[K_m ≈ 100 to 1000 μM (*17*)] and to *S. cerevisiae* hexose transporters' apparent affinity for glucose [K_m ≈ 1000 to 10,000 μM (*19*)]. Therefore, the cellodextrin transport systems should more effectively maintain soluble sugar levels below the concentration at which they inhibit fungal cellulases [inhibition constant (K_i) of cellobiose ≈ 19 to 410 μM (*20*)].

With little optimization, yeast expressing *cdt-1* and *gh1-1* fermented cellobiose with an ethanol yield of 0.441 ± 0.001 (grams of ethanol/grams of glucose ± SD), which is 86.3% of the theoretical value (Fig. 2A) (*21*). This yield is close to present industrial yields of ethanol from glucose of 90 to 93% (*22*). Yeasts expressing a cellodextrin transport system markedly improve the efficiency of SSF reactions by reducing the steady-state concentration of both cellobiose and glucose and by increasing the ethanol production rate (Fig. 2, B and C). The addition of a cellodextrin transport system to biofuel-producing strains of yeast (Fig. 3) overcomes a major bottleneck to fermentation of lignocellulosic feedstocks and probably will help to make cellulosic biofuels economically viable.

References and Notes

1. United States Department of Agriculture (USDA), "The economic feasibility of ethanol production from sugar in the United States" (USDA, Washington, DC, 2006).
2. E. M. Rubin, *Nature* **454**, 841 (2008).
3. G. Stephanopoulos, *Science* **315**, 801 (2007).
4. M. E. Himmel *et al.*, *Science* **315**, 804 (2007).
5. L. R. Lynd, P. J. Weimer, W. H. van Zyl, I. S. Pretorius, *Microbiol. Mol. Biol. Rev.* **66**, 506 (2002).
6. Y. Sun, J. Cheng, *Bioresour. Technol.* **83**, 1 (2002).
7. Z. Xin, Q. Yinbo, G. Peiji, *Enzyme Microb. Technol.* **15**, 62 (1993).
8. C. Tian *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 22157 (2009).
9. Materials and methods are available as supporting material on *Science* Online.
10. Y. Noguchi *et al.*, *Appl. Microbiol. Biotechnol.* **85**, 141 (2009).
11. F. Martin *et al.*, *Nature* **464**, 1033 (2010).
12. S. N. Freer, *Appl. Environ. Microbiol.* **57**, 655 (1991).
13. A. Vanden Wymelenberg *et al.*, *Appl. Environ. Microbiol.* **76**, 3599 (2010).
14. K. M. Bhat, R. Maheshwari, *Appl. Environ. Microbiol.* **53**, 2175 (1987).
15. J. Doran-Peterson *et al.*, *Methods Mol. Biol.* **581**, 263 (2009).
16. R. E. T. Drissen, R. H. W. Maas, J. Tramper, H. H. Beeftink, *Biocatalysis Biotransform.* **27**, 27 (2009).
17. M. Chauve *et al.*, *Biotechnol. Biofuels* **3**, 3 (2010).
18. K. A. Skinner, T. D. Leathers, *J. Ind. Microbiol. Biotechnol.* **31**, 401 (2004).
19. E. Reifenberger, E. Boles, M. Ciriacy, *Eur. J. Biochem.* **245**, 324 (1997).
20. S. P. Voutilainen *et al.*, *Biotechnol. Bioeng.* **101**, 515 (2008).
21. F. W. Bai, W. A. Anderson, M. Moo-Young, *Biotechnol. Adv.* **26**, 89 (2008).
22. L. C. Basso, H. V. de Amorim, A. J. de Oliveira, M. L. Lopes, *FEM. Yeast Res.* **8**, 1155 (2008).

23. We thank J. Doudna, M. Marletta, J. Taylor, T. Bruns, and C. Phillips for helpful discussions and comments on the manuscript; M. Toews for help with growth assays; S. Bauer and A. Ibanez for help with analytical methods; and C. Anderson for help with confocal microscopy. This work was supported by funding from the Energy Biosciences Institute to J.H.D.C. and N.L.G. The Regents of the University of California, the authors, and British Petroleum Technology Ventures (through the Energy Biosciences Institute) have submitted a patent for the use of cellodextrin transporters in fermenting organisms for the use of plant biomass.

