npg

ORIGINAL ARTICLE

Thorsellia anophelis is the dominant bacterium in a Kenyan population of adult Anopheles gambiae mosquitoes

Aurelio M Briones[1], Josephat Shililu[2], John Githure[2], Robert Novak[3] and Lutgarde Raskin[1]

[1]Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, MI, USA; [2]International Centre of Insect Physiology and Ecology, Nairobi, Kenya and [3]Illinois Natural History Survey, Champaign, IL, USA

Anopheles gambiae mosquitoes are not known to harbor endosymbiotic bacteria. Here we show, using nucleic acid-based methods, that 16S rRNA gene sequences specific to a recently described mosquito midgut bacterium, *Thorsellia anophelis*, is predominant in the midgut of adult *An. gambiae s.l.* mosquitoes captured in residences in central Kenya, and also occurs in the aquatic rice paddy environment nearby. PCR consistently detected *T. anophelis* in the surface microlayer of rice paddies, which is also consistent with the surface-feeding behavior of *A. gambiae s.l.* larvae. Phylogenetic analysis of cloned environmental 16S rRNA genes identified four major *Thorsellia* lineages, which are closely affiliated to an insect endosymbiont of the genus *Arsenophonus*. Physiological characterizations support the hypothesis that *T. anophelis* is well adapted to the female anopheline midgut by utilizing blood and tolerating the alkaline conditions in this environment. The results suggest that aquatically derived bacteria such as *T. anophelis* can persist through mosquito metamorphosis and become well-established in the adult mosquito midgut.

Introduction

The life cycle of the malaria parasite (*Plasmodium* spp) involves critical steps within the mosquito midgut, an environment it shares with bacteria. Various physical (for example, membranes, receptors), biochemical (for example, proteases, chitinases) and physiological (for example, immunological factors) barriers protect the insect against the parasite (for reviews see Sinden, 2002; Meister *et al.*, 2004), but the largest losses in parasite titer occur when *Plasmodium* zygotes develop into ookinetes within the adult mosquito midgut. During this transition, the parasite may undergo more than two orders of magnitude loss in abundance (Vaughan *et al.*, 1994). Although the negative impact of midgut bacteria on parasite development has been observed in the past (Pumpuni *et al.*, 1993; Lowenberger *et al.*, 1999; Gonzalez-Ceron *et al.*, 2003),

a symbiotic or consistent association between *Anopheles gambiae s.l.* and specific species of bacteria has not been described. Among other *Anopheles* species, *An. stephensi* has been reported to stably associate with bacteria of the genus *Asaia* (Favia *et al.*, 2007). Such a relationship would make use of genetically transformed bacteria in paratransgenesis more efficient (Beard *et al.*, 2002). Paratransgenics refers to the genetic transformation of a bacterium, typically a vector endosymbiont, to produce antiparasitic molecules, thus inhibiting parasite development (Conte, 1997; Beard *et al.*, 2002). Since *Anopheles* mosquitoes are currently considered a dead end host for midgut bacteria that are believed to 'perish when the adult mosquito dies' (Riehle and Jacobs-Lorena, 2005), the current paradigm for paratransgenesis in *Anopheles* focuses on bacteria that opportunistically infect adult mosquitoes (Lindh *et al.*, 2006; Favia *et al.*, 2007; Riehle *et al.*, 2007).

The most significant perturbation affecting mosquito midgut microbial communities occurs during metamorphosis from pupae to adults. For bacteria inhabiting the midgut, the process has been characterized as similar to 'gut sterilization' (Moll *et al.*, 2001), which, if correct, would imply

that environmentally derived bacteria that are present in larval midguts (through feeding) are greatly attenuated or are absent in adult midguts. It should be stressed, however, that microbial studies of the midgut of *Anopheles* are scarce, and have depended mainly on traditional culture-based techniques (Pumpuni *et al.*, 1996; Straif *et al.*, 1998; Gonzalez-Ceron *et al.*, 2003; Lindh *et al.*, 2005). One study has coupled culture- and PCR-based approaches to characterize *An. gambiae s.l.* midgut bacteria (Lindh *et al.*, 2005). Therefore, the application of other nucleic acid-based tools, such as denaturing gradient gel electrophoresis (DGGE) (Muyzer *et al.*, 1993), to study these systems is timely.

In this study, we provide evidence from PCR, DGGE and clone library analyses that adult field-captured *An. gambiae s.l.* mosquitoes in a Kenyan village are consistently associated with a *Thorsellia anophelis* lineage that was also detected in the surface microlayer of rice paddies. Unlike most members of the Enterobacteriaceae, this bacterium grows slowly in rich media, but is well adapted to the anopheline midgut because of its ability to utilize blood to enhance growth and its tolerance to alkaline conditions in the anopheline midgut.

Materials and methods

Study site and sample collection
The study site is situated within the Mwea Irrigation and Agricultural Development experimental station, located approximately 100 km northeast of Nairobi, Kenya (37° 20′ E and 0° 40′ S), at an elevation of 1159 m above sea level. The site falls within a region of low stable transmission of the malaria parasite, which infects around 20% of the human population (Mutero *et al.*, 2000). A detailed description of the site has been reported elsewhere (Muturi *et al.*, 2006). The experimental plots were laid out in a 0.4 ha area, which is situated within 100 m of Kariua, a village inhabited mostly by subsistence-level farmers living in houses with walls constructed of mud or wood. The experimental plots are arranged in eight blocks (rows) subdivided into eight plots (6.3 × 3.15 m), with each block hydrologically isolated using unidirectional inflow and outflow canals. Among the 64 plots, two adjacent rows (designated rows 1 and 2) were set aside for the activities described in this study. All the plots were exposed to natural colonization by mosquitoes. Previous fieldwork has shown that the most abundant anopheline species within Mwea Irrigation and Agricultural Development rice paddies is *An. arabiensis* (a sub-species of *An. gambiae s.l.*) (Mutero *et al.*, 2000).

