Tandem repeat markers as novel diagnostic tools for high resolution fingerprinting of *Wolbachia*

Markus Riegler1†, Iñaki Iturbe-Ormaetxe2,4†, Megan Woolfit2,4, Wolfgang J Miller3*, Scott L O'Neill2,4*

**Abstract**

Background: Strains of the endosymbiotic bacterium *Wolbachia pipientis* are extremely diverse both genotypically and in terms of their induced phenotypes in invertebrate hosts. Despite extensive molecular characterisation of *Wolbachia* diversity, little is known about the actual genomic diversity within or between closely related strains that group tightly on the basis of existing gene marker systems, including Multiple Locus Sequence Typing (MLST). There is an urgent need for higher resolution fingerprinting markers of *Wolbachia* for studies of population genetics, horizontal transmission and experimental evolution.

Results: The genome of the *wMel* *Wolbachia* strain that infects *Drosophila melanogaster* contains inter- and intragenic tandem repeats that may evolve through expansion or contraction. We identified hypervariable regions in *wMel*, including intergenic Variable Number Tandem Repeats (VNTRs), and genes encoding ankyrin (ANK) repeat domains. We amplified these markers from 14 related *Wolbachia* strains belonging to supergroup A and were successful in differentiating size polymorphic alleles. Because of their tandemly repeated structure and length polymorphism, the markers can be used in a PCR-diagnostic multilocus typing approach, analogous to the Multiple Locus VNTR Analysis (MLVA) established for many other bacteria and organisms. The isolated markers are highly specific for supergroup A and not informative for other supergroups. However, *in silico* analysis of completed genomes from other supergroups revealed the presence of tandem repeats that are variable and could therefore be useful for typing target strains.

Conclusions: *Wolbachia* genomes contain inter- and intragenic tandem repeats that evolve through expansion or contraction. A selection of polymorphic tandem repeats is a novel and useful PCR diagnostic extension to the existing MLST typing system of *Wolbachia*, as it allows rapid and inexpensive high-throughput fingerprinting of closely related strains for which polymorphic markers were previously lacking.

**Background**

*Wolbachia pipientis* (*α*-Proteobacteria) is an obligate endosymbionts of invertebrates, known to infect up to 70% of insect species, as well as spiders, terrestrial crustaceans and medically important filarial nematodes [1-5]. Many strains of *Wolbachia* found in insects manipulate their hosts by inducing feminisation, parthenogenesis, male killing or cytoplasmic incompatibility (CI) [6-9]; in contrast, the *Wolbachia* of nematodes are mutualists necessary for host reproduction [10]. Despite this great diversity of hosts and extended phenotypes, all strains of *Wolbachia* are currently recognised as the single species *W. pipientis*. Within this species, strains are clustered into at least eight divergent clades or ‘supergroups’, named A to K [11-15].

Several genes have been used for strain typing in *Wolbachia*. Initially, work focused on 16S rDNA[16], the genes encoding the cell division protein, *ftsZ* [11] and the *Wolbachia* surface protein, *wsp* [12]. Subsequent to the demonstration of widespread intra- and intergenic recombination between strains [17-19], two multi-locus sequence typing (MLST) systems were developed using different sets of a total of 14 *Wolbachia* genes [20,21]. The MLST approach uses partial nucleotide sequences of several ubiquitous loci with moderate rates of

---

† Contributed equally

2School of Biological Sciences, The University of Queensland, QLD 4072, Australia

3Center of Anatomy and Cell Biology, Medical University of Vienna, Währingerstr. 10, 1090 Vienna, Austria

Full list of author information is available at the end of the article

© 2012 Riegler et al; licensee BioMed Central Ltd. This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
evolution to generate an allelic profile for tested strains. These profiles can be used to type novel isolates, while the relationships between strains may be inferred on the basis of either the allelic profiles themselves or the nucleotide sequences underlying them. MLST data have been used for both strain typing and evolutionary analyses of horizontal transfer events between host species of Wolbachia (e.g. [22,23]). Since most MLST primer sets cover housekeeping genes that are under purifying selection, these markers often cannot differentiate between closely related strains. Such difficulties have been revealed in the comparisons between wMel, wMelCS and wMelPop [20] or wMel and wAu within the ST-13 complex which appear indistinguishable in MLST loci [21,24]. These strains induce different phenotypes in their hosts, i.e. wMel induces CI in Drosophila, but wAu does not [25] and wMelPop induces lifespan reduction in its hosts but not wMel [26-28]. The divergence between MLST typing and actual genomic diversity within ST-13 was also raised when these closely related strains were compared for presence or absence of Wolbachia prophage WO-A and WO-B [24] and other genomic differences such as a large chromosomal inversion and differential IS3 insertion sites between wMel, wMelPop and wMelCS [29,30]. Furthermore, MLST can be time consuming and expensive for large population genetic studies as it requires sequencing of all MLST loci for many individuals. Recently other typing systems have been developed for bacteria that build on markers that contain Variable Number Tandem Repeats (VNTR). VNTRs consist of units of DNA (periods) that are tandemly repeated and vary in copy number between different isolates. These loci can be used for a PCR-based typing system and are increasingly being utilised in bacterial strain typing such as Multi Locus VNTR Analysis (MLVA) (e.g. [31-35]). MLVA offers a number of advantages, including highly polymorphic markers that allow fine-scale typing of very closely related isolates, rapid, high-throughput screening that is not dependent on sequencing, and potentially the fingerprinting of multiply infected hosts. The modular structure and evolution of these sites through tandem expansion and contraction also allows cladistic and phylogenetic inference.

