

The Diversity and Evolution of *Wolbachia* Ankyrin Repeat Domain Genes

Stefanos Siozios^{1‡a}, Panagiotis Ioannidis^{1‡b}, Lisa Klasson², Siv G. E. Andersson², Henk R. Braig³, Kostas Bourtzis^{1,4,5*}

1 Department of Environmental and Natural Resources Management, University of Western Greece, Agrinio, Greece, **2** Department of Molecular Evolution, Uppsala University, Uppsala, Sweden, **3** School of Biological Sciences Bangor University, Bangor Gwynedd, United Kingdom, **4** Biomedical Sciences Research Center Al. Fleming, Vari, Greece, **5** Insect Pest Control Laboratory, Joint FAO/IAEA Division of Nuclear Techniques in Food and Agriculture, Vienna, Austria

Abstract

Ankyrin repeat domain-encoding genes are common in the eukaryotic and viral domains of life, but they are rare in bacteria, the exception being a few obligate or facultative intracellular Proteobacteria species. Despite having a reduced genome, the arthropod strains of the alphaproteobacterium *Wolbachia* contain an unusually high number of ankyrin repeat domain-encoding genes ranging from 23 in wMel to 60 in wPip strain. This group of genes has attracted considerable attention for their astonishing large number as well as for the fact that ankyrin proteins are known to participate in protein-protein interactions, suggesting that they play a critical role in the molecular mechanism that determines host-*Wolbachia* symbiotic interactions. We present a comparative evolutionary analysis of the wMel-related ankyrin repeat domain-encoding genes present in different *Drosophila-Wolbachia* associations. Our results show that the ankyrin repeat domain-encoding genes change in size by expansion and contraction mediated by short directly repeated sequences. We provide examples of intra-genic recombination events and show that these genes are likely to be horizontally transferred between strains with the aid of bacteriophages. These results confirm previous findings that the *Wolbachia* genomes are evolutionary mosaics and illustrate the potential that these bacteria have to generate diversity in proteins potentially involved in the symbiotic interactions.

Citation: Siozios S, Ioannidis P, Klasson L, Andersson SGE, Braig HR, et al. (2013) The Diversity and Evolution of *Wolbachia* Ankyrin Repeat Domain Genes. PLoS ONE 8(2): e55390. doi:10.1371/journal.pone.0055390

Editor: Richard Cordaux, University of Poitiers, France

Received: October 15, 2012; **Accepted:** December 21, 2012; **Published:** February 4, 2013

Copyright: © 2013 Siozios et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This research was supported in part by the European Community's Seventh Framework Programme CSA-SA_REGPROT-2007-1 (under grant agreement no 203590 (KB). LK, SA, HRB and KB benefitted by travel grants in the frame of the EU COST Action FA0701 "Arthropod Symbiosis: from fundamental studies to pest and disease management". The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: Kostas Bourtzis is a new member of the Editorial Board. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

* E-mail: K.Bourtzis@iaea.org

‡a Current address: Research and Innovation Center, Fondazione Edmund Mach, San Michelle all'Adige, Italy

‡b Current address: Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, Maryland, United States of America

Introduction

Wolbachia is a group of intracellular and maternally transmitted alphaproteobacteria, comprising strains with diverse symbiotic relationships with numerous arthropod as well as filarial nematode species [1,2]. These bacteria are quite widespread in insects and crustaceans [3]. Recent screens suggest that >40% of all terrestrial arthropod species are infected, rendering *Wolbachia* perhaps the most abundant symbiotic microorganism of the biosphere [4].

The successful spread of *Wolbachia* into insect host populations has been attributed to their unique ability to act as manipulators of host reproduction in order to ensure their own transmission. *Wolbachia* infections have been associated with the induction of feminization, thelytokous parthenogenesis, male killing and, most commonly, cytoplasmic incompatibility (CI) [1,2,5]. During the last decade, several studies have shown that *Wolbachia* infections can affect, in addition to reproduction, several additional aspects of host biology, physiology, ecology and evolution [1,2,5,6]. The molecular mechanisms that allow *Wolbachia* to establish symbiotic associations and to induce its extended phenotypes are yet to be unraveled.

Advances in genomics provided significant new information about *Wolbachia* biology. There are currently four complete genome sequences of *Wolbachia*, while several others are either available as permanent drafts or in progress [2,7–13]. A common feature of all insect *Wolbachia* genomes is the high percentage of repetitive elements such as insertion sequences, group II introns, duplicated segments of prophages and multi-gene families such as the ankyrin repeat genes [14–16].

Ankyrin repeat domains (ANK hereafter) consist of a tandem motif of about 33 amino acids and act as scaffolds that mediate protein-protein interactions [17,18]. Proteins with ANK domains are commonly found in eukaryotes and viruses, but they can also, although rarely, be found in the Bacteria and Archaea [19]. ANK proteins are known to be involved a multitude of functions, such as cell-cycle regulation, transcriptional regulation, cytoskeleton interactions, signal transduction, development, intracellular trafficking, sex differentiation, and they can also act as toxins [17,18,20]. Recent findings suggest that ANK proteins represent a new family of bacterial type IV effectors that play a major role in host-pathogen interactions and the evolution of infections [21,22]. Indeed, it was shown that ANK proteins of certain pathogenic

intracellular bacteria are secreted into the host cytoplasm and interact with host factors [22–25]. The intracellular pathogen *Anaplasma phagocytophilum*, for example, secretes AnkA through a type IV secretion system (T4SS). The protein interacts with specific regions of the host chromatin, resulting in a modulation of host gene transcription [23,24,26–28].

The potential role of ANK proteins in *Wolbachia* symbiosis and the manipulation of host reproduction has been investigated in *Drosophila* and mosquito species, but no direct correlation of the *Wolbachia* ankyrin gene repertoire with any bacterial phenotype has so far been established [10,29–34]. It is worth noting, however, that (a) the presence of certain ANK gene variants has been associated with crossing types in *Culex quinquefasciatus* [32] and (b) some ANK genes are under host sex-specific regulation [31]. A recent large-scale proteomic analysis of the excretory-secretory products of the *Wolbachia* infected filarial parasite *Brugia malayi* identified the presence of two *Wolbachia*-encoded ankyrin proteins [35]. These are probably the first *Wolbachia* ANK proteins shown to be secreted. *Wolbachia* also carries a functional T4SS [9,10,13,36], strengthening the hypothesis that the ANK genes play a role in functional and evolutionary processes of host-*Wolbachia* symbiosis.

Genomic analysis showed that the ANK genes account for up to 4% of the total number of genes in the insect *Wolbachia* strains *wMel*, *wRi* and *wPip* [9,10,13]. Additionally, comparative analyses between the ANK genes found in these genomes showed that they often evolve very rapidly, through for example gene duplication and contraction/expansion of repeated sequences within ANK domains [9,10]. Taken together, the above led to the suggestion that the ANK proteins may play a pivotal role in the molecular interaction between host and symbiont [9,10,13,30].

To learn more about the genetic diversity and the molecular mechanisms that underlie the evolution of the *Wolbachia* ANK gene family, we investigated the distribution of *wMel* ANK genes in several different *Drosophila*-*Wolbachia* associations, analyzed their genetic diversity by sequencing the identified orthologs and reconstructed their phylogenetic relationship. Our results indicate that both homologous and illegitimate recombination along with genomic flux provided by prophages and transposable elements are key factors in generating polymorphism and shaping the ANK gene pool in the *Wolbachia* strains studied.

Materials and Methods

Ethics Statement

N/A.

Fly Lines and *Wolbachia* Strains

The *Drosophila* lines and *Wolbachia* strains used in this study are listed in [37–45] Table 1. Flies were grown at 24°C ±1 on cornflour/sugar/yeast medium. *Wolbachia* strains were chosen according to their modification/rescue phenotype (mod⁺/resc⁺, mod⁻/resc⁺, mod⁻/resc⁻) and their embryonic localization pattern (posterior, anterior or uniform) as described by Veneti et al. [44]. The *Wolbachia* infection status of the *Drosophila* lines used was confirmed based on *wsp* and MLST typing as well as on CI properties [46–48].

Detection of ANK Genes with PCR and Southern Blot Analysis

Total genomic DNA was extracted from each *Drosophila* line according to the Holmes-Bonner method [49]. Specific primers for each of the 23 *wMel* ANK genes were designed (Table S1), based on the sequenced *Wolbachia* strain *wMel* and used to probe for the

homologs of the 23 *wMel* ANK genes in the 11 *Wolbachia* strains listed in Table 1. Reaction mixtures (final volume of 20 µl) contained 1x Taq reaction buffer (750 mM Tris-Cl pH 8.8, 200 mM (NH₄)₂SO₄, 0.1% Tween 20), 1.5 mM MgCl₂, 125 µM dNTPs, 10 pmol of each primer, 1 unit Taq polymerase (Promega) and 50 ng DNA. PCRs were run on a PTC-200 thermal cycler (MJ Research Inc.) with the following cycling conditions: an initial step at 94°C for 5 min, followed by 35 cycles of 94°C for 30 sec, 30 sec at a primer dependent annealing temperature, and 72°C for 3 min, plus a final extension at 72°C for 10 min.

For Southern blot analysis, 5 µg of total genomic DNA were digested with *EcoRI*. The DNA fragments were separated on a 0.8% agarose gel and blotted onto Immobilon-Ny+ filters (Millipore), according to the manufacturer's instructions and hybridized at 68°C for 18 h in Denhardt's solution, according to standard procedures [50]. Probes for the 23 *wMel* ANK genes were generated by PCR in 50 µl reactions, using specific primers amplifying part of each gene (see Table S1). PCR products were purified with the QIAquick PCR purification kit (QIAGEN) and radioactively labeled with [α-³²P] dATP (IZOTOP) using the Prime-a-gene labeling kit (Promega). A specific probe for the *Wolbachia* *ftsZ* cell division gene was generated (with primers *ftsZ*1:5'GTATGCCGATTGCAGAGCTTG and *ftsZ*2:5'GCCATGAGTATTCACTTGGCT [51]) and used as positive control.

Sequencing of *wMel*-like ANK Genes

The *wMel*-like ANK genes from the different *Wolbachia* strains were PCR amplified in 50 µl PCR reactions as described above, using the primers listed in Table S1. PCR products from three independent reactions were purified with the QIAquick PCR purification kit (QIAGEN), pulled together and sequenced by MacroGen (Korea). In cases of poor sequencing quality or the presence of unpredicted multiple products, PCR products were cloned into vector pGEM-T easy (Promega). Plasmid of at least three different clones was extracted from DH5α using the Qiaprep Miniprep kit (Qiagen) and directly sequenced with primers T7 and SP6. Sequence trace files from sequencing reactions were analyzed using the DNASTar 5.0 software package.