Supporting Online Material

www.sciencemag.org/cgi/content/full/science.1192838/DC1
Materials and Methods
Figs. S1 to S8
References

26 May 2010; accepted 25 August 2010
Published online 9 September 2010;
10.1126/science.1192838
Include this information when citing this paper.

Sequencing of *Culex quinquefasciatus* Establishes a Platform for Mosquito Comparative Genomics

Peter Arensburger,[1]* Karine Megy,[2] Robert M. Waterhouse,[3,4] Jenica Abrudan,[5] Paolo Amedeo,[6] Beatriz Antelo,[7] Lyric Bartholomay,[8] Shelby Bidwell,[9] Elisabet Caler,[6] Francisco Camara,[9] Corey L. Campbell,[10] Kathryn S. Campbell,[11] Claudio Casola,[12] Marta T. Castro,[13] Ishwar Chandramouliswaran,[6] Sinéad B. Chapman,[14] Scott Christley,[5] Javier Costas,[15] Eric Eisenstadt,[6] Cedric Feschotte,[16] Claire Fraser-Liggett,[17] Roderic Guigo,[9] Brian Haas,[14] Martin Hammond,[2] Bill S. Hansson,[18] Janet Hemingway,[19] Sharon R. Hill,[20] Clint Howarth,[14] Rickard Ignell,[20] Ryan C. Kennedy,[5] Chinnappa D. Kodira,[21] Neil F. Lobo,[5] Chunhong Mao,[22] George Mayhew,[23] Kristin Michel,[24] Akio Mori,[5] Nannan Liu,[25] Horacio Naveira,[26] Vishvanath Nene,[17,27] Nam Nguyen,[16] Matthew D. Pearson,[14] Ellen J. Pritham,[16] Daniela Puiu,[28] Yumin Qi,[22] Hilary Ranson,[19] Jose M. C. Ribeiro,[29] Hugh M. Roberston,[30] David W. Severson,[5] Martin Shumway,[29] Mario Stanke,[31] Robert L. Strausberg,[6] Cheng Sun,[16] Granger Sutton,[6] Zhijian (Jake) Tu,[22] Jose Manuel C. Tubio,[7] Maria F. Unger,[5] Dana L. Vanlandingham,[33] Albert J. Vilella,[2] Owen White,[17] Jared R. White,[14] Charles S. Wondji,[19] Jennifer Wortman,[17] Evgeny M. Zdobnov,[3,4,33] Bruce Birren,[14] Bruce M. Christensen,[23] Frank H. Collins,[5] Anthony Cornel,[32] George Dimopoulos,[35] Linda I. Hannick,[6] Stephen Higgs,[33] Gregory C. Lanzaro,[34] Daniel Lawson,[2] Norman H. Lee,[36] Marc A. T. Muskavitch,[14,37,38] Alexander S. Raikhel,[1] Peter W. Atkinson[1]

Culex quinquefasciatus (the southern house mosquito) is an important mosquito vector of viruses such as West Nile virus and St. Louis encephalitis virus, as well as of nematodes that cause lymphatic filariasis. *C. quinquefasciatus* is one species within the *Culex pipiens* species complex and can be found throughout tropical and temperate climates of the world. The ability of *C. quinquefasciatus* to take blood meals from birds, livestock, and humans contributes to its ability to vector pathogens between species. Here, we describe the genomic sequence of *C. quinquefasciatus*: Its repertoire of 18,883 protein-coding genes is 22% larger than that of *Aedes aegypti* and 52% larger than that of *Anopheles gambiae* with multiple gene-family expansions, including olfactory and gustatory receptors, salivary gland genes, and genes associated with xenobiotic detoxification.

