Medical and Veterinary Entomology

Medical and Veterinary Entomology (2014) **28** (Suppl. 1), 40–50

Second-generation sequencing of entire mitochondrial coding-regions (~15.4 kb) holds promise for study of the phylogeny and taxonomy of human body lice and head lice

H. XIONG[1,†], D. CAMPELO[1,†], R. J. POLLACK[2], D. RAOULT[3], R. SHAO[1,*], M. ALEM[4], J. ALI[5], K. BILCHA[6] and S. C. BARKER[1]

[1]Parasitology Section, School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, Queensland, Australia, [2]IdentifyUS LLC, 320 Needham Street, Suite 200, Newton, MA, U.S.A. , [3]Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes (URMITE), UMR CNRS 6236 IRD 198, Faculté de Médecine, Méditerranée Infection, Aix-MarseilleUniversité, Marseille, France , [4]Department of Immunology and Molecular Biology, College of Medicine and Health Sciences, University of Gondar, Gondar, Ethiopia, [5]Department of Microbiology, Immunology and Parasitology, College of Medicine and Health Sciences, University of Gondar, Gondar, Ethiopia and [6]Department of Dermatology and Venereology, College of Medicine and Health Sciences, University of Gondar, Gondar, Ethiopia

Abstract. The Illumina Hiseq platform was used to sequence the entire mitochondrial coding-regions of 20 body lice, *Pediculus humanus* Linnaeus, and head lice, *P. capitis* De Geer (Phthiraptera: Pediculidae), from eight towns and cities in five countries: Ethiopia, France, China, Australia and the U.S.A. These data (~310 kb) were used to see how much more informative entire mitochondrial coding-region sequences were than partial mitochondrial coding-region sequences, and thus to guide the design of future studies of the phylogeny, origin, evolution and taxonomy of body lice and head lice. Phylogenies were compared from entire coding-region sequences (~15.4 kb), entire *cox1* (~1.5 kb), partial *cox1* (~700 bp) and partial *cytb* (~600 bp) sequences. On the one hand, phylogenies from entire mitochondrial coding-region sequences (~15.4 kb) were much more informative than phylogenies from entire *cox1* sequences (~1.5 kb) and partial gene sequences (~600 to ~700 bp). For example, 19 branches had > 95% bootstrap support in our maximum likelihood tree from the entire mitochondrial coding-regions (~15.4 kb) whereas the tree from 700 bp *cox1* had only two branches with bootstrap support > 95%. Yet, by contrast, partial *cytb* (~600 bp) and partial *cox1* (~486 bp) sequences were sufficient to genotype lice to Clade A, B or C. The sequences of the mitochondrial genomes of the *P. humanus*, *P. capitis* and *P. schaeffi* Fahrenholz studied are in NCBI GenBank under the accession numbers KC660761-800, KC685631-6330, KC241882-97, EU219988-95, HM241895-8 and JX080388-407.

Key words. *Pediculus capitis*, *Pediculus humanus*, evolutionary genetics, mitochondrial coding-region.

Correspondence: Professor Stephen C. Barker, Parasitology Section, School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, Queensland 4072, Australia. Tel.: +61 7 336 53303; Fax: +61 7 33654699; E-mail: s.barker@uq.edu.au

*Present address: GeneCology Research Centre, School of Science, Health Education and Engineering, University of the Sunshine Coast, Maroochydore, Queensland, Australia.

†Both first authors contributed equally to this publication.

Introduction

Lice of the genus *Pediculus* (Phthiraptera) infect people worldwide. Lice that infect clothes are traditionally called body (or clothes) lice whereas lice that infect the scalp-hair are traditionally called head lice. Head lice mainly infect children and are thought to be unimportant as vectors of louse-borne infections (but see Robinson *et al.*, 2003; Boutellis *et al.*, 2013). By contrast, body lice mainly infect indigent adults and are vectors for the aetiological agents of louse-borne relapsing fever and louse-borne epidemic typhus in some regions of Africa, and trench fever throughout the world (Raoult *et al.*, 1998; Raoult & Roux, 1999; Ramos *et al.*, 2009; Angelakis *et al.*, 2011; Haitham *et al.*, 2012; Boutellis *et al.*, 2013). The taxonomy of body lice and head lice is remarkably contentious. Body lice and head lice are argued to be either species, *Pediculus humanus* and *P. capitis* (e.g. Freund, 1932; Busvine, 1978; Leo *et al.*, 2005) or subspecies, *P. humanus humanus* and *P. h. capitis* (e.g. Ferris, 1951; Light *et al.*, 2008a; Li *et al.*, 2010; Veracx *et al.*, 2012a).

The phylogeny, origin, evolution and taxonomy of body lice and head lice have intrigued scientists and clinicians since Linnaeus studied *Pediculus* lice 250 years ago. In spite of much scientific inquiry and the application of modern evolutionary genetic methods, the precise origin, in time and space, of body lice and the phylogeny of *Pediculus* lice from different geographical regions of the world are topics of considerable controversy (Leo *et al.*, 2002; Kittler *et al.*, 2003; Reed *et al.*, 2004; Leo & Barker, 2005; Light *et al.*, 2008a, 2008b; Light & Reed, 2009; Li *et al.*, 2010; Toups *et al.*, 2011; Veracx *et al.*, 2012a, 2012b). Persistent, unanswered questions are: (a) did body lice evolve from head lice once, or a few times, or are body lice evolving from head lice 'constantly' in the words of Veracx and Raoult (2012)? Are the body lice from different regions of the world monophyletic or polyphyletic?; (b) How are the three known clades (genotypes) of body lice and head lice related? Is Clade A (a cosmopolitan clade of body lice and head lice) the sister-group to Clade B (a clade of head lice from all continents except Asia) or Clade C (a clade of head lice from Ethiopia, Senegal and Nepal). Or is Clade B the sister-group to Clade C?; and (c) How much phylogenetic structure is there within Clades A, B and C (Veracx *et al.*, 2012b)?

Formal scientific enquiry into the phylogeny, origin, evolution and taxonomy of body lice and head lice, may be divided into two periods. From the 1700s until 2002, scientists studied the morphology, preferred habitat (clothes or scalp-hair) and preferred egg-laying sites (clothes or scalp-hair) of body lice and head lice, respectively. From 2002 (Leo *et al.*, 2002) until now, enquiry into the evolution of body lice and head lice has been almost entirely with nucleotide sequences of parts of mitochondrial genes, nuclear genes or both. Since Kittler *et al*. (2003) it has been apparent that the body lice and head lice comprise deeply divergent clades (Kittler *et al.*, 2003). We credit Kittler *et al.* (2003) with the discovery of Clade A (a cosmopolitan clade of body lice and head lice) and Clade C (a clade of head lice from Ethiopia, Senegal and Nepal). We credit Reed *et al.* (2004) with the discovery of Clade B (a clade of head lice from all continents except Asia). As yet, body lice have been found only in Clade A.

Whereas lice may occur on the scalp-hair regardless of a person's age, personal hygiene or economic resources, lice on the clothing tend to be restricted to persons who are indigent and have but one set of clothes and on people living traditional, rural lifestyles in cold, often high-altitude, regions. Veracx *et al*. (2012b) reported that body lice and head lice from different geographical regions of the world interbreed and that there was

Table 1. The 20 body lice and head lice studied + the chimpanzee louse, *Pediculus schaeffi*, and the crab louse, *Pthirus pubis*.

