The mosaic genome structure of the *Wolbachia w*Ri strain infecting *Drosophila simulans*

Lisa Klasson[a,1], Joakim Westberg[a,1], Panagiotis Sapountzis[b], Kristina Näslund[a], Ylva Lutnaes[a], Alistair C. Darby[a,2], Zoe Veneti[c], Lanming Chen[d,3], Henk R. Braig[e], Roger Garrett[d], Kostas Bourtzis[b,c], and Siv G. E. Andersson[a,4]

[a]Department of Molecular Evolution, Evolutionary Biology Centre, Uppsala University, SE-752 36 Uppsala, Sweden; [b]Department of Environmental and Natural Resources Management, University of Ioannina, 30100 Agrinio, Greece; [c]Institute of Molecular Biology and Biotechnology, FORTH, 71110 Heraklion, Crete, Greece; [d]Centre for Comparative Genomics, Department of Biology, University of Copenhagen, DK-2200 Copenhagen N, Denmark; and [e]School of Biological Sciences, Bangor University, Wales, LL57 2UW, United Kingdom

The obligate intracellular bacterium *Wolbachia pipientis* infects around 20% of all insect species. It is maternally inherited and induces reproductive alterations of insect populations by male killing, feminization, parthenogenesis, or cytoplasmic incompatibility. Here, we present the 1,445,873-bp genome of *W. pipientis* strain *w*Ri that induces very strong cytoplasmic incompatibility in its natural host *Drosophila simulans*. A comparison with the previously sequenced genome of *W. pipientis* strain *w*Mel from *Drosophila melanogaster* identified 35 breakpoints associated with mobile elements and repeated sequences that are stable in *Drosophila* lines transinfected with *w*Ri. Additionally, 450 genes with orthologs in *w*Ri and *w*Mel were sequenced from the *W. pipientis* strain *w*Uni, responsible for the induction of parthenogenesis in the parasitoid wasp *Muscidifurax uniraptor*. The comparison of these A-group *Wolbachia* strains uncovered the most highly recombining intracellular bacterial genomes known to date. This was manifested in a 500-fold variation in sequence divergences at synonymous sites, with different genes and gene segments supporting different strain relationships. The substitution-frequency profile resembled that of *Neisseria meningitidis*, which is characterized by rampant intraspecies recombination, rather than that of *Rickettsia*, where genes mostly diverge by nucleotide substitutions. The data further revealed diversification of ankyrin repeat genes by short tandem duplications and provided examples of horizontal gene transfer across A- and B-group strains that infect *D. simulans*. These results suggest that the transmission dynamics of *Wolbachia* and the opportunity for coinfections have created a freely recombining intracellular bacterial community with mosaic genomes.

horizontal transfer | recombination | ankyrin repeat gene | genome evolution | insect symbiosis

Wolbachia pipientis are intracellular α-proteobacteria of the order *Rickettsiales* that infect insects as well as isopods, spiders, scorpions, mites, and filarial nematodes (1, 2). These bacteria represent a single species, with strains classified into supergroups, of which the most abundant are supergroups A and B. A lack of concordance between host and bacterial strain phylogenies indicate frequent host shifts in addition to maternal inheritance within the individual host (1–4). In insect populations, *Wolbachia* induce reproductive manipulations to enhance their own spreading. The most frequently observed reproductive abnormality is cytoplasmic incompatibility (CI) (1, 2, 5), where uninfected females are unable to produce offspring with infected males, whereas infected females can produce offspring with both infected and uninfected males, thus creating a reproductive advantage for infected females. Other spectacular effects of *Wolbachia* infections are male embryo killing, feminization, and parthenogenesis induction (1, 2).

Three genomes of *Wolbachia* have been published to date. The A-group strain *w*Mel from *Drosophila melanogaster* (6) and the B-group strain *w*Pip from the mosquito *Culex quinquefasciatus* (7) are both reproductive parasites that cause CI, whereas the D-group strain *w*Bm is an obligate mutualist in the nematode *Brugia malayi* (8). The 1.27-Mb genome of *w*Mel and the 1.48-Mb genome of *w*Pip contain several prophages and high frequencies of repeated sequences, including many IS-elements. These genomes have a large repertoire of genes with ankyrin repeat motifs, 23 in *w*Mel and 60 in *w*Pip, several of which are associated with mobile elements (6, 7). In contrast, the 1.08-Mb genome of strain *w*Bm contains no prophage, only a few ankyrin repeat genes and a much lower fraction of repeated sequences, possibly reflecting its mutualistic adaptation to a single-host species.

Sequence comparisons of single genes from multiple strains have provided evidence for recombination (9–12) and horizontal transfer of prophages and insertion sequence (IS)-elements across *Wolbachia* strains (13–16). However, the distant relationship of the 3 sequenced *Wolbachia* genomes, manifested as an almost complete lack of gene-order conservation and synonymous substitution frequencies that are close to saturation, has precluded attempts to infer patterns and rates of recombination at the whole-genome level. Thus, with the exception of a few single-gene studies, the extent to which recombination distorts the evolutionary coherence of *Wolbachia* genomes is currently unknown.