Mosquitoes are the most important vectors of human disease and are responsible for the transmission of pathogens that cause malaria (*Anopheles*), yellow fever and dengue (*Aedes*), as well as lymphatic filariasis and encephalitis viruses (*Culex*, *Aedes*, *Anopheles*). Sequencing the *Anopheles gambiae* and *Aedes aegypti* genomes has provided important insights into the genomic diversity underlying the complexity of mosquito biology (*1*, *2*). We describe the sequencing of the *Culex quinquefasciatus* (the southern house mosquito) genome, which offers a reference genome from the third major taxonomic group of disease-vector mosquitoes. With more than 1200 described species, *Culex* is the most diverse and geographically widespread of these three mosquito genera. Apart from contributing to the spread of West Nile encephalitis, it also transmits St. Louis encephalitis and other viral diseases and is a major vector of the parasitic *Wuchereria bancrofti* nematode that has caused the majority of the 120 million current cases of lymphatic filariasis (*3*).

Taxonomy of the *Culex pipiens* species complex is the subject of a long-standing debate, an issue complicated by the occurrence of viable species hybrids in many geographic areas [reviewed in (*4*, *5*)]. We followed the standard set by the National Center for Biotechnology Information and refer to the species sequenced here as *C. quinquefasciatus*. The Johannesburg strain of *C. quinquefaciatus* was established from a single pond in Johannesburg, South Africa—an area where the two taxa, *C. quinquefasciatus* and *C. pipiens*, were found to be sympatric [(*5*) therein described as subspecies *C. pipiens quinquefasciatus* and *C. pipiens pipiens*] but have remained much more genetically distinct than the same two sympatric taxa found in California.

We were able to map 9% of the *C. quinquefasciatus* genes (1768 genes) on the three chromosomes with the use of published and new *C. quinquefasciatus* and *Ae. aegypti* markers (*6*). Of these mapped genes, 803 had *An. gambiae* orthologs and 641 had *Drosophila melanogaster*

[1]Center for Disease Vector Research, University of California Riverside, Riverside, CA 92521, USA. [2]European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK. [3]University of Geneva Medical School, 1 rue Michel-Servet, 1211 Geneva, Switzerland. [4]Swiss Institute of Bioinformatics, 1 rue Michel-Servet, 1211 Geneva, Switzerland. [5]University of Notre Dame, Notre Dame, IN 46556, USA. [6]J. Craig Venter Institute, Rockville, MD 20850, USA. [7]Complexo Hospitalario Universitario de Santiago, Santiago de Compostela 15706, Spain. [8]Iowa State University, Ames, IA 50011, USA. [9]Center for Genomic Regulation, Universitat Pompeu Fabra, E-08003 Barcelona, Catalonia, Spain. [10]Colorado State University, Fort Collins, CO 80523, USA. [11]Harvard University, Cambridge, MA 02138, USA. [12]Indiana University, Bloomington, IN 47405–3700, USA. [13]Programa d'Epigenètica i Biologia del Càncer, Hospital Duran i Reynals, 08907 Hospitalet de Llobregat, Barcelona, Spain. [14]The Broad Institute, Cambridge, MA 02142, USA. [15]Fundación Pública Galega de Medicina Xenómica–Servizo Galego de Saúde, Santiago de Compostela 15706, Spain. [16]University of Texas Arlington, Arlington, TX 76019, USA. [17]Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD 21201, USA. [18]Max Planck Institute for Chemical Ecology, 07749 Jena, Germany. [19]Liverpool School of Tropical Medicine, Liverpool L3 5QA, UK. [20]Swedish University of Agricultural Sciences, 230 53 Alnarp, Sweden. [21]454 Life Sciences, the Roche Group, Branford, CT 06405, USA. [22]Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA. [23]University of Wisconsin, Madison, WI 53706, USA. [24]Kansas State University, Manhattan, KS 66506, USA. [25]Auburn University, Auburn, AL 36849, USA. [26]Departamento de Bioloxía Celular e Molecular, Universidade da Coruña, 15071 A, Coruña, Spain. [27]International Livestock Research Institute, Nairobi, Kenya. [28]Center for Bioinformatics and Computational Biology, University of Maryland, College Park, MD 20742, USA. [29]National Institutes of Health, Bethesda, MD 20892, USA. [30]University of Illinois Urbana-Champaign, Urbana, IL 61801, USA. [31]University of Göttingen, 37077 Göttingen, Germany. [32]University of Texas Medical Branch, Galveston, TX 77555, USA. [33]Imperial College London, South Kensington Campus, London SW7 2AZ, UK. [34]University of California Davis, Parlier, CA 93648, USA. [35]Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD 21205, USA. [36]George Washington University Medical Center, Washington, DC 20037, USA. [37]Boston College, Chestnut Hill, MA 02467, USA. [38]Harvard School of Public Health, Boston, MA 02115, USA.