Analysis of Genetic Diversity

Multiple protein sequence alignments were performed and back-translated to nucleotide sequences with MUSCLE [52], as implemented in the Geneious package, version 4.0.3 [53]. Protein sequence alignment is usually preferable due to the larger protein alphabet and because it takes into account the redundancy of the amino acid codons resulting in more reliable nucleotide sequence alignment. Alignments were confirmed by visual inspection and edited manually. In order to avoid inflations, only unambiguously aligned sequences were used. Analysis of genetic diversity was performed with DNAsp, version 4.90.1 [54]. Substitution rates were estimated with Codeml, PAML 4.1 [55], using the codon substitution model of Goldman and Yang [56]. Ankyrin repeat domains were predicted by searching the sequences against the HMM profiles of the Pfam database (<http://pfam.sanger.ac.uk/>) and SMART (Simple Modular Architecture Research Tool; <http://smart.embl-heidelberg.de/>), using the default parameters [57,58]. Exact direct repeats larger than 8 nt within ANK genes were identified with the program REPFIND [59], using a P-value cutoff of 0.0001.

Phylogenetic Analysis

The phylogenetic relationships of the ANK genes were estimated using the maximum-likelihood (ML) method. Datasets

Table 1. *Drosophila* lines and *Wolbachia* strains.

<i>Wolbachia</i> supergroup	Host species	<i>Wolbachia</i> strain	Embryonic distribution	CI phenotype	Reference
A	<i>D. melanogaster</i>	wMel	posterior	mod ⁺ resc ⁺	[38,40,44]
	<i>D. melanogaster</i>	wMelPop	posterior	mod ⁺ resc ⁺	[43,44]
	<i>D. simulans</i>	wAu	posterior	mod ⁻ resc ⁻	[42,44]
	<i>D. teissieri</i>	wTei	posterior	mod ⁻ resc ⁺	[44,45]
	<i>D. yakuba</i>	wYak	posterior	mod ⁻ resc ⁺	[44,45]
	<i>D. santomea</i>	wSan	posterior	mod ⁻ resc ⁺	[44,45]
	<i>D. simulans</i>	wRi	uniform	mod ⁺ resc ⁺	[41,44]
	<i>D. simulans</i>	wHa	uniform	mod ⁺ resc ⁺	[41], Veneti pers. comm.
B	<i>D. simulans</i>	wNo	anterior	mod ⁺ resc ⁺	[41,44]
	<i>D. mauritiana</i>	wMau	anterior	mod ⁻ resc ⁺	[39,44]
	<i>D. simulans</i>	wMa	anterior	mod ⁻ resc ⁺	[37,44]

doi:10.1371/journal.pone.0055390.t001

with strong homology (>99% at nucleotide level) between the ankyrin gene orthologs were omitted from the analysis. ANK gene sequences from the A-supergroup strain *wUni* and the B-supergroup strain *wPip*, which infect the non-*Drosophila* hosts *Muscidifurax uniraptor* (Hymenoptera) and *Culex quinquefasciatus* (Diptera), respectively, were also included in our datasets for comparison. Prior to ML analysis, DNA substitution model parameters were estimated using Modeltest3.7 [60] and the Akaike information criterion (AIC): K81uf (WD0292, WD0766); TIM (WD0073); HKY (WD0035, WD0291, WD0441, WD0550, WD1213); TrN+I (WD0438, WD0498, WD0596); TVM+I (WD0191); TVM+G (WD0385, WD0566, WD0633, WD0754); TrN+G (WD0636); GTR (WD0147); GTR+I (WD0637). ML heuristic searches were performed using 100 random taxon addition replicates with tree bisection and reconnection (TBR) branch swapping. Bootstrap support was inferred using 100 bootstrap replicates. Searches were performed with PAUP, version 4.0b10 [61]. The ML method was also used to study the relationships between individual ankyrin repeat domains. Finally, the congruence between tree topologies was evaluated, using the SH-test [62], as implemented in PAUP 4.0b10. The SH-test compares the likelihood of score (lnL) of a given data set across its ML tree, versus the lnL of that data set across alternative topologies, which in this case are the ML phylogenies for the other ANK gene data sets [63]. The significance in differences among the likelihood scores was evaluated with a bootstrap test, using 1,000 permutations under full optimization.

Test for Recombination

For the purpose of recombination analysis, we discarded from the ANK gene datasets all but one sequence from groups of identical sequences. Furthermore, sequences that shared less than 70% identity were also discarded in order to avoid misalignment artifacts and to minimize the probability of false positive signals. In order to increase the possibility of including potential parental or daughter sequences, ankyrin repeat gene sequences from the A-group strain *wUni* and the B-group strain *wPip* were also added to our datasets. To identify potential recombination events, we used the recombination detection program RDP, version 3b27 [64], which implements different methods for detecting recombination signals. We primarily used the MaxChi method [65,66], but detected signals were considered significant only when they were confirmed by multiple methods, including Chimera [65,66],

Geneconv [67], RDP [68] and Bootscan [69]. The highest acceptable *P* value cutoff was set to 0.001, using a Bonferroni correction. Significance was evaluated with a permutation test based on 1000 permutations.

Nucleotide Sequence Accession Numbers

All ANK gene sequences generated in this study have been deposited into GenBank under accession numbers JX839306-JX839448.

Results

Distribution of *wMel* Like ANK Genes Using PCR and Southern Blot Analysis

The occurrence of the 23 *wMel* ANK genes in the eleven different *Drosophila-Wolbachia* associations listed in Table 1 was investigated by PCR and Southern blot analysis (Figure S1). The results of this analysis are summarized in Table S2. The data are largely in agreement with the work of Iturbe-Ormaetxe et al. [30], which examined the distribution of the 23 *wMel* ANK genes in nine *Wolbachia* strains [seven strains in common with the present study] (see Table S2). However, there are some differences, mostly between *Wolbachia* strains belonging to supergroup B (Table S2). Possible explanations for these variations are: (a) the different probes used in the two studies, (b) the different hybridization conditions and (c) the different primer sets used for the PCR analysis.

Only nine out of the twenty-three *wMel* ANK genes (WD0035, WD0191, WD0438, WD0441, WD0498, WD0636, WD0637, WD0766 and WD1213) are present in all supergroup A and B *Wolbachia* strains tested, based on Southern blot analysis. However, their presence was not for all of them confirmed by PCR, probably due to variability of the primer binding sites. All twenty-three *wMel* ANK genes were detected in the *Wolbachia* strains belonging to the *wMel* subgroup of supergroup A (*wMelPop*, *wAu*, *wTei*, *wYak*, *wSan*), with the exception of WD0514, which was absent from strain *wAu*. The distribution was different for the two more distantly related strains *wRi* and *wHa*. Both strains were tested positive for almost the same number of ANK genes (16 genes in *wRi* and 17 genes in *wHa*). Interestingly, the group of WO-A prophage-associated ANK genes (WD0285, WD0286, WD0291, WD0292, and WD0294) was found in *wHa*, whereas they were absent from *wRi*. However, the absence of these genes from *wRi*

does not imply the absence of prophage WO-A [10,14]. Lateral phage transfer, along with the activity of transposable elements could have resulted in the differential loss or independent acquisition of ANK genes between those strains. Although we could not detect a copy of the ANK gene WD0566, the published genome sequence of strain *w*Ri [10], confirmed the presence of a highly diverged variant (~52% pairwise identities at nucleotide level with *w*Mel). Finally, except for the nine universal genes, most ANK genes could not be detected in supergroup B strains (*w*No, *w*Mau, *w*Ma).

Sequence Analysis Revealed ANK Gene Polymorphism

Using internal and/or external primers, we obtained partial or, in some cases, full length sequence of *w*Mel-like ANK genes from the different *Wolbachia* strains. The sequence analysis showed that some ANK genes display, beyond single nucleotide polymorphisms, variations in the number and organization of the ANK repeat domains, as well as structural disruptions, including ORF disruption by frame shift mutations and insertion of transposable elements (Table 2).

Careful visual inspection of the alignments of the *w*Mel-related ANK genes that display domain number variations revealed that short or long identical direct repeats always flanked the duplicated and/or deleted segments (Figures 1, 2 and supporting Figure S2). For example, the identified deletions in the prophage-associated WD0294-related ANK genes are flanked by two groups of perfect repetitive elements of 26 bp

(AAAAGCAGAGATTAATGCAAAAAGATA) and 30 bp (CAGGGAAGGACTCCTTTACATTGGGCTGCT) long (Figure 1B). This striking observation suggests that the polymorphism of the number of ankyrin repeat domains depends on the presence of repetitive sequences scattered over the ANK clusters. In order to test the hypothesis that these short repetitive sequences are implicated in the expansion and/or contraction of the ankyrin repeat domains, the density of direct repeats (number of DRs/kb) was estimated for each ANK gene. The estimation was done on direct repeats larger than 8nt with the average length being between 10–20nt (supporting Figure S3). It was also investigated whether genes with ankyrin repeat domain number variations had greater DR density compared to genes that do not display variations. The former genes were indeed found to contain significantly more direct repeats (average of 14.3 and 1.4 repeats/kb, respectively, $P < 0.005$, Mann-Whitney test) (Figure 3).

Almost half of the *w*Mel-related ANK genes studied (WD0073, WD0147, WD0294, WD0385, WD0438, WD0514, WD0550, WD0566, WD0633, WD0754 and WD0766) display variations in the architecture of the ankyrin repeat domains in at least one of the studied *Wolbachia* strains (Table 2). This is particularly evident in the closely *w*Mel-related strains *w*MelPop, *w*Au, *w*Tei, *w*Yak and *w*San, in which the *w*Mel-related ANK genes share high sequence similarity at the nucleotide level; however display differences in the number of ankyrin repeat domains. The rest

Table 2. Major evolutionary events shaping *Wolbachia* ANK genes.

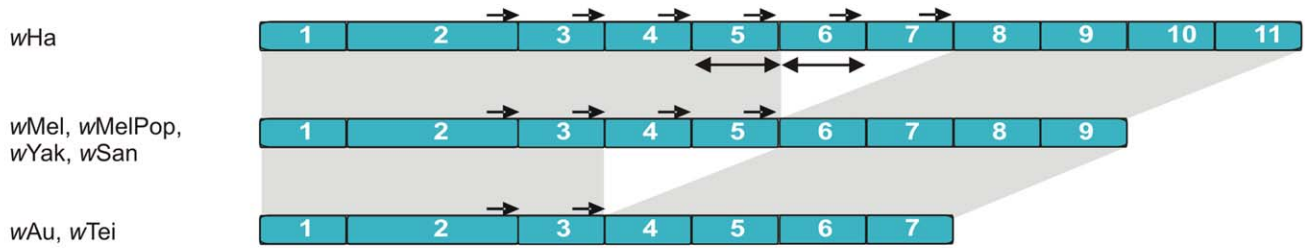
	Ankyrin genes ^a	Variation in the ankyrin repeat architecture	Frameshift mutations	Disruption by IS	Recombination (MaxChi, $P < 0.001$)	
homologs	WD0073	Yes			Yes (<i>w</i> Uni) ^b	
	WD0147	Yes				
	WD0294	Yes				
	WD0385	Yes	Yes	Yes		
	WD0438	Yes			Yes (<i>w</i> Ri)	
	WD0514	Yes				
	WD0550	Yes	Yes	Yes	Yes (<i>w</i> Uni)	
	WD0566	Yes				
	WD0633	Yes			Yes (<i>w</i> Au)	
	WD0636					
	WD0754	Yes				
	WD0766	Yes	Yes	Yes	Yes	
	WD1213			Yes		
	true orthologs	WD0035				
WD0191						
WD0285						
WD0286						
WD0291						
WD0292						
WD0441						
WD0498						
WD0596						
WD0637						

^aProphage associated ANK genes are presented in bold face.