Louse sample #	Country	City	Genbank #
B2470B1	China	Hotan	KC660761-KC660800, KC685631-KC686330
B2471H5	China	Hotan	KC660761-KC660800, KC685631-KC686330
B2557B	France	Marseille	KC660761-KC660800, KC685631-KC686330
B2557H	France	Marseille	KC660761-KC660800, KC685631-KC686330
B2558B	France	Marseille	KC660761-KC660800, KC685631-KC686330
B2558H	France	Marseille	KC660761-KC660800, KC685631-KC686330
B2516B2	Ethiopia	Asendabo	KC660761-KC660800, KC685631-KC686330
B2516H2	Ethiopia	Asendabo	KC660761-KC660800, KC685631-KC686330
B2517B	Ethiopia	Asendabo	KC660761-KC660800, KC685631-KC686330
B2517H	Ethiopia	Asendabo	KC660761-KC660800, KC685631-KC686330
B2560B	Ethiopia	Addis Ababa	KC660761-KC660800, KC685631-KC686330
B2560H	Ethiopia	Addis Ababa	KC660761-KC660800, KC685631-KC686330
B2563B	Ethiopia	Addis Ababa	KC660761-KC660800, KC685631-KC686330
B2563B2	Ethiopia	Addis Ababa	KC660761-KC660800, KC685631-KC686330
B2563H	Ethiopia	Addis Ababa	KC660761-KC660800, KC685631-KC686330
B2563H2	Ethiopia	Addis Ababa	KC660761-KC660800, KC685631-KC686330
B2663B	Ethiopia	Gondar	KC660761-KC660800, KC685631-KC686330
B2664H	Ethiopia	Gondar	KC660761-KC660800, KC685631-KC686330
Orlando lab strain	USA	—	KC660761-KC660800, KC685631-KC686330
B2135H	Australia	Brisbane	KC660761-KC660800, KC685631-KC686330
Pediculus schaeffi	—	—	KC241882-KC241897
Pthirus pubis	—	—	EU219988-95, HM241895-8, and JX080388-407

BL, body lice; HL, head lice.

an African sub-clade of head lice and body lice within Clade A. Veracx *et al.* (2012a) proposed, as further evidence of a recent divergence of head lice and body lice, that the head lice and body lice from five co-infected patients in the Madrague Ville Shelter for homeless people in Marseille, France, had the same nuclear genotype (Veracx *et al.*, 2012a).

Mitochondrial genome sequences have been used to address many questions about the phylogeny and evolutionary (population) genetics of arthropods (Shao & Barker, 2007; Cameron, 2014). Indeed, there are now entire mitochondrial (mt) genome sequences for over 600 species of arthropods (Cameron, 2014). Shao *et al.* (2009) discovered that the mt genomes of body lice and head lice comprise mini-chromosomes. Each mini-chromosome has 1–3 genes (0.06–1.7 kb) and a single non-coding region (1.4–4 kb) that has conserved sequence blocks that are identical in all of the mini-chromosomes of a species/sub-species and the close relatives of that species/sub-species. One pair of PCR primers, that target these conserved sequences blocks, allows all of the mini-chromosomes in an individual louse to be amplified, routinely, in one PCR reaction (~15.4 kb coding-region + 48 kb non-coding region sequence) (Shao *et al.*, 2009). All of the mini-chromosomes were amplified and then sequenced from 20 body lice and head lice from eight towns and cities in five countries: Ethiopia, France, Australia and the U.S.A. Then phylogenies were compared from 10 different data sets:

Table 2. Models of nucleotide substitution in Phylemon2 (Protest and Jmodeltest) that we used to infer phylogenetic trees.

Nucleotide data sets	Model	Amino acid ML datasets	Model
Entire mt coding-region (15.4 kb)	GTR + G:4	Entire mt coding-region (15.4 kb)	mtREV24 + G:4 + F
Entire *cox1* (1.5 kb)	SYM + G:4	Entire *cox1* (1.5 kb)	mtREV24 + G:4 + F
700 bp *cox1* + 600 bp *cytb*	SYM + G:4	700 bp *cox1* + 600 bp *cytb*	mtREV24 + G:4
600 bp *cytb*	K80 + G:4	600 bp *cytb*	mtREV24 + G:4
700 bp *cox1*	SYM + G:4	700 bp *cox1*	mtREV24 + G:4
13 mt protein coding genes	GTR + G:4	—	—
First codon of 13 protein-coding genes	GTR + G:4	—	—
Second codon of 13 protein-coding genes	SYM + G:4 + I	—	—
Third codon of 13 protein-coding genes	SYM + G:4 + I	—	—
2 rRNA genes	SYM + G:4	—	—
22 tRNA genes	SYM + G:4	—	—

Nucleotide model abbreviations: GTR, general time reversible; K80, kimura; SYM, symmetrical model, equal base frequencies.
Amino acid model abbreviations: GTR, general time reversible; mtREV24, general reversible mitochondrial.
F, fixed; G, gamma distribution; I, variant sites; mt, mitochondrial genome.

Table 3. Nucleotide composition of the entire coding-regions – the regions that are known to recombine (~15.1 kb) of the mitochondrial genomes of body lice, head lice, *Pediculus schaeffi* and *Pthirus pubis*; and the clade-identity of body lice and head lice.

Louse #	A (%)	C (%)	G (%)	T (%)	GC (%)	Clade identity
B2470B1^{BL}	28.4	14.4	17.2	40.0	31.6	A
B2471H5^{HL}	28.3	14.5	17.2	40.0	31.7	A
B2557B^{BL}	28.3	14.5	17.1	40.1	31.6	A
B2557H^{HL}	28.3	14.5	17.2	40.0	31.7	A
B2558B^{BL}	28.4	14.5	17.1	40.0	31.6	A
B2558H^{HL}	28.4	14.5	17.1	40.0	31.6	A
B2516B2^{BL}	28.2	14.5	17.3	40.0	31.8	A
B2516H2^{HL}	27.9	14.4	17.5	40.2	31.9	C
B2517B^{BL}	28.3	14.4	17.2	40.2	31.6	A
B2517H^{HL}	28.3	14.4	17.2	40.1	31.6	A
B2560B^{BL}	28.1	14.5	17.4	39.9	31.9	A
B2560H^{HL}	27.8	14.4	17.6	40.2	32.0	C
B2563B^{BL}	28.3	14.4	17.2	40.1	31.6	A
B2563B2^{BL}	28.1	14.5	17.4	40.0	31.9	A
B2563H^{HL}	28.3	14.4	17.2	40.1	31.6	A
B2563H2^{HL}	28.1	14.6	17.4	40.0	32.0	A
B2663B^{BL}	28.3	14.4	17.2	40.2	31.6	A
B2664H^{HL}	27.8	14.4	17.5	40.3	31.9	C
Orlando lab strain^{BL}	28.3	14.4	17.2	40.1	31.6	A
B2135H^{HL}	28.3	14.3	17.1	40.2	31.4	A
Pediculus schaeffi	30.4	13.1	15.0	41.4	28.1	—
Pthirus pubis	24.0	16.6	21.5	37.9	38.1	—

The Clade C lice (B2516H2, B2560H, B2664H, underline) had lower A% than the other body lice and head lice ($P < 0.05$).
BL, body louse; HL, head louse.