We report here the complete genome sequence of the supergroup-A *Wolbachia* strain *w*Ri that naturally infects *Drosophila simulans* and induces almost complete CI in its host (17, 18). Additionally, we present partial genome data of *Wolbachia* strain *w*Uni from *Muscidifurax uniraptor*, likewise a supergroup-A strain that induces parthenogenesis in its host (19). A comparison of these 2 A-group *Wolbachia* strains with the previously sequenced genome of the A-group strain *w*Mel reveals the most highly recombining obligate intracellular bacterial community examined to date.
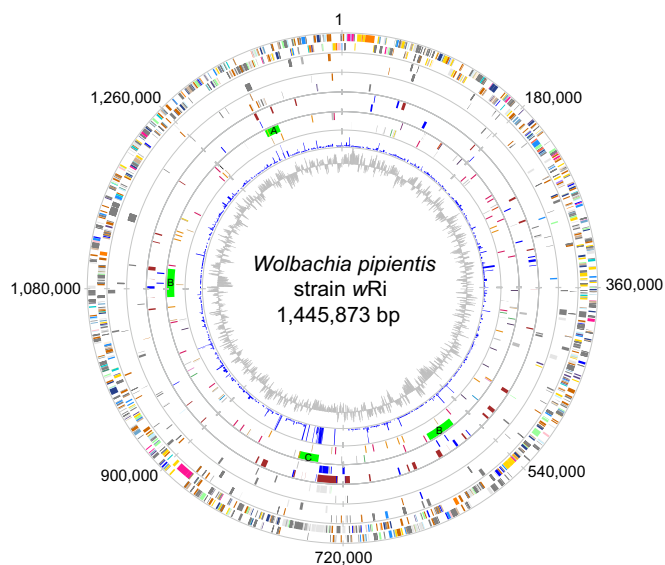
EVOLUTION

Fig. 1. Circular map of the *Wolbachia pipientis* *w*Ri genome. Each circle confined by the gray lines except for the 2 innermost circles illustrates different features on the plus (outer region) and minus (inner region) strands. Lines and boxes in the 3 outermost circles are colored according to the Clusters of Orthologous Groups (COG) categories. First (*Outer*) circle: protein-coding genes (CDSs). Second circle: pseudogenes. Third circle: unique CDSs compared to *w*Mel. Fourth circle: ANK genes in blue and gene synteny breakpoints compared to *w*Mel in red. Fifth circle: prophage regions in green (not affiliated to a strand) and IS elements color-coded as described in Table S5. Sixth circle: a diagram showing the synonymous substitution frequency (Ks) between *w*Ri and *w*Mel potential orthologs; the maximum cut-off was set to Ks = 1. Seventh circle: GC-skew of the *w*Ri genome.

Results and Discussion

General Features of the *Wolbachia* *w*Ri Genome. The complete genome sequence of *Wolbachia pipientis* *w*Ri from *Drosophila simulans* is a single circular chromosome of 1,445,873 bp, with 1,150 potential protein-coding sequences and 114 pseudogenes (Fig. 1). We identified 4 prophage segments, here called *w*Ri-WOA, *w*Ri-WOB (present in identical duplicates), and *w*Ri-WOC. Additionally, we found 35 genes coding for proteins containing one or more ankyrin repeat (ANK) domain [Tables S1–S3] as compared to 23 such genes annotated in the 1.27-Mb genome of *w*Mel (6). Of the genes solely present in either the *w*Ri or the *w*Mel genome for which a function or domain hit can be identified, the large majority encodes phage proteins, transposases and ANK proteins (Tables S4–S6).

Overall, the *w*Ri genome contains 22.1% repeated sequences, as compared to 8.9% in *w*Mel (> 200 bp, 95% sequence identity). About 10% of the *w*Ri genome is covered by IS-elements; these represent 11 different types, including 67 complete copies, 48 copies with frameshifts or internal stop codons, and 11 truncated copies. A total of 46 genes have been disrupted by IS-element insertions, of which 19 are hypothetical proteins, 10 are transposases, and 5 are ANK genes. Furthermore, more than 200 insertions of the *Wolbachia* palindromic element (20) or remnants thereof were found in the genome of *w*Ri and of these, 43 were inserted into genic sequences. The *Wolbachia* palindromic element is also present in *w*Mel and *w*Uni, but not always in the same genes or at the same locations. The most abundant 23-bp hairpin-loop structure is present in 79 copies in the *w*Ri genome, in 91 copies in *w*Mel, but only once in *w*Bm.

Genome Integrity Following Host Shifts. A comparison of the structures of the *w*Ri and *w*Mel genomes revealed 35 gene-order breakpoints (Fig. S1), 17 of which are flanked by IS-elements, 11

are located within or flanked by prophage sequences, and 6 are flanked by long repeats. The final breakpoint contains a 22-bp hairpin structure with a 4-bp loop. Among these breakpoints, 6 are close to genes encoding reverse trancriptase and 3 to genes encoding DNA recombinase. This shows that mobile genetic elements and repeated sequences are hot spots for rearrangements in *Wolbachia*. To examine the stability of these recombination hot spots, we used PCR to analyze all breakpoints from genomic DNA of *w*Ri isolated from naturally infected and transinfected symbiotic associations of *D. simulans*, *Drosophila yakuba*, *Drosophila teisseiri*, and *Drosophila santomea*, which have been kept in laboratory conditions from 8 to 15 years (Table S7). The results of this survey showed that the *w*Ri genome remains stable in structure over these sites and suggests that the *w*Ri genome does not oscillate between different genomic structures, nor do host switches trigger rearrangements at these sites. The observed short-term stability indicates that the recombination frequencies at these repeated sequences are lower than could be detected over this time period.