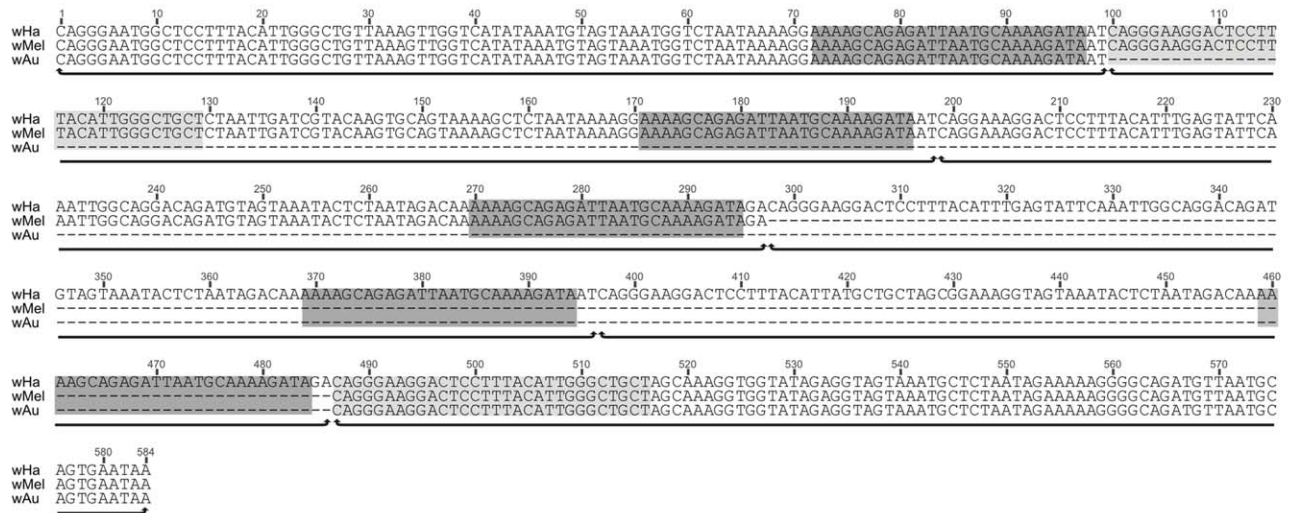
^bStrains potentially involved in recombination are shown in parenthesis.

doi:10.1371/journal.pone.0055390.t002

A WD0294 ANK clusters



B Repetitive elements and WD0294 ANK domain polymorphism



C WD0294 ANK domains relationships



Figure 1. Repetitive DNA sequences and ankyrin repeat domain number polymorphism. The example of WD0294. A) Only the ankyrin repeat domain containing regions are shown. Blue rectangles represent individual ankyrin repeat domains. The light gray shading between the ANK clusters indicates homologies between the different strains. Double arrows represent identical duplications. Small black arrows indicate direct repeats capable of engaging into illegitimate recombination. B) Detail of the nucleotide alignment around the deleted region. This region includes the ANK domains 3–8 from *wHa*, 3–6 from *wMel* and 3–4 from *wAu*. Gray rectangles show the position of repeated sites flanking the deletions. ANK repeats are underlined. C) ML phylogenetic relationships between individual ankyrin repeat domains.
doi:10.1371/journal.pone.0055390.g001

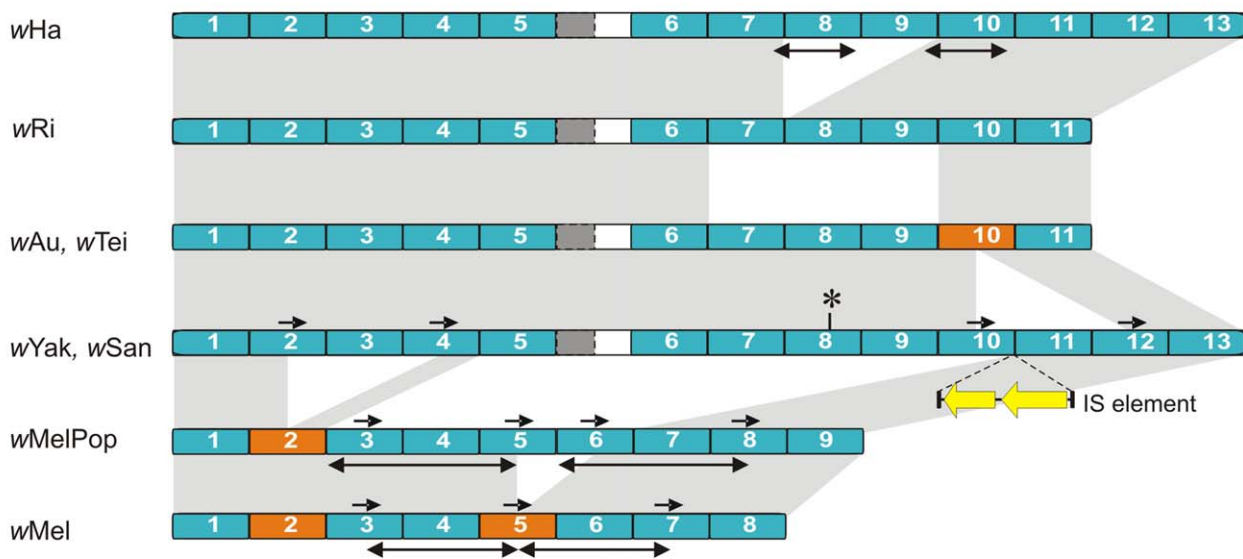
of the genes exhibit similar numbers and organization of ankyrin repeat domains in the different *Wolbachia* strains.

Duplications and/or deletions of ankyrin repeat domains, including shuffling between the structural elements of physically distant ankyrin repeat domains, are often the reason for copy number variations. For example, the prophage-associated WD0294-related ANK genes display high similarity at the nucleotide level in all strains which harbour it (Table S3). However, in the strains *wAu* and *wTei*, the genes code for two ankyrin repeat domains less than those present in the strains *wMel*, *wMelPop*, *wYak* and *wSan*. On the other hand, in the strain *wHa* the homolog of the same gene codes for two ankyrin repeat domains more (Figure 1A). Similarly, (a) the WD0514-related ANK genes present in the strains *wTei*, *wYak* and *wSan* code for

two ankyrin repeat domains less than those found in the strains *wMel* and *wMelPop* and (b) the WD0550-related ANK genes present in the strains *wAu* and *wMelPop* code for two ankyrin repeat domains more than those of the strains *wMel*, *wTei*, *wYak* and *wSan* (data not shown).

The present study confirmed that the WD0766-related ANK genes exhibit extensive variability in both number and organization of the ankyrin repeat domains (Figure 2A) as previously reported [30,70]. In addition, sequence analysis of the WD0766-related ANK genes in the closely related *Wolbachia* strains *wYak* and *wSan* showed that these genes are disrupted due to the insertion of a full-length IS5 element (a similar phenomenon was reported for the WD0385-related ANK gene in the strain *wAu* by Iturbe-Ormaetxe et al., 2005) [30] and the presence of a frame-

A WD0766 ANK clusters



B chimeric origin of WD0766 individual ANK domains

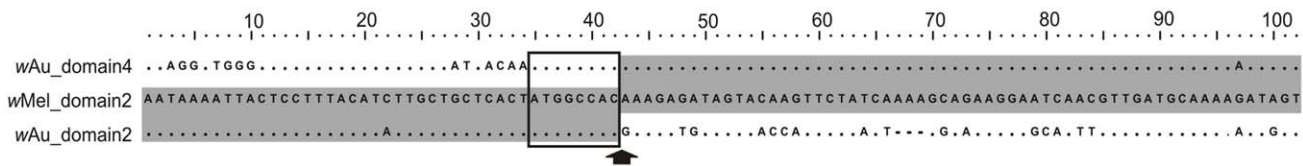


Figure 2. Repetitive DNA sequences and ankyrin repeat domain number polymorphism. The example of WD0766. A) Only the ankyrin repeat domain containing regions are shown. Blue rectangles represent individual ankyrin repeat domains. Dark gray rectangles with dotted outline represent ankyrin repeat domain remnants. The light gray shading between the ANK clusters indicates homologies between the different strains. Orange rectangles represent putative chimeric ankyrin repeat domains, and double arrows represent identical duplications. Small black arrows indicate direct repeats capable of engaging into illegitimate recombination. The reconstructed structure of disrupted *wYak* and *wSan* ANK homologs is also presented. The asterisk and the double yellow arrows correspond to a frame shift mutation and the position of the IS5 element, respectively. B) example of chimeric origin of *wMel* (and *wMelPop*) ankyrin repeat domains 2. Identities with the parental ankyrin repeat domains 2 and 4 from *wAu* are shaded. Box shows the position of the repeated site between the three sequences. The vertical arrow indicates the loop between the two α -helices of the ankyrin repeat domain.
doi:10.1371/journal.pone.0055390.g002

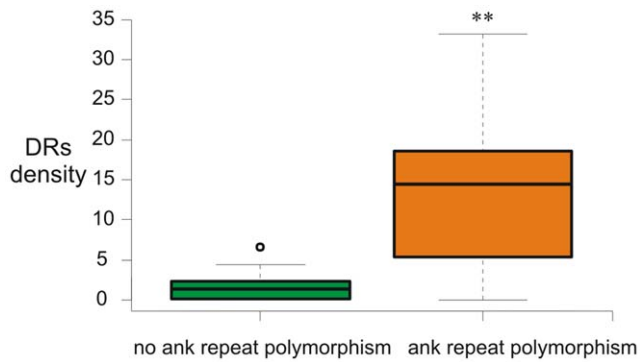


Figure 3. Density of direct repeats larger than 8nt per ANK gene. Box plot graph showing that ANK genes display ankyrin repeat number variations have greater DR density compared with genes that do not display variations (** $P < 0.005$, Mann-Whitney test). doi:10.1371/journal.pone.0055390.g003

shift upstream of the IS element. The reconstruction of these two ORFs showed that the WD0766-related ANK genes contain 13 complete ankyrin repeat domains, two repeats longer than the WD0766-related ANK gene of the closely related strain *wTei* and that of the *Wolbachia* strain *wAu*. It is worth noting that the 5' and the 3' regions of the WD0766-related ANK gene, including the first and the last ankyrin repeat domains, share high conservation between all *Wolbachia* strains studied. However, extensive duplications and deletions of the internal ankyrin repeat domains interrupt this homology.