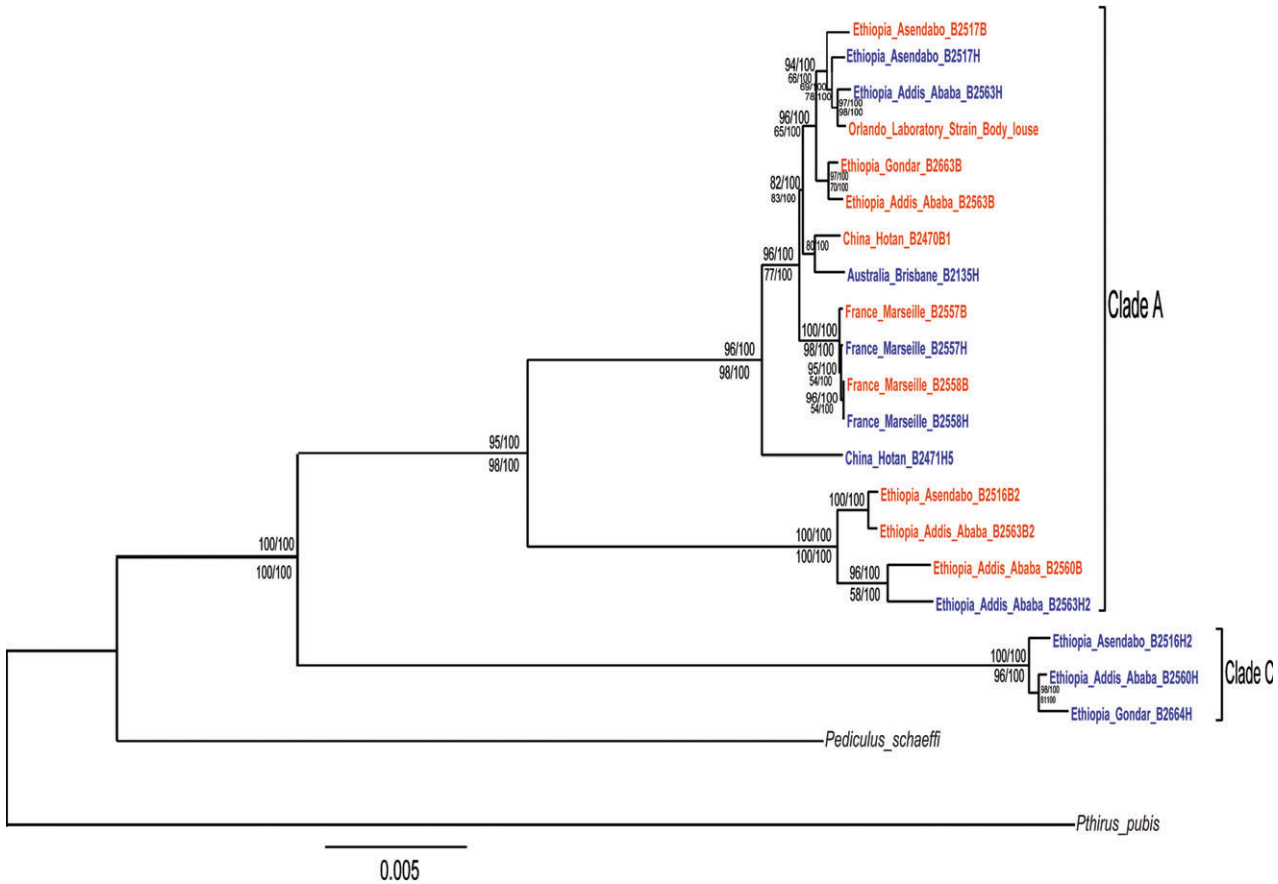

Fig. 1. Maximum likelihood trees inferred from the concatenated nucleotide sequences of entire mitochondrial coding-regions (13 protein coding genes, 2 rRNA genes and 22 tRNA genes; ~15.4 kb). Ten body lice (B2470B1, B2557B, B2558B, B2516B2, B2517B, B2560B, B2563B, B2563B2, B2663B and the Orlando laboratory strain of body lice), 11 head lice (B2470H5, B2557H, B2558H, B2516H2, B2517H, B2560H, B2563H, B2563H2, B2663H, B2664H and B2135H), *Pediculus schaeffi*, and *Pthirus pubis* (out-group). Numbers (GARLI support/Bayesian support) above branches are bootstrap support from nucleotide sequences whereas numbers (GARLI support/Bayesian support) below branches are bootstrap support for the putative amino acid sequences where support was > 50%. Bootstrap support < 50% not shown. Blue colour represents head lice and red colour represents body lice.

(a) entire mt coding-region nucleotide sequences, ~15.4 kb, and the putative amino acids; (b) entire *cox1* nucleotide sequences, ~1500 bp, and ~500 putative amino acid sequences; (c) 700 bp of *cox1* + 600 bp *cytb* nucleotide sequence, ~430 putative amino acid sequences; (d) 600 bp of *cytb* nucleotide sequence, ~200 putative amino acid sequences; and (e) 700 bp *cox1* nucleotide sequence, ~250 putative amino acid sequences. We found that 600 bp *cytb* or 700 bp *cox1* was sufficient to genotype body lice and head lice to either Clade A or Clade C. Entire mt coding-region sequences revealed, however, a well-supported (> 95% bootstrap support) phylogenetic structure within Clade A, that may well shed new light on the long and tangled evolutionary history of the body lice and head lice of humankind. Future studies with comprehensive samples of Clades A, B and C should reveal precisely how these three clades are related, how many times body lice evolved from head lice, whether or not the body lice in different regions of the world are monophyletic or polyphyletic, and how much phylogenetic structure is present in Clades A, B and C.

Materials and methods

Louse sampling and working definitions of body lice and head lice

Persons with body and/or head lice were identified at health clinics and hospitals. Lice removed for diagnostic purposes that would have otherwise been discarded were given to us. No personal identification information was collected or linked to the samples. Lice taken from the clothes and lice taken from the scalp-hair were put into separate tubes of 100% ethanol. Later, the lice in each tube were examined by S.C.B. with a hand-lens or a microscope. All of the lice from the scalp-hair at the three sites in Ethiopia and the one site in China were black; these lice were thereafter called head lice as they all lived in the scalp-hair. The vast majority of lice from the clothes at the three sites in Ethiopia and the one site in China were grey to white; these lice were thereafter called body lice as they all lived in the clothes. Occasionally, at the three sites in Ethiopia and the one site in China, S.C.B. found a single black louse in the clothes; these

