

# Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle

Ewen F. Kirkness<sup>a,1</sup>, Brian J. Haas<sup>a,2</sup>, Weilin Sun<sup>b</sup>, Henk R. Braig<sup>c</sup>, M. Alejandra Perotti<sup>d</sup>, John M. Clark<sup>e</sup>, Si Hyeock Lee<sup>f</sup>, Hugh M. Robertson<sup>b</sup>, Ryan C. Kennedy<sup>g,h</sup>, Eran Elhaik<sup>i</sup>, Daniel Gerlach<sup>j,k</sup>, Evgenia V. Kriventseva<sup>i,k</sup>, Christine G. Elsik<sup>l,3</sup>, Dan Graur<sup>i</sup>, Catherine A. Hill<sup>m</sup>, Jan A. Veenstra<sup>n</sup>, Brian Walenz<sup>a</sup>, José Manuel C. Tubío<sup>o</sup>, José M. C. Ribeiro<sup>p</sup>, Julio Rozas<sup>q</sup>, J. Spencer Johnston<sup>r</sup>, Justin T. Reese<sup>l</sup>, Aleksandar Popadic<sup>s</sup>, Marta Tojo<sup>t</sup>, Didier Raoult<sup>u</sup>, David L. Reed<sup>v</sup>, Yoshinori Tomoyasu<sup>w,4</sup>, Emily Krause<sup>w</sup>, Omprakash Mittapalli<sup>x</sup>, Venu M. Margam<sup>m</sup>, Hong-Mei Li<sup>b</sup>, Jason M. Meyer<sup>m</sup>, Reed M. Johnson<sup>b</sup>, Jeanne Romero-Severson<sup>g,y</sup>, Janice Pagel VanZee<sup>m</sup>, David Alvarez-Ponce<sup>z</sup>, Filipe G. Vieira<sup>q</sup>, Montserrat Aguadé<sup>q</sup>, Sara Guirao-Rico<sup>q</sup>, Juan M. Anzola<sup>l</sup>, Kyong S. Yoon<sup>e</sup>, Joseph P. Strycharz<sup>e</sup>, Maria F. Unger<sup>g,y</sup>, Scott Christley<sup>g,h</sup>, Neil F. Lobo<sup>g,y</sup>, Manfred J. Seufferheld<sup>z</sup>, NaiKuan Wang<sup>aa</sup>, Gregory A. Dasch<sup>bb</sup>, Claudio J. Struchiner<sup>cc</sup>, Greg Madey<sup>g,h</sup>, Linda I. Hannick<sup>a</sup>, Shelby Bidwell<sup>a</sup>, Vinita Joardar<sup>a</sup>, Elisabet Caler<sup>a</sup>, Renfu Shao<sup>dd</sup>, Stephen C. Barker<sup>dd</sup>, Stephen Cameron<sup>ee</sup>, Robert V. Bruggner<sup>g,h</sup>, Allison Regier<sup>g,h</sup>, Justin Johnson<sup>a</sup>, Lakshmi Viswanathan<sup>a</sup>, Terry R. Utterback<sup>a</sup>, Granger G. Sutton<sup>a</sup>, Daniel Lawson<sup>ff</sup>, Robert M. Waterhouse<sup>i,k</sup>, J. Craig Venter<sup>a</sup>, Robert L. Strausberg<sup>a</sup>, May R. Berenbaum<sup>b</sup>, Frank H. Collins<sup>g,y</sup>, Evgeny M. Zdobnov<sup>j,k,gg,1</sup>, and Barry R. Pittendrigh<sup>b,1,5</sup>

<sup>a</sup>J. Craig Venter Institute, Rockville, MD 20850; <sup>b</sup>Department of Entomology, University of Illinois, Urbana, IL 61801; <sup>c</sup>School of Biological Sciences, Bangor University, Bangor, Wales LL57 2UW, United Kingdom; <sup>d</sup>School of Biological Sciences, University of Reading, Reading RG6 6AS, United Kingdom; <sup>e</sup>Department of Veterinary and Animal Science, University of Massachusetts, Amherst, MA 01003; <sup>f</sup>Department of Agricultural Biotechnology, Seoul National University, Seoul, South Korea; <sup>g</sup>Eck Institute for Global Health, University of Notre Dame, Notre Dame, IN 46556; <sup>h</sup>Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN 46556; <sup>i</sup>Department of Biology and Biochemistry, University of Houston, Houston, TX 77204; <sup>j</sup>Department of Genetic Medicine and Development, University of Geneva Medical School, 1211 Geneva, Switzerland; <sup>k</sup>Swiss Institute of Bioinformatics, 1211 Geneva, Switzerland; <sup>l</sup>Department of Animal Science, Texas A&M University, College Station, TX 77843; <sup>m</sup>Department of Entomology, Purdue University, West Lafayette, IN 47907; <sup>n</sup>Centre National de la Recherche Scientifique, Centre Neurosciences Intégratives et Cognitives, University of Bordeaux, 33405 Talence Cedex, France; <sup>o</sup>Servicio de Hematología, Complejo Hospitalario Universitario de Santiago de Compostela, 15706 Santiago de Compostela, Spain; <sup>p</sup>Laboratory of Malaria and Vector Research, National Institutes of Health, Bethesda, MD 20892-8132; <sup>q</sup>Departament de Genètica, Universitat de Barcelona, 08028 Barcelona, Spain; <sup>r</sup>Department of Entomology, Texas A&M University, College Station, TX 77843; <sup>s</sup>Department of Biological Sciences, Wayne State University, Detroit, MI 48202; <sup>t</sup>Servicio de Anatomía Patológica, Complejo Hospitalario Universitario de Santiago de Compostela, 15706 Santiago de Compostela, Spain; <sup>u</sup>Unité des Rickettsies, 13385 Marseille Cedex 05, France; <sup>v</sup>Florida Museum of Natural History, University of Florida, Gainesville, FL 32611; <sup>w</sup>Division of Biology and Department of Entomology, Kansas State University, Manhattan, KS 66502; <sup>x</sup>Department of Entomology, Ohio Agriculture Research and Development Center/Ohio State University, Wooster, OH 44691; <sup>y</sup>Department of Biological Sciences, University of Notre Dame, Notre Dame, IN 46556; <sup>z</sup>Department of Crop Sciences, Urbana, IL 61801; <sup>aa</sup>Chung Hwa College of Medical Technology, Jen-Te Hsiang, Tainan 700, Taiwan; <sup>bb</sup>Centers for Disease Control and Prevention, Atlanta, GA 30333; <sup>cc</sup>Escola Nacional de Saúde Pública Sergio Arouca/Fundação Oswaldo Cruz and Instituto de Medicina Social Universidade do Estado do Rio de Janeiro, 4365 Rio de Janeiro, Brazil; <sup>dd</sup>School of Chemistry and Molecular Biosciences, University of Queensland, St. Lucia, Brisbane, Queensland 4072, Australia; <sup>ee</sup>Australian National Insect Collection and Commonwealth Scientific and Industrial Research Organization Entomology, Canberra ACT 2601, Australia; <sup>ff</sup>European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SD, United Kingdom; and <sup>gg</sup>Imperial College London, London SW7 2AZ, United Kingdom.