Inspection of the multiple alignment of WD0766-related ANK gene sequences reveals that deletion events within the ANK cluster may have resulted in the shuffling of physically distant ankyrin repeat domains. One such putative case is the formation of the second ankyrin repeat domain of the *wMel* and *wMelPop* genes. According to the alignment, this ankyrin repeat domain is chimeric. The first 42 bp, which encode the first α -helix of the ankyrin repeat domain (amino acid residues 1–14), are very similar to the corresponding region of the second ankyrin repeat domain of the WD0766-related ANK genes of strains *wAu*, *wTei*, *wRi* and *wHa*, while the last 57 bp, which encode for the second α -helix (amino acid residues 15–33), are very similar to the fourth ankyrin repeat domain of the same strains (Figure 2B). Similarly, the 10th ankyrin repeat domain of the *wAu* and *wTei* WD0766-related ANK genes is probably the result of shuffling between the 10th and the 12th ankyrin repeat domains of the corresponding *wYak* and *wSan* genes, while the 5th ankyrin repeat domain of the *wMel* gene (WD0766) seems to be the result of a shuffling event between the 5th and the 6th ankyrin repeat of the *wMelPop* WD0766-related ankyrin gene (Figure 2A).

The WD0147-, WD0438-, WD0566- and WD0754-related ANK genes present in the *Wolbachia wMel* subgroup display no ankyrin-repeat polymorphism. It is also worth noting that the partial sequences of the WD0385-related ANK genes of the *wMel* subgroup strains, which correspond to the last two ankyrin repeat domains, have identical sequences. However, the corresponding five genes of the *wRi* strain display extensive variations in number and/or organization of the ankyrin repeat domains, suggesting that they have undergone extensive duplications and rearrangements [10].

The prophage-associated WD0636-related ANK genes present in *Wolbachia* strains *wAu* and *wTei* contain a premature stop codon, due to the insertion of G between position 283 and 284 of the *wMel* sequence, eliminating the last ankyrin repeat domain of

the gene together with the 3'-end of the gene (data not shown). The WD1213-related ANK genes present in the strains *wNo*, *wMau* and *wMa* (all B supergroup strains) are identical to each other; however, they carry eleven indels, three of which cause frame shifts, compared to the WD1213 ANK gene in *wMel* (data not shown).

Based on the above, the studied *wMel*-related ANK genes were classified into two major groups: (a) the true orthologs, which are of the same length and have the same number of ankyrin repeat domains and (b) the homologs which differ in the length, copy number and the organization of the ankyrin repeat domains and may also carry structural disruptions, such as frame shifts, deletions and insertions.

Genetic Diversity of ANK Genes

Table S3 summarizes the features of the ANK genes detected in the studied *Wolbachia* strains, the great majority of which belong to supergroup A. The G+C content ranged from 30.8% (chromosomal ANK gene WD0438) to 42% (phage ANK gene WD0636), with an average of 35%. Overall, the genetic diversity analysis revealed high conservation at the nucleotide level across the ANK genes, as indicated by the synonymous substitutions (K_s) (Figure 4 and Table S3). In general, the majority of the ANK genes of *wMel*-subgroup strains (*wMel*, *wMelPop*, *wAu*, *wTei*, *wYak* and *wSan*) are highly conserved (almost 100% identity at the nucleotide level) compared to the corresponding genes of the more distantly related strains *wRi* and/or *wHa* (distantly related based on MLST analysis [46,47]), which display the highest degree of genetic variability and in some cases act as outliers.

It is worth noting that although the WD0294, WD0514, WD0550 and WD0766 homologs differ in ankyrin repeat domain architecture, they show limited sequence polymorphism (SNPs). However, the WO-B prophage-associated homologs WD0633 and WD0636 exhibit significant levels of genetic diversity ($K_s > 0.05$) (Figure 4 and Table S3), including the two copies of the WD0636 gene detected in the *Wolbachia* strain *wSan* (95% identity at nucleotide level - $K_s = 0.082$).

The sequences of only five (WD0441, WD0498, WD0636, WD0637 and WD01213) of the nine universally occurring ANK genes were retrieved from the three B-super group *Wolbachia* strains *wNo*, *wMau* and *wMa* (Table S2) and showed almost 100% identity. The prophage-associated ANK gene WD0636 was again the exception, exhibiting a high level of genetic diversity ($K_{s,avg} = 0.07535$) (Table S3). Interestingly, two copies of the WD0636 gene were detected in the strain *wNo*, which were 89% identical at nucleotide level ($K_s = 0.15$). It should also be noted that we sequenced a partial fragment (356 bp) of a highly diverged WD0191 ortholog from strain *wMau*, which exhibited 67.2% identity to A-super group orthologs. Overall, there is a greater degree of genetic divergence between supergroups A and B, as indicated by the patterns of synonymous substitutions (K_s) ranging from $K_{s,avg} = 0.238$ (WD0636) to $K_{s,avg} = 0.495$ (WD0498; Table S3).

Phylogenetic Analysis

The relationships of the ANK genes were also studied with ML phylogenetic analysis. ANK orthologs with almost 100% identity at nucleotide level (like the WO-A prophage ankyrin genes WD0285 and WD0286) were omitted from the analysis. Most trees clearly reflect the evolutionary divergence of the two major *Wolbachia* supergroups A and B; however, branch lengths separating the two supergroups are often in disagreement between the different ANK gene-based trees (Figure 5 and supporting

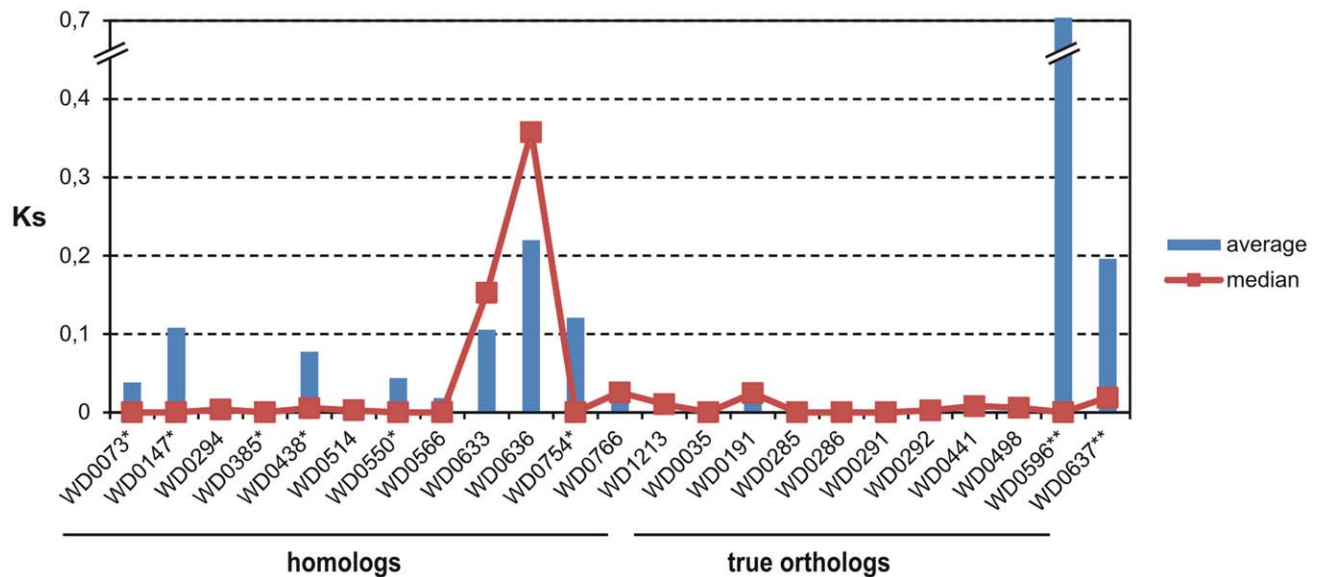


Figure 4. Comparative analysis of genetic diversity. The graph represents the patterns of synonymous substitutions within A-supergroup strains. (*: *w*Ri is an outlier, **: *w*Ha is an outlier). doi:10.1371/journal.pone.0055390.g004

Figure S4). The discordance suggests that the rate of evolution varies for different genes, and this is supported by the different levels of genetic diversity observed between these genes. Within supergroup A, the topologies were not significantly different between the datasets, placing all strains in a single clade. However, in some cases the more distantly related strains *w*Ri and *w*Ha branched at different positions, being separated by long genetic distances (WD0498 and WD0754 gene-based trees in Figure 5). This is supported by the highly heterogeneous pattern of genetic diversity observed in different ANK genes in different *Wolbachia* strains and could reflect different evolutionary rates or even different times of gene acquisition.

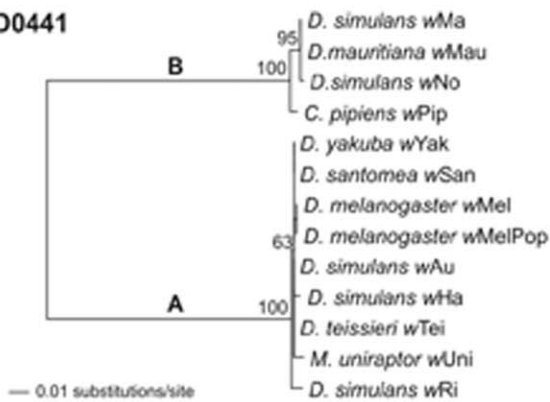
Perhaps the most interesting observation was that the two WO-B prophage-associated ANK genes WD0636 and WD0637 showed significant topological conflicts with the chromosomal ANK genes and also with other WO-B prophage-associated ankyrin genes (Figure 5). The WD0636 and WD0637 orthologs-based phylogenetic trees strongly support grouping of the two A-group strains *w*Yak and *w*San with *w*Ha and the B-supergroup strains *w*No, *w*Mau and *w*Ma. To statistically evaluate the topological incongruence, the ML phylogenies of the two prophage-associated ANK genes were compared with the ML phylogenies of the chromosomal ANK genes (WD0441, WD0498 and WD1213), using the SH test. The analysis was restricted to eight *Wolbachia* strains (*w*Mel, *w*Au, *w*Tei, *w*San, *w*Ri, *w*Ha, *w*No and *w*Mau). The likelihood-based SH test for significance of topological differences supports the discordances among topologies of the WD0636 and WD0637 gene-based phylogenies with the topologies of chromosomal ANK genes (Table 3). Interestingly, the phylogenies of WD0636 and WD0637 also showed topological incongruence (SH test $P < 0.05$, data not shown) with other WO-B prophage-associated ANK genes like WD0596 and WD0633. That could be explained either by the presence of multiple prophage elements in the genomes of *w*Yak and *w*San, harbouring a different ANK gene repertoire, or due to recombination events between different prophages.