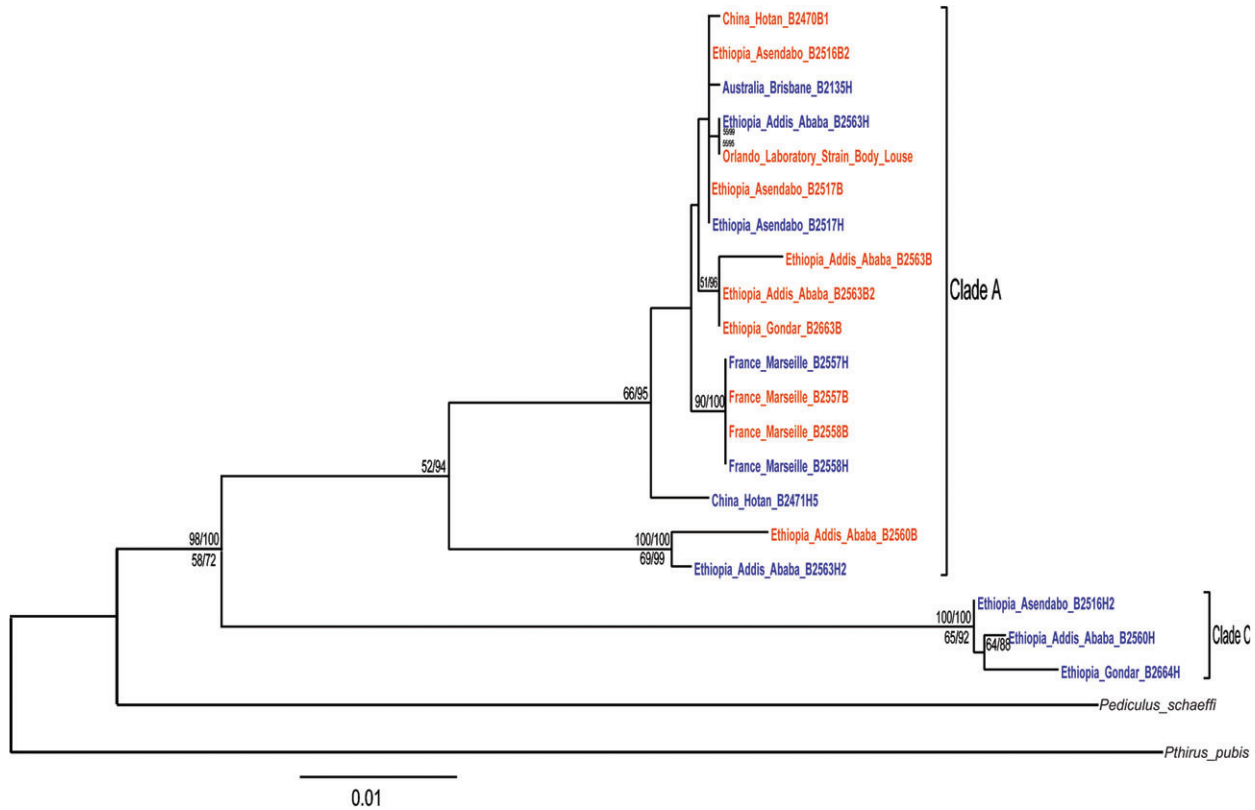

Fig. 2. Maximum likelihood trees inferred from the entire *cox1* coding-region (~1500 bp) of the 10 body lice, 10 head lice, *Pediculus schaeffi* and *Pthirus pubis* (outgroup). Numbers (GARLI support/Bayesian support) above branches are bootstrap support from nucleotide sequences whereas numbers (GARLI support/Bayesian support) below branches are bootstrap support for the putative amino acid sequences where support was > 50%. Bootstrap support < 50% not shown. Blue colour represents head lice and red colour represents body lice.

lice were thereafter called head lice as all the other black lice lived in the scalp-hair but they were set aside and not studied in this project. By contrast to lice from the scalp-hair in Ethiopia and China, lice from scalp-hair from the Madrague Ville Shelter for homeless people, Marseille, had a similar, often identical, greyish colour to lice from the clothes. Thus, the colour of the lice did not help to classify the lice at Marseille as body lice and head lice. Those lice from Marseille from the scalp-hair were therefore called head lice whereas lice from the clothes were called body lice. One head louse and one body louse were chosen at random for genetic analysis, from lice collected from each of the nine subjects. S.C.B. collected the lice from Marseille (France), Addis Ababa and Gondar (Ethiopia). R.J.P. collected the lice from Asendabo (Ethiopia). R.S. collected the lice from Hotan (China). Only lice from Clade A and C were studied; lice from Clade B were not available to us at the time of this effort.

DNA extraction, PCR amplification and nucleotide sequencing

Total-DNA was extracted from nine body lice and nine head lice from Ethiopia, China and France with a DNeasy Tissue Kit (QIAGEN Pty Ltd, Doncaster, Vic, Australia) (Table 1). One pair of PCR primers, forward USF1PH1: 5′-GAA ATT AAA ATT TCA ACA AAT CTC AAC TCG-3′) and reverse PHR10: 5′-CCC CCC AAG CTA TTT ATA GCT TGGAGTATTAAC GG-3′, was designed from the two conserved sequence blocks in the non-coding-regions of the mini-chromosomes to amplify 20 mini-chromosomes (Shao *et al.*, 2012). La Taq DNA polymerase kits (Takara; Scientifix Pty Ltd, Clayton, Vic, Australia) were used for all the amplifications. The PCR conditions were: 96 °C 2 min; 39 cycles: 98 °C 10 s, 62 °C 30 s, 72 °C 2.5 min; 72 °C 5 min; and 25 °C 2 min. The Wizard® SV Gel and PCR Clean-Up System (Promega, Sydney, NSW, Australia) were used to purify and concentrate the PCR products for sequencing. Purified PCR products were sent to the Beijing Genome Institute (BGI, Hong Kong, China) for second-generation sequencing. The purified PCR products were sheared randomly to generate DNA fragments 140–800 bp long, end-repaired and adenine (A) bases were added to the 3′ ends at the BGI. Sequencing adapters were then ligated to the modified DNA fragments. Methods for the enrichment of adapter-ligated DNA fragments for Illumina Hiseq2000 sequencing are described at http://bgitechsolutions.com/service-solutions/services/genomics/de-novo-sequencing.

Sequences assembly and phylogenetic analyses

The raw Illumina sequence-reads were ‘cleaned up’ in two steps: first, the adapter sequences were removed, and then the

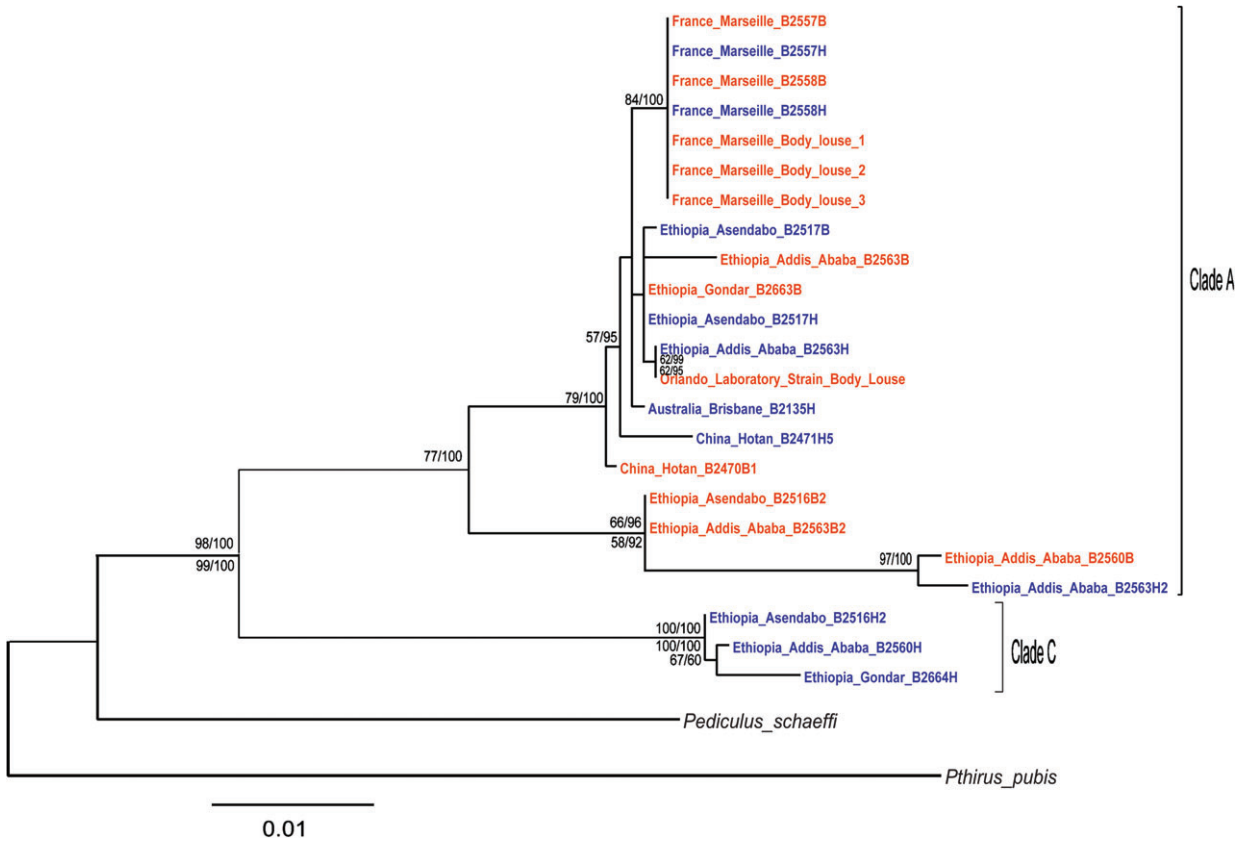

Fig. 3. Maximum likelihood tree inferred from the concatenated coding-region nucleotide sequences of part of *cox1* (~700 bp) and part of *cytb* (~600 bp) genes of 10 body lice, 10 head lice, *Pediculus schaeffi*, and *Pthirus pubis* (out-group). Numbers (GARLI support/Bayesian support) above branches are bootstrap support from nucleotide sequences whereas numbers (GARLI support/Bayesian support) below branches are bootstrap support for the putative amino acid sequences where support was > 50%. Bootstrap support < 50% not shown. Blue colour represents head lice and red colour represents body lice.

sequence-reads with more than 50% low-quality nucleotides (quality value ≤ 5) were removed. Second, the modified sequence-reads, 90 bp in length, were assembled into contigs according to the mt genome sequences of *P. humanus* described in Shao *et al.* (2009).