The A-Group *Wolbachia* Strains are Evolutionary Genome Mosaics. We estimated the nonsynonymous (K_a) and synonymous (Ks) substitution frequency per site to quantify sequence divergences across strains, although these values may not correspond to actual substitutions if genes evolve mainly by recombination. For 851 positional homologs in the *w*Ri and *w*Mel genomes identified by reciprocal BLAST searches (excluding phage genes), the median K_a and Ks values were estimated to be 6.2×10^{-3} and 3.2×10^{-2} substitutions per site, respectively, with a more than 500-fold variation in Ks-values across genes (Fig. S2*a*). A similar spectrum of divergences between *w*Ri and *w*Mel was observed for 343 core genes that are conserved across 3 *Wolbachia* strains (*w*Ri, *w*Mel, and *w*Bm), *Orientia tsutsugamushi*, and 8 *Rickettsia* species (as defined in ref. 21) (Fig. S2*b*), suggesting that these estimates are not inflated by the inadvertent inclusion of inactivated gene fragments.

To investigate the mechanisms underlying the remarkable variation in Ks-values across genes, we sequenced orthologous genes in *w*Uni, another A-group *Wolbachia* strain, using *w*Mel as the reference genome. A 3-strain comparison of 450 sequenced orthologs showed a broad variability of Ks-values (Fig. 2*A*) that did not correlate with functional categories (Mann-Whitney and Kolmogorov-Smirnov tests using Bonferroni-Holm correction, $P > 0.05$ for all COG categories) (Fig. S2*c*). One-third of the genes were most similar in *w*Mel and *w*Ri, including 52 orthologs with no synonymous sustitutions. Another one-third indicated the highest sequence similarity for *w*Mel and *w*Uni, of which 26 have no synonymous substitutions. Finally, one-fifth of the genes were most similar in *w*Ri and *w*Uni, with 22 lacking synonymous substitutions. Only 40 genes showed no substitutions at synonymous sites in any of the 3 pairs. Such a complex pattern of sequence divergences indicates extensive recombination.

To obtain a simple measure of the relative levels of recombination for comparisons across species, we calculated the spread of the relative Ks-values as the median distance to the average relative Ks-values in the ternary plots (see Fig. 2). Notably, the level of recombination thus inferred was higher in *Wolbachia* (spread = 0.42) (see Fig. 2*A*) than in *Neisseria meningitidis* (spread = 0.34) (see Fig. 2*C*), which is naturally competent for transformation and highly recombining (22). The level was twice as high as in its close relative *Rickettsia* (spread = 0.19) (see Fig. 2*B*), where the lowest Ks-values were typically associated with 1 pair of strains, as expected in nonrecombining bacteria.

Additionally, of the 450 orthologous genes found in *w*Ri, *w*Mel, and *w*Uni, intragenic recombination was detected in 129 genes by at least 2 different methods implemented in RDP3 (23). Genes for which intragenic recombination was detected by recombination detection program (RDP) were concentrated to
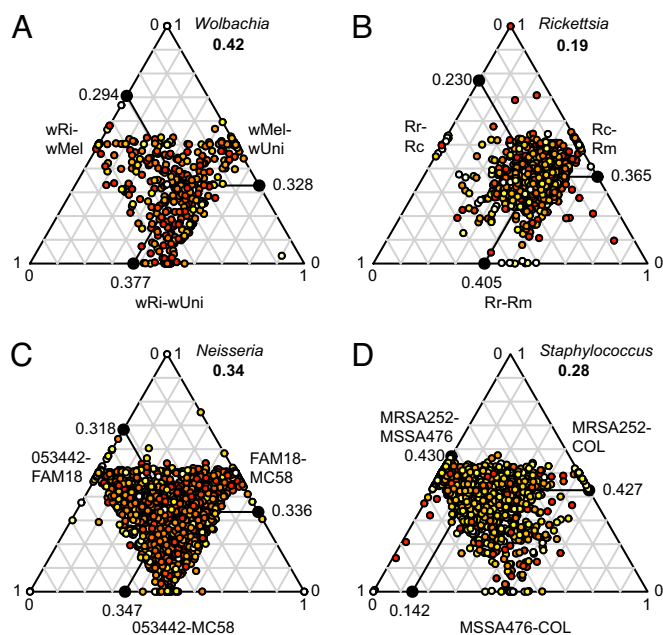
Fig. 2. Ternary plot showing substitution frequency variation across (*A*) 410 genes in A-group *Wolbachia* (40 genes with no synonymous substitutions in all 3 strains were excluded); (*B*) 818 genes in Spotted-Fever Group *Rickettsia*; (*C*) 1,529 genes in *Neisseria meningitidis*; and (*D*) 2,207 genes in *Staphylococcus aureus*. Each dot in the diagram represents 1 gene. Absolute Ks-values have been transformed to relative values between 0 and 1. The mean relative Ks-value for each pair is shown on each axis. Numbers in bold represent the median distance to the average point. The color of each dot represents the maximum absolute Ks-value among the 3 pairs, ranging from light yellow (low values) to red (high values). Median Ks-values for all pairs are reported in Table S8. Rr, *Rickettsia rickettsii*; Rc, *Rickettsia conorii*; Rm, *Rickettsia massiliae*.