Amplon size polymorphic markers have previously been identified in Wolbachia genomes and include transposable element insertion sites [30,36,37], VNTRs [30,38-40] and genes encoding ankyrin repeat domains [36], but their efficiency for strain typing has not yet been compared. In this paper, we used some of these markers in order to estimate the feasibility of a MLVA system for Wolbachia. We isolated markers with tandem repeats from the wMel genome [41] and applied them to a number of Wolbachia strains from supergroups A, B and C to assess their applicability and resolution for Wolbachia strain typing. We chose two types of loci containing tandem repeats, two intergenic VNTR loci and two genes encoding proteins containing ankyrin repeats. The two VNTR loci, VNTR-105 and VNTR-141 were originally isolated from supergroup A strain wMel and were polymorphic between wMel, wMelCS and wMelPop isolates from different D. melanogaster lines [30]. VNTRs are also polymorphic between the closely related wAu from D. simulans and wWil from Drosophila willistoni [38], and serve as highly diagnostic marker sets for fingerprinting conspecific Wolbachia strains in the Drosophila paulistorum species cluster [39]. Recently, a polymorphic VNTR locus was isolated from supergroup B strain wPip [40]. Ankyrin repeat genes are abundant in the genomes of Wolbachia and a number of other intracellular bacteria [42,43]. The number and distribution of these repeats varies substantially between strains that induce different host phenotypes, suggesting that they may be involved in host manipulation [36]. We extended our analysis to include a wider range of Wolbachia strains from supergroup A, B and C in order to evaluate the usefulness of the four markers VNTR-105, VNTR-141, WD0550 and WD0766, originally isolated from wMel, in discriminating between Wolbachia strains.

**Methods**

### Wolbachia strains and hosts

We used 14 supergroup A Wolbachia isolates from 8 different Drosophila species and 2 tephritid species, Rhagoletis cerasi, a host that is naturally infected, and Ceratitis capitata, microinjected with Wolbachia originating from R. cerasi (Table 1). Based on previous strain typing using 16S rRNA, ftsZ, wsp and some MLST loci, these 14 strains are moderately or closely related, yet they reveal different phenotypic characteristics, such as varying levels of CI induction (strong, weak, or non-CI inducers), and different CI rescue phenotypes (reviewed in [44]). Wolbachia DNA was isolated from Drosophila fly stocks reared on standard corn-flour-sugar-yeast medium at 25°C. Wolbachia-free controls D. melanogaster yw^{67c23}T and D. simulans Riverside-DSRT were established by tetracycline treatment using standard techniques [45]. Wolbachia of R. cerasi was isolated from field collected samples from Austria and Hungary [46]. Wolbachia from C. capitata was isolated from the WolMed 88.6 lab line that was artificially infected with wCer2 from R. cerasi [47]. We also included strains from B (wNo, wBo1, wMau) and C (wDim) supergroups. wNo and wMau were isolated from D. simulans, wBo1 from Hypolimnas bolina [48] and wDim from dog heart worm Dirofilaria immitis [49].
**DNA extraction, PCR amplification and sequencing of molecular markers**

Total genomic DNA was extracted from either freshly collected specimens or specimens stored in pure ethanol in a -20°C freezer. Extraction was carried out on pools of *Drosophila* flies and single individuals of *Rhagoletis, Ceratitis, Hypolimnas* and *Dirofilaria*. Flies were homogenized and extracted following either the Holmes-Bennon protocol [50] or the STE extraction method [16]. *Wolbachia* markers were amplified from total genomic DNA using specific primers (Table 2). The *wsp* gene was used as a quality control for DNA extraction and was amplified using the primers 81F and 691R, described in [12]. PCR cycling conditions were as follows: 94°C 3 min, (94°C 30 s, 50°C 30 s, 72°C 3 min) x 35 cycles, then 72°C 10 min. The reaction mixture contained 500 nM of each primer, 200 µM dNTPs, 1.5 mM MgCl₂, 100 ng of DNA and 1 unit of Taq Polymerase in a final volume of 20 µl. The reaction buffer contained 10 mM Tris pH 9.0, 50 mM KCl and 0.1% Triton X-100. PCR products were separated in 1% agarose gels, stained with ethidium bromide and gel-purified using gel extraction kits (QIAGEN). Purified DNA was cloned into the pGEM®-T-easy plasmid (Promega) and sequenced by Macrogen, in Korea, using T7, M13R, and M13F primers.

Table 2 List of primers designed according to the *wMel* genome sequence to amplify VNTRs and ANK genes.

| Locus/primer | 5’ sequence | Reference |
|--------------|-------------|-----------|
| VNTR-141 for | ggaattgattgtgtgctg | [30] |
| VNTR-141 rev | gacctagggtctgtct | [30] |
| VNTR-105 for | gcaagagatattgtgctc | [30] |
| VNTR-105 rev | atgacacctacttacgc | [30] |
| RO550F | ggctcctctctctctctct | [82] |
| RO550R | gatctctctctctctctct | [82] |
| RO766F | gcctcctctctctctctct | [82] |
| RO766R | tcttctctctctctctct | [82] |