Mosquito and water samples were collected in June and July 2005. Mosquito larvae and rice paddy water were collected during a fallow period when the plots had been continuously flooded for 6 days prior to sampling. Surface microlayer (SML) samples of two randomly selected plots from each row (designated plots A and B) were collected using a stainless steel mesh screen measuring 400 cm², with a mesh size of 1.25 and 0.36 mm wire diameter as described previously (Agogue *et al.*, 2004). Mosquito larvae and bulk water from four plots (plots A, B, C and D; two plots per row) were sampled with a 350 ml plastic dipper, using standard dipping techniques (Service, 1993). Since the larval abundance was low during the time of sampling, mosquito larvae from the remaining 12 plots were also collected and pooled according to the row. Among the pooled water samples from each plot, 300 ml (from 30 pooled dips of SML water) and 500 ml (from 20 pooled dips of bulk water) aliquots were placed in sterile Nalgene bottles (Nalge Nunc International, Rochester, NY, USA) and transported on ice immediately to the laboratory. All water and larval samples were processed within 24 h of collection. Female *An. gambiae s.l.* fourth instar larvae were selected morphologically (Gillies and Coetzee, 1987) and rinsed three times in filter-sterilized phosphate-buffered saline (137 mM sodium chloride, 2.7 mM potassium chloride and 10 mM sodium phosphate dibasic/potassium phosphate monobasic, pH 7.4) prior to dissection. The larvae were dissected using aseptic techniques, with the aid of a dissecting microscope (Zeiss Stemi 1000; Carl Zeiss Inc., Jena, Germany). Each midgut sample subjected to DNA extraction consisted of 30 larval midguts pooled together in sterile phosphate-buffered saline and corresponded to larvae collected from two adjacent rows of plots.

Indoor resting mosquitoes from two houses (designated house 1 and house 2) approximately 100 m apart were captured in the morning hours starting around 0700 hours using the pyrethrum spray collection method (Service, 1993). Immediately after collection, mosquitoes were transported to the laboratory and adult female *An. gambiae s.l.* mosquitoes were selected based on morphology (Gillies and Coetzee, 1987). The samples were stored at −20 °C in phosphate-buffered saline solution containing 30% glycerol and 1% Tween 80. Dissection and midgut collection of adult mosquitoes was carried out as described above. Each pooled sample of 30 midguts corresponded to adult mosquitoes collected from one house. Greater than 50% of the extracted midguts were derived from blood-fed anophelines.

DNA extraction, Bacterial PCR and DGGE conditions
DNA from all samples was extracted using the UltraClean soil DNA isolation kit (MO BIO Laboratories Inc., Carlsbad, CA, USA) following the manufacturer's instructions. Each water sample consisted of pellet obtained from a 250 ml volume centrifuged for 20 min at 5100 g. Each midgut extract consisted of 30 pooled midguts suspended in 100 μl

76

phosphate-buffered saline. DNA extractions were performed at the International Centre for Insect Physiology and Ecology laboratory in Nairobi, Kenya. DNA samples were transported to the USA for subsequent analyses. PCR reactions were performed in a Mastercycler ep thermocycler (Eppendorf North America, Inc., Westbury, NY, USA). PCR targeting the domain *Bacteria* and DGGE conditions were based on optimized conditions reported previously using primers 954f(GC) and 1369r (Yu and Morrison, 2004), except for the following modifications: PCR reactions (50 μl) contained 1 × PCR buffer, 200 μM deoxyribonucleotide triphosphates, 500 nM of each primer, 1.5 mM MgCl$_2$, 1.25 U of Ex *Taq* DNA polymerase (Hotstart version; Takara Mirus Bio, Madison, WI, USA) and 10 ng template DNA. The thermocycling program included an initial denaturation step at 94 °C for 2 min followed by 10 cycles of touchdown PCR (denaturation at 94 °C for 30 s, annealing at 61 °C for 30 s with 0.5 °C decrement per cycle and extension at 72 °C for 1 min). The touchdown cycles were followed by 25 cycles of regular PCR using 56 °C as the annealing temperature. Amplicons of the expected size (456 bp, excluding the GC-clamp) were gel purified (MinElute gel extraction kit; Qiagen Inc., Valencia, CA, USA) following the manufacturer's instructions and resolved by DGGE (Dcode Universal Mutation detection system; Bio-Rad Laboratories, Hercules, CA, USA) using previously described conditions (Yu and Morrison, 2004). All DGGE gels included two lanes of standards (DGGE Marker II; Wako Chemicals USA Inc., Richmond, VA, USA). The DGGE bands were stained using SYBR Gold (Molecular Probes Inc., Eugene, OR, USA) and the images were captured using a Gel Logic 100 imaging system (Kodak Molecular Imaging Systems, New Haven, CT, USA). All PCR-DGGE steps were performed twice to determine the reproducibility of the results.

DGGE bands were detected and analyzed using Quantity One software (version 4.6.2; Bio-Rad Laboratories, Hercules, CA, USA). We considered only bands with intensities exceeding 2.0% of the most intense band in each lane for subsequent analyses. Lanes were compared using Dice's similarity coefficient and cluster analysis was performed using unweighted pair group method with arithmetic means. Taxon richness was determined from the number of bands in each lane, while the Shannon-Weaver diversity index, *H*, was calculated from $H = -\sum (p_i)(\log p_i)$, where p is the proportion of an individual band's intensity (peak height) relative to the cumulative peak height (Atlas and Bartha, 1998). The identities of the most intense DGGE bands were determined by excising the bands, reamplifying the PCR products using primers 954f and 1369r and sequencing the purified PCR products through the University of Michigan's DNA Sequencing Core facilities.

