2010

Sequencing of Culex quinquefasciatus establishes a platform for mosquito comparative genomics

Peter Arensburger
University of California - Riverside

Karine Megy
European Bioinformatics Institute

Robert M. Waterhouse
Imperial College London

Jenica Abrudan
University of Notre Dame

Paolo Amedeo
J. Craig Venter Institute

See next page for additional authors
Follow this and additional works at: http://lib.dr.iastate.edu/ent_pubs

Part of the Entomology Commons, and the Genomics Commons

The complete bibliographic information for this item can be found at http://lib.dr.iastate.edu/ent_pubs/159. For information on how to cite this item, please visit http://lib.dr.iastate.edu/howtocite.html.
Sequencing of Culex quinquefasciatus establishes a platform for mosquito comparative genomics

Abstract
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.

Disciplines
Entomology | Genomics

Comments
This is an author’s manuscript of an article from Science 330 (2010)L 88, doi:10.1126/science.1191864.

Authors

This article is available at Iowa State University Digital Repository: http://lib.dr.iastate.edu/ent_pubs/159
Sequencing of *Culex quinquefasciatus* establishes a platform for mosquito comparative genomics


1University of California Riverside, Riverside, CA 92521, USA 2European Bioinformatics Institute, Hinxton, Cambridge CB10 1SD, UK 3University of Geneva Medical School, 1 rue Michel-Servet, 1211 Geneva, Switzerland 4Swiss Institute of Bioinformatics, 1 rue Michel-Servet, 1211 Geneva, Switzerland 5University of Notre Dame, Notre Dame, IN 46556, USA 6J. Craig Venter institute, Rockville, MD 20850, USA 7Complexo Hospitalario Universitario de Santiago, Santiago de Compostela 15706, Spain 8Iowa State University, Ames, IA 50011, USA 9Center for Genomic Regulation, Universitat Pompeu Fabra, E-08003 Barcelona, Catalonia, Spain 10Colorado State University, Fort Collins CO 80523, USA 11Harvard University, Cambridge, MA 02138, USA 12Indiana University, Bloomington, IN 47405-3700, USA 13PEBC. Hospital Duran i Reynals. 08907 Hospital de Llobregat, Barcelona, Spain 14The Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA 15Fundación Pública Galega de Medicina Xenómica-SERGAS, Santiago de Compostela 15706, Spain 16University of Texas Arlington, Arlington, TX 76019, USA 17Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD, USA 18Max Planck Institute Chem. Ecol., D-07749 Jena, Germany 19Liverpool School of Tropical Medicine, Liverpool L3 5QA, UK 20Swedish University of Agricultural Sciences. 230 53 Alnarp, Sweden 21454 Life Sciences, a member of the Roche Group, Branford, CT 06405, USA 22Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA 23University of Wisconsin, Madison, WI 53706, USA 24Kansas State University, Manhattan, KS 66506, USA 25Auburn University, Auburn, AL 36849, USA 26Univ. A, Coruña, Spain 27International Livestock Research Institute, Nairobi, Kenya 28J. Craig Venter Institute, Rockville, MD 20850, USA 29Rockefeller University, New York, NY 10016, USA 30Auburn, Alabama 36849, USA 31University of California, Los Angeles, CA 90095, USA 32University of California, Berkeley, CA 94720, USA 33Institute for Genomic Research, Rockville, MD 20850, USA 34School of Biological Sciences, University of Oxford, Oxford, UK 35INRA, UMR Genoma, 78352, Jouy-en-Josas, France 36University of California, Berkeley, CA 94720, USA 37School of Agricultural and Environmental Sciences, University of Reading, Reading, Berks, RG6 6AJ, UK 38J. Craig Venter Institute, Rockville, MD 20850, USA

*To whom correspondence should be addressed. arensburger@gmail.com.*

Supporting Online Material
www.sciencemag.org
Materials and Methods
Figs. S1 – S12
Tables S1 – S14
Abstract