Fig. 3. Recombination within *leuRS* and *virB10*. (*A*) The diagram shows the Ks-values calculated for the *leuRS* gene pair segments consisting of a 99-base window sliding over the gene alignment with 15-base steps. The boxes under the *x*-axis indicate regions where *w*Ri (*red box*) and *w*Uni (*black box*) are highly diverged from the other two strains. (*B* and *C*) Inference of sequence relationships based on segments in the *virB10* gene between (*B*) positions 72 and 892 and (*C*) positions 967 and 1526, respectively, of the gene alignment. The gray circle shows the position of *w*Ri in the tree. The strains *w*KueYo, *w*Mel, *w*Uni, *w*Atab3, and *w*Ri belong to supergroup A (*names in red*), whereas strains *w*Pip and *w*Tai belong to supergroup B (*names in blue*).

the middle of the inner triangle (spread = 0.27), whereas genes with no detected recombination were more spread out (spread = 0.56) in the ternary plot (Fig. S2*d*). Genes located near different corners of the inner triangle in the ternary plot provide the strongest evidence for inconsistency in the phylogenetic signal. As no conflict in signal was detected by RDP within these genes, it is likely that recombination has occurred over the complete sequences in many cases. Taken together, this indicates that at least three-fourths of the genes that were analyzed are affected by recombination.

The switches in sequence similarity patterns within gene alignments is here illustrated with the gene for Leucyl-tRNA synthetase (LeuRS), where 1 region was found to be identical in *w*Mel and *w*Uni, but contained many substitutions in *w*Ri, followed by a divergent segment in *w*Uni that was identical in *w*Ri and *w*Mel (Fig. 3*A*). Another example is the *virB10* gene, where the 5′-part of the *w*Ri gene clustered with the A-group strain *w*Atab3 from the braconid wasp *Asobara tabida* (Fig. 3*B*), whereas the 3′-part of the same gene clustered with the B-group strains *w*Tai, from the Taiwan cricket, *Teleogryllus taiwanemma*, and *w*Pip, from the mosquito *Culex quinquefasciatus* (Fig. 3*C*).

Rapid Diversification by Expansion-Contraction of Tandem Repeats. We classified the 35 ANK genes identified in the *w*Ri genome into 3 different groups based on the extent of sequence divergence from their homologs in *w*Mel (see Tables S1–S3). One-third, 13 genes, are highly conserved in length, domain organization, and nucleotide sequence (Ks <0.05). Another 12 genes showed variability in the number of ANK domains and gene length plus high sequence divergence (Ks 0.1 to >1.0), although
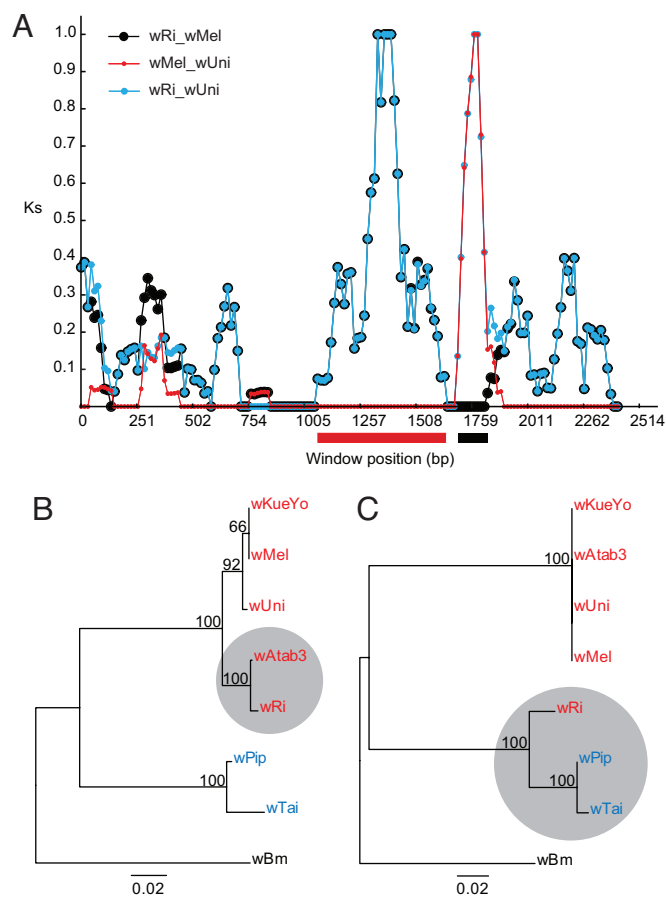
they are located in segments with otherwise conserved gene-order structures. The final 10 genes lacked homologs in *w*Mel.

To study the relationships of the ANK domains within and among genomes, we performed a Bayesian phylogenetic analysis of 319 ANK domains identified in the *w*Ri, *w*Mel, and *w*Uni genomes (Fig. 4). The ANK domains found in the highly conserved genes clustered with the corresponding domains in the other strains. In contrast, several ANK domains in the variable gene set clustered with other domains in the same gene. In these cases, variability in domain numbers and organization is generated by expansion or contraction of repeated sequences, with the repeated unit often spanning across the borders of the ANK domains (Fig. S3). The ANK protein WRi_003070 is particularly interesting; the N-terminal repeats are most similar to *w*Mel and the repeats in the middle part to *w*Uni, with recent duplications of ANK domains in both *w*Ri and *w*Uni (see Fig. S3). Additionally, this is the only *w*Ri ANK gene that is solely transcribed in female adults and ovaries, but not in male adults and testes (Tables S9 and S10).