T. anophelis clone libraries, sequencing and phylogenetic analysis
Partial sequences of the *T. anophelis* 16S rRNA gene were amplified from the environmental DNA extracts described above using primers designated as Tha226f (5′-TTCGGGCCTCACGCACTAG-3′) and Tha1293r (5′-ATCCGGACTTTGAGG-3′), based on *Escherichia coli* numbering. The expected product size including the primers is 1101 bp. The primer sequences perfectly matched the 16S rRNA gene of *T. anophelis* CCUG 49520^T and DGGE band-derived *Thorsellia* sequences, while each primer contained at least three mismatches to closely related Gammaproteobacteria sequences (see below). PCR was carried out as described above, except the extension times were increased to 1 min and 15 s.

Four clone libraries, each consisting of 48 clones, derived from *T. anophelis* 16S rRNA gene amplified by Tha226f and Tha1293r were constructed using a cloning kit (TOPO TA cloning kit for sequencing, Invitrogen, Carlsbad, CA, USA) and, respectively, were derived from pooled amplicons generated from SML water, bulk water and adult midguts from houses 1 and 2. The clones were initially screened by colony PCR (Gussow and Clackson, 1989), followed by digestion of cloned sequences using the restriction enzyme *Mvn*I (Roche Diagnostics Corp., Indianapolis, IN, USA). Digests were loaded on 2% agarose gels to evaluate restriction patterns. Inserts from 41 clones redundantly representing seven unique restriction pattern types were sequenced bi-directionally through the University of Michigan's DNA Core Sequencing facilities. The 33 unique sequences generated from 41 clones were deposited in GenBank under accession numbers EF434748–EF434780.

DNA sequences were assembled and aligned using BioEdit software version 7.0 (Hall, 1999) freely available at http://www.mbio.ncsu.edu/BioEdit/bioedit.html. The sequences selected for phylogenetic analysis consisted of the 33 unique *T. anophelis* sequences generated in this study, *T. anophelis* CCUG 49520^T and 42 other insect-associated and free-living representatives of the Gammaproteobacteria. Compressed branches in Figure 3 represent the following genera: insect-associated *Arsenophonus*: a symbiont of *Wahlgreniella nervata*, a symbiont of *Bemisia tabaci*, a symbiont of *Austaliococcus greville*, a symbiont of *Diaphorina citri* and a symbiont of *Glycaspis brimblecombei*; free-living *Proteus*: *P. vulgaris* and *P. hauseri*; representatives of *Vibrio*: *V. splendidus* and *V. cholerae*; and *Pseudomonas aeruginosa*. The compressed branches in Figure 3 corresponding to the Pasteurellaceae consist of *Haemophilus* sp oral clone BJ021, *Actinobacillus indolicus* and *Pasteurella multocida*. The accession numbers of these taxa as well as other insect-associated and free-living representatives of the Enterobacteriaceae are detailed in Supplementary Information.

Phylogenetic affiliations based on the neighbor-joining algorithm were evaluated using three different methods for distance matrix calculations: Kimura-2-parameter, Tajima-Nei and Tamura-Nei methods. All methods assumed different substitution rates among sites requiring an estimate of the γ parameter, which was calculated using the software package Diverge 2.0 (Gu, 1999) freely available at http://xgu.zool.iastate.edu. Distance matrix calculations and phylogenetic tree construction were performed using the software package MEGA 3.1 (Kumar *et al.*, 2001) freely available at http://www.megasoftware.net. The robustness of each resulting tree was evaluated by bootstrap analysis based on 1000 replications. In addition to distance-based neighbor-joining trees, we also compared trees based on maximum parsimony algorithm in MEGA 3.1 and the maximum likelihood algorithm in Phylip 3.66 (Felsenstein, 2006). We have included the bootstrapped Tajima-Nei and maximum parsimony trees in Supplementary Figures.

T. anophelis *growth measurements*

T. anophelis CCUG 49520^{T} cultures were maintained on agar slants containing trypticase soy agar supplemented with 5% sheep's blood (Blood agar, Remel, Lenexa, KS, USA). *T. anophelis* was grown overnight with shaking (all incubations were done at 30 °C) in trypticase soy broth (TSB; Difco Laboratories Livonia MI, USA) supplemented with 5% sheep's blood (Hemostat Laboratories, Dixon, CA, USA). This culture was used to inoculate duplicate flasks containing 150 ml of TSB and TSB supplemented with 5% sheep's blood. The growth of the shaken cultures was monitored by counting colony forming units from duplicate serial dilutions plated onto 5% sheep's blood agar plates (Remel, Lenexa, KS, USA). A similar protocol was followed to monitor the growth of *T. anophelis* in TSB maintained in either pH 7.4 or 9.1. The media were buffered at the appropriate pH using a final concentration of 0.1 M sodium carbonate-bicarbonate buffer following the proportions recommended by Delory (1945).

Results and discussion

Since the bulk of the current global burden of malaria is borne by countries in Africa and since flooded rice paddies serve as larval habitats and are a known risk factor for malaria transmission (Koudou *et al.*, 2005), we selected a rice-growing region in central Kenya (Mwea) as our study site. Rice paddy water and *An. gambiae s.l.* larvae were collected from experimental rice plots. Since anopheline larvae are known to feed mainly on the water SML (Merritt *et al.*, 1992), we collected both bulk water (from four plots) and water SML samples (from two of these four plots). *An. gambiae*

s.l. female larvae (all identifications were accomplished morphologically) were collected from the same plots as well as from adjacent plots. The experimental rice plots selected for sampling were adjacent to a village, from where we selected two residences within 100 m of the rice paddies to collect adult female *An. gambiae s.l.* mosquitoes.