*To whom correspondence should be addressed. E-mail: arensburger@gmail.com

orthologs, consistent with the established species phylogeny (Fig. 1A). Examining correlations between chromosomal arms indicated whole-chromosome conservation between *C. quinquefasciatus*, *An. gambiae*, and *D. melanogaster* (Fig. 1B) (*6*), whereas—and as suggested from earlier work (*7*)—*Ae. aegypti* appears to have experienced an arm exchange between the two longest chromosomes after the *Aedes*/*Culex* divergence (fig. S1).

A significant fraction of the assembled *C. quinquefasciatus* genome (29%) was composed of transposable elements (TEs) (fig. S2). This amount is less than the TE fraction of *Ae. aegypti* (42 to 47%), but greater than that of *An. gambiae* (11 to 16%) (*1*, *2*, *6*), suggesting an increased level of TE activity and/or reduced intensity of selection against TE insertions in the two culicinae lineages since their divergence from the *An. gambiae* lineage. A comparative analysis of the age distribution of the different TE types in the three sequenced mosquito genomes revealed that retrotransposons have consistently been the dominant TE type in the *Ae. aegypti* lineage over time (fig. S3). More recently, retrotransposons have become the predominant type of TEs active in all three species.

The *C. quinquefasciatus* repertoire of 18,883 protein-coding genes is 22% larger than that of *Ae. aegypti* (15,419 genes) and 52% larger than that of *An. gambiae* (12,457 genes) (Fig. 1C). Our estimated gene number combines ab initio and similarity-based predictions from three independent automated pipelines, optimizing gene identification (*6*). The relative increase in *C. quinquefasciatus* gene number is explained in part by the presence of substantially more expanded gene families, including olfactory and gustatory receptors, immune-related genes, and genes with possible xenobiotic detoxification functions (table S1). Expert curation of selected gene families revealed expansions in cytosolic glutathione transferases and a substantial expansion of cytochrome P450s. A large cytochrome P450 repertoire may reflect adaptations to polluted larval habitats and may have played a role in rendering this species particularly adaptable to evasion of insecticide-based control programs, with several *C. quinquefasciatus* P450s being associated with resistance (*8*, *9*).

Mosquitoes are the subject of intense efforts aimed at designing novel vector control methods that are often based on the ability of the insect to sense its environment (*10*, *11*). *C. quinquefasciatus* has the largest number of olfactory-receptor–related genes (180) of all dipteran species examined to date (table S1). This expansion may reflect culicine olfactory behavioral diversity, with particular regard to host and oviposition site choice. *C. quinquefasciatus* females are opportunistic feeders, being able to detect and feed on birds, humans, and livestock, depending on their availability. This plasticity in feeding behavior contributes to the ability of *C. quinquefasciatus* to vector pathogens, such as West Nile virus and St. Louis encephalitis virus, from birds to humans. The repertoire of gustatory receptors, which are known to mediate perception of both odorants and tastants (*12*), has also expanded in *C. quinquefasciatus*, primarily through a large alternatively spliced gene locus.

The saliva of blood-sucking arthropods contains a complex cocktail of pharmacologically active components that disarm host hemostasis (*13*). The ability of *C. quinquefasciatus* to feed on birds, humans, and livestock would suggest that it contains an expanded number of proteins that would increase its ability to imbibe blood from multiple host species. Consistent with this idea, a large protein family

