or up-regulation in the absence of a specific challenge: for example, *SRPN9* in pupae and *SRPN10* in early larvae, *CLIP15* primarily in early larvae, *CLIP6* in late larvae, and *CLIP14* and *PGRPLB* in adults. Developmental regulation of *FBN* family members was prominent, with three members expressed preferentially in embryos and early larvae, one in late larvae and pupae, and two in the adults. *FBN9*, which was strongly inducible both by bacteria and during *Plasmodium* penetration of the midgut, proved to be adult specific.

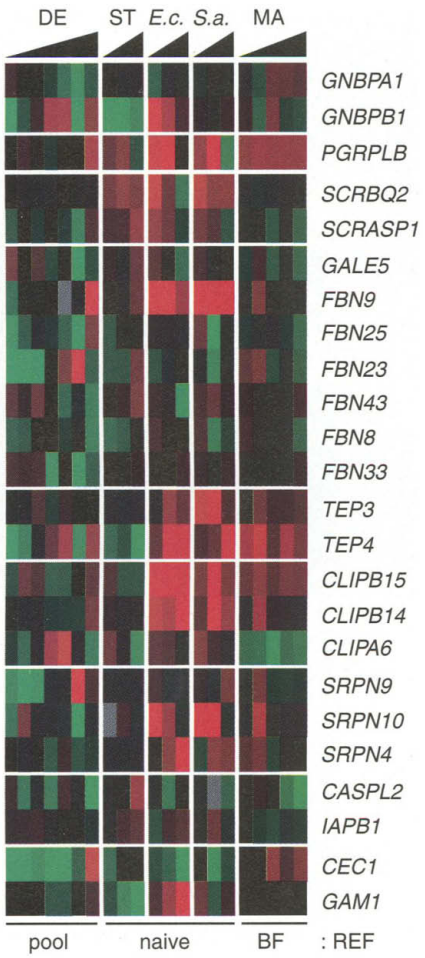

Fig. 5. Expression profiles of immunity gene family members. From left to right: Developmental (DE) expression profiles were examined at embryonic, 1st, 2nd, 3rd, 4th instar larval, pupal, and adult stages. Adult female mosquitoes were pricked with a sterile needle (ST) or infected with *E. coli* (*E. c.*) or *S. aureus* (*S. a.*), and assayed at 6, 12, and 24 hours after treatment. Mosquitoes were infected with malaria (MA), and expression profiles were examined at 24 hours, 28 hours, 6 days, 11 days, and 16 days after infection. Green- and red-colored data points of increasing intensity indicate up- and down-regulation relative to the reference (REF) samples, respectively. Regulation values below 1.5-fold are masked in black; gray represents missing points.

Concluding Remarks

The newly available genome sequence has created unprecedented opportunities for mosquito research. Genomic expression profiling will be facilitated by a consortium that is developing standarized whole-genome microarrays. Tools for reverse genetic analysis will be critically important. Hemocyte-like cell lines (*33*), coupled with in vitro transient and stable, transposable element–mediated transfection/transformation, are already in place (*38*). Germline transformation has been accomplished for both *A. stephensi* (*39*) and *A. gambiae* (*40*), and more sophisticated methodologies for gene disruption and conditional gain- and loss-of-function analysis are becoming available (*41*). Most recently, a convenient RNA interference–mediated approach for functional gene disruption by direct injection of double-stranded RNA has been developed (*42*). Phenotypic as well as genome-scale analysis of immune-related genes is now feasible for the malaria mosquito.