**Selection of size variable markers**

Polyomviric loci were previously identified from the sequenced genome of *wMel* of *D. melanogaster* ([41], GenBank reference sequence NC_002978) *in silico* by using Tandem Repeats Finder TRF (http://tandem.bu.edu/trf/trf.html) [51]. Two VNTR regions of interest, VNTR-105 and VNTR-141 were found to be
polymorphic between different lines of *D. melanogaster* [30]. The TRF analysis also detected more candidate loci, including some genes encoding ANK domain repeats that can also contain tandemly repeated DNA, and are hence candidate markers for MLVA. Genes encoding ANK domain repeats were previously annotated [41] and variability was found in supergroup A and B *Wolbachia* strains [36]. All of the tandem repeats analysed here were amplified by using primers designed for the conserved flanking regions (single copy coding genes) of the repeats within *wMel*. We further extended the TRF analysis to other completed *Wolbachia* genomes, *wRi* ([52] NC_012416), *w Pip* ([53] NC_010981) and *wBm* ([54] NC_006833) in order to highlight the potential of MLVA for more distantly related *Wolbachia* strains *in silico*. The TRF analysis also included the genomes of *Anaplasma marginalia* strain St. Maries (CP_000030) and *Ehrlichia ruminantium* strain Welgevonden (NC_005295) and *Neorickettsia risticii* strain Illinois (NC_013009), the closest relatives of the genus *Wolbachia* [55], as well as a comparison with free living *Escherichia coli* K12 strain MG1655 (NC_000913). The bacterial genomes were analysed in the basic mode of TRF (version 4.04), with alignment parameters for match, mismatch and indels set at 2, 7 and 7, respectively. The minimum alignment score to report repeats was set at 50, with a maximum period size of 500bp (Table 4).

Sequence analysis

The analysis and assembly of the sequences was done using the *EditSeq*, *SeqMan* and *MegAlign* components of the Lasergene sequence analysis software package (DNASTar Inc., Madison, Wis.). The sequenced VNTR loci of the *Wolbachia* strains had to be manually aligned because of their long period length, internal repeats, SNPs and indels within individual VNTR periods. VNTR periods were searched for internal direct repeats, palindromic (dyad) repeats and secondary structures by using DNA Strider [56]. For ANK proteins, domain architecture was predicted using SMART v3.5 (Simple Modular Architecture Research Tool) (http://smart.embl-heidelberg.de/) [57,58] and TMHMM2 (http://www.cbs.dtu.dk/services/TMHMM/). We analysed the phylogenetic relationships between individual ANK repeats from *WD0766* and their orthologs to investigate the mode of evolution of these repeats. All ANK repeats were extracted from the full length sequences of each gene and translated into amino acids. Gaps were inserted where necessary to correct for frameshifts. Sequences were aligned using T_coffee [59]. Maximum likelihood phylogenetic analysis of this alignment was performed using PhyML [60], with a JTT model of amino acid substitution, and a gamma model of rate heterogeneity with four rate classes and the gamma parameter estimated from the data. 1000 bootstrap replicates were performed.

Results and discussion

**VNTR variability between strains of A-group Wolbachia**

We isolated sequences for two *Wolbachia* VNTR loci, VNTR-141 and VNTR-105, with tandemly repeated periods of 141 and 105bp, respectively, for representative supergroup A *Wolbachia* strains. The loci had previously produced size polymorphic PCR fragments in isolates of *wMel* and *wMelCS/wMelPop* when amplified using primers that were designed to the flanking regions of the two VNTR loci of the sequenced *wMel* genome [30]. VNTR-141 is positioned between *WD0096* and *WD0098*, and VNTR-105 is between *WD1129* and *WD1131* of the final *wMel* genome annotation (NCBI accession NC_002978, [41]). The basic 141bp period of VNTR-141 consists of the internal 15bp direct repeat A, a 23bp hairpin with a 9bp palindromic stem, an 18bp insertional and the internal 15bp direct repeat B (Figure 1 of this paper, and Figure 2E of [38]). Diagnostic VNTR-141 PCRs were run on DNA obtained from different *Wolbachia* hosts known to harbour very closely related strains of the symbiont that were not clearly distinguishable by using MLST [20,21,24]. The VNTR-141 fragments were sequenced and compared to the 141bp period of *wMel*. The shortest VNTR-141 alleles were amplified from *wWil* and *wCer1*: they contained only one single period consisting of a 108bp core period without the 18bp insertion, and missing the downstream 15bp A repeat. All other supergroup A strains produced VNTR-141 alleles containing different copy numbers of the 141bp period (Figure 1), i.e. 0.8 (*wWil*, amplicon size using the locus specific primers 387bp, *wCer1*...
388bp), 1.7 (wAu 530bp), 2.3 (wSpt 643bp), 4.3 (wSan 889bp, wPro 925bp), wYak and wTei had similar amplicon sizes to wSan but were not sequenced), 6.3 (wMelCS 1189bp, wMelPop 1189bp) and 7.3 (wMel 1330bp, wCer2 1348bp) for both original host R. cerasi and novel host C. capitata (Figure 1). These polymorphic amplicons in VNTR-141 were visualised by standard PCR as different amplicon sizes on an agarose gel (Figure 2). Multiply infected R. cerasi [46,61] revealed two bands, with amplicons representing wCer1 and wCer2 (Figure 2). The VNTR alleles of wCer2 were assigned through comparisons with the isolates from the microinjected novel hosts D. simulans [62] and C. capitata [47]. Besides the internal deletions in the wWil and wCer1 periods, and variation in copy numbers, the sequence composition of the VNTR-141 periods are almost identical (i.e. 99%) within wMel and other strains, and hence highly conserved. For this reason a phylogenetic sequence analysis, other than the analysis of repeat numbers in cladistical approaches, is not informative.