Recombination within ANK Genes

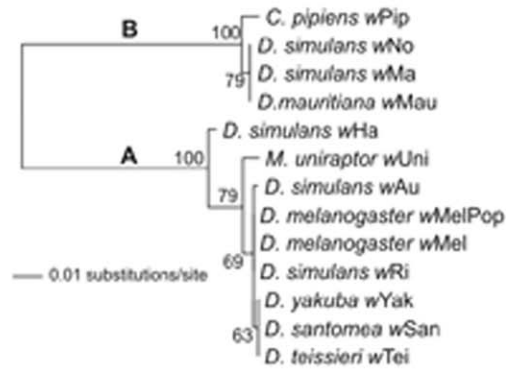
The role of recombination in the evolution of *Wolbachia* ANK genes was investigated with the program MaxChi [65,66]. Statistically significant recombination signals were also confirmed by the programs Chimera [65,66], Geneconv [67], RDP [68] and Bootscan [69]. As summarized in Table 2, five ANK genes, the prophage-associated ANK gene WD0633 and the chromosomal ANK genes WD0073, WD0438, WD0550 and WD0766, exhibited significant evidence of intragenic recombination (MaxChi, $P < 0.0001$ based on 1000 permutations). We present in detail only a single example, that of the prophage-associated ANK gene WD0633. A significant recombination event was detected by MaxChi, as well as by Chimera, Geneconv, RDP, and Bootscan ($P < 0.001$ based on 1000 permutations), at position 620 of the nucleotide sequence alignment. The breakpoint detected by MaxChi divides the alignment into two parts, the first of which encodes the 3rd and the 4th ankyrin repeat domains, demonstrating exchange of the entire ANK clusters between *Wolbachia* strains *w*Ri and *w*Au (Figure 6A). The sequence of the *w*Au homolog 5' region with respect to the predicted breakpoint is more similar to the *w*Ri (as well as to the *w*Mel and *w*MelPop) homologs, while the sequence of the 3' region of the breakpoint is more similar to the *w*Yak and *w*San homologs. Phylogenetic trees reconstructed separately for the two regions (Figure 6B & C) clearly show the shifted position of strain *w*Au.

Our results indicate that *Wolbachia* ANK genes undergo recombination rather frequently, as it was possible to detect significant recombination signals in our small dataset. The possibility that more ANK genes have been involved in recombination events could not be excluded, as the detection of recombination signals might have been masked by extensive sequence variability and the sampling bias. This is supported by the fact that some ANK genes displayed local variations in nucleotide divergences between different strains. For example, the chromosomal ANK gene WD0441 displayed in a comparison between A- and B-supergroup strains, local conservation in the 3' region of the gene (including the two ankyrin repeat domains) with

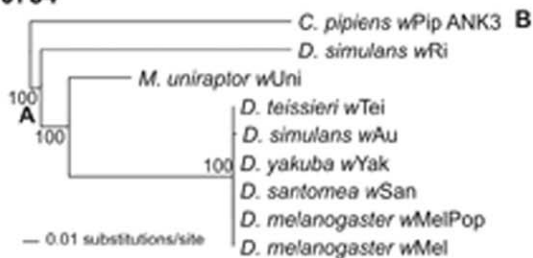
WD0441



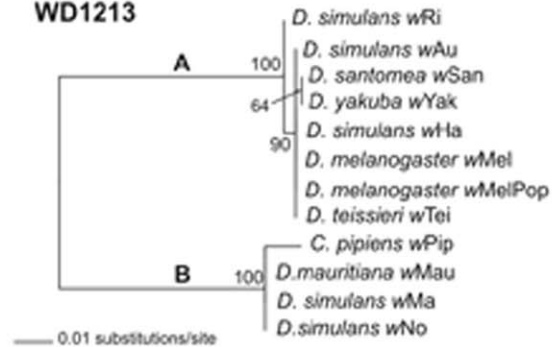
WD0498



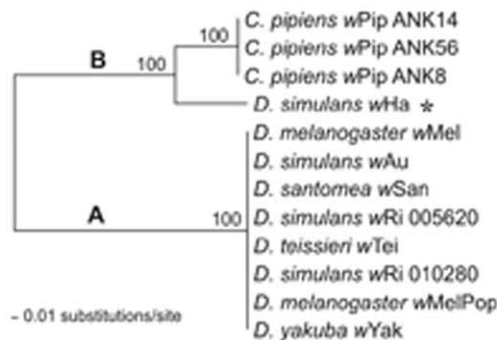
WD0754



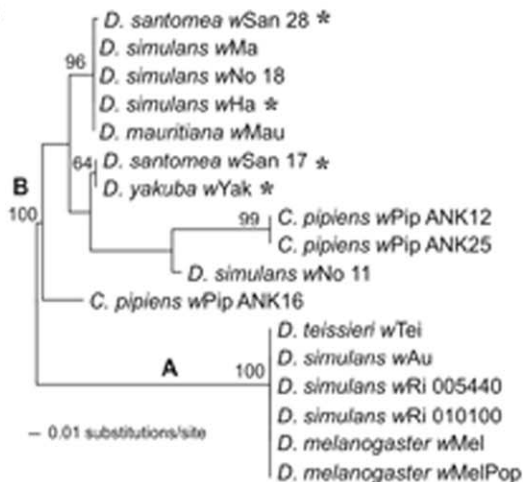
WD1213



WD0596



WD0636



WD0637

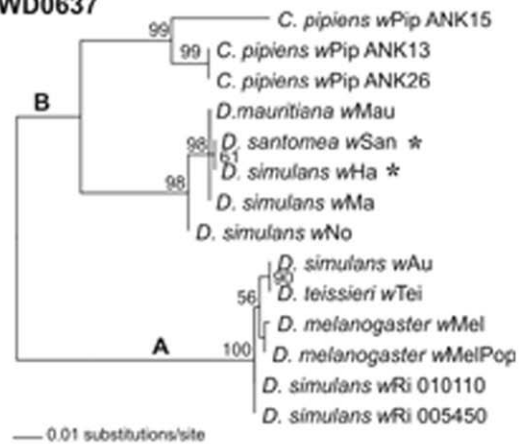


Figure 5. Phylogeny of four chromosomal- (WD0441, WD0498, WD0754 and WD1213) and three prophage-associated (WD0596, WD0636, WD0637) ANK genes. The trees are midpoint-rooted and inferred using maximum likelihood. ML bootstrap support values inferred from 100 replicates are also presented. Bootstrap values lower than 50 are omitted. The discordant positions of strains *wYak*, *wSan* and *wHa* between the chromosomal- and prophage-associated ANK gene phylogenies are highlighted with asterisks. Evolutionary model parameters were estimated with Modeltest under the Akaike Information Criterion: HKY (WD0441,WD1213); TrN+I (WD0498, WD0596); TVM+G (WD0754); TrN+G (WD0636); GTR+I (WD0637).
doi:10.1371/journal.pone.0055390.g005

~90% identity, while the 5' region of the gene displayed extensive variability with less than 50% identity (Figure S5).

Discussion

ANK proteins are involved in numerous and diverse processes and have been suggested to play an important role in host-symbiont interactions [17,70–73]. This study investigated the presence, diversity and evolution of ANK genes in different *Drosophila-Wolbachia* associations. Our data show that the *Wolbachia* ANK genes form a rapidly evolving gene family and the plausible mechanisms of their evolution are classical homologous recombination, illegitimate recombination and genomic flux mediated by prophages. These data further confirm that recombination represents a powerful mechanism that accelerates and shapes genomic evolution [74–79].

Sequence and phylogenetic analysis of the *Wolbachia* ANK genes revealed that one of the major causes for the observed sequence polymorphism is recombination, both homologous recombination and illegitimate recombination. Intragenic recombination was observed in both chromosomal and prophage-associated ANK genes. In several cases, intragenic recombination events resulted in the exchange of entire ANK clusters.

An important finding of the present study was the positive correlation of the presence of short DRs scattered over the ankyrin repeat domains with ankyrin repeat domain number polymorphism providing an example of the mechanisms affecting ANK gene evolution. Such short repetitive sequence elements are known to play a major role in DNA deletion and duplication events in both prokaryotes and eukaryotes. Illegitimate recombination events are independent of RecA and are thought to occur by at least two different mechanisms: replication slippage and single-strand annealing [80,81]. The presence of DRs in the same location with respect to ankyrin repeat domains is important in order to maintain reading frame, structure and hence the function of the ankyrin repeats. Thus, the direct repeats involved in

illegitimate recombination, duplication, deletion as well as recombination between distant ankyrin domains, likely play an important role in evolution of *Wolbachia* ANK genes. Recombination events mediated by short repeated sequences scattered over genes may have resulted in gene deterioration and the production of species-specific orphans, including a putative ANK gene, in the closely related *Rickettsia* species, *R. conori*, *R. rickettsii* and *R. montana* [82].

It is also important to note that stress-response genes in prokaryotes are known to have a higher than average number of short or large repeats capable of engaging in recombination, probably as a strategy to cope with unstable environmental conditions [83]. One could speculate that the ANK genes may have a similar role in *Wolbachia* or they may determine the host range and/or tissue tropism, like the ANK proteins of eukaryotic viruses [84]. In addition, it has been suggested that the evolutionary success of the eukaryotic ANK protein family is in part due to their ability to bind to multiple targets by adapting their binding sites through duplication, deletion and shuffling, as a result of alternative exon splicing [71,72,85,86]. It is also believed that modular proteins, like the ANK proteins, evolve faster than non-modular proteins through recombination [72,86]. Illegitimate recombination, as well as homologous recombination, can act within one genome (including the ANK genes) providing a source of genetic variability needed for *Wolbachia* strains to rapidly adapt to new environmental conditions.

Bacteriophage Flux and Evolution of *Wolbachia* ANK Genes

Bacteriophages are major determinants in bacterial genome evolution [87–89]. *Wolbachia* prophage elements are abundant and widespread and can laterally transfer between different *Wolbachia* strains co-inhabiting the same arthropod host [15,63,90–93]. A recent study reported that the prophage-associated ANK genes form one of the most divergent groups of *Wolbachia* genes [14]. Our analysis supports this finding and indicates that the ANK gene “cargo” of a given prophage may differ between strains, as is the case for the five WO-A associated ANK genes that seem to be an independent acquisition in *wHa* and the *wMel*-like strains. Furthermore, there is an absence of congruence between the phylogenies of prophage WO-B associated ANK genes not only with the chromosomal but also with other WO-B associated ANK genes. This incongruence is indicative of an active phage, which is able to move horizontally between different *Wolbachia* strains sharing a common host. This mechanism of genetic exchange was previously suggested for *Wolbachia* strains including *wHa* and is known as the “intracellular arena” hypothesis [63,91]. Lateral phage transfer coupled to intra- and intergenic recombination events could account for this rapid exchange and spreading of ANK genes.