DNA extracted from water (bulk and SML) and midgut (larvae and adults) samples served as sources of templates for PCR targeting the 16S rRNA gene of the domain *Bacteria*. The results of PCR followed by DGGE on the different water and mosquito midgut samples are shown in Figure 1. The most similar bacterial community profiles within sample types (defined as water, larval midguts and adult midguts) were observed in the rice paddy water samples; there were no clear distinctions between bulk and SML water community profiles. Our bulk water samples consisted of the entire water column (including the SML), which may partially explain the similarities between bulk water and the SML. On the other hand, the least similar profiles were observed between two pooled samples of larval midguts, corresponding to larvae collected from two adjacent rows of plots. Cluster analysis grouped one sample of larval midguts (designated LM-2 in Figure 1) to rice paddy water samples, while the other sample (LM-1) was more similar to the adult midgut community profiles. These results are consistent with the bacterial communities in larval midguts as an intermediate stage between the aquatic habitat and the adult stage. Bacterial diversity as measured by the Shannon-Weaver diversity index, *H*, in the SML samples ranged from 2.48 to 2.72, which was slightly higher than those observed for bulk water (1.32–2.42). Bacterial diversity indices in all midgut samples were within the range of *H* values observed for water (larvae, $H = 2.26–2.63$; adults, $H = 2.16–2.52$).

The most significant finding of PCR-DGGE was the consistent detection of 16S rRNA of populations that appears predominant across two sets of adult midguts collected from different houses (the most intense bands in lanes AM-1 and AM-2, Figure 1). Based on their relative band intensities, these populations constituted around 40% of the total bacterial community. Alignments of sequences obtained from gel-excised DGGE bands from lanes AM-1 and AM-2 to sequences available in GenBank (http://www.ncbi.nlm.nih.gov/BLAST/) and the Ribosomal Database Project (RDP-II) (http://rdp.cme.msu.edu/) revealed the closest match as a recently described bacterium with proposed genus and species names, *T. anophelis* CCUG 49520^{T} (Kampfer *et al.*, 2006). Since no similar *T. anophelis* DGGE bands were evident in the water and larval midgut samples, the results of PCR-DGGE would appear to support the prevailing view that adult mosquito midguts are populated mostly by bacteria obtained opportunistically (Riehle and Jacobs-Lorena, 2005). We sought to confirm this by assaying
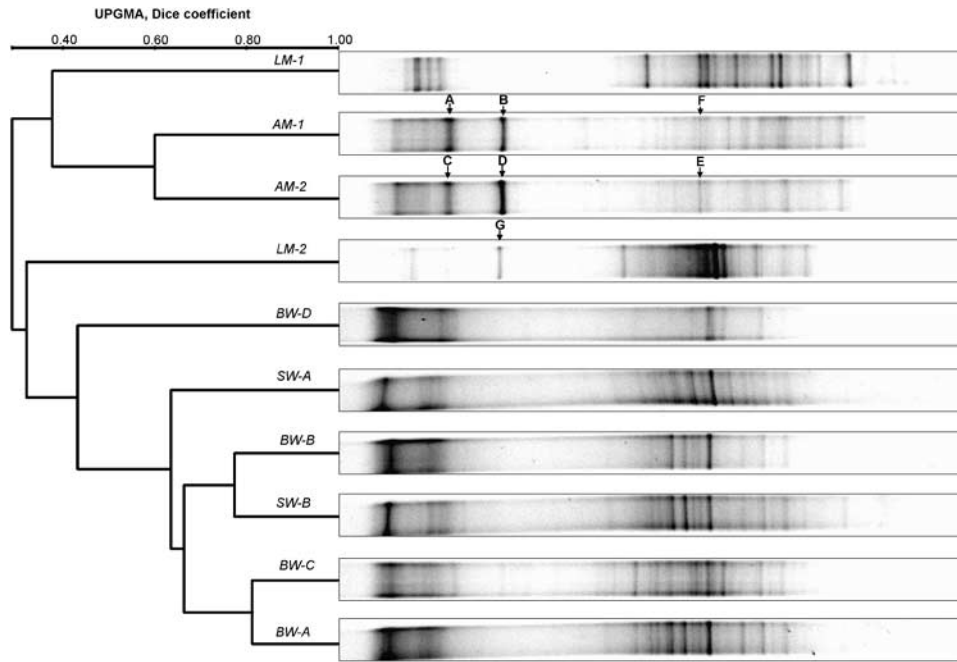
Figure 1 Cluster analysis of 16S rRNA gene-targeted DGGE profiles of bacterial communities in rice paddies and mosquito midguts. LM = larval midgut; AM = adult midgut; BW = bulk water; SW = surface microlayer water. The bands labeled A–E represent DNA that were sequenced and identified as *T. anophelis*. Band F failed to generate a reliable sequence, while band G in lane LM-2 had similar mobility to bands B and D but was identified as a species of *Chitinophaga*. The sources of the water, larvae and adult mosquitoes are the same as those described in Figure 2.

specifically for *T. anophelis* 16S rRNA genes in the various environments using *T. anophelis*-specific PCR primers (Tha226f and Tha1293r), which were designed in this study (see Materials and methods). Because the reverse primer Tha1293r hybridizes to an internal region within the fragment amplified by the DGGE primers (954f and 1369r), we obtained only partial alignments (82%) between the DGGE sequences and the sequences obtained from *Thorsellia* primers.

The results of PCR using the *T. anophelis*-specific PCR primers are shown in Figure 2. *T. anophelis* DNA was consistently detected in adult midguts and water SML, suggesting that these midgut populations are derived through ingestion by larvae from their aquatic environments, and are not acquired post-metamorphosis. Their absence in larval midguts suggests that their levels here are below the detection limit of the PCR, due to the fast gut passage rate of material through mosquito larvae (Nilsson, 1986; Merritt *et al.*, 1992), as well as the slow growth rate of *T. anophelis* (Kampfer *et al.*, 2006), which we confirmed in this study (see below). It should be noted that the 30 midguts processed per mosquito sample correspond to less than 100 μl volume of sample compared to the 250 ml of water samples that were processed.