In contrast to VNTR-141, the basic period of VNTR-105 is 105bp long containing two 23bp hairpins with 9bp palindromic stem structures and one internal short repeat of 10bp (Figure 3). VNTR-105 of wMel contains four complete 105bp periods, and two with internal deletions of 25bp each. wMelCS and wMelPop lack one of the complete 105bp periods, i.e. contain three complete 105bp copies and two with internal deletions of 32bp (Figure 3). The tested supergroup A strains display different alleles in the VNTR-105 locus with amplicon sizes ranging from 3x0.5 copies (wCer1, amplicon size using the locus specific primers 998bp), 2.5 copies (wWil 1065bp, wAu 1065bp), 3+2x0.5 copies (wMelCS and wMelPop 1241bp), 4x2x0.5 copies (wMel 1347bp), 3+4x0.5 copies (wSpt 1408bp) and 5+2x0.5 copies (wSan, 1476bp; wYak and wTei had similar amplicon sizes to wSan but were not sequenced). wCer2 had a large amplicon for this VNTR locus and difficulties were experienced with accurately sequencing these large loci because of restrictions with read lengths, as well as problems in detecting an accurate overlap between forward and reverse sequences. VNTR-105 amplicon size differences can be easily resolved on agarose gels (data not shown). In comparison to VNTR-141, the structure of the VNTR-105 locus is less conserved within and between strains because of internal deletions, yet the period sequences are almost identical (i.e. 98%) within wMel and between other strains. For this reason a phylogenetic analysis of period sequence data is not appropriate, whereas the analysis of diagnostic characters such as copy numbers are more informative (Figure 3).

We extended our PCR analysis to a wider range of Wolbachia strains, including wRi and wHa, both supergroup A strains that are distantly related to wMel, as well as strains from supergroup B (wNo, wBo1, wMau) and C (wDim). None of these strains yielded PCR products for the tested VNTR primers, probably because of sequence divergence within the primer region or genome rearrangements [52-54]. Because of the latter it was not attempted to design primers of conserved coding regions in distantly related strains.

### Evolution of repeats in VNTR loci

The individual periods of VNTR-141 and VNTR-105 respectively display high sequence conservation within and between strains, with variability in the copy numbers and internal deletions within some of the repeated periods. Two evolutionary processes may be shaping these loci with high variability in repeat copy numbers yet small sequence divergence. The accumulation of tandemly repeated periods may be facilitated through slippage and mispairing in the process of Wolbachia DNA replication and repair. Slipped-strand mispairing has previously been identified as a source for generation of repeat copies in general [63-65] and in E. ruminantium.
in particular, a genome with an elevated number of tandem repeats [66]. Palindromic sequences with the strong potential of forming secondary stem loops are well known to cause slipped-strand mispairing [67]. Hence we assume that the hairpins present in both Wolbachia VNTRs may trigger slippage in both these loci. The second evolutionary mechanism in action could be concerted evolution between different periods within the two loci, a phenomenon that has previously been observed in members of gene families that tend to be more similar within a species than between species because of the elimination or fixation of new point mutations [68]. The high structural turnover, triggering expansions and/or contractions of copy numbers in
both VNTR loci of wMel-like Wolbachia, can thus be applied for simple and rapid but highly informative symbiont fingerprinting by standard PCR (Figure 2). We cannot infer directionality between expansion and contractions in the evolution of both loci. It is hence impossible to determine whether low copy numbers within the intergenic loci manifest an ancestral or derived state. It has been suggested though that tandem repeats go through cycles of gradual expansion followed by collapse [69]. It is hence adequate to state that closely related strains are more likely to have similar copy numbers, e.g. wMel and wMelCS. Interestingly, the CI inducing strains wCer2, wMel and wMelCS contain larger VNTR loci when compared to the non CI inducing wWil and wAu, with larger VNTR loci in wMel than wMelCS that coincide with stronger CI induction in wMel than wMelCS [70]. Furthermore increased copy numbers in one locus correspond with increased copy numbers in the second. Such a coincidence of intergenic tandem repeat variation with CI phenotype was also observed for supergroup B Wolbachia in C. pipiens[40]. Yet, these observations are not sufficiently supported by replication to conclude about any potential links between genotypes and phenotypes, but they warrant further structural and functional studies of the VNTR repeat expansions.

ANK gene variability between strains of A-group Wolbachia

Unlike most bacteria, genes that encode proteins with ANK repeats are extremely abundant in Wolbachia, representing up to 2-4% of the total number of genes in wMel [41], wRi [52] and wPip [53,71]. Some of the variability in these genes appears to correlate with crossing types in mosquitoes [72]. Several of the 23 ANK genes initially annotated in the wMel genome are highly variable between the CI-inducing strain wMel and the non-CI inducing related strain wAu [36]. These differences included point mutations, frameshifts and premature stop codons, presence/absence of transmembrane domains, disruption by insertion elements and variability in the number of predicted ANK repeats in the encoded proteins. Based on earlier work [36], we performed an initial PCR screening (data not shown) using the most variable wMel ANK genes (WD0035, WD0294, WD0385, WD0498, WD0514, WD0550, WD0636, WD0766 and WD1213- also see results of TRF analysis below) in order to look for size differences across the Wolbachia strains used in this study. Some of the ANK genes could not be amplified in all strains, probably due to sequence divergence. For the ones that could be amplified, the non-phage related ANK genes WD0550 and in