Multiple Phages vs Multiple Infections

Our analysis of the 23 *wMel*-like ANK genes in different *Drosophila-Wolbachia* associations revealed striking differences between the strains, as well as between the genes studied. This was particularly evident for the prophage-associated ANK genes,

Table 3. Shimodaira-Hasegawa test for the statistical significance of the topological incongruence between alternative chromosomal- and prophage-associated ankyrin genes.

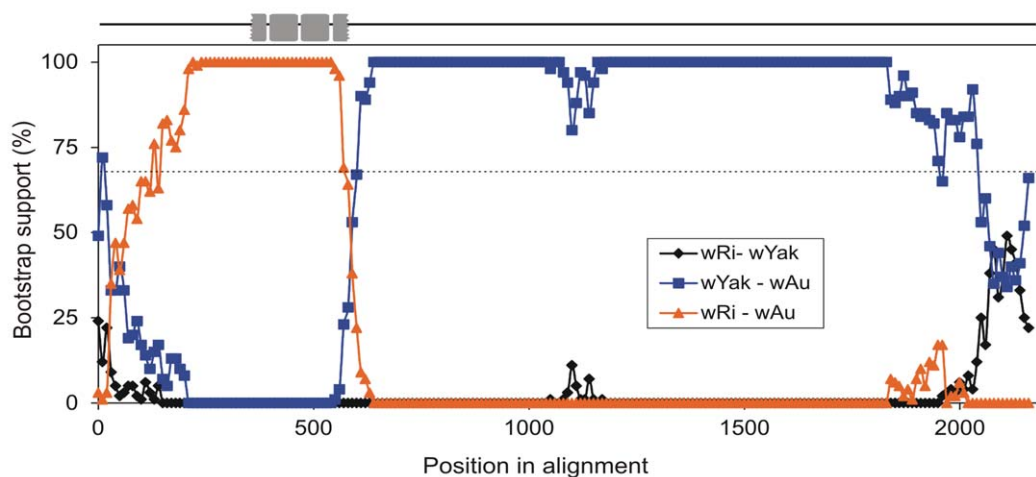
Topology	Data Set				
	WD0441	WD0498	WD0636	WD0637	WD1213
WD0441	1228.38	1054.82	497.64***	1599.62***	1196.26
WD0498	1228.56	1046.17	454.18***	1365.17***	1200.87
WD0636	1624.91***	1141.92***	399.14	1118.92	1665.61***
WD0637	1624.91***	1145.01***	409.40	1091.34	1653.28***
WD1213	1228.38	1054.82	497.64***	1565.64***	1188.06

Values are the likelihood score (-ln L) of a given data set across its own ML tree (boldface), as well as across the alternative tree topologies. Significance levels are based on full optimization.

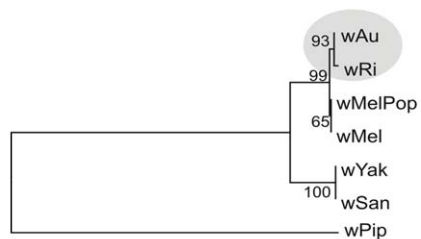
*** $P < 0.0001$.

doi:10.1371/journal.pone.0055390.t003

A recombination within WD0633



B WD0633 sequence before position 620



C WD0633 sequence after position 620

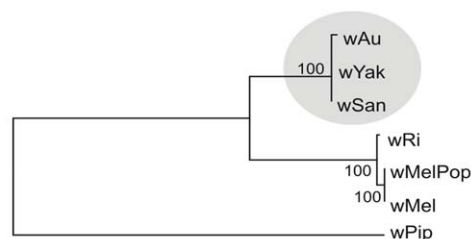


Figure 6. Recombination within the prophage-associated ANK gene WD0633. A) The relative bootstrap support values (1000 bootstrap resampling) are shown, calculated for a moving 200 bp window with a 10bp step size across the alignment of the WD0633 homologs. For each alignment-window, nucleotide distances and phylogenetic trees were produced using the neighbor joining method. The dotted line indicates 70% cutoff. Gray rectangles represent the positions of the ankyrin repeats in the alignment. B, C) ML phylogeny of *Wolbachia* strains reconstructed separately for the 5' and 3' regions of WD0633 supports the group shift of the putative recombinant strain (wAu). doi:10.1371/journal.pone.0055390.g006

suggesting that the two ANK groups (chromosomal and phage) have different evolutionary histories. However, conclusions about the evolutionary history of some genes may not be easily drawn. For example, the *Wolbachia* strains *wSan* (supergroup A) and *wNo* (supergroup B) contain two copies of the prophage-associated ANK gene WD0636. It is not clear if the two copies are the products of a gene duplication event or belong to different copies of the WO-B prophage. The existence of multiple WO-B prophages has already been described in strain *wRi*, which harbors two identical copies of a WO-B-like prophage [10], as well as in strain *wPip*, which harbors five WO-B-like prophage regions [9].

It was recently shown that another prophage-associated gene, the DNA adenine methyltransferase gene *met2*, is present in two copies in the symbiotic association between *D. teissieri* and *wTei*; *wTei* being a strain closely related to *wSan*. Molecular analysis indicated that the two *met2* gene copies are present in two different *Wolbachia* strains which co-exist in the host *D. teissieri* [94]. Coinfection may thus explain multiple copies of prophage-associated ANK gene WD0636 in the hosts *D. santomea* and in *D. simulans* Noumea. Although the presence of a double infection is rather a speculation for *D. santomea* (but see also below), it may indeed be the case for *D. simulans* Noumea. This host was originally doubly infected with two *Wolbachia* strains, *wNo* and *wHa*, before

it was established as a *wNo* mono-infected line through selection [95]. However, recent sequencing of the *Wolbachia* strain(s) infecting *D. simulans* Noumea also suggested the presence of *wHa* in low quantities (supergroup A) [Ellegaard, Klasson, Näslund, Bourtzis and Andersson, unpublished data]. A PCR analysis may thus report genes present in *wHa* but not in *wNo* as present in *D. simulans* Noumea because of the slight contamination with *wHa*. Indeed, the identification of WD0285 in *wNo* (Table S2) is an artifact of the presence of this gene in *wHa*, which is co-infecting at a low concentration.

Riegler et al. [70] recently proposed that polymorphic variable number tandem repeats and ANK genes can be used as a new diagnostic tool for genotyping *Wolbachia* strains. They also suggested two variable ANK genes (WD0766 and WD0550) for fingerprinting and discrimination between closely related *Wolbachia* strains belonging to supergroup A, including strains analyzed in the present study. Although the results of the two studies are largely in agreement, two differences deserve clarification. According to Riegler and colleagues, the WD0766-like ANK gene in the *wTei* strain is disrupted by an IS5 element, as is also the case for strains *wYak* and *wSan*. This observation is based on PCR amplicon size similarities between the three strains. Only the *wSan* gene copy was sequenced. Our sequence analysis clearly demonstrates that the WD0766-like ANK gene of *wTei* is identical

to that of *wAu*. Furthermore, the number of ankyrin repeat domains of the WD0550-like ANK gene of *wSan* differs between the two studies (eight ankyrin repeats in Riegler et al. vs six in the present study). While the first difference in WD0766 between the two studies was confirmed, it was raised that for WD0550 *wSan* was erroneously listed with 8 ankyrin repeat domains in Riegler et al. [70] when it only has 6 (personal communication M Riegler, I Iturbe-Ormaetxe, WJ Miller). There are two possible explanations for the discrepancy regarding WD0776. First, the presence of hidden multiple infections with different infection levels in the *Drosophila* stocks could account for these differences. As discussed above, there is evidence supporting the presence of more than one *Wolbachia* strain in *D. santomea*, as recently documented for the closely related species *D. teissieri* [94]. An alternative explanation could be different evolutionary events in the two stocks, which could account for the observed differences. Also, *Wolbachia* IS elements are quite active, and frequent IS5 polymorphisms have been documented for *wMel* strains [96], *wPip* strains [97] and across a range of different A-group *Wolbachia* strains [30], which could also play a role [30,96–98]. These hypotheses can be tested by further analyzing the WD0766-like ankyrin gene in *wTei* and related strains used in both studies.

Wolbachia Ankyrin Genes: Unknown Origin and Function

Earlier studies also highlighted the genetic diversity between *Wolbachia* ANK genes, but restricted themselves, however, to an attempt to correlate the observed genetic variability with different CI patterns [30,32]. According to our results, a correlation between *Wolbachia* phenotypes and distribution or genetic polymorphism of the different ANK genes is not obvious. Papafotiou et al. [31] showed that two ANK genes, WD0438 and WD1213, show higher expression levels in testes than in ovaries; however, the authors did not detect any evidence that this sex-specific expression is related to CI.

The origin of the *Wolbachia* ANK genes remains unclear. It has been suggested that prokaryotic ANK genes were acquired from a eukaryotic host rather than evolved independently [99]. However, the discovery of ANK genes in archaea and free-living bacterial species suggests a more ancient origin [19]. The largest numbers of ANK genes in prokaryotes were found in the genomes of *Coxiella burnetii*, *Legionella pneumophila*, *Rickettsia bellii*, *Rickettsia felis*, *Orientia tsutsugamushi* and sponge symbiotic bacteria residing within eukaryotic host cells [100–106]. The presence of a large number of ANK proteins within the genomes of these bacteria may be related to their unique lifestyle. This may also be the case for *Wolbachia*. Despite the fact that several studies investigated the direct or indirect association of the ANK genes with *Wolbachia*-induced reproductive phenotypes, mainly with cytoplasmic incompatibility, a functional correlation remains elusive [29–32,70]. The present study does not shed more light on this either. However, our results strongly indicate that phage transfer, homologous recombination and illegitimate recombination have provided *Wolbachia* with a unique repertoire of ANK genes. Their role for the lifestyle of *Wolbachia* remains to be established.

References

1. Saridakis A, Bourtzis K (2010) *Wolbachia*: more than just a bug in insects genitals. *Curr Opin Microbiol* 13: 67–72.
2. Werren JH, Baldo L, Clark ME (2008) *Wolbachia*: master manipulators of invertebrate biology. *Nat Rev Microbiol* 6: 741–751.
3. Cordaux R, Pichon S, Ben Afia Hatira H, Doublet V, Grève P, et al. (2012) Widespread *Wolbachia* infection in terrestrial isopods and other crustaceans. *Zookeys* 176: 123–131.