References and Notes

1. G. Dimopoulos, H. M. Muller, E. A. Levashina, F. C. Kafatos, *Curr. Opin. Immunol.* **13**, 79 (2001).
2. F. H. Collins *et al.*, *Science* **234**, 607 (1986).
3. J. A. Hoffmann, J.-M. Reichhart, *Nature Immunol.* **3**, 121 (2002).
4. E. M. Zdobnov, *Science* **298**, 149 (2002).
5. Y. S. Han, J. Thompson, F. C. Kafatos, C. Barillas-Mury, *EMBO J.* **19**, 6030 (2000).
6. C. A. Janeway Jr., R. Medzhitov, *Annu. Rev. Immunol.* **20**, 197 (2002).
7. T. Werner *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 13772 (2000).
8. K. M. Choe, T. Werner, S. Stoven, D. Hultmark, K. V. Anderson, *Science* **296**, 359 (2002).
9. M. Gottar *et al.*, *Nature* **416**, 640 (2002).
10. M. Ramet, P. Manfruelli, A. Pearson, B. Mathey-Prevot, R. A. Ezekowitz, *Nature* **416**, 644 (2002).
11. E. A. Levashina *et al.*, *Cell* **104**, 709 (2001).
12. F. Oduol, J. Xu, O. Niare, R. Natarajan, K. D. Vernick, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11397 (2000).
13. M. Lagueux, E. Perrodou, E. A. Levashina, M. Capovilla, J. A. Hoffmann, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11427 (2000).
14. Y. S. Kim *et al.*, *J. Biol. Chem.* **275**, 32721 (2000).
15. G. Dimopoulos, A. Richman, H. M. Muller, F. C. Kafatos, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 11508 (1997).
16. N. C. Franc, J. L. Dimarcq, M. Lagueux, J. Hoffmann, R. A. Ezekowitz, *Immunity* **4**, 431 (1996).
17. G. Dimopoulos, D. Seeley, A. Wolf, F. C. Kafatos, *EMBO J.* **17**, 6115 (1998).
18. P. Ligoxygakis, N. Pelte, J. A. Hoffmann, J. M. Reichhart, *Science* **297**, 114 (2002).
19. H. Jiang, M. R. Kanost, *Insect Biochem. Mol. Biol.* **30**, 95 (2000).
20. E. A. Levashina *et al.*, *Science* **285**, 1917 (1999).
21. A. Danielli, unpublished data.
22. C. Barillas-Mury, Y. S. Han, D. Seeley, F. C. Kafatos, *EMBO J.* **18**, 959 (1999).
23. C. Luna, X. Wang, Y. Huang, J. Zhang, L. Zheng, *Insect Biochem. Mol. Biol.* **32**, 1171 (2002).
24. B. Shen, J. L. Manley, *Development* **125**, 4719 (1998).
25. S. Tauszig-Delamasure, H. Bilak, M. Capovilla, J. A. Hoffmann, J. L. Imler, *Nature Immunol.* **3**, 91 (2002).
26. E. De Gregorio, P. T. Spellman, P. Tzou, G. M. Rubin, B. Lemaitre, *EMBO J.* **21**, 2568 (2002).
27. C. Barillas-Mury *et al.*, *EMBO J.* **15**, 4691 (1996).
28. S. C. Lai, C. C. Chen, R. F. Hou, *J. Med. Entomol.* **39**, 266 (2002).
29. B. T. Beerntsen, A. A. James, B. M. Christensen, *Microbiol. Mol. Biol. Rev.* **64**, 115 (2000).
30. T. Nagai, S. Kawabata, *J. Biol. Chem.* **275**, 29264 (2000).
31. H. Jiang, Y. Wang, S. E. Korochkina, H. Benes, M. R. Kanost, *Insect Biochem. Mol. Biol.* **27**, 693 (1997).
32. W. J. Lee *et al.*, *Insect Mol. Biol.* **7**, 41 (1998).
33. H. M. Muller, G. Dimopoulos, C. Blass, F. C. Kafatos, *J. Biol. Chem.* **274**, 11727 (1999).
34. H. M. Muller, unpublished data.
35. J. Vizioli *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 12630 (2001).
36. F. Leulier, A. Rodriguez, R. S. Khush, J. M. Abrams, B. Lemaitre, *EMBO Rep.* **1**, 353 (2000).
37. S. Y. Vernooy *et al.*, *J. Cell Biol.* **150**, F69 (2000).
38. F. Catteruccia *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 6236 (2000).
39. F. Catteruccia *et al.*, *Nature* **405**, 959 (2000).
40. G. L. Grossman *et al.*, *Insect Mol. Biol.* **10**, 597 (2001).
41. G. Lycett, F. C. Kafatos, T. G. Loukeris, unpublished data.
42. S. Blandin, L. F. Moita, F. C. Kafatos, E. A. Levashina, *EMBO Rep.* **3**, 852 (2002).
43. T. Michel, J. M. Reichhart, J. A. Hoffmann, J. Royet, *Nature* **414**, 756 (2001).

44. We acknowledge the collaborative spirit of the *Anopheles* Genome Project and the importance of the genome sequence determination performed by the Celera Genomics, Genoscope, and TIGR teams. We are grateful for the support provided by the major funders: National Institutes of Health, the National Institute of Allergy and Infectious Disease (USA), National Science Foundation (USA), the French Ministry of Research, our institutions, and the European Union. G.K.C. and D.V. were supported by Marie-Curie fellowships. The long-standing support by Tropical Disease Research–TDR (WHO) and the John D. and Catherine T. MacArthur Foundation was also essential for the project. F.C.K. dedicates this article to the memory of Anne Gruner Schlumberger.

Supporting Online Material

www.sciencemag.org/cgi/content/full/298/5591/159/DC1
Methods
SOM Text
Figs. S1 to S10
Table S1
References and Notes

8 August 2002; accepted 3 September 2002