Figure 2 Diagnostic size difference for the VNTR-141 locus of Wolbachia. Lane 1: wCer1 and wCer2 doubly infected R. cerasi from Austria (the two arrows indicate the two faint bands for wCer1 and wCer2); 2-4: wWil infected D. willistoni from populations collected recently in Panama (Pan98), Mexico (Apa), and Ecuador (JS); lane 5-6: wAu infected D. simulans strain Coffs Harbor and Yaunde 6; lane 7: uninfected (tetracycline treated) controls = D. melanogaster yw^P{ Cyc1}; lane 8: wIei infected D. teissieri GNS3, lane 9: wMel infected D. melanogaster yw^P{ Cyc1}, lane 10: wSpt infected D. septentriocollarsus; lane 11: wCer1 singly infected R. cerasi from Hungary; lane 12: uninfected (tetracycline treated) control = D. melanogaster line yw^P{ Cyc1}; lane 13: wMel infected D. melanogaster yw^P{ Cyc1}; lane 14: wMelCS infected D. melanogaster Canton S.
particular WD0766 were found to be the most variable in terms of size difference among the Wolbachia strains and they were selected for further analysis, with sequence data reported for WD0766 only.

In wMel, WD0766 encodes a 51.8kDa protein containing eight ANK repeats and two transmembrane domains (TMDs) in the C-terminus. When this gene was sequenced in several Wolbachia strains, the number of predicted ANK repeats was found to be quite different among them, ranging from eight repeats in wMel to 14 in wCer1 (Figure 4). The wAu, wWil and wRi strains contained 11 ANK repeats, but the proteins were truncated by a premature stop codon that resulted in the elimination of the predicted TMDs in wAu and wWil. WD0766 in wSan is disrupted by a premature stop after the seventh ANK domain and
contains a 918bp IS5 insertion element in the middle of its 10th ANK repeat (Figure 4). PCR results (data not shown) suggest that this IS5 insertion is also present in the orthologous gene in \( w_{\text{Yak}} \) and \( w_{\text{Tei}} \), but these amplicons were not sequenced. The sequence of the \( w_{\text{San}} \) IS5 element is identical to that of the 13 IS5 elements present in the \( w_{\text{Mel}} \) genome \[41\]. Disruption of a \( \text{Wolbachia} \) ANK gene by an IS5 insertion element has previously been observed in the WD0385 gene from \( w_{\text{Au}} \) (GenBank AY664873) \[36\], although in this case the insertion sequence differs by 5 nucleotides from the \( w_{\text{Mel}} \) and \( w_{\text{San}} \) IS5 elements. \( w_{\text{Spt}} \), \( w_{\text{Cer2}} \) and \( w_{\text{Ha}} \) strains had the same structure for the WD0766 proteins (13 ANK domains + 2 TMDs), whereas the \( w_{\text{Cer1}} \) protein contained 14 ANK domains and 2 TMDs.

\[WD0550\] was also found to be variable among the strains analysed, although it was not as informative as \( WD0766 \). For this reason only a subset of strains was analysed for this locus in more detail. \( WD0550 \) codes for a 36.4kDa protein containing six predicted ANK repeats and has no TMDs. The protein contains six ANK repeats in \( w_{\text{Mel}} \) and \( w_{\text{Spt}} \), and eight repeats in \( w_{\text{Mel-CS}}, w_{\text{San}}, w_{\text{Cer2}}, w_{\text{Au}} \) and \( w_{\text{Wil}} \) (data not shown).

**Evolution of repeats in WD0766**

Orthologs of \( WD0766 \) encode for proteins containing different numbers of ANK repeats in different \( \text{Wolbachia} \) strains. Additional repeat copies may be gained by the duplication or loss of single or multiple repeats, and genes containing these repeats may also diverge due to
loss or shuffling of repeat periods. To investigate the patterns of change in the number and order of ANK repeats in these proteins, we aligned the amino acid sequences of all individual repeats and performed a maximum likelihood analysis of the phylogenetic relationships between them (Figure 5). The tree shows clusters of typically six to ten repeats, separated by relatively long internal branches. Despite the large ratio of internal to tip branch lengths, bootstrap values on this tree are almost all extremely small, probably due to the short length of the alignment (34 residues). However, a clear pattern is observed wherein repeats in similar positions within multiple orthologs cluster together. For example, the first ANK repeat present in every ortholog clusters in a single clade, marked in yellow in Figures 4 and 5. A similar clustering is seen for the last repeat of every ortholog (marked in green), and for the second repeat in wMel and wMelPop/wMelCS with the fourth repeat of all other orthologs (marked in blue). Figure 4 shows the structure of each ortholog, with repeats that cluster...
together in the tree coloured in the same shade. Similar to VNTRs, ANK loci of *Wolbachia* provide highly informative and strain-specific marker sets that allow easy separation via PCR and high-resolution diagnosis of host infections (Figure 6).

A number of inferences about the evolution of the ANK repeats in these genes can be drawn from the tree in Figure 5 and the mapping of the phylogenetic data onto the modular structure of the genes. First, it is likely that the ancestral copy of this gene at the base of supergroup A already contained most of the repeats seen today, probably in a very similar linear order. Most of the clusters in the tree contain repeats from 7 or more of the orthologs, and the order of these orthologous repeats along the genes is highly similar. There is only one clear example of repeat shuffling: the eighth and ninth repeats in the *wPro/wSan/wAu* groups occur in the reverse order in *wCer1* (as repeat periods 10 and 9), while *wHa* may represent an intermediate stage, with the repeats orthologous to *wPro* 8 and 9 followed by a second copy of a repeat orthologous to *wPro* 8. Secondly, at least some variation in repeat number is due to lineage-specific tandem duplication of a single repeat (e.g. repeats 7 and 8 in *wCer1*) or of multiple repeats (repeats 3-4 and 5-6 in *wMel*).