Gene Transfer Across Supergroups. Of the combined 58 ANK genes in the *w*Ri and *w*Mel genomes, 31 are located near to prophages and many of these are unique to the individual strain and may
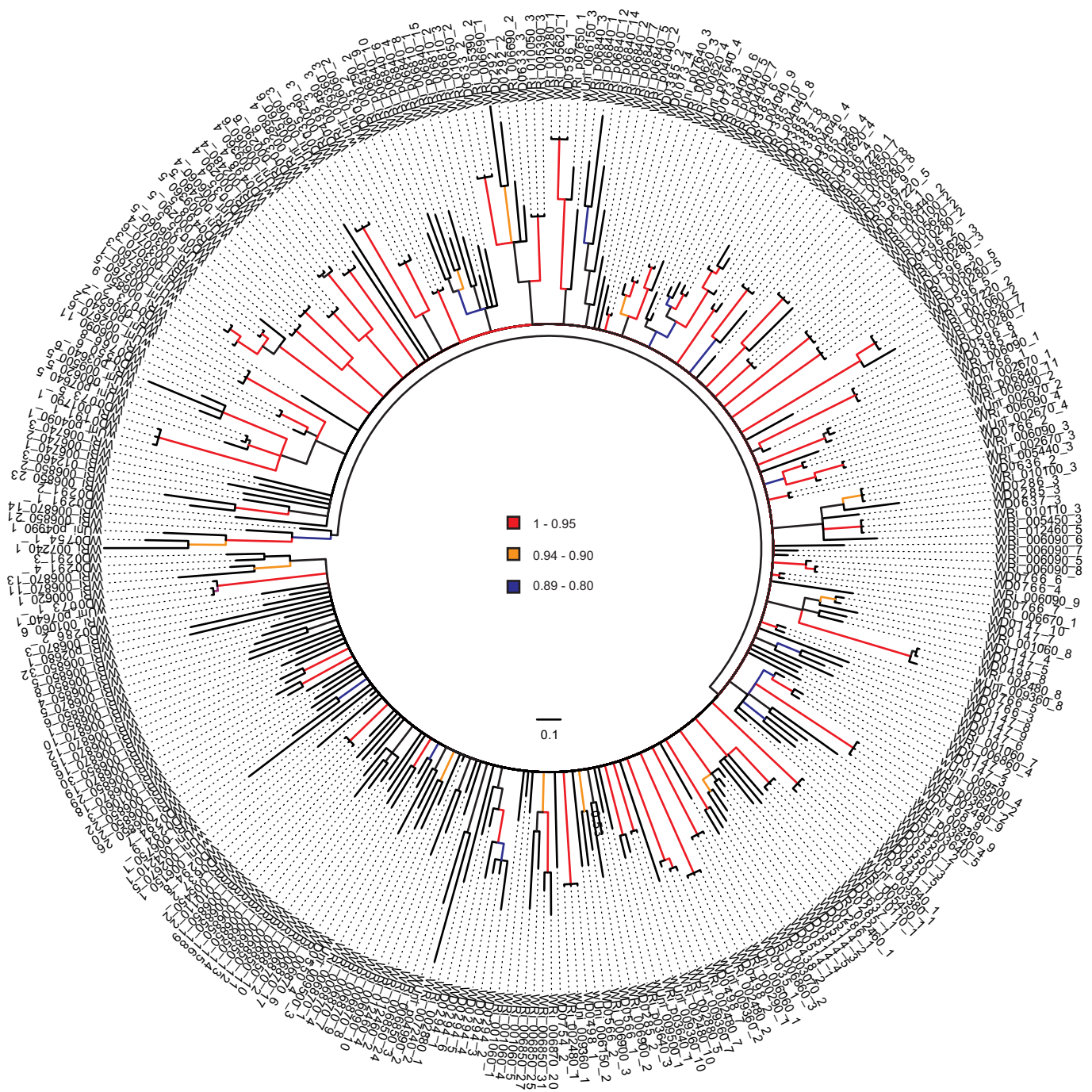
Fig. 4. Clustering of ankyrin repeat domains from *w*Ri, *w*Mel, and *w*Uni using Bayesian phylogenetic inference. The colors on the branches indicate different intervals of posterior probability: (*red*) 0.95–1.0, (*yellow*) 0.90–0.94, and (*blue*) 0.89–0.80 (nonsignificant), and (*black*) <0.80.

have been acquired recently. One such gene, WRi_006870, is located near phage *w*Ri-WOC. Although the gene is absent from the other 2 A-group strains, homologs were identified in *w*No, *w*Ma, and *w*Mau B-group strains that like *w*Ri use *D. simulans* as their natural host. The gene is not transcribed in early or late (overnight) embryos in *w*Ri, not in testes and early embryos in *w*No and not in adult males in *w*Mau: that is, this ANK gene exhibits stage-specific expression patterns in both A- and B-group genomic backgrounds (see Tables S9 and S10). A phylogeny based on the minor capsid protein shows that *w*Ri-WOC clusters with one of the prophages in *w*No (Fig. S4), indicating that prophage *w*Ri-WOC, along with some of its associated ANK

genes, may have been acquired from the B-group strains. Transmission of phages between *Wolbachia* strains of different supergroups that infect the same host has been seen in moths (14) and has also been shown to occur between *w*No and *w*Ha during coinfections of *D. simulans* in the laboratory (15).

In contrast, no examples of horizontal gene transfer between the A-group strains and the D-group *Wolbachia* strain *w*Bm were observed, consistent with the absence of prophages and low levels of recombination in mutualistic *Wolbachia* strains (24). Likewise, mutualistic endosymbionts of aphids show low recombination frequencies and have the most stable bacterial genomes identified to date (25). This dramatic difference in recombina-

tion may reflect variability in host-adaptation strategies, access to mobile elements, and different sets of recombination genes (15), as well as increased possibilities for rare genome variants to reach fixation in such populations (26, 27).

Wolbachia Sequences in Drosophila Genome Assemblies. *Wolbachia* sequences were previously identified in the trace archive files of both the *D. simulans* and the *Drosophila ananassae* genome projects and assumed to originate from contaminating bacterial DNA (28), a discovery that was followed by a debate about whether any corresponded to *w*Ri (29, 30). The more recent finding that *Wolbachia* genes have been transferred into the nuclear genomes of *D. ananassae* and other hosts (31–34) further added complexity with regards to the origin of these sequences.