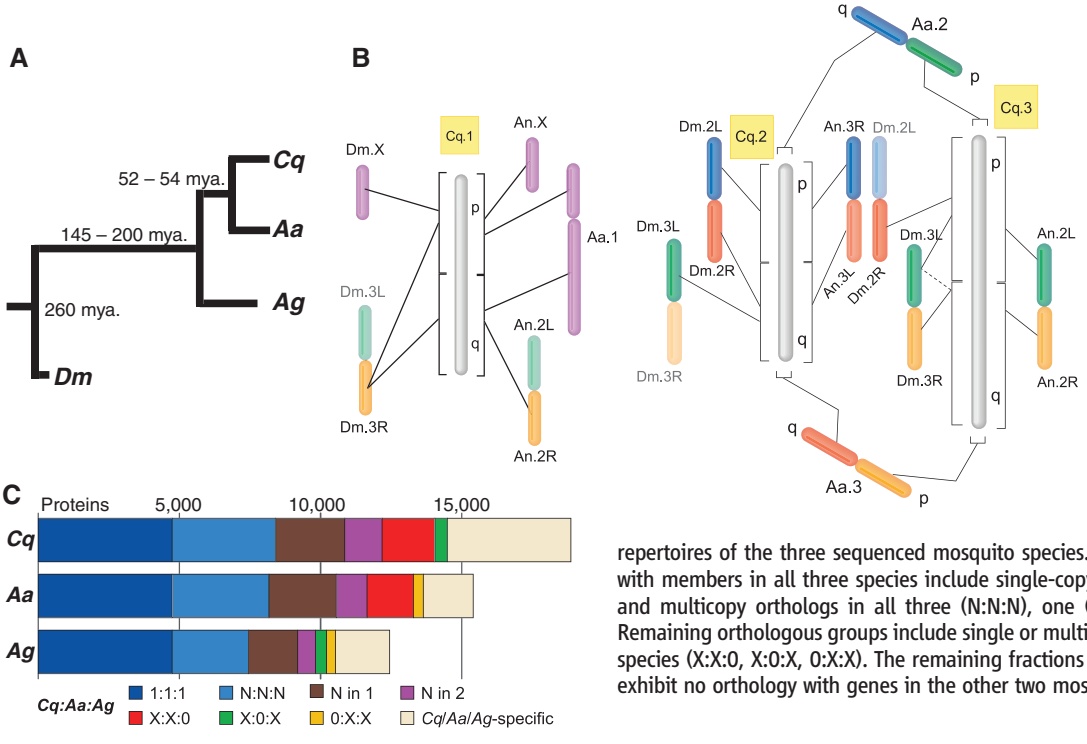

Fig. 1. (**A**) Codon-based estimates of DNA substitutions along the mosquito phylogeny: *C. quinquefasciatus* (*Cq*), *Ae. aegypti* (*Aa*), and *An. gambiae* (*Ag*) with *D. melanogaster* (*Dm*) as an outgroup. Dates of divergence were taken from previous studies (*6*). mya, million years ago. (**B**) Chromosomal synteny between *C. quinquefasciatus*, *Ae. aegypti*, *An. gambiae*, and *D. melanogaster*. Solid lines indicate main orthologous chromosomes; the dashed line denotes secondary orthologous chromosomes. Colors indicate syntenic chromosome arms. Chromosomes are not drawn to scale. (**C**) Orthology delineation among the protein-coding gene repertoires of the three sequenced mosquito species. Categories of orthologous groups with members in all three species include single-copy orthologs in each species (1:1:1) and multicopy orthologs in all three (N:N:N), one (N in 1), or two (N in 2) species. Remaining orthologous groups include single or multicopy groups with genes in only two species (X:X:0, X:0:X, 0:X:X). The remaining fractions in each species (*Cq*/*Aa*/*Ag*-specific) exhibit no orthology with genes in the other two mosquitoes.

unique to the *Culex* genus, the 16.7 kD family, was previously discovered after salivary transcriptome analysis (*13*). The genome of *C. quinquefasciatus* revealed 28 additional members of this family.

We have outlined and quantified general similarity and differences at the chromosomal and genomic levels between three disease-vector mosquito genomes, building a foundation for more in-depth future analyses. We found substantial differences in the relative abundance of TE classes among the three mosquitoes with sequenced genomes. Most unexpectedly, this study revealed numerous instances of expansion of *C. quinquefasciatus* gene families compared with *An. gambiae* and the more closely related *Ae. aegypti*. The consequent diversity in many different genes may be an important factor that led to the wide geographic distribution of *C. quinquefasciatus*.