**Extension of MLVA markers to other *Wolbachia* supergroups**

In comparison to the MLST markers, the highly polymorphic markers used here have a major trade-off in the loss of universal applicability for all *Wolbachia* strains. Here we have focused on *Wolbachia* supergroup A and tested the primers of these markers in other supergroups but primers did not amplify the loci or the loci were not informative. The presence of VNTR loci was restricted to subsets of supergroup A while genes containing ANK domain repeats were found in all supergroup A strains. *In silico* analysis of three other completed genomes, *wRi*, *wPip* and *wBm* of supergroups A, B and D, respectively, revealed though that tandem repeated regions occur throughout these supergroups and may be of relevance for MLVA in other supergroups. As further genome data become available it will be possible to extend this to an even larger group of *Wolbachia* isolates. A TRF analysis of *wMel* revealed 93 sites with direct tandem repeats of periods ranging from 10bp to 291bp, with internal match percentages from 68% to 100% (Table 4). The larger *wRi* genome has a similar number of tandem repeats while *wPip* has a smaller set of tandem repeats. The tandem repeats of *wMel*, *wRi* and *wPip* have similar characteristics such as

---

**Figure 6 Diagnostic size polymorphism of the WD0766 gene.** Isolates include *Wolbachia* of *D. melanogaster* (*wMel*, *wMelCS*), *D. willistoni* (*wWil*), *D. prosaltans* (*wPro*), *D. septentrionsaltans* (*wSpt*) and *D. simulans* transinfected with *Wolbachia* from *R. cerasi* (*wCer2*).
comparable period sizes, copy numbers as well as internal match ratios (Table 4). The number of tandem repeats in wBm is reduced by a factor of 10 when compared with the supergroup A and B Wolbachia, and the tandem periods appear to be shorter. This reduction in wBm is in accordance with the earlier described higher rate of secondary genome reduction in this strain [54]. Within the group of the closest relatives of the genus Wolbachia, the sequence of *E. ruminantium* revealed the highest content of tandem repeats for bacteria reported so far (Table 4), with size polymorphism in tandem repeats within the isolate that was used for genome sequencing the genome [66]. Our *in silico* analysis predicted the presence of variable tandem repeat markers in supergroup A strains and could hence readily be developed and tested on *Wolbachia* isolates from other supergroups. Highly polymorphic markers will be useful in population dynamic and population genetic studies similar to the ones undertaken in wMel-like strains [30,38,39]. We have not analysed the unfinished genome data sets of *Wolbachia* (e.g. [73]). A large proportion of tandem repeats are located in intergenic regions that tend to be assembled in genome sequencing projects last, yet their conserved flanking regions are required for the isolation of VNTR markers from total genomic extracts. A polymorphic VNTR locus has recently been reported for a supergroup B strain after applying a similar approach to wPip isolated from different *C. pipiens* populations [40].

Interestingly, our TRF analysis only detected five ANK repeat regions (WD0294, WD0385, WD0514, WD0550 and WD0766) of the 23 annotated genes encoding ANK repeat domains. Coincidentally, this group of genes includes the most variable genes encoding ANK repeat domains, suggesting that repeat extension/contraction is a strong diversifying mechanism in these genes.

Most of the primers designed for wMel ANK genes amplified expected PCR amplicons from supergroup A *Wolbachia*, but not from the majority of supergroup B, probably due to sequence divergence [36]. ANK domain genes are known to be present in other *Wolbachia* groups. In the B group mosquito strain wPip that infects mosquitoes there are 60 genes encoding ANK repeats, some of them also variable [53,71,72], whereas the fully sequenced D group wBm strain that infects the nematode *Brugia malayi* contains 5 ANK genes and 7 related pseudogenes [54]. Although wMel ANK genes were used as a reference in our study, another A group *Wolbachia* strain, wRi, contains 35 ANK genes, some of them very distinct from the wMel genes, probably as a result of duplications and recombination events [52]. Partial sequences of other A group strains have also revealed high numbers of ANK genes [73]. Thus, it seems clear that ANK genes are a signature feature in *Wolbachia* that can be potentially utilised to fingerprint closely related strains in A and other groups.

**Conclusion**

The identification of amplicon size polymorphic markers of *Wolbachia* provides a valuable addition to existing typing systems such as MLST, for the following three reasons: (1) The MLVA markers presented here display higher rates of evolution than the MLST loci, which are conserved protein encoding genes. Using MLVA, *Wolbachia* strains clustered in the same groups as in MLST typing, yet with a higher resolution that could be useful for different types of questions that MLST has not yet been able to target. These questions include the study of *Wolbachia* population genetics within infected species [30,38,39], and will further extend studies of horizontal transmission between host species for which MLST was originally developed [22]. Highly polymorphic markers will also be useful for experimental evolution of *Wolbachia* in order to track small genomic changes in short time frames. This higher resolution comes with the cost though, that markers are not universally applicable to the entire diversity of *Wolbachia*. (2) The majority of *Wolbachia* genomes are dotted with many different repeat regions which are highly appropriate to be targeted for the isolation of possible polymorphic markers. Tandem repeat markers such as the ones developed here can be tailored to individual studies. (3) MLVA markers are ideal for rapid and high-throughput DNA fingerprinting, as no sequencing is required. The markers are ideal to detect multiple infections in single PCR reactions if strains contain alleles with variable amplicon sizes. Our analysis of the evolution of the tandem repeat regions shows that they evolve by gain or loss of repeats. The variability in the number of ANK repeats, generally constituted by 33 amino acids each, creates size differences that are multiples of 99bp and, like VNTRs consisting of >100bp periods, can be clearly identified following simple PCR screenings without the need of initial sequencing or RFLP analyses as in the case of point mutations. The use of 2-3 highly variable markers per strain can generate easily readable fingerprints.