Contributed by May R. Berenbaum, April 14, 2010 (sent for review February 10, 2010)

**As an obligatory parasite of humans, the body louse (*Pediculus humanus humanus*) is an important vector for human diseases, including epidemic typhus, relapsing fever, and trench fever. Here, we present genome sequences of the body louse and its primary bacterial endosymbiont *Candidatus Riesia pediculicola*. The body louse has the smallest known insect genome, spanning 108 Mb. Despite its status as an obligate parasite, it retains a remarkably complete basal insect repertoire of 10,773 protein-coding genes and 57 microRNAs. Representing hemimetabolous insects, the genome of the body louse thus provides a reference for studies of holometabolous insects. Compared with other insect genomes, the body louse genome contains significantly fewer genes associated with environmental sensing and response, including odorant and gustatory receptors and detoxifying enzymes. The unique architecture of the 18 minicircular mitochondrial chromosomes of the body louse may be linked to the loss of the gene encoding the mitochondrial single-stranded DNA binding protein. The genome of the obligatory louse endosymbiont *Candidatus Riesia pediculicola* encodes less than 600 genes on a short, linear chromosome and a circular plasmid. The plasmid harbors a unique arrangement of genes required for the synthesis of pantothenate, an essential vitamin deficient in the louse diet. The human body louse, its primary endosymbiont, and the bacterial pathogens that it vectors all possess genomes reduced in size compared with their free-living close relatives. Thus, the body louse genome project offers unique information and tools to use in advancing understanding of coevolution among vectors, symbionts, and pathogens.**

Author contributions: E.F.K., B.J.H., J.M.C., H.M.R., R.C.K., E.E., C.G.E., D. Graur, C.A.H., B.W., J.M.C.T., J.M.C.R., J.R., J.S.J., J.T.R., M.T., D.L.R., O.M., V.M.M., J.M.M., J.R.-S., J.P.V., M.A., J.M.A., K.S.Y., J.P.S., M.F.U., S. Christley, N.F.L., G.A.D., C.J.S., G.M., S. Cameron, A.R., G.G.S., J.C.V., R.L.S., F.H.C., E.M.Z., and B.R.P. designed research; B.J.H., W.S., J.M.C., S.H.L., R.C.K., E.E., D. Gerlach, E.V.K., C.G.E., D. Graur, C.A.H., J.A.V., B.W., J.M.C.R., J.R., J.S.J., J.T.R., A.P., Y.T., M.T., D.R., D.L.R., E.K., O.M., H.-M.L., J.M.M., R.J., J.R.-S., J.P.V., D.A.-P., F.G.V., M.A., S.G.-R., J.M.A., K.S.Y., J.P.S., M.F.U., S. Christley, M.J.S., N.W., C.J.S., R.S., S.C.B., A.R., J.J., L.V., T.R.U., D.L., R.M.W., M.R.B., E.M.Z., and B.R.P. performed research; J.M.C., R.C.K., E.E., C.G.E., D. Graur, C.A.H., J.M.C.T., J.S.J., J.T.R., M.T., J.M.M., J.R.-S., J.M.A., K.S.Y., J.P.S., M.F.U., S. Christley, L.I.H., V.J., E.C., R.S., S.C.B., R.V.B., L.V., R.M.W., and E.M.Z. contributed new reagents or analytic tools; E.F.K., W.S., H.R.B., M.A.P., S.H.L., H.M.R., R.C.K., D. Gerlach, E.V.K., J.A.V., B.W., J.M.C.T., J.M.C.R., J.R., A.P., Y.T., M.T., D.L.R., H.-M.L., R.J., J.R.-S., D.A.-P., F.G.V., M.A., S.G.-R., K.S.Y., J.P.S., M.F.U., S. Christley, M.J.S., N.W., G.A.D., C.J.S., L.I.H., S.C.B., V.J., E.C., R.S., S. Cameron, R.V.B., A.R., J.J., D.L., R.M.W., M.R.B., F.H.C., E.M.Z., and B.R.P. analyzed data; and E.F.K., B.J.H., W.S., H.R.B., M.A.P., S.H.L., H.M.R., R.C.K., B.W., J.M.C.T., J.R., A.P., Y.T., M.T., D.R., D.L.R., O.M., R.J., D.A.-P., F.G.V., M.A., S.G.-R., J.M.A., K.S.Y., J.P.S., M.F.U., S. Christley, G.M., S. Cameron, A.R., R.M.W., M.R.B., F.H.C., E.M.Z., and B.R.P. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AAZO00000000, NC\_014109, and NC\_013962).

<sup>1</sup>E.F.K., E.M.Z., and B.R.P. contributed equally to this work.

<sup>2</sup>Present address: Broad Institute, 7 Cambridge Center, MA 02142.

<sup>3</sup>Present address: Department of Biology, Georgetown University, Washington, DC 20057.

<sup>4</sup>Present address: Department of Zoology, Miami University, 252 Pearson Hall, Oxford, OH 45056.

<sup>5</sup>To whom correspondence should be addressed. E-mail: pittendr@illinois.edu.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1003379107/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1003379107/-DCSupplemental).