Consensus sequences of the coding-regions of the mt mini-chromosomes of body lice and head lice, the chimpanzee louse (*Pediculus schaeffi*) and the crab louse (*Pthirus pubis* Linnaeus) were generated by the 'Generate Consensus Sequence' function in Geneious 5.6.2 (Kearse *et al.*, 2012) with the following two settings: threshold –50%; and assign quality –Total (Sum quality of contributing bases – non-contributing bases). The Orlando laboratory strain of the body louse (Kirkness *et al.*, 2010), a head louse from Brisbane (B2135H), *P. schaeffi* (K. Herd, S.C. Barker & R. Shao, unpublished data, 2013) and *Pt. pubis* (Shao *et al.*, 2012) were sequenced by us previously. Putative open reading frames (ORFs) were found by Geneious 5.6.2 for each of the 13 mt protein-coding genes. The tRNA genes were aligned in Geneious 5.6.2 with the global alignment option, with the settings: (a) free end gaps, with 65% similarity in the cost matrix; and (b) gap open penalty and the gap extension penalty of 3. The ORFs of the 13 protein-coding genes were aligned and trimmed with PAL2NAL (http://www.bork.embl.de/pal2nal/) according to putative-protein alignments from Geneious. The two rRNA genes of each of the body lice, head lice, *P. schaeffi* and *Pt. pubis* were joined to generate concatenated sequences in Geneious, and aligned and trimmed with TrimAl 1.3 in Phylemon 2.0 (Sánchez *et al.*, 2011). The tRNA genes were modified as per the rRNA genes. Then, the 37 mt genes (13 protein-coding genes, 2 rRNA genes, 22 tRNA genes) were joined together in Geneious 5.6.2 to make a concatenated sequence of the entire mt coding-region. We partitioned our entire coding-regions (~15.4 kb) into seven datasets: (a) the 13 protein-coding genes; (b) the first codon position of the 13 mt coding genes; (c) the second codon position of the 13 mt coding genes; (d) the third position codon of the 13 mt coding genes; (e) the two rRNA genes; (f) the 22 tRNA genes; and (g) the coding-regions from each of the 20 mini-chromosomes (mini-chromosomes *atp6-atp8*, *cytb*, *cox1*, *trnY-cox2*, *cox3-trnA*, *nad1-trnQ*, *trnP-nad2-trnI*, *trnR-nad3*, *trnK-nad4*, *trnG-nad4L-trnV*, *nad5*, *trnF-nad6*, $trnL_{1(tag)}$-*rrnL*, $trnL_{2(taa)}$-*rrnL*, $trnL_{1(tag)}$-*rrnS-trnC*, $trnL_{2(taa)}$-*rrnS-trnC*, *trnE-trnN-*$trnS_{1(tct)}$, *trnH-trnD-trnT*, *trnM* and $trnS_{2(tga)}$-*trnW*). We excluded the 345 bp of the mt coding-region that is known to recombine in body lice and head lice (Shao *et al.*, 2009, 2012): (a) 127 bp of *nad 5* and *nad 4*; (b)

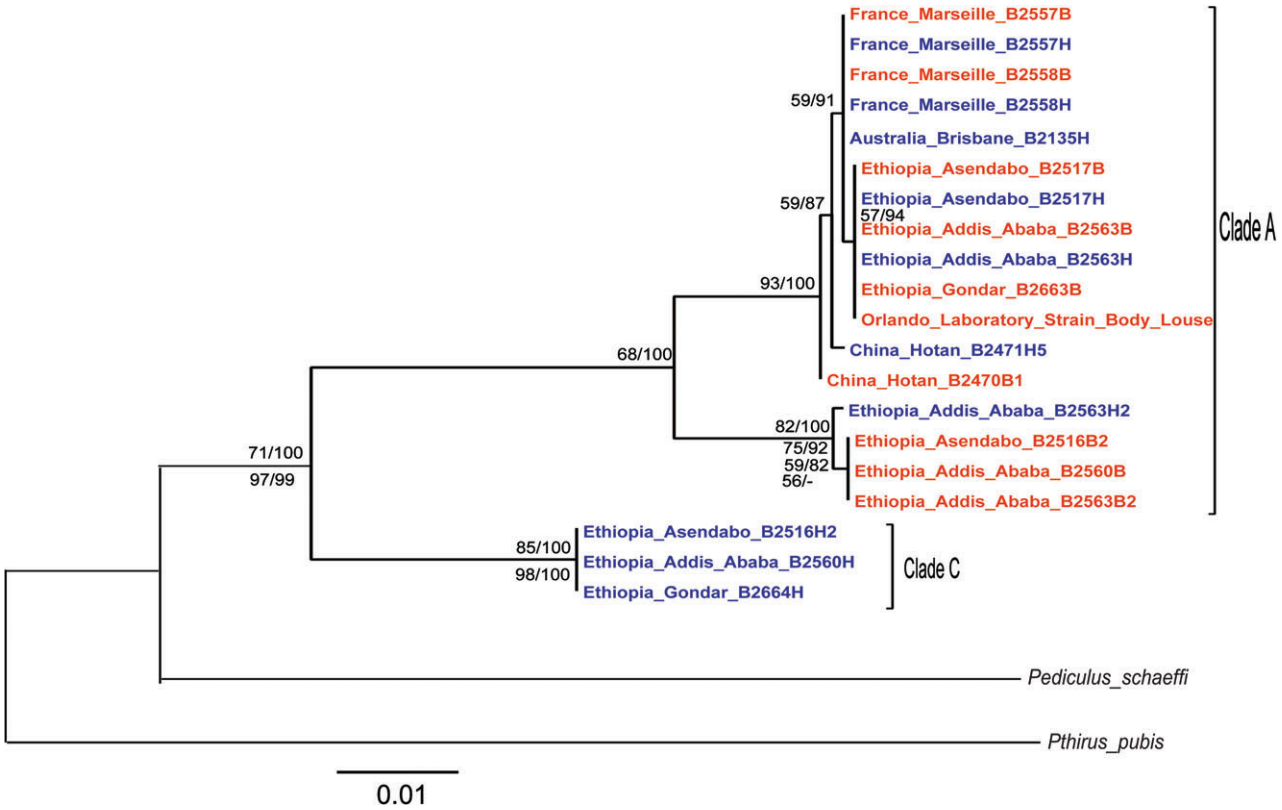

Fig. 4. Maximum likelihood trees inferred from nucleotide sequences from part of the *cytb* (~600 bp) of the 10 body lice, 10 head lice, *Pediculus schaeffi*, and *Pthirus pubis* (outgroup). Numbers (GARLI support/Bayesian support) above branches are bootstrap support from nucleotide sequences whereas numbers (GARLI support/Bayesian support) below branches are bootstrap support for the putative amino acid sequences where support was > 50%. Bootstrap support < 50% not shown. Blue colour represents head lice and red colour represents body lice.