Supporting Information

Figure S1 Example of Southern blot analysis. Each membrane was hybridized simultaneously with two probes: a probe specific for the ANK gene under study and a probe specific for the *ftsZ* gene, which was used as a positive control. A) WD0441 and B) WD0294. 1: *wMel*, 2: *wMelPop*, 3: *wAu*, 4: *wTei*, 5: *wYak*, 6: *wSan*, 7: *wRi*, 8: *wHa*, 9: *wNo*, 10: *wMa*, 11: *wMau*. (TIFF)

Figure S2 Repetitive DNA sequences. Alignments of partial fragments of ANK genes present ankyrin repeat domain polymorphism (A) WD0385, (B) WD0514 and (C) WD0550. Gray rectangles show the position of repeated sites flanking the deletions. ANK repeats are underlined. (TIF)

Figure S3 Abundance of DRs within ANK genes. (TIF)

Figure S4 Phylogeny of the ANK genes. The trees are midpoint rooted and inferred using maximum likelihood. ML bootstrap support values inferred from 100 replicates are also presented. Bootstrap values lower than 50 are omitted. Evolutionary model parameters were estimated with Modeltest under the Akaike Information Criterion: K81uf (WD0292, WD0766); TIM (WD0073); HKY (WD0035, WD0291, WD0550); TrN+I (WD0438); TVM+I (WD0191); TVM+G (WD0385, WD0566, WD0633); GTR (WD0147). (TIF)

Figure S5 Local variation in nucleotide divergences within WD0441-like ANK gene sequences. A sliding window analysis of genetic distance between A- and B- supergroup strains indicates that the two supergroups share more similarities in the 3' end, which includes the ankyrin repeat domains. (TIFF)

Table S1 Primers used in this study. (DOCX)

Table S2 Distribution of *wMel*-like ANK genes in different *Drosophila-Wolbachia* associations. (DOCX)

Table S3 Comparison of *Wolbachia* ankyrin repeat genes. (XLSX)

Acknowledgments

The authors thank Dr. Stefan Oehler for his critical comments on an earlier version of the manuscript. The authors also thank two anonymous reviewers for valuable comments and suggestions which greatly improved the manuscript.

Author Contributions

Conceived and designed the experiments: KB HRB SS. Performed the experiments: SS PI. Analyzed the data: KB SS PI LK SA HRB. Contributed reagents/materials/analysis tools: KB HRB SA LK. Wrote the paper: KB SS LK SA HRB.

7. Doudoumis V, Alam U, Aksoy E, Abd-Alla AMM, Tsiamis G, et al. (2012) Tsetse-*Wolbachia* symbiosis: Comes of age and has great potential for pest and disease control. *J Invertebr Pathol*.
8. Foster J, Ganatra M, Kamal I, Ware J, Makarova K, et al. (2005) The *Wolbachia* genome of *Brugia malayi*: Endosymbiont evolution within a human pathogenic nematode. *PLoS Biol* 3: 599–614.
9. Klasson L, Walker T, Sebahia M, Sanders MJ, Quail MA, et al. (2008) Genome evolution of *Wolbachia* strain *wPip* from the *Culex pipiens* group. *Mol Biol Evol* 25: 1877–1887.
10. Klasson L, Westberg J, Sapountzis P, Näslund K, Lutnaes Y, et al. (2009) The mosaic genome structure of the *Wolbachia* *wRi* strain infecting *Drosophila simulans*. *Proc Natl Acad Sci U S A* 106: 5.
11. Salzberg SL, Hotopp JCD, Delcher AL, Pop M, Smith DR, et al. (2005) Serendipitous discovery of *Wolbachia* genomes in multiple *Drosophila* species. *Genome Biol* 6.
12. Salzberg SL, Puiu D, Sommer DD, Nene V, Lee NH (2009) Genome sequence of the *Wolbachia* endosymbiont of *Culex quinquefasciatus* JHB. *J Bacteriol* 191: 1725–1725.
13. Wu M, Sun LV, Vamathevan J, Riegler M, Deboy R, et al. (2004) Phylogenomics of the reproductive parasite *Wolbachia pipientis wMel*: A streamlined genome overrun by mobile genetic elements. *PLoS Biol* 2: 327–341.
14. Ishmael N, Dunning Hotopp JC, Ioannidis P, Biber S, Sakamoto J, et al. (2009) Extensive genomic diversity of closely related *Wolbachia* strains. *Microbiology* 106: 5725–5730.
15. Kent BN, Bordenstein SR (2010) Phage WO of *Wolbachia*: lambda of the endosymbiont world. *Trends Microbiol* 18: 173–181.
16. Leclercq S, Giraud I, Cordaux R (2011) Remarkable abundance and evolution of mobile group II Introns in *Wolbachia* bacterial endosymbionts. *Mol Biol Evol* 28: 685–697.
17. Mosavi LK, Cammett TJ, Desrosiers DC, Peng ZY (2004) The ankyrin repeat as molecular architecture for protein recognition. *Protein Science* 13: 1435–1448.
18. Sedgwick SG, Smerdon SJ (1999) The ankyrin repeat: a diversity of interactions on a common structural framework. *Trends Biochem Sci* 24: 311–316.
19. Al-Khodori S, Price CT, Kalia A, Abu Kwaik Y (2010) Functional diversity of ankyrin repeats in microbial proteins. *Trends Microbiol* 18: 132–139.
20. Bork P (1993) Hundreds of ankyrin-like repeats in functionally diverse proteins - mobile modules that cross phyla horizontally. *Proteins* 17: 363–374.
21. Habyarimana F, Al-Khodori S, Kalia A, Graham JE, Price CT, et al. (2008) Role for the Ankyrin eukaryotic-like genes of *Legionella pneumophila* in parasitism of protozoan hosts and human macrophages. *Environ Microbiol* 10: 1460–1474.
22. Pan X, Luhrmann A, Satoh A, Laskowski-Arce MA, Roy CR (2008) Ankyrin repeat proteins comprise a diverse family of bacterial Type IV effectors. *Science* 320: 1651–1654.
23. Lin MQ, den Dulk-Ras A, Hooykaas PJJ, Rikihisa Y (2007) *Anaplasma phagocytophilum* Anka secreted by type IV secretion system is tyrosine phosphorylated by Abl-1 to facilitate infection. *Cell Microbiol* 9: 2644–2657.
24. Park J, Kim KJ, Choi K, Grab DJ, Dumler JS (2004) *Anaplasma phagocytophilum* Anka binds to granulocyte DNA and nuclear proteins. *Cell Microbiol* 6: 743–751.
25. Rikihisa Y, Lin M (2010) *Anaplasma phagocytophilum* and *Ehrlichia chaffeensis* type IV secretion and Ank proteins. *Curr Opin Microbiol* 13: 59–66.
26. Caturegli P, Asanovich KM, Walls JJ, Bakken JS, Madigan JE, et al. (2000) *anka*: an *Ehrlichia phagocytophila* group gene encoding a cytoplasmic protein antigen with ankyrin repeats. *Infect Immun* 68: 5277–5283.
27. Garcia-Garcia JC, Rennoll-Bankert KE, Pelly S, Milstone AM, Dumler JS (2009) Silencing of host cell CYBB gene expression by the nuclear effector AnkaA of the intracellular pathogen *Anaplasma phagocytophilum*. *Infect Immun* 77: 2385–2391.
28. Ijdo JW, Carlson AC, Kennedy EL (2007) *Anaplasma phagocytophilum* AnkaA is tyrosine-phosphorylated at EPIYA motifs and recruits SHP-1 during early infection. *Cell Microbiol* 9: 1284–1296.
29. Duron O, Boureux A, Echaubard P, Berthomieu A, Berticat C, et al. (2007) Variability and expression of ankyrin domain genes in *Wolbachia* variants infecting the mosquito *Culex pipiens*. *J Bacteriol* 189: 4442–4448.
30. Iturbe-Ormaetxe I, Burke GR, Riegler M, O'Neill SL (2005) Distribution, expression, and motif variability of ankyrin domain genes in *Wolbachia pipientis*. *J Bacteriol* 187: 5136–5145.
31. Papafotiou G, Oehler S, Savakis C, Bourtzis K (2011) Regulation of *Wolbachia* ankyrin domain encoding genes in *Drosophila* gonads. *Res Microbiol* 162: 764–772.
32. Sinkins SP, Walker T, Lynd AR, Steven AR, Makepeace BL, et al. (2005) *Wolbachia* variability and host effects on crossing type in *Culex* mosquitoes. *Nature* 436: 257–260.
33. Walker T, Klasson L, Sebahia M, Sanders MJ, Thomson NR, et al. (2007) Ankyrin repeat domain-encoding genes in the *wPip* strain of *Wolbachia* from the *Culex pipiens* group. *BMC Biol* 5.
34. Yamada R, Iturbe-Ormaetxe I, Brownlie JC, O'Neill SL (2010) Functional test of the influence of *Wolbachia* genes on cytoplasmic incompatibility expression in *Drosophila melanogaster*. *Insect Mol Biol* 20: 75–85.
35. Bennuru S, Semnani R, Meng Z, Ribeiro JMC, Veenstra TD, et al. (2009) *Brugia malayi* excreted/secreted proteins at the host/parasite interface: stage- and gender-specific proteomic profiling. *PLoS Negl Trop Dis* 3: e410.
36. Félix C, Pichon S, Braquart-Varnier C, Braig H, Chen L, et al. (2008) Characterization and transcriptional analysis of two gene clusters for type IV secretion machinery in *Wolbachia* of *Armadillidium vulgare*. *Res Microbiol* 159: 481–485.
37. Bourtzis K, Dobson SL, Braig HR, O'Neill SL (1998) Rescuing *Wolbachia* have been overlooked. *Nature* 391: 852–853.
38. Bourtzis K, Nirgianaki A, Onyango P, Savakis C (1994) A prokaryotic *dnaA* sequence in *Drosophila melanogaster*: *Wolbachia* infection and cytoplasmic incompatibility among laboratory strains. *Insect Mol Biol* 3: 131–142.
39. Giordano R, O'Neill SL, Robertson HM (1995) *Wolbachia* infections and the expansion of cytoplasmic incompatibility in *Drosophila sechellia* and *D. Mauritiana*. *Genetics* 140: 1307–1317.
40. Hadfield SJ, Axton JM (1999) Germ cells colonized by endosymbiotic bacteria. *Nature* 402: 482.
41. Hoffmann AA, Turelli M, Simmons GM (1986) Unidirectional incompatibility between populations of *Drosophila simulans*. *Evolution* 40: 692–701.
42. Hoffmann AA, Clancy D, Duncan J (1996) Naturally-occurring *Wolbachia* infection in *Drosophila simulans* that does not cause cytoplasmic incompatibility. *Heredity* (Edinb) 76: 1–8.
43. McGraw EA, Merritt DJ, Droller JN, O'Neill SL (2001) *Wolbachia*-mediated sperm modification is dependent on the host genotype in *Drosophila*. *Proc R Soc Lond B Biol Sci* 268: 2565–2570.
44. Veneti Z, Clark ME, Karr TL, Savakis C, Bourtzis K (2004) Heads or tails: Host-parasite interactions in the *Drosophila-Wolbachia* system. *Appl Environ Microbiol* 70: 5366–5372.
45. Zabalou S, Charlat S, Nirgianaki A, Lachaise D, Mercot H, et al. (2004) Natural *Wolbachia* infections in the *Drosophila yakuba* species complex do not induce cytoplasmic incompatibility but fully rescue the *wRi* modification. *Genetics* 167: 827–834.
46. Baldo L, Hotopp JCD, Jolley KA, Bordenstein SR, Biber SA, et al. (2006) Multilocus sequence typing system for the endosymbiont *Wolbachia pipientis*. *Appl Environ Microbiol* 72: 7098–7110.
47. 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.
48. Zabalou S, Apostolaki A, Pattas S, Veneti Z, Paraskevopoulos C, et al. (2008) Multiple rescue factors within a *Wolbachia* strain. *Genetics* 178: 2145–2160.
49. Holmes DS, Bonner J (1973) Preparation, molecular-weight, base composition, and secondary structure of giant nuclear ribonucleic-acid. *Biochemistry* 12: 2330–2338.
50. Sambrook SL FE, Maniatis T (1989) *Molecular Cloning - A Laboratory Manual* (2nd. Ed.): Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
51. Holden PR, Brookfield JFY, Jones P (1993) Cloning and characterization of an *ftsZ* homolog from a bacterial symbiont of *Drosophila melanogaster*. *Mol Gen Genet* 240: 213–220.
52. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32: 1792–1797.
53. Drummond AJ, Cheung M, Heled J, Kearse M, Moir R, Stones-Havas S, Thierer T, Wilson A (2008) Geneious v4.0. Available from: <http://www.geneious.com>.
54. Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19: 2496–2497.
55. Yang Z (2007) PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol Biol Evol* 24: 1586–1591.
56. Goldman N, Yang ZH (1994) Codon-based model of nucleotide substitution for protein-coding DNA-sequences. *Mol Biol Evol* 11: 725–736.
57. Finn RD, Tate J, Mistry J, Coghill PC, Sammut SJ, et al. (2008) The Pfam protein families database. *Nucleic Acids Res* 36: D281–D288.
58. Schultz J, Milpetz F, Bork P, Ponting CP (1998) SMART, a simple modular architecture research tool: Identification of signaling domains. *Proc Natl Acad Sci U S A* 95: 5857–5864.
59. Betley JN, Frith MC, Graber JH, Choo S, Deshler JO (2002) A ubiquitous and conserved signal for RNA localization in chordates. *Curr Biol* 12: 1756–1761.
60. Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14: 817–818.
61. Swofford DL (2002) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4: Sinauer Associates, Sunderland, Massachusetts.
62. Shimodaira H, Hasegawa M (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol Biol Evol* 16: 1114–1116.
63. Bordenstein SR, Wernegreen JJ (2004) Bacteriophage flux in endosymbionts (*Wolbachia*): Infection frequency, lateral transfer, and recombination rates. *Mol Biol Evol* 21: 1981–1991.
64. Martin DP, Williamson C, Posada D (2005) RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics* 21: 260–262.
65. Posada D, Crandall KA (2001) Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. *Proc Natl Acad Sci U S A* 98: 13757–13762.
66. Smith JM (1992) Analyzing the mosaic structure of genes. *J Mol Evol* 34: 126–129.