Using BLAST, we recovered 73 scaffolds from the FlyBase assembly of the *D. ananassae* genome containing 177 kb (excluding gaps) that partially or fully matched *w*Ri sequences, some of which are currently annotated as *D. ananassae* genes in both GenBank and FlyBase. We did not identify any *w*Ri sequences that are shared with *w*Mel, presumably because the *w*Mel genome was used to filter out *Wolbachia* reads. A comparison to the *w*Ri genome revealed the absence of 2 IS-elements within scaffolds of otherwise conserved gene-order structures, suggesting that the *Wolbachia* sequences in the *D. ananassae* genome are similar but not identical to the *w*Ri genome. In contrast, no *w*Ri sequences were identified in the FlyBase assembly of the *D. simulans* genome. Furthermore, amplification of *w*Ri ANK and VIR sequences by PCR from *D. simulans* treated with tetracycline failed. We conclude that a corresponding transfer of *Wolbachia* genes into the nuclear genome of *D. simulans* is not likely.

Patchy Wolbachia Populations. Pervasive recombination in parasitic *Wolbachia* destroys the anticipated correlation between gene history, genome history, and strain phenotype. The *wsp* surface protein has been extensively used for genotyping but was found to be especially prone to recombination (9, 35) and 2 different sets of housekeeping genes, *gatB*, *coxA*, *hcpA*, *fbpA*, *ftsZ* (11) and *aspC*, *atpD*, *sucB*, and *pdhB* (12) were proposed as an alternative. However, not even these genes are protected from recombination events (11) and our comparisons of *w*Uni, *w*Mel, and *w*Ri show divergences at synonymous sites ranging from $K_s = 0$ in *aspC* to $K_s = 0.1–0.2$ for *gatB*, with different genes and segments of genes supporting different strain relationships. Hence, no single gene sequence will accurately describe the relationships of these A-group *Wolbachia* strains.

The global *Wolbachia* population is likely to consist of many subpopulations, or patches (36), where the boundaries are defined by, for example, geography or host specificity. Recombination among strains is expected to be frequent within patches, but less so between patches. For example, mutualistic adaptations to a single host may lead to the isolation and evolution of a subpopulation with limited recombination, as is possibly the case in nematode *Wolbachia*. While multilocus sequence typing may be useful to characterize supergroups, the intense recombination seen between A-group strains indicates that characterization of genotypes might require analysis at the whole genome level. However, as selection is expected to act on traits involved in host-adaptation processes within a patch, such genes may be useful to identify "fitness types," although not conferring any information that is meaningful in a phylogenetic sense.

Future Perspectives. The availability of complete genome sequence data for the model organisms *D. melanogaster* and *D. simulans*, as well as for their respective *Wolbachia* endosymbionts, offers an excellent opportunity to study host-adaptation processes by monitoring the coevolution of host and endosymbiont gene interactions in natural and transinfected hosts. Rapid

diversification of the ANK genes by segmental gene duplication may reflect diversifying selection to match a divergent set of target molecules in different cells, tissues, and hosts. To test this hypothesis, target proteins should be searched for among host genes that are also rapidly evolving, with prime candidates in gametogenesis, meiosis, reproduction (37), and innate immunity responses (38). An exciting avenue for future research is to identify the interacting endosymbiont-host proteins and determine whether these evolve by purifying, positive, or diversifying selection within *Wolbachia* subpopulations.

The association between *w*Ri and *D. simulans* is one of a few *Wolbachia* infections that have been studied in natural populations. Using *w*Ri as the reference genome, it is now possible to initiate comparative studies of *w*Ri genomes extracted from natural *D. simulans* populations with different phenotypes. Although we have demonstrated stability over the breakpoints during a period of 15 years in *w*Ri strains kept in the laboratory, other genetic changes, such as transposition of IS-elements, gene inactivation by IS-element insertions, and novel gene acquisition could occur rapidly in natural populations. For example, changes in fecundity have been observed during a 20-year period in a natural *D. simulans* population infected with *w*Ri in southern California (18). The genetic basis of this and other rapidly changing phenotypes can now be investigated.

Materials and Methods

Sequencing Strategies. *w*Ri: *Drosophila simulans* Riverside eggs were collected after 2 h and 1 to 2 ml of embryos were homogenized. A continuous renografin gradient (28%–45%) was used to concentrate *Wolbachia* cells. The 28% to 32% zone was collected and placed in agarose plugs that were treated with bacterial cell lysis and proteinase K solution. To remove contaminating host DNA, the plugs were run on a 1% Seakem Gold agarose (FMC BioProducts) gel for 24 h and the isolated DNA was subsequently used for library construction in a modified M13 vector as described previously (39). From the M13 library, 34,322 reads were sequenced, of which 19,727 were present in the final assembly, resulting in an overall 8.2-times coverage. An additional 18,031 reads were generated during gap closure and finishing.

*w*Uni: DNA was isolated from 300 dissected ovaries from adult females *Muscidifurax uniraptor* using a CTAB protocol, followed by lysozyme treatment and chloroform extraction. Primers were designed based on the genome sequence of *Wolbachia* strain *w*Mel, to amplify 1,100-bp products with 300-bp overlap on both ends to the adjacent product. Primers that successfully amplified short PCR products were selected and combined to generate long-range PCR products, where short products did not amplify. Next, 26,834 reads were sequenced and assembled into 287 contigs, of which 106 were longer than 2 kb. Short PCR-products were sequenced directly and long products were sheared by nebulization and cloned into the pSMART-HCKan vector before sequencing.