Like their primate relatives, humans have had a long evolutionary association with parasitic sucking lice. Contact between sucking lice and primate hosts dates back at least 25 million years (1). Chimpanzee lice (*Pediculus schaeffi*) and human lice (*Pediculus humanus*) diverged from their common ancestors, as did chimpanzees (*Pan troglodytes*) and humans (*Homo sapiens*), 5–7 million years ago (2, 3). The two subspecies—the human body louse (*Pediculus humanus humanus* L.) and the head louse (*P. h. capitis* DG.)—are closely related obligate parasites that feed exclusively on human blood. Body lice likely evolved from head louse ancestors when humans began to wear clothing, which is required for egg deposition by body lice (4).

*P. h. humanus* has been of tremendous medical and social importance throughout human history. Of the two forms, only the body louse has been implicated as a vector of human disease and is the principal vector of epidemic typhus (*Rickettsia prowazekii*), relapsing fever (*Borrelia recurrentis*), and trench fever (*Bartonella quintana*) (5–9). In the United States as well as the rest of the world, body lice are primarily a concern in transient homeless populations, whereas head lice tend to infest populations of elementary school-aged children. Historically, epidemic typhus has been responsible for massive mortality in wartime (9); in contemporary times, major outbreaks of epidemic typhus are found primarily among refugees [e.g., in Burundi in 1996 (8)], but sporadic cases have also been observed in general populations in Russia (10), Peru, Algeria, and France (11).

Like all hematophagous lice, body lice depend on obligate endosymbionts to supplement their nutritionally deficient blood diet (12). The primary endosymbiont of *P. h. humanus* has been given the provisional name *Candidatus* Riesia pediculicola (13) (hereafter, Riesia). The body louse maintains organs called mycetomes that house the primary endosymbiont, except during passage to the ovaries for transovarial transmission (14). The tripartite interdependency of this bacterial endosymbiont, its body louse host, and the human host of the body louse seems to have coevolved over several million years (15).

Here, we present the genome sequences of the body louse and its coevolved primary endosymbiont. This genome, the smallest known insect genome, encodes a remarkably complete gene repertoire and thus, provides a robust phylogenetic outgroup for understanding the evolution of holometabolous insects. The striking reduction in genome size is particularly notable in gene families associated with environmental sensing and response; this reduction befits a monophagous permanent parasite with a substantially reduced need to seek out food sources and detect and avoid enemies relative to free-living species.

## Results and Discussion

**Genome Features. Genome sequencing, assembly, and annotation.** The genome of the body louse was sequenced to 8.5× average coverage using a whole genome shotgun approach with 1.3 million paired-end reads from plasmid libraries. The assembled contigs and scaffolds, spanning 108 Mb and 110 Mb, respectively, confirmed previous estimates based on flow cytometry data (103–109 Mb) that the body louse has the smallest known genome size of any insect (16, 17). The 300 longest scaffolds span more than 95% of the assembled genome sequence (scaffold N50 size of 488 kb). A range of automated and manual methods (18) yielded 10 tentative superscaffolds of up to 9 Mb each, spanning a total of 49 Mb. This effort provided large chromosomal segments, which were close to continuous, with only a few remaining clone gaps, usually involving simple-sequence gene deserts.

The remarkable compactness of the genome greatly facilitated accurate gene annotation. Predictions using multiple gene-modeling approaches resulted in consensus annotation (Table 1) of 10,773 protein-coding genes, 161 transfer ribonucleic acids (tRNAs) for all 20 amino acids, and 57 microRNAs (Table S14). Comparing predicted protein lengths with their *Drosophila melanogaster* orthologs (the best experimentally studied insect that drives comparative gene annotation) revealed greater consistency with body louse genes (concordance = 0.91; identical with *Anopheles gambiae*) than with the honey bee *Apis mellifera* (concordance = 0.89) or the red flour beetle *Tribolium castaneum* (concordance = 0.88), despite greater evolutionary divergence (Fig. S14).

**GC content.** Compared with other sequenced insect genomes, the body louse genome has the highest abundance of small homogeneous GC-content domains (7–30 kb with GC content between 18% and 63%). The average GC content of the *P. h. humanus* genome is 28%, which is similar to that of the *A. mellifera* genome (33%), making these two genomes unusually AT-rich. However, the *A. mellifera* genome harbors more extremes. Only 77% of homogeneous domains have a GC content between 20% and 60% in *A. mellifera* compared with 94% in *P. h. humanus*, which is more similar in this respect to the genome of *T. castaneum* (99%) (Fig. S2 A and B).

**Telomeres.** Unlike *A. mellifera* telomeres (19), none of the body louse telomeres appeared to be assembled completely at the ends of long superscaffolds. Therefore, we sought candidate telomere sequences with the strategy used for *T. castaneum* (20). The body louse is diploid, and it has a haploid complement of five metacentric chromosomes and one telocentric chromosome for a total of 11 putative telomeres (21). Although we were unable to reconstruct an entire telomere because of its highly repetitive nature, we identified a long subtelomeric repeat region that was partially assembled on at least 9 of 11 putative telo-

**Table 1. Summary of the genome features of *Pediculus humanus humanus* compared with *Drosophila melanogaster***

Genome feature	Count	Nucleotides (Mb)	Genome fraction (%)
<i>P. h. humanus</i> ( <i>D. melanogaster</i> )	6 chromosomes (4 chromosomes)	110 (169)	100 (100)
Gene-rich clusters* containing 95% of genes	1,110 (1,130)	55 (70)	50 (41)
Protein-coding genes			
Total [multi-exon]	10,773, [10,424]; (13,794, [11,458])	33.8 (82.6)	31 (49)
Coding exons	69,261 (54,606)	16.6 (22.3)	15 (13)
Introns	58,522 (44,698)	17.2 (48.6)	15 (29)
Non-protein-coding genes			
tRNAs	161 (292)	0.012 (0.022)	<1
miRNAs	57 (90)	0.005 (0.008)	<1
Transposable elements	3,558 (9,409)	1.1 (11.6)	1 (7)
Tandem repeats	130,608 (25,904)	6.9 (6.1)	6 (4)

*D. melanogaster* values were obtained from FlyBase release 5.23 with the same parameters used to obtain, parse, and count the *P. h. humanus* genome. The more numerous body louse exons and introns suggest intron loss in *D. melanogaster* but with an increase in their sizes.