107 bp of *nad 5* and *rrnL*; (c) the entire $trnL_1$ and $trnL_2$ (70 bp); and (d) 28 and 13 bp of *trnR* and *trnG*. The 'best DNA/Protein model' (Protest and Jmodeltest in Phylemon 2.0), with the default parameters, was used to infer maximum likelihood trees and bayesian trees from: (a) entire mt coding-regions (~15.4 kb); (b) partial *cox 1* (700 bp); (c) partial *cytb* (600 bp); (d) concatenated sequences of the 700 bp *cox1* and 600 bp *cytb* sequence; and (e) entire *cox1* sequences (~1500 bp), from the consensus sequences of the 10 body lice, 10 head lice, *P. schaeffi* and *Pt. pubis* (see Supporting information, File S1–S10, for the nucleotide and putative amino acid sequence alignments). MrBayes v 3.2.0 (Ronquist *et al.*, 2012) and GARLI v 2.0 (Zwickl, 2006; Bazinet & Cummings, 2011) were used to measure bootstrap support. In MrBayes, the MCMC runs were sampled and printed every 100 generations; diagnostics were computed every 1000 generations. Analyses stopped when the average standard deviation of split frequencies was below 0.01. The prime settings in GARLI were: (a) analysis type: bootstrap; (b) number of replicates: 1000; and (c) model type: nucleotide/amino acid (Table 2). The other settings in GARLI were the default settings. *Pthirus pubis* was used for out-group reference as the genus *Pthirus* is apparently the sister-group to the genus *Pediculus* (Barker *et al.*, 2003; Johnson *et al.*, 2004; Murrell & Barker, 2005). The bootstrap trees were summed to generate consensus trees with SumTrees v 3.3.1 (Sukumaran & Holder, 2010). Bootstrap values were added to the bootstrap consensus tree of each dataset with SumTrees v 3.3.1.

Results

The GC ratio was substantially lower than the AT ratio in the body lice, head lice and chimpanzee louse, *P. schaeffi*, compared with the crab louse, *Pt. pubis* (Table 3). Whereas the nucleotide composition of the entire mt coding-regions of the 17 body lice and head lice from Clade A were similar to one another, the three head lice from Clade C had statistically significantly lower A% than the other lice ($P < 0.05$) (Table 3).

Pediculus schaeffi, the chimpanzee louse, was the sister-group to the 20 body lice and head lice in all 10 of our phylogenetic trees with 71–100% bootstrap support (Figs 1–5). The two main clades in our trees were: (a) a head louse from each of the three towns and cities in Ethiopia, Asendabo, Gondar and Addis Ababa (Clade C); and (b) the other head lice and body lice from Asendabo, Gondar and Addis Ababa (Ethiopia) + the body lice and head lice from China and France, the head louse from Australia, and the Orlando laboratory strain of the body louse (Clade A) (Figs 1–5).

The number of branches that had > 95% bootstrap support indicates the strength of the phylogenetic signal in trees of the

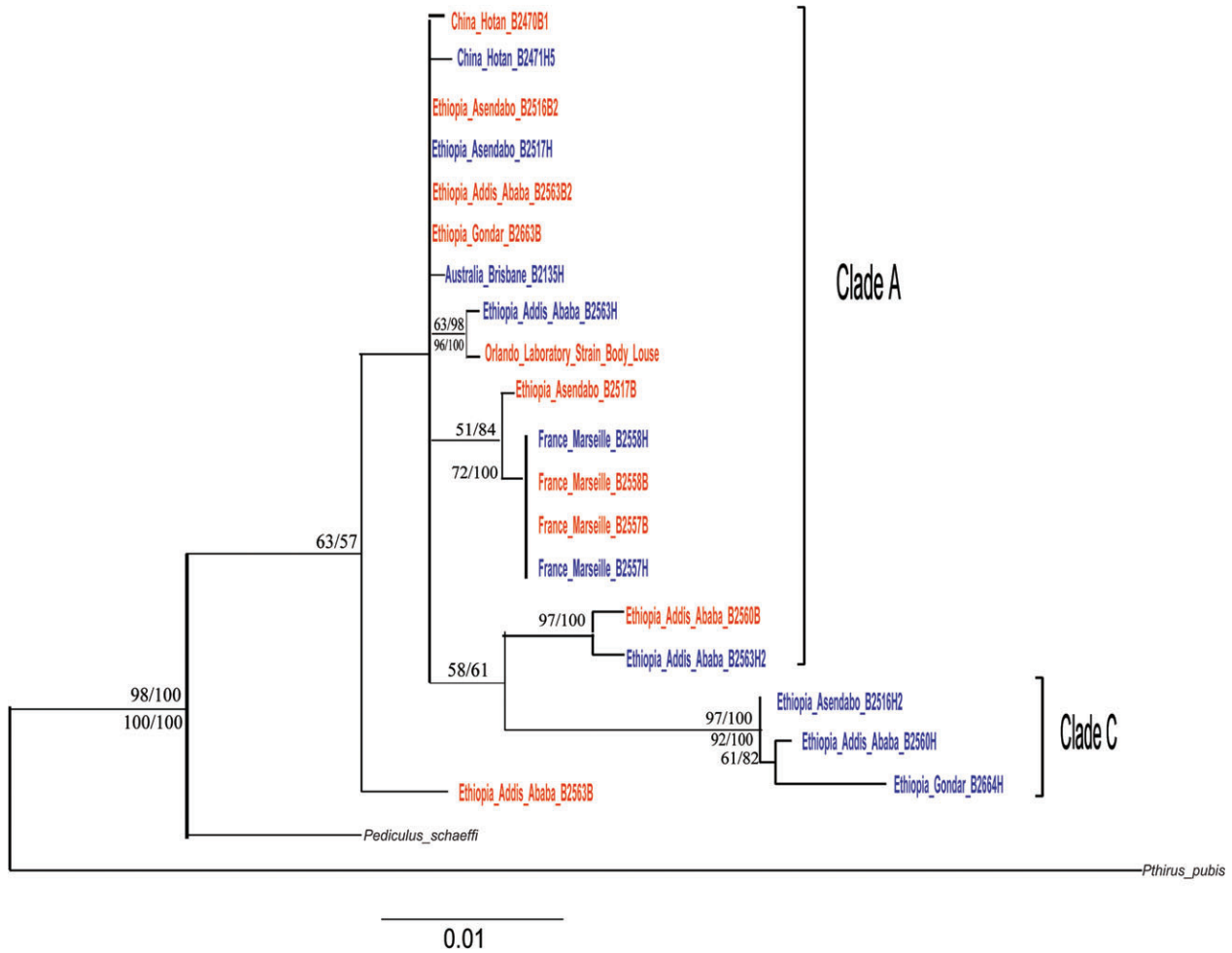

Fig. 5. Maximum likelihood trees inferred from nucleotide sequences from part of *cox1* (~700 bp) of the 10 body lice, 10 head lice, *Pediculus schaeffi*, and *Pthirus pubis* (out-group). Numbers (GARLI support/Bayesian support) above branches are bootstrap support from nucleotide sequences whereas numbers (GARLI support/Bayesian support) below branches are bootstrap support for the putative amino acid sequences where support was > 50%. Bootstrap support < 50% not shown. Blue colour represents head lice and red colour represents body lice.

current study (Table 4). On the one hand, the trees from the nucleotide sequences and putative amino acid sequences of the entire mt coding-regions had many branches with > 95% bootstrap support (Fig. 1). Indeed all 19 of the 19 branches in the trees from entire mt coding-regions had > 95% bootstrap support from the nucleotides and/or putative amino acids. By contrast, fewer than a third of the branches had > 95% support from the nucleotide sequences of entire *cox1* (~1.5 kb), the concatenated nucleotide sequences of partial *cox1* (~700 bp) and the partial *cytb* (~600 bp); these data sets allowed body lice and head lice to be genotyped to Clade A or Clade C, but revealed little else.