Verifying the wRi Genome Assembly. The *w*Ri assembly was confirmed over each IS-element or inferred breakpoint using genomic DNA from *w*Ri isolated from *D. simulans* and other infected hosts using PCR with specific primers. The size of the assembled genome *w*Ri is slightly lower than the 1.66 Mb previously estimated from pulse-field gel electrophoresis (40). However, the relative order of the observed restriction fragments matches those predicted from the genome sequence, except that the sizes of the individual fragments appear to have been systematically overestimated in the pulsed-field gel electrophoretic -analysis.

Informatics. Assembly was performed with PHRED-PHRAP-CONSED (41–43). Protein-coding genes were identified with GLIMMER (44) and CRITICA (45) and tRNA genes by tRNAscan-SE (46). Putative functions were inferred using BLAST against the National Center for Biotechnology Information databases and InterProScan (47). Repeat identification was made using MUMmer (48). Codeml, PAML 3.14 (49) was used to calculate substitution rates. Orthologs used for Ks calculations were retrieved by reciprocal best blast with additional cutoffs. RDP3 (23) was used to check nucleotide alignments for intragenic recombination using 6 methods, RDP, Geneconv, Bootscan, MaxChi, Chimaera, and 3Seq, with default settings except for window and step sizes. Sequences of the minor capsid gene were aligned with CLUSTALW (50) on the protein level and back-translated to nucleotide sequences. The phylogeny was reconstructed using MrBayes 3.12 (51) with the GTR+G model and run for

EVOLUTION

10,000,000 generations. Ankyrin repeats were found with the ANK HMM from PFAM (52) running HMMER 2.0 (53). An amino acid alignment was produced with hmmalign and then back-translated to nucleotides. The phylogeny was reconstructed using MrBayes3.12 (51) under the GTR+I+G model and run for 27,000,000 generations. For both trees, sampling was made every one-hundredth generation with 2 runs of 4 chains and default priors and a consensus trees were constructed using a ''burnin'' of 25%.

Transcription Analyses. For each of the tested *Wolbachia-Drosophila* associations, 300 testis and 150 ovaries were dissected from adults (1-day-old males and 3-day-old females). Embryos were collected every 2 h and late embryos every 16 h. Total RNA was extracted using TRIzol (Invitrogen) and treated with RNase-free DNase (Invitrogen). First-strand cDNA was synthesized from 5 μg of total RNA using reverse transcriptase (SuperScript III; Invitrogen) and

random primers (Promega), and thereafter treated with RNase H. For each gene, specific primers were designed based on the corresponding *w*Ri gene nucleotide sequence and used for PCR amplification