Four clone libraries each consisting of 48 cloned *T. anophelis* 16S rRNA gene amplicons derived from water (bulk and SML) and adult midguts (houses 1 and 2) were screened based on banding patterns generated by digestion with the restriction enzyme
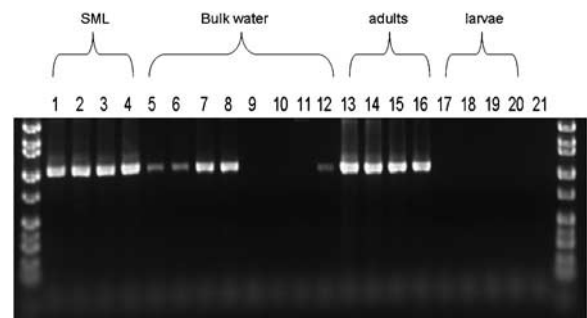

Figure 2 16S rRNA gene PCR products generated using *T. anophelis*-specific primers. Lanes 1,2, plot A SML, lanes 3,4, plot B SML; lanes 5,6, plot A bulk water; lanes 7,8, plot B bulk water; lanes 9,10, plot C bulk water; lanes 11,12, plot D bulk water; lanes 13,14, adult midguts from house 1; lanes 15,16, adult midguts from house 2; lanes 17,18, larval midguts collected from row 1 containing eight plots that includes plots A and C; lanes 19–20, larval midguts from adjacent row 2 of plots that includes plots B and D; lane 21, no template.

*Mvn*I. Seven unique restriction types were identified and 41 clones were selected to redundantly represent each restriction type. Subsequent sequencing found that 33 of the 41 clones (80%) represented unique sequences. All unique sequences, including *T. anophelis* CCUG 49520[T], exhibited pairwise sequence similarities ranging from 97.1% to 99.9%, based on an alignment of 1053 nucleotides. At this point, it is not possible to conclude whether all the obtained sequences correspond to a single species. However, this remarkable diversity of sequences suggests at least the existence of numer-
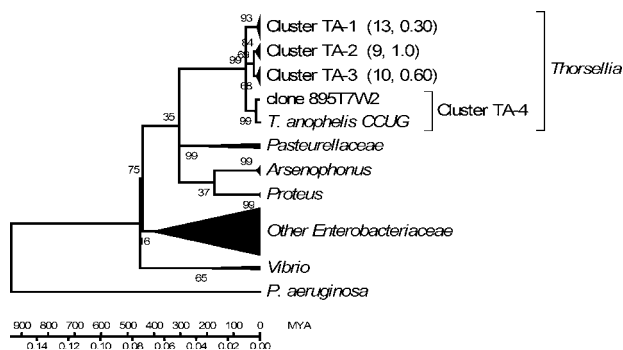
Figure 3 Phylogenetic analysis representing 76 sequences of Gammaproteobacteria. Individual taxa belonging to *Thorsellia*, *Arsenophonus*, *Proteus*, *Vibrio*, *Pseudomonas* and Pasteurellaceae are listed in the Methods while all accession numbers and descriptions of other taxa are provided in the Supplementary Information. The numbers in parentheses after three of the four *T. anophelis* clusters represent the number of unique sequences in each cluster and the proportion of water sequences in the cluster. DNA distance and treeing methods used were the Tamura–Nei model and neighbor-joining method, respectively. The numbers are bootstrap values from 1000 replicates. Molecular clock scale was calibrated by the divergence between *E. coli* and *S. typhimurium*. Two trees offering expanded views of the Gammaproteobacteria lineages are presented in the Supplementary Information.

ous strains of *T. anophelis* that can be resolved into at least four major lineages designated TA-1 to TA-4 in Figure 3. Sequences derived from rice paddy water comprised the majority of cluster TA-3 and the entirety of cluster TA-2, while the bulk of cluster TA-1 comprised midgut sequences. Excluding the single *Thorsellia* sequence available in GenBank, the highest BLAST alignment scores for all members of cluster TA-1 correspond to members of the genus *Arsenophonus*. Members of clusters TA-2 and TA-3 were also closely affiliated (by highest alignment scores) to members of the genus *Pectobacterium* or other unidentified Gammaproteobacteria. Cluster TA-4 consisted of a single water clone that was most closely affiliated with *T. anophelis* CCUG 49520^{T}, which was originally isolated from *Anopheles* midguts in Lwanda, Kenya, approximately 550 km northwest of our study site (Lindh *et al.*, 2005). Therefore, *Thorsellia* appears to be widely distributed in Kenya, although the geographic separation may influence the composition and distribution of strains. In Mwea, the more diverse collection of *T. anophelis* sequences occurred in rice paddy water: 24 clones representing all clusters TA-1, TA-2, TA-3 and TA-4 were detected at 21%, 46%, 29% and 4%, respectively. The midgut samples were less diverse, consisting of 17 clones represented by only two clusters, TA-1 (71%) and TA-3. The sequences derived from DGGE gels and clone libraries overlap only partially (82%) and should therefore be compared cautiously. However, based on these partial alignments, bands designated as B and D in Figure 1 are most closely similar to cluster TA-1 (zero to one base mismatch). The presence of

ambiguous bases in the sequences derived from the other DGGE bands precludes reliable affiliation with the other TA clusters.

Since mosquito larvae likely do not discriminate between these different taxa during feeding, these results are consistent with the following scenario: only certain *T. anophelis* strains (mostly cluster TA-1 and to a lesser extent TA-3) are able to survive and proliferate during and after metamorphosis, respectively. Based on its high frequency in the aquatic environment, cluster TA-2 represents a group that is most likely ingested, but does not survive during subsequent growth and development, because of digestive, excretory or gut remodeling processes.