\*Supporting documentation is in Fig. S4F.

meres between unique flanking DNA and telomeric TTAGG repeats. This subtelomeric region consists of various satellite-like repeats in addition to pseudogenes and simple sequences, and it varies considerably in length. The TTAGG repeats commonly contain sequence associated repeat telomeric (SART)-like retrotransposons, which are also characteristic of the telomeres from *T. castaneum* and *Bombyx mori* (domestic silkworm). This combination might represent the basal insect situation. If so, the simple TTAGG telomeres of *A. mellifera* would represent a derived condition in which most retrotransposons have been lost rather than the ancestral condition (19). Alternatively, insect telomeres may have repeatedly been invaded as a safe harbor by non-LTR retrotransposons of the R-element family that belongs to the SART group (20).

**Transposable elements.** Both class I and class II mobile elements are present in the genome of *P. h. humanus*, yet they represent only 1% of the genome (Table S1B), which is markedly lower than any sequenced insect genome. Interestingly, the body louse genome size is near the hypothesized 100 Mb critical threshold at which transposable elements can be established in eukaryote genomes (22).

**Mitochondrial genome.** The mitochondrial genome of *P. h. humanus* contains the full complement of 37 genes organized in an unusual architecture of 18 minicircular chromosomes (23). It is possible that multiple minicircular chromosomes promote recombination between genes on different chromosomes. Indeed, there is evidence in the genome sequence data for at least two chimeric minicircular chromosomes that have arisen from such recombination (Fig. S1B).

Of 305 mitochondrial-targeted, nuclear-encoded genes known in *D. melanogaster*, 282 have louse orthologs. This finding suggests that the basic mitochondrial functions (e.g., oxidative phosphorylation, membrane transport, and protein synthesis) are unimpeded by the reorganized mitochondrial genome. The body louse genome revealed the apparent loss of the mitochondrial single-stranded binding protein (mtSSB), a factor required for optimal initiation and processivity during mitochondrial genome replication in both insects and mammals (24, 25). In the absence of mtSSB, complete replication of a full-sized mitochondrial genome may not be possible (25); the loss of mtSSB function in *D. melanogaster* is lethal at the late third instar/pupal stages because of a loss of mtDNA content (26). It is not yet known if the mtSSB function can be replaced by an endosymbiont homolog or if the multiple minicircles render the mtSSB unnecessary.

**Endosymbiont Genome. Genome sequencing, assembly, and annotation.** Like many other sucking lice (Anoplura, Rhynchophthirina), the body louse has mycetozoa that harbor the primary endosymbiotic bacteria (p-endosymbionts). The genome of the *Pediculus* symbiont, *Riesia*, was sequenced to an average coverage of 50 $\times$  and is composed of a single linear chromosome of at least 574,526 bp with palindromic termini and a single circular plasmid of 7,628 bp. The chromosome contains 557 ORFs, 33 tRNAs, 6 ribosomal RNAs, and 1 other structural RNA.

**Comparisons with other endosymbionts.** We compared the genome of *Riesia* with the genomes of other endosymbionts and the infectious plague pathogen *Yersinia pestis* (Fig. S3). This genome-wide sequence comparison revealed a core of 237 genes common to all bacteria examined; only 24 genes were unique to *Riesia*, and 30 genes were present in all except *Riesia* (Table S2 A and B). Several genes unique to *Riesia* code for transport and binding proteins as well as for enzymes involved in lipopolysaccharide biosynthesis. Conversely, the enzymes missing from *Riesia* are mainly exonucleases, which are required for conjugation, and enzymes involved in energy metabolism. The *Riesia*-specific transport and binding proteins and the lack of energy metabolism genes may reflect the dependence of *Riesia* on its louse host for nutrients. Lipopolysaccharides might be important for cell-wall

stability when *Riesia* migrate extracellularly through the louse to reach filial mycetozoa in the ovaries (14) (Table S2B).

*Riesia* is required by lice for the production of pantothenic acid (vitamin B5). Without *Riesia*, nymphs die during their first molt (27). Surprisingly, the genes for three key enzymes in the synthesis of pantothenic acid, *panB*, *panC*, and *panE*, are missing from the linear chromosome of *Riesia*. These genes are, instead, found together on the plasmid. Similar cases are known from evolutionarily more ancient endosymbionts (e.g., *Buchnera*) in which essential genes are also extrachromosomal (28). Having these genes on a multicopy plasmid could represent a mechanism that reduces the risk of genome degradation and increases expression levels to secure synthesis of pantothenic acid at required amounts. Interestingly, there is preliminary evidence that endosymbiont replacement may be commonplace in sucking lice (29), possibly facilitated by the acquisition of plasmids that harbor genes essential to the host.

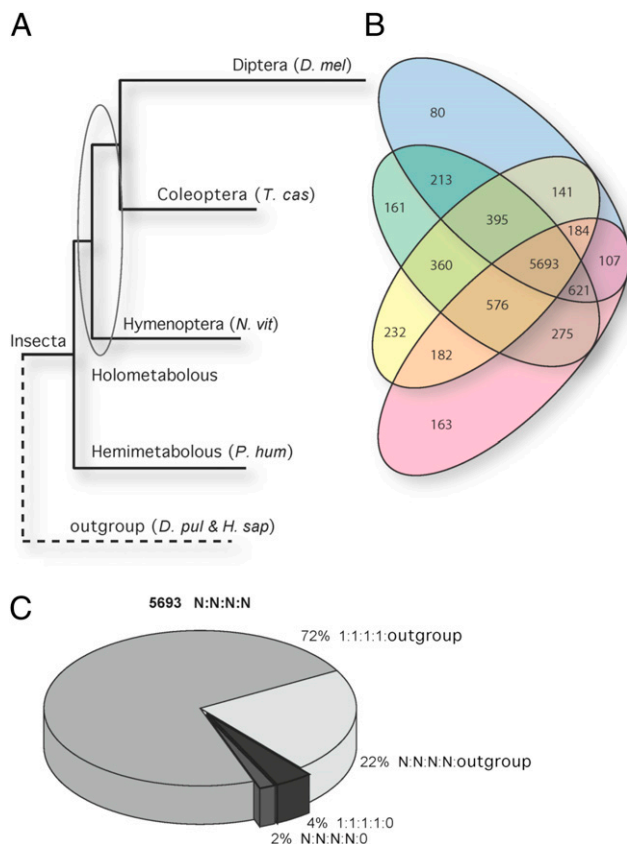
Nakabachi et al. (30) proposed that integration of essential genes from the p-endosymbiont into the host genome might be an important mechanism for the host to overcome the consequences of genome degradation of its endosymbiont. *Riesia* in the human body louse and *Buchnera* in the pea aphid (*Acyrtosiphon pisum*) represent cases where the genomes of both symbiotic partners are available to test this hypothesis. The body louse genome does not appear to contain any genes of prokaryotic origin, suggesting the absence of transfers from *Riesia*. In the pea aphid, there is also no gene transfer from the endosymbiont, but there is evidence of gene transfer from other bacteria (31).