References and Notes

1. R. A. Holt *et al.*, *Science* **298**, 129 (2002).
2. V. Nene *et al.*, *Science* **316**, 1718 (2007); published online 17 May 2007 (10.1126/science.1138878).
3. World Health Organization, *WHO Wkly. Epidemiol. Rec.* **84**, 437 (2009).
4. E. B. Vinogradova, Culex pipiens pipiens *Mosquitoes: Taxonomy, Distribution, Ecology, Physiology, Genetics, Applied Importance and Control* (Pensoft, Sofia, Bulgaria, Moscow, Russia, 2000).
5. A. J. Cornel *et al.*, *J. Med. Entomol.* **40**, 36 (2003).
6. Materials and methods are available as supporting material on *Science* Online.
7. A. Mori, D. W. Severson, B. M. Christensen, *J. Hered.* **90**, 160 (1999).
8. S. Kasai, I. S. Weerashinghe, T. Shono, M. Yamakawa, *Insect Biochem. Mol. Biol.* **30**, 163 (2000).
9. O. Komagata, S. Kasai, T. Tomita, *Insect Biochem. Mol. Biol.* **40**, 146 (2010).
10. L. Alphey *et al.*, *Vector-Borne Zoonotic Dis.* **10**, 295 (2010).
11. M. Benedict *et al.*, *Vector-Borne Zoonotic Dis.* **8**, 127 (2008).
12. E. A. Hallem, A. Dahanukar, J. R. Carlson, *Annu. Rev. Entomol.* **51**, 113 (2006).
13. J. M. C. Ribeiro, B. Arcà, *Adv. Insect Physiol.* **37**, 59 (2009).

14. This work was supported by NIH grant HHSN266200400039C and by the National Institute of Allergy and Infectious Diseases, NIH, Department of Health and Human Services under contract numbers N01-AI-30071 and HHSN266200400001C. The assembled genome was deposited in the GenBank database with accession number AAWU00000000.

Supporting Online Material

www.sciencemag.org/cgi/content/full/330/6000/86/DC1
Materials and Methods
Figs. S1 to S12
Tables S1 to S14
References

5 May 2010; accepted 27 August 2010
10.1126/science.1191864

Pathogenomics of *Culex quinquefasciatus* and Meta-Analysis of Infection Responses to Diverse Pathogens

Lyric C. Bartholomay,[1]* Robert M. Waterhouse,[2,3,15]* George F. Mayhew,[4] Corey L. Campbell,[5] Kristin Michel,[6] Zhen Zou,[7] Jose L. Ramirez,[8] Suchismita Das,[8] Kanwal Alvarez,[7] Peter Arensburger,[9] Bart Bryant,[6,7] Sinead B. Chapman,[10] Yuemei Dong,[8] Sara M. Erickson,[4] S. H. P. Parakrama Karunaratne,[11,12] Vladimir Kokoza,[7] Chinnappa D. Kodira,[13] Patricia Pignatelli,[11] Sang Woon Shin,[7] Dana L. Vanlandingham,[14] Peter W. Atkinson,[9] Bruce Birren,[10] George K. Christophides,[15] Rollie J. Clem,[6] Janet Hemingway,[11] Stephen Higgs,[14] Karine Megy,[16] Hilary Ranson,[11] Evgeny M. Zdobnov,[2,3,15] Alexander S. Raikhel,[7] Bruce M. Christensen,[4] George Dimopoulos,[8] Marc A. T. Muskavitch[10,17,18]†

The mosquito *Culex quinquefasciatus* poses a substantial threat to human and veterinary health as a primary vector of West Nile virus (WNV), the filarial worm *Wuchereria bancrofti*, and an avian malaria parasite. Comparative phylogenomics revealed an expanded canonical *C. quinquefasciatus* immune gene repertoire compared with those of *Aedes aegypti* and *Anopheles gambiae*. Transcriptomic analysis of *C. quinquefasciatus* genes responsive to WNV, *W. bancrofti*, and non-native bacteria facilitated an unprecedented meta-analysis of 25 vector-pathogen interactions involving arboviruses, filarial worms, bacteria, and malaria parasites, revealing common and distinct responses to these pathogen types in three mosquito genera. Our findings provide support for the hypothesis that mosquito-borne pathogens have evolved to evade innate immune responses in three vector mosquito species of major medical importance.