T. anophelis sequences formed a distinct and robust clade within the class Gammaproteobacteria (Figure 3); it is currently classified under the family Enterobacteriaceae in both NCBI and RDP classification schemes. Although cluster TA-1 aligned consistently with members of *Arsenophonus* (above), phylogenetic analysis provided only weak bootstrap support for affiliating the entire *Thorsellia* clade with this genus. Surprisingly, our analyses (distance based, maximum parsimony and maximum likelihood) consistently grouped members of the Pasteurellaceae with *Thorsellia*, *Arsenophonus* and *Proteus*. While 16S rRNA gene sequences alone should not be relied upon to fully resolve the phylogenetic affiliations within the Gammaproteobacteria, it is interesting to note that some members of the Pasteurellaceae are α-hemolytic when grown on blood agar plates (Donachie *et al.*, 1995), a trait that we report here to be exhibited by *T. anophelis* (see below). Additionally, we found *T. anophelis* CCUG 49520^{T} to be oxidase positive (Oxidase test discs, Sigma-Aldrich, St Louis, MO, USA), a characteristic shared by members of the Vibrionaceae and Pasteurellaceae, but not Enterobacteriaceae.

The genus *Arsenophonus* is comprised mostly of members that are endosymbiotic to a variety of arthropod hosts, including insects and ticks (Gherna *et al.*, 1991; Grindle *et al.*, 2003). Artificial culture has been achieved only on insect cell lines (Hypsa and Dale, 1997; Dale *et al.*, 2006). Different species of *Arsenophonus* from different hosts share >99% 16S rRNA gene sequence similarity, suggesting recent host acquisition (Dale *et al.*, 2006). The divergences between members of the *T. anophelis* clusters are substantially greater, suggesting a more ancient time of divergence between sublineages. By calibrating the phylogenetic tree in Figure 3 to the divergence between *E. coli* and *Salmonella typhimurium* at 130 ± 10 million years ago (Ochman and Wilson, 1987), we estimate the *T. anophelis* divergence to have occurred around 50–70 million years ago (the lower limit is based on calibrated maximum likelihood and neighbor joining trees, while the upper limit is based on minimum evolution trees). This matches the estimated divergence between the lineages of *Anopheles* at 70–85 million years ago

(Krzywinski *et al.*, 2006). This concordance raises the possibility that mosquitoes have played a role in the evolutionary diversification of *Thorsellia* and vice versa. Based on the scenario we describe above, we speculate that a lineage such as cluster TA-1, which predominates in adult *An. gambiae s.l.* midguts, is more widely disseminated geographically by being transported from and deposited back to the aquatic environment as a direct or indirect consequence of oviposition on water. Whether cluster TA-1 represents an *Anopheles* symbiont will require further study, but its affiliation with *Arsenophonus* suggests this possibility.

The possibility of a close association between *T. anophelis* and *Anopheles* mosquitoes is further strengthened by two characteristics of strain CCUG 49520[T], that is, its alkali tolerance and its ability to utilize blood to enhance growth (Figure 4). The
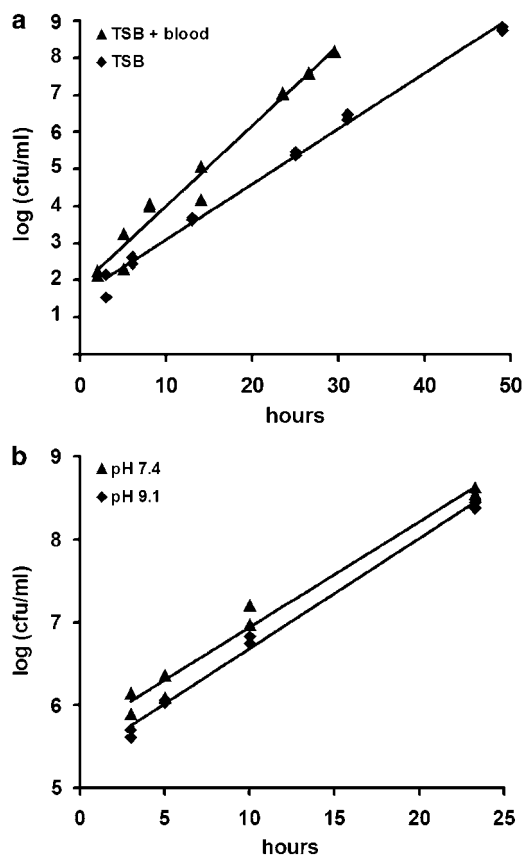

Figure 4 Exponential phase of growth curves of *T. anophelis* grown in trypticase soy broth (TSB) with and without blood (**a**) and in TSB maintained at two different pH values (**b**). Each point represents one of the two independent replicates per sampling time (some points not visible because of overlapping symbols) and the curves were fitted by nonlinear regression using an exponential growth equation. Comparisons of curves from log-transformed data showed that the slopes in (**a**) are very significantly different ($P < 0.0001$) while those in (**b**) are equal ($P = 0.5164$). The difference in elevation (significant, $P = 0.0017$) of curves in (**b**) was due to a longer lag phase in the pH 9.1 treatment. Curves and statistical tests were performed using Prism 4 for Windows (GraphPad Software Inc., San Diego, CA, USA).

alkaline conditions in the mosquito midgut (maximum pH of 10.5 in larvae; pH of 8.0–9.5 in adults) have long been recognized (Dadd, 1975; Corena *et al.*, 2005). Therefore, a mosquito midgut association would explain the evolutionary significance of alkali tolerance in *T. anophelis*, a trait that would be useless in rice paddy environments. In contrast to most Enterobacteriaceae, *T. anophelis* grows slowly in various media and carbon sources (Kampfer *et al.*, 2006). In trypticase soy agar supplemented with 5% sheep's blood, we observed α-hemolysis and growth enhancement by a factor of 1.4 (generation time of 1.4 h with blood supplementation vs 2.0 h without blood). This enhanced growth rate is still relatively low compared to most Enterobacteriaceae grown in nutrient rich media. Therefore, although the physiological adaptations we describe here help explain the association of *T. anophelis* with adult mosquitoes that have acquired a blood meal, other factors may be involved in their apparent predominance in this environment. Research on *Arsenophonus* also suggests that the basis of symbiosis with its host is probably not based on nutrition but on still unknown, but critical, function(s) (Dale *et al.*, 2006).