The dramatic reduction in genome size and high AT bias suggest a long association between *Riesia* and its host insect, and like some other ancient gammaproteobacterial symbiotic associations, the *Riesia* genome is free of mobile elements. However, *Riesia*'s association with its host is only 13–25 million years old, making *Riesia* one of the youngest known endosymbionts (31).

**Comparative Genomics. Hemimetabolous outgroup.** The human body louse is among the first sequenced representatives of hemimetabolous insects (32), a group distinguished by progressive intermediate development as nymphal instars rather than larva-pupa-adult transformations. The louse genome is, therefore, an important outgroup reference for comparative analyses of sequenced holometabolous insects (Fig. 1A). The complete metamorphosis of holometabolous insects is a highly successful evolutionary strategy, whereby larvae and adults can take advantage of different ecological niches. The molecular innovations that have contributed to the success of holometabolous insects can now be viewed in the context of a hemimetabolous outgroup genome sequence that is largely complete.

In addition to being the smallest genome of any insect studied to date, the body louse genome is, as far as can be determined, functionally complete. Of 10,773 body louse protein-coding genes, 90% share homology to genes known in other species, enabling orthology delineation for 80% of louse genes (33). This level is comparable with results from initial analyses from *A. mellifera* (34) and *T. castaneum* (20). The phylogenetic tree reconstructed using single-copy orthologs (Fig. 1A) confirms the basal position of Hemimetabola compared with Holometabola within Insecta. This suggests an average rate of molecular evolution in the lineage of lice that is comparable with that of Hymenoptera and Coleoptera.

Microsynteny analysis (35) between genomes of the body louse and hymenopteran honey bee *A. mellifera* or *Nasonia* parasitoid wasp species suggests that about 20% of single-copy orthologs are retained in their ancestral arrangements (Table S3A). This percentage is similar to microsynteny conservation levels between *A. mellifera* (Hymenoptera) and *T. castaneum* (Coleoptera), and it is substantially greater than their conservation with dipterans (<15%) (36), highlighting the derived state of Diptera.



**Fig. 1.** The *Pediculus humanus humanus* (*P. hum*) genome reveals a basal insect gene repertoire. The encoded *P. hum* proteome is compared with sequenced representatives of the orders Diptera, Coleoptera, and Hymenoptera and outgroup species beyond Insecta. *D. mel*, *Drosophila melanogaster*; *T. cas*, *Tribolium castaneum*; *N. vit*, *Nasonia vitripennis*; *D. pul*, *Daphnia pulex*; *H. sap*, *Homo sapiens*. (A) The Maximum-Likelihood phylogenetic tree was reconstructed using the superalignment of protein sequences of universal single-copy orthologs. The obtained tree confirms the basal position of Hemimetabola compared with Holometabola within Insecta. The branch lengths are proportional to the accumulated number of substitutions, suggesting an average rate of molecular evolution in lice that is comparable with that in Hymenoptera and Coleoptera. (B) The Venn diagram shows the numbers of orthologous groups of genes shared among the four insects (a lower estimate of the ancestral number of genes). It depicts the phylogenetic distribution of orthologs, highlighting the completeness of the gene repertoire encoded in the body louse genome. Pink, *P. hum*; yellow, *N. vit*; green, *T. cas*; blue, *D. mel*. (C) The pie chart partitions the largest fraction of core body louse proteins with orthologs in three holometabolous insect orders and the outgroup species beyond Insecta with respect to single- (1:1:1:1) and multiple- (N:N:N:N) copy orthologs. Of 5,693 groups of single- and multiple-copy orthologs common across Insecta, 94% are shared across Bilateria as single-copy (72%) or multiple-copy (22%) orthologs, and only 6% are insect-specific orthologous groups (4% as single copies and 2% as multiple copies).

**Ancestral insect gene repertoire.** Contrary to the expectations of reductive evolution common in obligate parasites, the body louse has retained a remarkably complete repertoire of both protein-coding and non-protein-coding genes (Table 1). The distribution of orthologous genes across four representative insect species (Fig. 1B and C) shows that Hymenoptera and Coleoptera share more orthologs with the body louse than they do with the fruit fly *D. melanogaster*. Relative to the well-studied *D. melanogaster* model, the louse genome may be used as a robust outgroup to Holometabola.

Examining microRNA gene families shared among crustaceans and insects revealed that mir-315, mir-283, mir-33, and mir-29 were lost from the body louse genome (37) (Table S14

and Fig. S4A–D) (mir-iab-4 and mir-46 have been found in the trace archive). Because all true lice are wingless, it is intriguing to note that mir-315 has been identified as a potent activator of wingless signaling in *D. melanogaster* (38).

**Evolution of Gene Families in Relation to the Life History of the Body Louse.** The body louse has maintained many genes important for basic physiological processes, losing only a few of these roles to its endosymbiont *Riesia*. Because the expansion and contraction of gene families may indicate functional adaptation and evolution, we compared the body louse gene repertoire with those of the honey bee and red flour beetle. Comparisons were made both at the level of protein families, which could be generally defined using InterPro domain signatures (Table S3B–D), and at a finer scale at the level of orthologous groups of genes (Fig. S4E). On both scales, the body louse genome seems to have several gene families with fewer members than those found in other invertebrates.