**List of abbreviations used**

- CI: cytoplasmic incompatibility
- MLVA: multiple locus variable number tandem repeat analysis
- MLST: multiple locus sequence typing
- VNTR: variable number tandem repeat
- ANK: ankyrin domain
- TRF: tandem repeats finder

**Acknowledgements**

We thank Sylvain Charlat, Kostas Bourtzis and the School of Veterinary Science, UQ, for supplying biological material, i.e. *H. bolina*, *C. capitata* and *D. immitis*, respectively. We thank the special edition editor Greg Hurst and two anonymous reviewers for their valuable comments. The research was supported by grants of the Australian Research Council ARC to MR, IIO, MW.
and SLO, and from COST Action FA-0701 and the research grant P22634-B17 of the Austrian Science Fund FWF to WJM. This article has been published as part of BMC Microbiology Volume 11 Supplement 1, 2012. Arthropod symbioses: from fundamental studies to pest and disease management. The full contents of the supplement are available online at http://www.biomedcentral.com/1471-2180/11/S1.

Author details
1 Hawkesbury Institute for the Environment, University of Western Sydney, Locked Bag 1797, Penrith NSW 2751 Australia. 2School of Biological Sciences, The University of Queensland, QLD 4072, Australia. 3Center of Anatomy and Cell Biology, Medical University of Vienna, Währingr. 10, 1090 Vienna, Austria. 4School of Biological Sciences, Monash University, Clayton, VIC 3800, Australia.

Authors’ contribution
MR, IIO, WJM and SLO had the initial idea for this manuscript. MR, IIO, WJM and SLO designed the study. MR, IIO and WJM performed laboratory work. MR, IIO, WJM, MW performed data analysis. MR, IIO, WJM, MW and SLO wrote the manuscript. All authors approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Published: 18 January 2012