Another group has recently described an association between *An. stephensi* and an Alphaproteobacterium of the genus *Asaia* (Favia *et al.*, 2007). High levels (41% of 16S rRNA gene copies using quantitative PCR; 90% of 16S rRNA gene clones) of this bacterium associated primarily with insectary-reared *An. stephensi*, although it was also detected at lower levels (5% of 16S rRNA gene clone libraries) in *An. gambiae* captured from villages in Burkina Faso,West Africa. None of the major DGGE bands sequenced in our study were affiliated to *Asaia*. On the other hand, *Thorsellia* has been detected by at least two independent studies in Kenya using culture-(Lindh *et al.*, 2005; Kampfer *et al.*, 2006) and PCR-based techniques (this study). We recognize the need to further evaluate the association between *T. anophelis* and *An. gambiae* (or other possible mosquito hosts). In this respect, the primers and environmental sequences generated in this study will be helpful in developing a quantitative molecular assay (for example, real-time PCR) that should supplement further strain/ species collection and characterization. The identification of a *T. anophelis* lineage that is associated with the adult midgut, but can also be detected in the aquatic environments where larvae live provides a promising starting point to determine the feasibility of using this bacterium for paratransgenic control of *Plasmodium*. Ingestion of genetically transformed *T. anophelis* by mosquito larvae in the rice paddy environment, where larvae are concentrated, relatively immobile and accessible to human intervention, should provide a more efficient paratransgenic system compared to opportunistic acquisition of bacteria by adult mosquitoes.

Acknowledgements

References

Agogue H, Casamayor EO, Joux F, Obernosterer I, Dupuy C, Lantoine F *et al.* (2004). Comparison of samplers for the biological characterization of the sea surface microlayer. *Limnol Oceanogr Methods* **2**: 213–225.

Atlas RM, Bartha R. (1998). *Microbial Ecology fundamentals and applications*, 4th edn. Benjamin/Cummings Science Publishing: Menlo Park, CA.

Beard CB, Cordon-Rosales C, Durvasula RV. (2002). Bacterial symbionts of the triatominae and their potential use in control of Chagas disease transmission. *Annu Rev Entomol* **47**: 123–141.

Conte JE. (1997). A novel approach to preventing insect-borne diseases. *N Engl J Med* **337**: 785–786.

Corena MD, VanEkeris L, Salazar MI, Bowers D, Fiedler MM, Silverman D *et al.* (2005). Carbonic anhydrase in the adult mosquito midgut. *J Exp Biol* **208**: 3263–3273.

Dadd RH. (1975). Alkalinity within the midgut of mosquito larvae with alkaline-active digestive enzymes. *J Insect Physiol* **21**: 1847–1853.

Dale C, Beeton M, Harbison C, Jones T, Pontes M. (2006). Isolation, pure culture, and characterization of 'Candidatus Arsenophonus arthropodicus,' an intracellular secondary endosymbiont from the hippoboscid louse fly *Pseudolynchia canariensis*. *Appl Environ Microb* **72**: 2997–3004.

Delory GE. (1945). A sodium carbonate-bicarbonate buffer for alkaline phosphatases. *Biochem J* **39**: 245.

Donachie W, Lainson FA, Hodgson JC. (1995). *Haemophilus, Actinobacillus and Pasteurella*. Plenum Press: New York.

Favia G, Ricci I, Damiani C, Raddadi N, Crotti E, Marzorati M *et al.* (2007). Bacteria of the genus *Asaia* stably associate with *Anopheles stephensi*, an Asian malarial mosquito vector. *Proc Nat Acad Sci USA* **104**: 9047–9051.

Felsenstein J. (2006). *PHYLIP (Phylogeny Inference Package) version 3.66*. (distributed by the author.) Department of Genome Sciences, University of Washington, Seattle, WA, USA.

Gherna RL, Werren JH, Weisburg W, Cote R, Woese CR, Mandelco L *et al.* (1991). *Arsenophonus-Nasoniae* Gen-Nov, Sp-Nov, the Causative Agent of the Son-Killer Trait in the Parasitic Wasp *Nasonia-Vitripennis*. *Int Syst Bacteriol* **41**: 563–565.

Gillies MT, Coetzee M. (1987). *A supplement to the Anophelinae of Africa south of the Sahara. Publication no. 55*. South African Institute of Medical Research: Johannesburg, South Africa, 143pp.

Gonzalez-Ceron L, Santillan F, Rodriguez MH, Mendez D, Hernandez-Avila JE. (2003). Bacteria in midguts of field-collected *Anopheles albimanus* block *Plasmodium vivax* sporogonic development. *J Med Entomol* **40**: 371–374.

Grindle N, Tyner JJ, Clay K, Fuqua C. (2003). Identification of Arsenophonus-type bacteria from the dog tick *Dermacentor variabilis*. *J Invertebr Pathol* **83**: 264–266.

Gu X. (1999). Statistical methods for testing functional divergence after gene duplication. *Mol Biol Evol* **16**: 1664–1674.

Gussow D, Clackson T. (1989). Direct clone characterization from plaques and colonies by the Polymerase chain-reaction. *Nucleic Acids Res* **17**: 4000.

Hall TA. (1999). BioEdit, a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* **41**: 95–98.