**Fewer Genes Are Associated with Environmental Sensing and Response.**

**G protein-coupled receptors.** With 104 nonsensory G protein-coupled receptors (GPCRs) and 3 opsins (visual receptors) (Table S4), *P. h. humanus* has the smallest repertoire of GPCRs identified in any sequenced insect genome to date (20, 34, 39–41). The louse genome has orthologs for ~80% of nonsensory GPCRs identified in *D. melanogaster*. These GPCRs seemingly represent a minimal suite of receptors needed to maintain conserved GPCR-mediated signaling pathways common to diverse insect taxa (42). The relatively small number of louse opsins likely reflects its simple visual system. Moreover, the body louse lacks a putative short (blue)-wavelength sensitive opsin typically found in other insects (43), a feature that might have evolved during its adaptation to the obligate parasitic lifestyle.

**Odorant-, gustatory-, and chemosensory-related genes.** The genome sequence revealed just 10 odorant receptor (Or) genes, fewer than any other insect examined to date by almost an order of magnitude. The gustatory receptor (Gr) family is comparably small with just six loci encoding eight proteins through alternative splicing of the N terminus of one locus. There are no orthologs of the otherwise highly conserved carbon dioxide heterodimer Gr receptors (40, 41, 44, 45) or the putative sugar receptors (46, 47). *P. h. humanus* contains five and seven putative functional odorant-binding proteins (OBPs) and chemosensory proteins (CSPs), respectively (Table S4), and this number is dramatically less than that found in other insects (48). These aforementioned sensory genes and their resultant proteins are presumably not necessary for host location and selection. Furthermore, lice do not need to avoid the many bitter xenobiotic toxins to which most insect Grs seem to be tuned (46).

**Insulin/Target of Rapamycin (TOR) pathway genes.** The insulin/TOR signal transduction pathway plays a central role in multiple and critical biological processes, including organismal growth, anabolic metabolism, cell survival, fertility, and lifespan determination (49, 50). This pathway has been well-characterized in multiple organisms, including *D. melanogaster* (51). Both the structure of the pathway and the molecular function of its components are well-conserved across metazoans. The body louse genome encodes a complete insulin/TOR signaling pathway. However, these genes are reduced in number in the body louse in contrast with *D. melanogaster*, where some genes have multiple copies (Table S4D). Remarkably, the louse has a single insulin-like peptide (*ilp*) gene. Given that there is some evidence for differential expression of *ilp* genes under different dietary conditions in insects (52, 53), the presence of a single *ilp* gene in the body louse genome might reflect its restricted and homogeneous diet.

**Detoxification enzymes.** The louse genome encodes the smallest number of detoxification enzymes observed in any insect, reflecting its obligate parasite lifestyle in which it is sheltered from xenobiotic challenges faced by free-living insects (e.g., plant

secondary compounds). There are notably few cytochrome P450s and only 12 genes within the *CYP3* clade which is closely associated with xenobiotic metabolism. In contrast, *D. melanogaster* and *A. mellifera* have 36 and 28 *CYP3* clade genes, respectively. Among the 13 glutathione-S-transferases (GST) (Table S4E), none belong to the Epsilon class that has been shown to contribute to insect adaptation to environmental selection pressures (54). The Epsilon class was also missing in the pea aphid genome. In contrast, the relative abundance of Delta class GSTs (more than *A. mellifera*) suggests that *P. h. humanus* still possesses some capacity for detoxification of xenobiotics, including insecticides (55).

**Body louse coevolution and allopatric speciation.** With their characteristic extreme host specificity, pediculi lice provide dramatic examples of host–parasite coevolution and allopatric speciation (56). One consequence of this specificity is the difficulty encountered when adapting human lice to novel experimental hosts (8). Body lice have reduced genomes and harbor specific bacterial symbionts and pathogens that also exhibit genome reduction (57–64). These combined observations support the hypothesis that *P. h. humanus* has become highly specialized since its divergence from the chimpanzee louse 5–7 million years ago. Such extreme specializations in the endosymbiont, associated with dramatic genome reductions, may have resulted from a lack of gene exchange after allopatric speciation. This association of an insect host, its symbionts, and its bacterial pathogens coevolving and showing congruent reductive genome evolution provides a dramatic example of the evolutionary consequences of genome interactions and interdependency over time.

## Conclusions

The body louse genome provides a unique repository of data that has considerable basic and practical significance. The availability of sequence data will facilitate molecular studies of a vector for diseases that continue to afflict human populations around the world. The

louse relies on *Riesia*, an obligatory louse bacterial endosymbiont that lacks antibiotic resistance genes, for survival; thus, the development of louse-control strategies targeting this symbiont may be possible. With respect to understanding the evolution of multigene families mediating responses to environmental selective forces, the body louse genome, with its drastically reduced inventories in the context of its exceptionally homogeneous environment, provides extraordinary prospects for characterizing the functionalities of these rapidly evolving proteins. As well, further studies focusing on the smaller repertoire of detoxification genes and olfactory receptors in the body louse may guide the development of pediculicides and repellents with negligible impacts on human hosts. Moreover, the remarkable completeness of this genome, despite its small size, will serve as a key evolutionary reference point for studies of all sequenced insect species in characterizing the fundamental prerequisites for insect growth and development. Finally, the body louse genome will provide an opportunity for the scientific community to gain greater insights into host–parasite–symbiont tripartite coevolution and speciation.

## Materials and Methods

Lice were obtained from an inbred colony derived from the Culpepper strain (65) that has been maintained on rabbits since 1999 at the University of Massachusetts, Amherst, MA. Total DNA was extracted from ~100 first instar nymphs before their first blood meal and was used to construct libraries in the plasmid, pHOS2 (3- to 4-kb and 10- to 12-kb inserts), or the fosmid, pCCFOS1 (35- to 40-kb inserts). End sequencing of clones from each library was conducted using a standard capillary platform (ABI 3730), and it yielded 1.30 million good traces (96% paired) with a mean clear read length of 656 bases. All traces were deposited in the National Center for Biotechnology Information (NCBI) trace archive (<http://www.ncbi.nlm.nih.gov/Traces/trace.cgi>). The reads were assembled with Celera Assembler (<http://wgs-assembler.sourceforge.net>) (66–68) and deposited with NCBI (accession no. AAZ000000000). The details of the assembly and annotation are given in *SI Text*. Additional analyses of other aspects of the body louse genome are given in the *SI Text*.