*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 of nematodes that cause lymphatic filariasis. It is one species within the *Culex pipiens* species complex and enjoys a distribution 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. We describe the genomic sequence of *C. quinquefasciatus*, its repertoire of 18,883 protein-coding genes is 22% larger than *Ae. aegypti* and 52% larger than *An. 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, responsible for the transmission of pathogens that cause malaria (*Anopheles*), yellow fever and dengue (*Aedes*), lymphatic filariasis (*Culex, Aedes, Anopheles*), and West Nile encephalitis (*Culex*). 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 over 1,200 described species, *Culex* is the most diverse and geographically widespread of these three mosquito genera, and apart from contributing to the spread of West Nile encephalitis, also transmits St. Louis encephalitis and other viral diseases, and is a major vector of the parasitic *Wuchereria bancrofti* nematode which causes the majority of 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 National Center for Biotechnology Information and refer to the species sequenced here as *C. quinquefasciatus*. The Johannesburg strain (JHB) of *C. quinquefasciatus* was established from a single pond in Johannesburg, South Africa, an area where two subspecies *C. pipiens quinquefasciatus* and *C. pipiens pipiens* are sympatric but have remained much more genetically distinct than the same two sympatric subspecies found in California (5).

We were able to map 9% of the *C. quinquefasciatus* genes (1,768 genes) on the three chromosomes using published and new *C. quinquefasciatus* and *Ae. aegypti* markers (6). Of these mapped genes, 803 had *An. gambiae* orthologs and 641 had *Drosophila melanogaster* 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 following the *Aedes/Culex* divergence (Fig. S1).
A significant fraction of the assembled \textit{C. quinquefasciatus} genome (29\%) was composed of transposable elements (TEs) (Fig. S2); less than the TE fraction of \textit{Ae. aegypti} (42\% – 47\%) but greater than that of \textit{An. gambiae} (11\% – 16\%) (1, 2, 6). This suggests 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 \textit{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 \textit{Ae. aegypti} lineage over time (Fig. S3). More recently, retrotransposons have become the predominant type of TEs active in all three species.

The \textit{C. quinquefasciatus} repertoire of 18,883 protein coding genes is 22\% larger than that of \textit{Ae. aegypti} (15,419) and 52\% larger than that of \textit{An. gambiae} (12,457) (Fig. 1C). Our estimated gene number combines \textit{ab initio} and similarity-based predictions from three independent automated pipelines, optimizing gene identification (6). The relative increase in \textit{C. quinquefasciatus} gene number is explained in part by the presence of significantly more expanded gene families including olfactory and gustatory receptors, immune-related genes, as well as 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 have played a role in rendering this species particularly adaptable to evasion of insecticide-based control programs, with several \textit{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). \textit{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. \textit{C. quinquefasciatus} females are opportunistic feeders, being able to detect and feed upon birds, humans and livestock depending on their availability. This plasticity in feeding behavior contributes to the ability of \textit{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 \textit{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 \textit{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, a large protein family unique to the \textit{Culex} genus, the 16.7 kDa family, was previously discovered following salivary transcriptome analysis (13). The genome of \textit{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 \textit{C. quinquefasciatus} gene families compared to \textit{An. gambiae} and the more closely related \textit{Ae. aegypti}. The consequent diversity in many different genes may be an important factor that led to the wide geographic distribution of \textit{C. quinquefasciatus}. 

\textit{Science. Author manuscript; available in PMC 2013 August 12.}
**Supplementary Material**

Refer to Web version on PubMed Central for supplementary material.

**Acknowledgments**

This work was supported by the National Institutes of Health grant HHSN266200400039C, and by the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services under contract numbers N01-AI-30071 and HHSN266200400001C. The assembled genome was deposited under Genbank accession AAWU00000000.

**References**

6. Information on materials and methods is available on Science Online.
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). B) Chromosomal synteny between *C. quinquefasciatus*, *Ae. aegypti*, *An. gambiae* and *D. melanogaster*. Plain lines indicate main orthologous chromosomes and dashed lines secondary orthologous chromosomes. Colors indicate syntenic chromosome arms. Chromosomes not 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 multi-copy orthologs in all three (N:N:N), one (N in 1), or two (N in 2) species. Remaining orthologous groups include single or multi-copy 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.