Hypsa V, Dale C. (1997). *In vitro* culture and phylogenetic analysis of 'Candidatus Arsenophonus triatominarum,' an intracellular bacterium from the triatomine bug, *Triatoma infestans*. *Int J Syst Bacteriol* **47**: 1140–1144.

Kampfer P, Lindh JM, Terenius O, Haghdoost S, Falsen E, Busse H-J *et al.* (2006). *Thorsellia anophelis* gen. nov., sp. nov., a new member of the Gammaproteobacteria. *Int J Syst Evol Microbiol* **56**: 335–338.

Koudou BG, Tano Y, Doumbia M, Nsanzabana C, Cisse G, Girardin O *et al.* (2005). Malaria transmission dynamics in central Cote d'Ivoire: the influence of changing patterns of irrigated rice agriculture. *Med Vet Entomol* **19**: 27–37.

Krzywinski J, Grushko OG, Besansky NJ. (2006). Analysis of the complete mitochondrial DNA from *Anopheles funestus*: an improved dipteran mitochondrial genome annotation and a temporal dimension of mosquito evolution. *Mol Phylogenet Evol* **39**: 417–423.

Kumar S, Tamura K, Jakobsen IB, Nei M. (2001). MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17**: 1244–1245.

Lindh JM, Terenius O, Eriksson-Gonzales K, Knols BGJ, Faye I. (2006). Re-introducing bacteria in mosquitoes—a method for determination of mosquito feeding preferences based on coloured sugar solutions. *Acta Tropica* **99**: 173–183.

Lindh JM, Terenius O, Faye I. (2005). 16S rRNA gene-based identification of midgut bacteria from field-caught *Anopheles gambiae sensu lato* and *A. funestus* mosquitoes reveals new species related to known insect symbionts. *Appl Environ Microb* **71**: 7217–7223.

Lowenberger CA, Kamal S, Chiles J, Paskewitz S, Bulet P, Hoffmann JA *et al.* (1999). Mosquito–plasmodium interactions in response to immune activation of the vector. *Exp Parasitol* **91**: 59–69.

Meister S, Koutsos AC, Christophides GK. (2004). The *Plasmodium* parasite—a 'new' challenge for insect innate immunity. *Int J Parasitol* **34**: 1473–1482.

Merritt RW, Dadd RH, Walker ED. (1992). Feeding behavior, natural food, and nutritional relationships of larval mosquitoes. *Annu Rev Entomol* **37**: 349–376.

Moll RM, Romoser WS, Modrakowski MC, Moncayo AC, Lerdthusnee K. (2001). Meconial peritrophic membranes and the fate of midgut bacteria during mosquito (Diptera: Culicidae) metamorphosis. *J Med Entomol* **38**: 29–32.

Mutero CM, Blank H, Konradsen F, van der Hoek W. (2000). Water management for controlling the breeding of *Anopheles* mosquitoes in rice irrigation schemes in Kenya. *Acta Tropica* **76**: 253–263.

Muturi EJ, Shililu J, Jacob B, Gu WD, Githure J, Novak R. (2006). Mosquito species diversity and abundance in relation to land use in a riceland agroecosystem in Mwea, Kenya. *J Vector Ecol* **31**: 129–137.

Muyzer G, de Waal EC, Uitterlinden AG. (1993). Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl Environ Microbiol* **59**: 695–700.

Nilsson C. (1986). Energetics of the suspension feeding mosquito larva culex torrentium. *Holarctic Ecol* **9**: 267–271.

Ochman H, Wilson AC. (1987). Evolutionary history of enteric bacteria. In: Neidhardt F, Ingraham JL, Low KB, Magsanik B, Schaechter M, and Umbarger HE (eds). *Escherichia coli and Salmonella Typhimurium: Cellular and Molecular Biology*. American Society for Microbiology: Washington, DC, pp 1615–1649.

Pumpuni CB, Beier MS, Nataro JP, Guers LD, Davis JR. (1993). *Plasmodium falciparum*: inhibition of sporogonic development in *Anopheles stephensi* by Gram-negative bacteria. *Exp Parasitol* **77**: 195–199.

Pumpuni CB, DeMaio J, Kent M, Davis JR, Beier JC. (1996). Bacterial population dynamics in three anopheline species: the impact on *Plasmodium* sporogonic development. *Am J Trop Med Hyg* **54**: 214–218.

Riehle MA, Jacobs-Lorena M. (2005). Using bacteria to express and display anti-parasite molecules in mosquitoes: current and future strategies. *Insect Biochem Mol Biol* **35**: 699–707.

Riehle MA, Moreira CK, Lampe D, Lauzon C, Jacobs-Lorena M. (2007). Using bacteria to express and display anti-*Plasmodium* molecules in the mosquito midgut. *Int J Parasitol* **37**: 595–603.

Service MW. (1993). *Field sampling methods*, 2nd edn. Elsevier: Essex, UK, 954pp.

Sinden RE. (2002). Molecular interactions between *Plasmodium* and its insect vectors. *Cell Microbiol* **4**: 713–724.

Straif SC, Mbogo CNM, Toure AM, Walker ED, Kaufman M, Toure YT *et al.* (1998). Midgut bacteria in *Anopheles gambiae* and An-funestus (Diptera: Culicidae) from Kenya and Mali. *J Med Entomol* **35**: 222–226.

Vaughan JA, Noden BH, Beier JC. (1994). Sporogonic development of cultured *Plasmodium falciparum* in six species of laboratory-reared *Anopheles* mosquitoes. *Am J Trop Med Hyg* **51**: 233–243.

Yu Z, Morrison M. (2004). Comparisons of different hypervariable regions of *rrs* genes for use in fingerprinting of microbial communities by PCR-denaturing gradient gel electrophoresis. *Appl Environ Microbiol* **70**: 4800–4806.

Supplementary Information accompanies the paper on The ISME Journal website (http://www.nature.com/ismej)