**ACKNOWLEDGMENTS.** We thank Georg Jander and Marian Goldsmith for helpful discussion and manuscript review.

- Reed DL, Light JE, Allen JM, Kirchman JJ (2007) Pair of lice lost or parasites regained: The evolutionary history of anthropoid primate lice. *BMC Biol* 5:7.
- Reed DL, Smith VS, Hammond SL, Rogers AR, Clayton DH (2004) Genetic analysis of lice supports direct contact between modern and archaic humans. *PLoS Biol* 2:e340.
- Light JE, Reed DL (2009) Multigene analysis of phylogenetic relationships and divergence times of primate sucking lice (Phthiraptera: Anoplura). *Mol Phylogenet Evol* 50:376–390.
- Kittler R, Kayser M, Stoneking M (2003) Molecular evolution of *Pediculus humanus* and the origin of clothing. *Curr Biol* 13:1414–1417.
- Eremeeva ME, Madan A, Shaw CD, Tang K, Dasch GA (2005) New perspectives on rickettsial evolution from new genome sequences of *Rickettsia*, particularly *R. canadensis*, and *Orientia tsutsugamushi*. *Ann N Y Acad Sci* 1063:47–63.
- Rotz LD, Khan AS, Lillibridge SR, Ostroff SM, Hughes JM (2002) Public health assessment of potential biological terrorism agents. *Emerg Infect Dis* 8:225–230.
- Andersson JO, Andersson SG (2000) A century of typhus, lice and *Rickettsia*. *Res Microbiol* 151:143–150.
- Raoult D, Roux V (1999) The body louse as a vector of reemerging human diseases. *Clin Infect Dis* 29:888–911.
- Raoult D, et al. (2006) Evidence for louse-transmitted diseases in soldiers of Napoleon's Grand Army in Vilnius. *J Infect Dis* 193:112–120.
- Tarasevich I, Rydkina E, Raoult D (1998) Outbreak of epidemic typhus in Russia. *Lancet* 352:1151.
- Bechah Y, Capo C, Mege JL, Raoult D (2008) Epidemic typhus. *Lancet Infect Dis* 8: 417–426.
- Buchner P (1965) *Endosymbiosis of Animals with Plant Microorganisms* (Interscience Publishers, New York), pp 909.
- Sasaki-Fukatsu K, et al. (2006) Symbiotic bacteria associated with stomach discs of human lice. *Appl Environ Microbiol* 72:7349–7352.
- Perotti MA, Allen JM, Reed DL, Braig HR (2007) Host-symbiont interactions of the primary endosymbiont of human head and body lice. *FASEB J* 21:1058–1066.
- Allen JM, Reed DL, Perotti MA, Braig HR (2007) Evolutionary relationships of "*Candidatus* *Riesia* spp.," endosymbiotic Enterobacteriaceae living within hematophagous primate lice. *Appl Environ Microbiol* 73:1659–1664.
- Pittendrigh BR, et al. (2006) Sequencing of a new target genome: The *Pediculus humanus humanus* (Phthiraptera: Pediculidae) genome project. *J Med Entomol* 43: 1103–1111.
- Johnston JS, Yoon KS, Strycharz JP, Pittendrigh BR, Clark JM (2007) Body lice and head lice (Anoplura: Pediculidae) have the smallest genomes of any hemimetabolous insect reported to date. *J Med Entomol* 44:1009–1012.
- Robertson HM, et al. (2007) Manual superscaffolding of honey bee (*Apis mellifera*) chromosomes 12–16: Implications for the draft genome assembly version 4, gene annotation, and chromosome structure. *Insect Mol Biol* 16:401–410.
- Robertson HM, Gordon KH (2006) Canonical TTAGG-repeat telomeres and telomerase in the honey bee, *Apis mellifera*. *Genome Res* 16:1345–1351.
- Richards S, et al. (2008) The genome of the model beetle and pest *Tribolium castaneum*. *Nature* 452:949–955.
- Hindle E, Pontecorvo G (1942) Mitotic divisions following meiosis in *Pediculus corporis* males. *Nature* 149:668.
- Lynch M, Conery JS (2003) The origins of genome complexity. *Science* 302:1401–1404.
- Shao R, Kirkness EF, Barker SC (2009) The single mitochondrial chromosome typical of animals has evolved into 18 minichromosomes in the human body louse, *Pediculus humanus*. *Genome Res* 19:904–912.
- Farr CL, Matsushima Y, Lagina AT, 3rd, Luo N, Kaguni LS (2004) Physiological and biochemical defects in functional interactions of mitochondrial DNA polymerase and DNA-binding mutants of single-stranded DNA-binding protein. *J Biol Chem* 279: 17047–17053.
- Korhonen JA, Pham XH, Pellegrini M, Falkenberg M (2004) Reconstitution of a minimal mtDNA replisome *in vitro*. *EMBO J* 23:2423–2429.
- Maier D, et al. (2001) Mitochondrial single-stranded DNA-binding protein is required for mitochondrial DNA replication and development in *Drosophila melanogaster*. *Mol Biol Cell* 12:821–830.
- Perotti MA, Kirkness EF, Reed DL, Braig HR (2009) Endosymbionts of lice. *Insect Symbiosis 3*, ed Bourtzis KMT (Taylor & Francis, Boca Raton, FL), pp 205–220.
- Ding H, Hynes MF (2009) Plasmid transfer systems in the rhizobia. *Can J Microbiol* 55: 917–927.
- Hypsa V, Krizek J (2007) Molecular evidence for polyphyletic origin of the primary symbionts of sucking lice (phthiraptera, anoplura). *Microb Ecol* 54:242–251.
- Nakabachi A, et al. (2006) The 160-kilobase genome of the bacterial endosymbiont *Carsonella*. *Science* 314:267.
- Allen JM, Light JE, Perotti MA, Braig HR, Reed DL (2009) Mutational meltdown in primary endosymbionts: Selection limits Muller's ratchet. *PLoS One* 4:e4969.
- The International Aphid Genomics Consortium (2010) Genome sequence of the pea aphid *Acyrtosiphon pisum*. *PLoS Biol* 8:e1000313.