The trees from four of the six partitioned datasets (13 mt protein-coding genes, first, second codon positions, and 22 tRNA genes; Figures S1 and S3–S5) had strong support (71–100% bootstrap) for Clade A and Clade C whereas the other two data sets (22tRNA genes, first codon position) had four Ethiopian lice (B2516B2, B2560B, B2563B2, B2563H2) + the Clade C head lice (B2516H2, B2560H and B2664H) in a weak to well-supported clade (59–100% bootstrap support) (Figures S2 and S6). The trees from the entire mt coding-regions, the regions that are known to recombine (~15.1 kb, Figure S7), were similar to the tree from the entire mt coding region (~15.4 kb, Fig. 1), apart for slight differences in the bootstrap support of branches.

Trees from the coding-regions of 12 of the 13 individual mini-chromosomes (mini-chromosomes *atp6-atp8*, *cox1*, *cox3-TrnA*, *cytb*, *TrnK-nad4*, *TrnG-nad4L-TrnV*, *nad5*, *TrnF-nad6*, *TrnL$_1$-rrnL*, *TrnL$_2$-rrnS*, *TrnP-nad2-TrnI* and *TrnY-cox2*) had Clade A and Clade C (99–100% bootstrap support); only the tree from mini-chromosome *TrnR-nad3* did not have Clade A and Clade C (data not shown).

Discussion

In previous studies of the phylogeny of the lice of humankind, the current study and those of our colleagues used a single pair of PCR primers to amplify fragments of mt genes that were

Table 4. Numbers of branches with > 95% bootstrap support in our phylogenetic trees.

Nucleotide dataset	Number of branches with (> 95%)	Putative amino acid sequence dataset	Bootstrap support for branches (> 95%)
Entire mt coding-region (~15.4 kb)	14	Entire mt coding-region (~15.4 kb)	6
Entire *cox1* (~1.5 kb)	3	Entire *cox1* (~1.5 kb)	0
700 bp *cox1* + 600 bp *cytb*	3	700 bp *cox1* + 600 bp *cytb*	2
600 bp *cytb*	0	600 bp *cytb*	2
700 bp *cox1*	3	700 bp *cox1*	2

mt, mitochondrial genome.

hundreds of nucleotides long. In the present study, however, a single pair of PCR primers was used to amplify entire mt coding regions (~15.4 kb). Routine PCR-sequencing of entire mt coding-regions with just one pair of PCR primers is straightforward if second-generation sequencing platforms such as Illumina Hiseq 2000 and 454 are used, as just one pair of PCR primers will amplify all 20 mini-chromosomes of the lice of humankind (Shao *et al.*, 2009; Shao *et al.*, 2012; Herd *et al.*, 2012; and Materials and methods section in the present paper). If sequencing entire mt genomes is not logistically feasible and only partial mt genomes can be studied, it is advocated that sequencing of the entire cox1 gene (~1500 bp) be done as an additional advantage of studying entire *cox1* sequences is that *cox1* sequenced may be used by people in the DNA barcode of life enterprise (Hebert *et al.*, 2003; Hajibabaei *et al.*, 2006; Chen *et al.*, 2009). Thus, data from the *Pediculus* lice of humankind might contribute to this enterprise and vice versa.

Clades A and C in the present study are consistent with Clades A and C of previous studies (Kittler *et al.*, 2003; Reed *et al.*, 2004; the unpublished analyses of the *cytb* sequences from Kittler *et al.*, 2003 and Reed *et al.*, 2004 with our sequences). Phylogenies from entire coding-region sequences (~15.4 kb) were much more informative than phylogenies from entire *cox1* sequence (~1.5 kb) and partial gene sequences (~600 to ~700 bp). For example, 19 of 19 branches had > 95% bootstrap support in our maximum likelihood tree from the entire coding-regions (~15.4 kb) whereas the tree from 700 bp *cox1* had only two branches with bootstrap support > 95%. Partial *cytb* (~600 bp) sequences were sufficient to genotype lice to either Clade A, B or C. It is becoming increasingly clear, however, that there is much phylogenetic structure within Clade A [(Veracx *et al.*, 2012b); and the present study]. Moreover, it is likely that there is an informative phylogenetic structure within Clade B and Clade C as well. Entire mt coding-region sequences are almost certainly needed to resolve the phylogenetic structure within Clades A, B and C. Doubtless, resolution of this phylogenetic structure will shed new light on the tangled phylogeny of the body lice and head lice of humankind.

Finally, the current study acknowledges that mt sequences give mt genome history; nuclear gene histories will continue to be important. In addition the current study notes that entire mt coding-regions hold promise for dating the divergence of the three Clades (A, B, C) and sub-clades of the *Pediculus* lice of humankind, e.g. the African cluster of body lice and head lice (Veracx *et al.*, 2012b).

Supporting Information

Additional Supporting Information may be found in the online version of this article under the DOI reference: DOI: 10.1111/mve.12076

Figure S1. Maximum likelihood tree inferred from the first partition dataset: the 13 mt protein coding-genes (~11 kb). Ten body lice (B2470B1, B2557B, B2558B, B2516B2, B2517B, B2560B, B2563B, B2563B2, B2663B and the Orlando laboratory strain of body lice), 11 head lice (B2470H5, B2557H, B2558H, B2516H2, B2517H, B2560H, B2563H, B2563H2, B2663H, B2664H and B2135H), *Pediculus schaeffi*, and *Pthirus pubis* (out-group). Numbers above branches are bootstrap support from nucleotide sequences whereas numbers below branches are bootstrap support for the putative amino acid sequences where support was > 80%. Bootstrap support < 80% not shown.

Figure S2. Maximum likelihood tree inferred from the fourth partition dataset: first codon of 13 protein-coding gene. Samples and tree details as per Figure S1.

Figure S3. Maximum likelihood tree inferred from the fifth partition dataset: second codon of 13 protein-coding gene. Samples and tree details as per Figure S1.

Figure S4. Maximum likelihood tree inferred from the sixth partition dataset: third codon of 13 protein-coding gene. Sample and tree details as per Figure S1.

Figure S5. Maximum likelihood tree inferred from the second partition dataset: the two rRNA genes. Samples of body lice and head lice, body lice in red and head lice in blue, and tree details as per Figure S1.

Figure S6. Maximum likelihood tree inferred from the third partition dataset: the 22 mt tRNA genes. Samples and tree details as per Figure S1.

Figure S7. Maximum likelihood tree inferred from the concatenated nucleotide sequences of entire mitochondrial coding-regions – the regions that recombine; these regions have identical nucleotide sequences. Sequence length: ~15.1 kb. Samples and tree details as per Figure S1.

File S1. Nucleotide sequence alignment used to infer Fig. 1 (fasta format).

File S2. Nucleotide sequence alignment used to infer Fig. 2 (fasta format).

File S3. Nucleotide sequence alignment used to infer Fig. 3 (fasta format).

File S4. Nucleotide sequence alignment used to infer Fig. 4 (fasta format).

File S5. Nucleotide sequence alignment used to infer Fig. 5 (fasta format).

File S6. Putative amino acid sequence alignment used to infer Fig. 1.

File S7. Putative amino acid sequence alignment used to infer Fig. 2.

File S8. Putative amino acid sequence alignment used to infer Fig. 3.

File S9. Putative amino acid sequence alignment used to infer Fig. 4.

File S10. Putative amino acid sequence alignment used to infer Fig. 5.

Acknowledgements

We thank Tom Burger (University of Queensland) for showing us how to partition our data and for advice on phylogenetics; Aurelie Veracx (Aix-Marseille Université) for helping S.C.B. to collect the four lice from Marseille; and Mr Dell Kittey and Olit Olijira for helping R.J.P. to collect the four lice from Asendabo.

Author's declaration of interests

No competing interests have been declared.

References

Angelakis, E., Diatta, G., Abdissa, A. *et al.* (2011) Altitude-dependent *Bartonella quintana* genotype C in head lice, Ethiopia. *Emerging Infectious Diseases*, **17**, 2357–2359.