67. Padidam M, Sawyer S, Fauquet CM (1999) Possible emergence of new geminiviruses by frequent recombination. *Virology* 265: 218–225.
68. Martin D, Rybicki E (2000) RDP: detection of recombination amongst aligned sequences. *Bioinformatics* 16: 562–563.
69. Salminen MO, Carr JK, Burke DS, McCutchan FE (1995) Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. *AIDS Res Hum Retroviruses* 11: 1423–1425.
70. Riegler M, Iturbe-Ormaetxe I, Woolfit M, Miller W, O'Neill S (2012) Tandem repeat markers as novel diagnostic tools for high resolution fingerprinting of *Wolbachia*. *BMC Microbiol* 12: S12.
71. Kobe B, Kajava AV (2000) When protein folding is simplified to protein coiling: the continuum of solenoid protein structures. *Trends Biochem Sci* 25: 509–515.
72. Marcotte EM, Pellegrini M, Yeates TO, Eisenberg D (1998) A census of protein repeats. *J Mol Biol* 293: 151–160.
73. McGraw EA, O'Neill SL (2004) *Wolbachia pipiensis*: intracellular infection and pathogenesis in *Drosophila*. *Curr Opin Microbiol* 7: 67–70.
74. Baldo L, Bordenstein S, Wernegreen JJ, Werren JH (2006) Widespread recombination throughout *Wolbachia* genomes. *Mol Biol Evol* 23: 437–449.
75. Baldo L, Lo N, Werren JH (2005) Mosaic nature of the *Wolbachia* surface protein. *J Bacteriol* 187: 5406–5418.
76. Ioannidis P, Hotopp JCD, Sapountzis P, Siozios S, Tsiamis G, et al. (2007) New criteria for selecting the origin of DNA replication in *Wolbachia* and closely related bacteria. *BMC Genomics* 8.
77. Jiggins FM (2002) The rate of recombination in *Wolbachia* bacteria. *Mol Biol Evol* 19: 1640–1643.
78. Posada D, Crandall KA, Holmes EC (2002) Recombination in evolutionary genomics. *Annu Rev Genet* 36: 75–97.
79. Werren JH, Bartos JD (2001) Recombination in *Wolbachia*. *Curr Biol* 11: 431–435.
80. Bzymek M, Lovett ST (2001) Instability of repetitive DNA sequences: The role of replication in multiple mechanisms. *Proc Natl Acad Sci U S A* 98: 8319–8325.
81. Rocha EPC (2003) An appraisal of the potential for illegitimate recombination in bacterial genomes and its consequences: From duplications to genome reduction. *Genome Res* 13: 1123–1132.
82. Amiri H, Davids W, Andersson SGE (2003) Birth and death of orphan genes in *Rickettsia*. *Mol Biol Evol* 20: 1575–1587.
83. Rocha EPC, Matic I, Taddei F (2002) Over-representation of repeats in stress response genes: a strategy to increase versatility under stressful conditions? *Nucleic Acids Res* 30: 1886–1894.
84. Werden SJ, McFadden G (2008) The role of cell signaling in poxvirus tropism: The case of the M-T5 host range protein of myxoma virus. *Biochim Biophys Acta* 1784: 228–237.
85. Cai X, Zhang Y (2006) Molecular evolution of the ankyrin gene family. *Mol Biol Evol*.
86. Tripp KW, Barrick D (2004) The tolerance of a modular protein to duplication and deletion of internal repeats. *J Mol Biol* 344: 169–178.
87. Casjens S (2003) Prophages and bacterial genomics: what have we learned so far? *Mol Microbiol* 49: 277–300.
88. Moran NA, Degnan PH, Santos SR, Dunbar HE, Ochman H (2005) The players in a mutualistic symbiosis: Insects, bacteria, viruses, and virulence genes. *Proc Natl Acad Sci U S A* 102: 16919–16926.
89. Wagner PL, Waldor MK (2002) Bacteriophage control of bacterial virulence. *Infect Immun* 70: 3985–3993.
90. Chafee ME, Funk DJ, Harrison RG, Bordenstein SR (2010) Lateral phage transfer in obligate intracellular bacteria (*Wolbachia*): verification from natural populations. *Mol Biol Evol* 27: 501–505.
91. Gavotte L, Henri H, Stouthamer R, Charif D, Charlat S, et al. (2006) A survey of the bacteriophage WO in the endosymbiotic bacteria *Wolbachia*. *Mol Biol Evol* 24: 427–435.
92. 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.
93. Metcalf JA, Bordenstein SR (2012) The complexity of virus systems: the case of endosymbionts. *Curr Opin Microbiol* 15: 546–552.
94. Saridaki A, Sapountzis P, Harris HL, Batista PD, Biliske JA, et al. (2011) *Wolbachia* prophage DNA adenine methyltransferase genes in different *Drosophila-Wolbachia* associations. *PLoS One* 6: e19708.
95. Merçot H, Llorente B, Jacques M, Atlan A, Montchamp-Moreau C (1995) Variability within the Seychelles cytoplasmic incompatibility system in *Drosophila simulans*. *Genetics* 141: 1015–1023.
96. Riegler M, Sidhu M, Miller WJ, O'Neill SL (2005) Evidence for a global *Wolbachia* replacement in *Drosophila melanogaster*. *Curr Biol* : CB 15: 1428–1433.
97. Duron O, Lagnel J, Raymond M, Bourtzis K, Fort P, et al. (2005) Transposable element polymorphism of *Wolbachia* in the mosquito *Culex pipiens*: evidence of genetic diversity, superinfection and recombination. *Mol Ecol* 14: 1561–1573.
98. Cordaux R, Pichon S, Ling A, Pérez P, Delaunay C, et al. (2008) Intense transpositional activity of insertion sequences in an ancient obligate endosymbiont. *Mol Biol Evol* 25: 1889–1896.
99. Ponting CP, Aravind L, Schultz J, Bork P, Koonin EV (1999) Eukaryotic signalling domain homologues in archaea and bacteria. Ancient ancestry and horizontal gene transfer. *J Mol Biol* 289: 729–745.
100. Cazalet C, Rusniok C, Bruggemann H, Zidane N, Magnier A, et al. (2004) Evidence in the *Legionella pneumophila* genome for exploitation of host cell functions and high genome plasticity. *Nat Genet* 36: 1165–1173.
101. Chien MC, Morozova I, Shi SD, Sheng HT, Chen J, et al. (2004) The genomic sequence of the accidental pathogen *Legionella pneumophila*. *Science* 305: 1966–1968.
102. Nakayama K, Yamashita A, Kurokawa K, Morimoto T, Ogawa M, et al. (2008) The whole-genome sequencing of the obligate intracellular bacterium *Orientia tsutsugamushi* revealed massive gene amplification during reductive genome evolution. *DNA Res* 15: 185–199.
103. Ogata H, La Scola B, Audic S, Renesto P, Blanc G, et al. (2006) Genome sequence of *Rickettsia bellii* illuminates the role of amoebae in gene exchanges between intracellular pathogens. *PLoS Genet* 2: 733–744.
104. Ogata H, Renesto P, Audic S, Robert C, Blanc G, et al. (2005) The genome sequence of *Rickettsia felis* identifies the first putative conjugative plasmid in an obligate intracellular parasite. *PLoS Biol* 3: 1391–1402.
105. Seshadri R, Paulsen IT, Eisen JA, Read TD, Nelson KE, et al. (2003) Complete genome sequence of the Q-fever pathogen *Coxiella burnetii*. *Proc Natl Acad Sci U S A* 100: 5455–5460.
106. Thomas T, Rusch D, DeMaere MZ, Yung PY, Lewis M, et al. (2010) Functional genomic signatures of sponge bacteria reveal unique and shared features of symbiosis. *ISME J* 4: 1557–1567.