The Southern house mosquito, *Culex quinquefasciatus*, is a geographically widespread, often abundant mosquito that is an epidemiologically important vector for an exceptionally diverse array of pathogens, including multiple arboviruses, filarial worms, and protozoa. *C. quinquefasciatus* transmits West Nile virus (WNV), St. Louis encephalitis virus, and other arboviruses, and acts as the most important vector of the causative agent of lymphatic filariasis, *Wuchereria bancrofti*, and *Plasmodium relictum*, an avian malaria parasite. Despite the public health importance of *C. quinquefasciatus*, knowledge of the insect's response capacities to this diverse array of pathogens is limited.

Availability of the *C. quinquefasciatus* genome sequence (*1*) enabled comparative phylogenomic analyses with *Aedes aegypti* (*2*), *Anopheles gambiae* (*3*), and *Drosophila melanogaster* (*4*) that identified 500 *C. quinquefasciatus* immunity genes from 39 (sub)families or processes (table S1). Conservation of *C. quinquefasciatus* gene family members follows the species phylogeny, showing greatest similarities with *A. aegypti*. Expansions of C-type lectins (CTLs), fibrinogen-related proteins (FREPs), and serine protease inhibitors (SRPNs) account for much of the 20 to 30% increase in *C. quinquefasciatus* immunity gene number compared with *A. aegypti* (417 genes) and *A. gambiae* (380 genes) (figs. S1 to S4). This apparent diversification in immune surveillance and immune signal amplification processes seems consistent with selection driven by polluted, microbially complex habitats in which *C. quinquefasciatus* oviposits and develops (*5*).

Whole genome microarray analysis revealed dynamic changes in infection response gene (IRG) transcription in WNV-infected mosquitoes (fig. S5). Significant changes are observed for 22 transcripts in the midgut and 309 in the carcass (i.e., the remainder of the body) at 7 days postinfection (dpi), with the greater number of IRGs in the latter apparently reflecting the diversity of infected cell and tissue types in the carcass. At 14 dpi, more IRGs are modulated in midgut (539) and carcass (490) when WNV infection has spread in midgut cells and has disseminated to the salivary glands (*6*). Few canonical immunity genes are represented among *C. quinquefasciatus* WNV IRGs (fig. S5). Five CTL genes within a *C. quinquefasciatus*–specific gene expansion (fig. S3) are up-regulated. Several genes related to the

[1]Department of Entomology, Iowa State University, Ames, IA 50011, USA. [2]Department of Genetic Medicine and Development, University of Geneva Medical School, 1 Rue Michel-Servet, 1211 Geneva, CH, Switzerland. [3]Swiss Institute of Bioinformatics, 1 Rue Michel-Servet, 1211 Geneva, CH, Switzerland. [4]Department of Pathobiological Sciences, University of Wisconsin, Madison, WI 53706, USA. [5]Microbiology, Immunology, and Pathology Department, Colorado State University, Fort Collins, CO 80523, USA. [6]Division of Biology, Arthropod Genomics Center, Molecular and Cellular Developmental Biology Program, Kansas State University, Manhattan, KS 66506, USA. [7]Department of Entomology, University of California, Riverside, CA 92521, USA. [8]W. Harry Feinstone Department of Molecular Microbiology and Immunology, Malaria Research Institute, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD 21205, USA. [9]Department of Entomology, Center for Disease Vector Research, University of California, Riverside, CA 92521, USA. [10]The Broad Institute, Cambridge MA 02142, USA. [11]Liverpool School of Tropical Medicine, Liverpool, L3 5QA, UK. [12]Faculty of Science and Department of Zoology, University of Peradeniya, Peradeniya 20400, LK, Sri Lanka. [13]454 Life Sciences, Branford, CT 06405, USA. [14]Pathology Department, University of Texas Medical Branch, Galveston, TX 77555, USA. [15]Division of Cell and Molecular Biology, Department of Life Sciences, Imperial College London, Exhibition Road, London SW7 2AZ, UK. [16]European Bioinformatics Institute (EMBL), Hinxton CB10 1SD Cambridge, UK. [17]Biology Department, Boston College, Chestnut Hill, MA 02467, USA. [18]Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, MA 02115, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: marc.muskavitch@bc.edu