33. Kriventseva EV, Rahman N, Espinosa O, Zdobnov EM (2008) OrthoDB: The hierarchical catalog of eukaryotic orthologs. *Nucleic Acids Res* 36:D271–D275.
34. Weinstock GM, et al. (2006) Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature* 443:931–949.
35. Zdobnov EM, et al. (2002) Comparative genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster*. *Science* 298:149–159.
36. Zdobnov EM, Bork P (2007) Quantification of insect genome divergence. *Trends Genet* 23:16–20.
37. Gerlach D, Kriventseva EV, Rahman N, Vejnar CE, Zdobnov EM (2009) miROrtho: Computational survey of microRNA genes. *Nucleic Acids Res* 37:D111–D117.
38. Klingensmith J, Nusse R (1994) Signaling by wingless in *Drosophila*. *Dev Biol* 166:396–414.
39. Adams MD, et al. (2000) The genome sequence of *Drosophila melanogaster*. *Science* 287:2185–2195.
40. Robertson HM, Warr CG, Carlson JR (2003) Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster*. *Proc Natl Acad Sci USA* 100:(Suppl2):14537–14542.
41. Benton R, Sachse S, Michnick SW, Vosshall LB (2006) Atypical membrane topology and heteromeric function of *Drosophila* odorant receptors *in vivo*. *PLoS Biol* 4:e20.
42. Wistrand M, Kall L, Sonnhammer EL (2006) A general model of G protein-coupled receptor sequences and its application to detect remote homologs. *Protein Sci* 15:509–521.
43. Briscoe AD, Chittka L (2001) The evolution of color vision in insects. *Annu Rev Entomol* 46:471–510.
44. Jones WD, Cayirlioglu P, Kadow IG, Vosshall LB (2007) Two chemosensory receptors together mediate carbon dioxide detection in *Drosophila*. *Nature* 445:86–90.
45. Frommer WB (2010) CO<sub>2</sub> common sense. *Science* 327:275–276.
46. Marella S, et al. (2006) Imaging taste responses in the fly brain reveals a functional map of taste category and behavior. *Neuron* 49:285–295.
47. Chyb S, Dahanukar A, Wickens A, Carlson JR (2003) *Drosophila* Gr5a encodes a taste receptor tuned to trehalose. *Proc Natl Acad Sci USA* 100:(Suppl2):14526–14530.
48. Sanchez-Gracia A, Vieira FG, Rozas J (2009) Molecular evolution of the major chemosensory gene families in insects. *Heredity* 103:208–216.
49. Goberdhan DC, Wilson C (2003) The functions of insulin signaling: Size isn't everything, even in *Drosophila*. *Differentiation* 71:375–397.
50. Oldham S, Hafen E (2003) Insulin/IGF and target of rapamycin signaling: A TOR de force in growth control. *Trends Cell Biol* 13:79–85.
51. Alvarez-Ponce D, Aguade M, Rozas J (2009) Network-level molecular evolutionary analysis of the insulin/TOR signal transduction pathway across 12 *Drosophila* genomes. *Genome Res* 19:234–242.
52. Wheeler DE, Buck N, Evans JD (2006) Expression of insulin pathway genes during the period of caste determination in the honey bee, *Apis mellifera*. *Insect Mol Biol* 15:597–602.
53. Arsic D, Guerin PM (2008) Nutrient content of diet affects the signaling activity of the insulin/target of rapamycin/p70 S6 kinase pathway in the African malaria mosquito *Anopheles gambiae*. *J Insect Physiol* 54:1226–1235.
54. Ranson H, et al. (2002) Evolution of supergene families associated with insecticide resistance. *Science* 298:179–181.
55. Enayati AA, Ranson H, Hemingway J (2005) Insect glutathione transferases and insecticide resistance. *Insect Mol Biol* 14:3–8.
56. Page RD, Lee PL, Becher SA, Griffiths R, Clayton DH (1998) A different tempo of mitochondrial DNA evolution in birds and their parasitic lice. *Mol Phylogenet Evol* 9:276–293.
57. Blanc G, et al. (2007) Reductive genome evolution from the mother of *Rickettsia*. *PLoS Genet* 3:e14.
58. Andersson SG, et al. (1998) The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* 396:133–140.
59. Ogata H, et al. (2001) Mechanisms of evolution in *Rickettsia conorii* and *R. prowazekii*. *Science* 293:2093–2098.
60. Lescot M, et al. (2008) The genome of *Borrelia recurrentis*, the agent of deadly louse-borne relapsing fever, is a degraded subset of tick-borne *Borrelia duttonii*. *PLoS Genet* 4:e1000185.
61. Alsmark CM, et al. (2004) The louse-borne human pathogen *Bartonella quintana* is a genomic derivative of the zoonotic agent *Bartonella henselae*. *Proc Natl Acad Sci USA* 101:9716–9721.
62. Fournier PE, Suhre K, Fournous G, Raoult D (2006) Estimation of prokaryote genomic DNA G+C content by sequencing universally conserved genes. *Int J Syst Evol Microbiol* 56:1025–1029.
63. Fournier PE, et al. (2006) Comparative genomics of multidrug resistance in *Acinetobacter baumannii*. *PLoS Genet* 2:e7.
64. Vallenet D, et al. (2008) Comparative analysis of Acinetobacters: Three genomes for three lifestyles. *PLoS One* 3:e1805.
65. Culpepper GH (1944) The rearing and maintenance of a laboratory colony of the body louse. *Am J Trop Med Hyg* 24:327–329.
66. Levy S, et al. (2007) The diploid genome sequence of an individual human. *PLoS Biol* 5:e254.
67. Myers EW, et al. (2000) A whole-genome assembly of *Drosophila*. *Science* 287:2196–2204.
68. Venter JC, et al. (2001) The sequence of the human genome. *Science* 291:1304–1351.