1. Werren JH, Baldo L, Clark ME (2008) *Wolbachia*: master manipulators of invertebrate biology. *Nat Rev Microbiol* 6:741–751.
2. Stouthamer R, Breeuwer JA, Hurst GD (1999) *Wolbachia pipientis*: microbial manipulator of arthropod reproduction. *Annu Rev Microbiol* 53:71–102.
3. Werren JH, Zhang W, Guo LR (1995) Evolution and phylogeny of *Wolbachia*: reproductive parasites of arthropods. *Proc Biol Sci* 261:55–63.
4. McGraw EA, O'Neill SL (1999) Evolution of *Wolbachia pipientis* transmission dynamics in insects. *Trends Microbiol* 7:297–302.
5. Bourtzis K, Braig HR, Karr TL (2003) in *Insect Symbiosis*, eds. Bourtzis K, Miller TA (CRC Press, Boca Raton), pp. 217–246.
6. Wu M, et al. (2004) Phylogenomics of the reproductive parasite *Wolbachia pipientis w*Mel: a streamlined genome overrun by mobile genetic elements. *PLoS Biol* 2:E69.
7. Klasson L, et al. (2008) Genome evolution of *Wolbachia* strain *w*Pip from the *Culex pipiens* group. *Mol Biol Evol* 25:1877–1887.
8. Foster J, et al. (2005) The *Wolbachia* genome of *Brugia malayi*: endosymbiont evolution within a human pathogenic nematode. *PLoS Biol* 3:e121.
9. Baldo L, Lo N, Werren JH (2005) Mosaic nature of the *Wolbachia* surface protein. *J Bacteriol* 187:5406–5418.
10. Baldo L, Bordenstein S, Wernegreen JJ, Werren JH (2006) Widespread recombination throughout *Wolbachia* genomes. *Mol Biol Evol* 23:437–449.
11. Baldo L, et al. (2006) Multilocus sequence typing system for the endosymbiont *Wolbachia pipientis*. *Appl Environ Microbiol* 72:7098–7110.
12. Paraskevopoulos C, Bordenstein SR, Wernegreen JJ, Werren JH, Bourtzis K (2006) Toward a *Wolbachia* multilocus sequence typing system: discrimination of *Wolbachia* strains present in *Drosophila* species. *Curr Microbiol* 53:388–395.
13. Cordaux R, et al. (2008) Intense transpositional activity of insertion sequences in an ancient obligate endosymbiont. *Mol Biol Evol* 25:1889–1896.
14. Masui S, Kamoda S, Sasaki T, Ishikawa H (2000) Distribution and evolution of bacteriophage WO in *Wolbachia*, the endosymbiont causing sexual alterations in arthropods. *J Mol Evol* 51:491–497.
15. Bordenstein SR, Wernegreen JJ (2004) Bacteriophage flux in endosymbionts (*Wolbachia*): infection frequency, lateral transfer, and recombination rates. *Mol Biol Evol* 21:1981–1991.
16. Gavotte L, et al. (2004) Diversity, distribution and specificity of WO phage infection in *Wolbachia* of four insect species. *Insect Mol Biol* 13:147–153.
17. Hoffmann AA, Turelli M, Simmons GM (1986) Unidirectional incompatibility between populations of *Drosophila simulans*. *Evolution* 40:692–701.
18. Weeks AR, Turelli M, Harcombe WR, Reynolds KT, Hoffmann AA (2007) From parasite to mutualist: rapid evolution of *Wolbachia* in natural populations of *Drosophila*. *PLoS Biol* 5:e114.
19. Stouthamer R, Breeuwert JA, Luck RF, Werren JH (1993) Molecular identification of microorganisms associated with parthenogenesis. *Nature* 361:66–68.
20. Ogata H, Suhre K, Claverie JM (2005) Discovery of protein-coding palindromic repeats in *Wolbachia*. *Trends Microbiol* 13:253–255.
21. Fuxelius HH, Darby AC, Cho NH, Andersson SG (2008) Visualization of pseudogenes in intracellular bacteria reveals the different tracks to gene destruction. *Genome Biol* 9:R42.
22. Jolley KA, Wilson DJ, Kriz P, McVean G, Maiden MC (2005) The influence of mutation, recombination, population history, and selection on patterns of genetic diversity in *Neisseria meningitidis*. *Mol Biol Evol* 22:562–569.
23. Martin DP, Williamson C, Posada D (2005) RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics* 21:260–262.
24. Jiggins FM (2002) The rate of recombination in *Wolbachia* bacteria. *Mol Biol Evol* 19:1640–1643.
25. Tamas I, et al. (2002) 50 million years of genomic stasis in endosymbiotic bacteria. *Science* 296:2376–2379.
26. Cho NH, et al. (2007) The *Orientia tsutsugamushi* genome reveals massive proliferation of conjugative type IV secretion system and host-cell interaction genes. *Proc Natl Acad Sci USA* 104:7981–7986.
27. Darby AC, Cho NH, Fuxelius HH, Westberg J, Andersson SG (2007) Intracellular pathogens go extreme: genome evolution in the Rickettsiales. *Trends Genet* 23:511–520.
28. Salzberg SL, et al. (2005) Serendipitous discovery of *Wolbachia* genomes in multiple *Drosophila* species. *Genome Biol* 6:R23.
29. Iturbe-Ormaetxe I, Riegler M, O'Neill SL (2005) New names for old strains? *Wolbachia w*Sim is actually *w*Ri. *Genome Biol* 6:401; author reply 401.
30. Salzberg SL, et al. (2005) Correction: Serendipitous discovery of *Wolbachia* genomes in multiple *Drosophila* species. *Genome Biol* 6:402.
31. Kondo N, Nikoh N, Ijichi N, Shimada M, Fukatsu T (2002) Genome fragment of *Wolbachia* endosymbiont transferred to X chromosome of host insect. *Proc Natl Acad Sci USA* 99:14280–14285.
32. Fenn K, et al. (2006) Phylogenetic relationships of the *Wolbachia* of nematodes and arthropods. *PLoS Pathog* 2:e94.
33. Hotopp JC, et al. (2007) Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science* 317:1753–1756.
34. Nikoh N, et al. (2008) *Wolbachia* genome integrated in an insect chromosome: evolution and fate of laterally transferred endosymbiont genes. *Genome Res* 18:272–280.
35. Baldo L, Werren JH (2007) Revisiting *Wolbachia* supergroup typing based on WSP: spurious lineages and discordance with MLST. *Curr Microbiol* 55:81–87.
36. Berg OG, Kurland CG (2002) Evolution of microbial genomes: sequence acquisition and loss. *Mol Biol Evol* 19:2265–2276.
37. Begun DJ, et al. (2007) Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. *PLoS Biol* 5:e310.
38. Sackton TB, et al. (2007) Dynamic evolution of the innate immune system in *Drosophila*. *Nat Genet* 39:1461–1468.
39. Andersson B, Wentland MA, Ricafrente JY, Liu W, Gibbs RA (1996) A ''double adaptor'' method for improved shotgun library construction. *Anal Biochem* 236:107–113.
40. Sun LV, et al. (2001) Determination of *Wolbachia* genome size by pulsed-field gel electrophoresis. *J Bacteriol* 183:2219–2225.
41. Ewing B, Green P (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* 8:186–194.
42. Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 8:175–185.
43. Gordon D, Abajian C, Green P (1998) Consed: a graphical tool for sequence finishing. *Genome Res* 8:195–202.
44. Salzberg SL, Delcher AL, Kasif S, White O (1998) Microbial gene identification using interpolated Markov models. *Nucleic Acids Res* 26:544–548.
45. Badger JH, Olsen GJ (1999) CRITICA: coding region identification tool invoking comparative analysis. *Mol Biol Evol* 16:512–524.
46. Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25:955–964.
47. Zdobnov EM, Apweiler R (2001) InterProScan–an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17:847–848.
48. Kurtz S, et al. (2004) Versatile and open software for comparing large genomes. *Genome Biol* 5:R12.
49. Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* 13:555–556.
50. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680.
51. Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574.
52. Bateman A, et al. (2004) The Pfam protein families database. *Nucleic Acids Res* 32:D138–D141.
53. Eddy SR (1998) Profile hidden Markov models. *Bioinformatics* 14:755–763.