Barker, S.C., Whiting, M.F., Johnson, K.P. *et al.* (2003) Phylogeny of the lice (Insecta, Phthiraptera) inferred from small subunit rRNA. *Zoologica Scripta*, **32**, 407–414.

Bazinet, A.L. & Cummings, M.P. (2011) Computing the tree of life: Leveraging the power of desktop and service grids. *Parallel and Distributed Processing Workshops and Phd Forum (IPDPSW), 16–20 May, 2011, Shanghai*, pp. 1896–1902. IEEE, Shanghai, China.

Boutellis, A., Mediannikov, O., Bilcha, K.D. *et al.* (2013) *Borrelia recurrentis* in head lice, Ethiopia. *Emerging Infectious Diseases*, **19**, 796–798.

Busvine, J.R. (1978) Evidence from double infestations for the specific status of human head lice and body lice (Anoplura). *Systematic Entomology*, **3**, 1–8.

Cameron, S.L. (2014) Insect mitochondrial genomics: implications for evolution and phylogeny. *Annual Review of Entomology*, **59**, 95–117.

Chen, W., Seifert, K.A. & Lévesque, C.A. (2009) A high density *COX1* barcode oligonucleotide array for identification and detection of species of *Penicillium* subgenus *Penicillium*. *Molecular Ecology Resources*, **9**, 114–129.

Ferris, G.F. (1951) The sucking lice. *Memoirs of the Pacific Coast Entomological Society*, **1**, 1–320.

Freund, L. (1932) Anoplua. *Die Tierwelt Mittleuropas*. Quelle anal Meyer, Leipzig, Germany.

Haitham, E., Gimenez, G., Sokhna, C. *et al.* (2012) Multispacer sequence typing relapsing fever borreliae in Africa. *PLoS Neglected Tropical Diseases*, **6**, e1652.

Hajibabaei, M., Janzen, D.H., Burns, J.M. *et al.* (2006) DNA barcodes distinguish species of tropical Lepidoptera. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 968–971.

Hebert, P.D., Cywinska, A., Ball, S.L. *et al.* (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences*, **270**, 313–321.

Herd, K., Barker, S.C. & Shao, R.F. (2012) High-level of heteroplasmy in the mitochondrial *cox1*-minichromosome of the human body louse, *Pediculus humanus*, and the human head louse, *Pediculus capitis*. *The Open Genomics Journal*, **5**, 14–17.

Johnson, K.P., Yoshizawa, K. & Smith, V.S. (2004) Multiple origins of parasitism in lice. *Proceedings of the Royal Society of London B*, **271**, 1771–1776.

Kearse, M., Moir, R., Wilson, A. *et al.* (2012) Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, **28**, 1647–1649.

Kirkness, E.F., Haas, B.J., Sun, W.L. *et al.* (2010) Genomec sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 12168–12173.

Kittler, R., Kayser, M. & Stoneking, M. (2003) Molecular evolution of *Pediculus humanus* and the origin of clothing. *Current Biology*, **13**, 1414–1417.

Leo, N.P. & Barker, S.C. (2005) Unraveling the origins of the head lice and body lice of humans. *Parasitology Research*, **98**, 44–47.

Leo, N.P., Campbell, N.J.H., Yang, X. *et al.* (2002) Evidence from mitochondrial DNA that head lice and body lice of humans (Phthiraptera: Pediculidae) are conspecific. *Journal of Medical Entomology*, **394**, 662–666.

Leo, N.P., Hughes, J.M., Yang, X. *et al.* (2005) The head and body lice of humans are genetically distinct (Insecta: Phthiraptera:Pediculidae): evidence from double infestations. *Heredity*, **95**, 34–40.

Li, W., Ortiz, G., Fournier, P.E. *et al.* (2010) Genotyping of human lice suggests multiple emergencies of body lice from local head louse populations. *PLoS Neglected Tropical Diseases*, **4**, e641.

Light, J.E. & Reed, D.L. (2009) Multigene analysis of phylogenetic relationships and divergence times of primate sucking lice (Phthiraptera: Anoplura). *Molecular Phylogenetics and Evolution*, **50**, 376–390.

Light, J.E., Toups, M.A. & Reed, D.L. (2008a) What's in a name: the taxonomic status of human head and body lice. *Molecular Phylogenetics and Evolution*, **47**, 1203–1216.

Light, J.E., Allen, J.M., Long, L.M. *et al.* (2008b) Geographic distributions and origins of human head lice (*Pediculus humanus capitis*) based on mitochondrial data. *The Journal of Parasitology*, **94**, 1275–1281.

Murrell, A. & Barker, S.C. (2005) Multiple origins of parasitism in lice: phylogenetic analysis of SSU rDNA indicates that the Phthiraptera and Psocoptera are notmonophyletic. *Parasitology Research*, **97**, 274–280.

Ramos, J.M., Malmierca, E., Reyes, F. *et al.* (2009) Louse-borne relapsing fever in Ethiopian children: experience of a rural hospital. *Tropical Doctor*, **39**, 34–36.

Raoult, D. & Roux, V. (1999) The body louse as a vector of reemerging human diseases. *Clinical Infectious Diseases*, **29**, 888–911.

Raoult, D., Ndihokubwayo, J., Tissot-Dupont, H. *et al.* (1998) Outbreak of epidemic typhus associated with trench fever in Burundi. *The Lancet*, **352**, 353–358.

Reed, D.L., Smith, V.S., Hammond, S.L. *et al.* (2004) Genetic analysis of lice supports direct contact between modern and archaic humans. *PLoS Biology*, **2**, e340.

Robinson, D., Leo, N., Prociv, P. *et al.* (2003) Potential role of head lice, *Pediculus humanus capitis*, as vectors of *Rickettsia prowazekii*. *Parasitology Research*, **90**, 209–211.

Ronquist, F., van der Teslenko, M., Marker, P. *et al.* (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, **61**, 539–542.

Sánchez, R., Serra, F., Tárraga, J. *et al.* (2011) Phylemon 2.0: a suite of web-tools for molecular evolution, phylogenetics, phylogenomics and hypotheses testing. *Nucleic Acids Research*, **39**, W470–W474. doi: 10.1093/nar/gkr408.

Shao, R. & Barker, S.C. (2007) Mitochondrial genomes of parasitic arthropods: implications for studies of population genetics and evolution. *Parasitology*, **134**, 153–167.

Shao, R., Kirkness, E.F. & Barker, S.C. (2009) The single mitochondrial chromosome typical of animals has evolved into 18 minichromosomes in the human body louse, *Pediculus humanus*. *Genome Research*, **19**, 904–912.

Shao, R., Zhu, X.Q., Barker, S.C. & Herd, K. (2012) Evolution of extensively fragmented mitochondrial genomes in the lice of humans. *Genome Biology and Evolution*, **4**, 1088–1101.

Sukumaran, J. & Holder, M.T. (2010) DendroPy: a Python library for phylogenetic computing. *Bioinformatics*, **26**, 1569–1571.

Toups, M.A., Kitchen, A., Light, J.E. & Reed, D.L. (2011) Origin of clothing lice indicates early clothing use by anatomically modern humans in Africa. *Molecular Biology and Evolution*, **28**, 29–32.

Veracx, A. & Raoult, D. (2012) Biology and genetics of human head and body lice. *Trends in Parasitology*, **28**, 563–571.

Veracx, A., Rivet, R., McCoy, K.D. *et al.* (2012a) Evidence that head and body lice on homeless persons have the same genotype. *PLoS One*, **7**, e45903.

Veracx, A., Boutellis, A., Merhej, V. *et al.* (2012b) Evidence for an African cluster of human head and body lice with variable colors and interbreeding of lice between continents. *PLoS One*, **7**, e37804.

Zwickl, D.J. (2006) Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. PhD Dissertation. The University of Texas at Austin, Austin, TX.

Accepted 28 January 2014