References
1. Werren JH, Windsor D, Guo LR: Distribution of Wolbachia among neotropical arthropods. Proceedings of the Royal Society of London Series B-Biological Sciences 1995, 262(1364):197-204.
2. West SA, Cook JM, Werren JH, Godfray HCJ: Wolbachia in two insect host-parasitoid communities. Molecular Ecology 1998, 7(11):1457-1645.
3. Jeyaprakash A, Hoy MA: Long PCR improves Wolbachia DNA amplification: wsp sequences found in 76% of sixty-three arthropod species. Insect Molecular Biology 2000, 9(4):393-405.
4. Hilgenboecker K, Hammerstein P, Schlattmann P, Telschow A, Werren JH: How many species are infected with Wolbachia? - A statistical analysis of current data. FEMS Microbiology Letters 2008, 281(2):215-220.
5. Arthur W, Riegler M, Avits DN, Stauf F: Criterion for low-titre infections in insect symbiosis: Wolbachia in the bark beetle Phytophaga chalcographus (Coleoptera, Scolytinae). Environmental Microbiology 2009, 11(8):1923-1933.
6. O’Neill SL, Hoffmann AA, Werren JH: Influenzal passengers. Inherited microorganisms and arthropod reproduction. Oxford: Oxford University Press; 1997.
7. Stouthamer R, Breeuwer JA: Wolbachia pipientis: microbial manipulator of arthropod reproduction. Annual Review of Microbiology 1999, 53:71-102.
8. Werren JH, Baldo L, Clark ME: Wolbachia: Master manipulators of invertebrate biology. Nature Reviews Microbiology 2008, 6(10):741-751.
9. Schneider D, Miller WJ, Riegler M: Arthropods shopping for Wolbachia. In Manipulative Tants Bacteria associated with arthropods. Boca Raton: CRC Press/ZcHorn Fe; Bourzou K 2011:149-173.
10. Taylor ML, Bandi C, Horaena A: Wolbachia bacterial endosymbionts of filarial nematodes. Adv Parasitol 2005, 60:245-284.
11. Werren JH, Baldo L, SLO: Evolution and phylogeny of Wolbachia - reproductive parasites of arthropods. Proceedings of the Royal Society of London Series B-Biological Sciences 1995, 261(1360):55-63.
12. Zhou WG, Roussel F, O’Neill S: Phylogeny and PCR-based classification of Wolbachia strains using wsp gene sequences. Proceedings of the Royal Society of London Series B-Biological Sciences 1998, 265(1399):509-515.
13. Lo N, Paraksopeoulou C, Bourzou K, O’Neill SL, Werren JH, Bordenstein SR, Bandi C: Taxonomic status of the intracellular bacterium Wolbachia pipientis. International Journal of Systematic and Evolutionary Microbiology 2007, 57(3):654-657.
14. Baldo L, Werren JH: Revisiting Wolbachia supergroup typing based on wsp: spurious lineages and discordance in MLST. Current Microbiology 2007, 55(5):81-87.
15. Ros VID, Fleming VM, Fei EJ, Breeuwer JA: How diverse is the genus Wolbachia? Multiple-gene sequencing reveals a putatively new Wolbachia supergroup recovered from spider mites (Acari: Tetranychidae). Applied and Environmental Microbiology 2009, 75(4):1036-1043.
16. O’Neill SL, Giordano R, Colbert AME, Karr TL, Robertson HM: 16S ribosomal RNA phylogenetic analysis of the bacterial endosymbionts associated with cytoplasmic incompatibility in insects. Proceedings of the National Academy of Sciences of the United States of America 1992, 89(7):2699-2702.
17. Jiggins FM, von der Schulenburg JHG, Hurst GDD, Majerus MEN: Recombination confounds interpretations of Wolbachia evolution. Proceedings of the Royal Society of London Series B-Biological Sciences 2001, 268(1474):1423-1427.
18. Jiggins FM: The rate of recombination in Wolbachia bacteria. Molecular Biology and Evolution 2002, 19(9):1640-1643.
19. Baldo L, Bordenstein SR, Werren JH: Widespread recombination throughout Wolbachia genomes. Molecular Biology and Evolution 2006, 23(2):437-449.
20. Paraksopeoulou C, Bordenstein SR, Werren JH, Wilkes J, Bourtzis K: Toward a Wolbachia multilocus sequence typing system: Discrimination of Wolbachia strains present in Drosophila species. Current Microbiology 2006, 53(5):388-395.
21. Baldo L, Hotopp JCD, Jolly KA, Bordenstein SR, Ribier SA, Choudhury RR, Hayashi C, Moiden MCI, Tettelin H, Werren JH: Multilocus sequence typing system for the endosymbiont Wolbachia pipientis. Applied and Environmental Microbiology 2006, 72(11):7098-7110.
22. Baldo L, Ayoub NA, Hayashi C, Russell JA, Stahlhut JK, Werren JH: Insight into the routes of Wolbachia invasion: High levels of horizontal transfer in the spider genus Agelenopsis revealed by Wolbachia strain and mitochondrial DNA diversity. Molecular Ecology 2008, 17(2):357-369.
23. Zabala J, Apostolaki A, Patras S, Veneti Z, Paraksopeoulou C, Livadaras I, Markakis G, Bissias T, Merpat H, Bourzou K: Multiple rescue factors within a Wolbachia strain. Genetics 2006, 178(4):2149-2160.
24. Ishmael N, Hotopp JCD, Loaandis P, Ribier S, Sakamoto J, Sioso S, Nene V, Werren J, Bourzou K, Bordenstein SR, et al: Extensive genomic diversity of closely related Wolbachia strains. Microbiology 2009, 155(7):2211-2222.
25. Hoffmann AA, Clancy D, Duncan J: Naturally-occurring Wolbachia infection in Drosophila simulans that does not cause cytoplasmic incompatibility. Heredity 1999, 76:1-8.
26. Min KT, Benzer S: Wolbachia, normally a symbiont of Drosophila, can be virulent, causing degeneration and early death. Proceedings of the National Academy of Sciences of the United States of America 1997, 94(20):10792-10796.
27. McGraw EA, Merritt DJ, Drollier JN, O’Neill SL: Wolbachia density and virulence attenuation after transfer into a novel host. Proceedings of the National Academy of Sciences of the United States of America 2002, 99(5):2918-2923.
28. McNimman CJ, Lane RV, Cass BN, Forng AFC, Sithu M, Wang YF, O’Neill SL: Stable introduction of a life-shortening Wolbachia infection into the mosquito Aedes aegypti. Science 2009, 323(5910):141-144.
29. Sun LV, Riegler M, O’Neill SL: Development of a physical and genetic map of the virulent Wolbachia strain wMelPop. Journal of Bacteriology 2003, 185(24):7077-7084.
30. Riegler M, Sithu M, Miller WJ, O’Neill SL: Evidence for a global Wolbachia replacement in Drosophila melanogaster. Current Biology 2005, 15:1428-1433.
31. Achtman M, Moellii G, Zhu P, Wirth T, Diehl I, Kuseke B, Vogler AJ, Wagner DM, Allender CJ, Eastern WR, et al: Microevolution and history of the plague bacillus, Yersinia pestis. Proceedings of the National Academy of Sciences of the United States of America 2001, 101(15):7183-71842.
32. Pourcel C, Andr-Mazeaud F, Neubauer H, Raman S, Vergnaud F, Vergnaud G: Tandem repeats analysis for the high resolution phylogenetic analysis of Yersinia pestis. BMC Microbiology 2004, 4.
33. Johansson A, Farlow J, Larsson P, Dukerich M, Chambers E, Byström M, Fox J, Chu M, Forsman M, Sjöstedt A, et al: Worldwide genetic relationships among Francisella tularensis isolates determined by multiple-locus variable-number tandem repeat analysis. Journal of Bacteriology 2008, 190(17):5580-5589.
34. Yazdankhah SP, Lindstedt BA: Multiple locus variable number of tandem repeats analysis. Methods in molecular biology (Clifton, N J) 2009, 531:141-158.
36. Iturbe-Ormaetxe I, Burke GR, Riegler M, O'Neill SL. Distribution, expression, and motif variability of ankyrin domain genes in Wolbachia pipientis. Journal of Bacteriology 2005, 187:5136-5145.

37. Duron O, Lagner J, Raymond M, Bourtzis K, Fort P, Welli M. Transposable element polymorphism of Wolbachia in the mosquito Culex pipiens: evidence of genetic diversity, superinfection and recombination. Molecular Ecology 2005, 14(1):1561-1573.

38. Miller WI, Riegler M. Evolutionary dynamics of wAu-like Wolbachia variants in neotropical Drosophila spp. Applied and Environmental Microbiology 2006, 72(1):826-835.

39. Miller WI, Ehman L, Schneider D. Infectious speciation revisited: Impact of symbiont-depletion on female fitness and mating behavior of drosophila paulistorum. Plos Pathogens 2010, 6(12).

40. Petridis M, Chatzidimitriou D. Characterization of an intergenic polymorphic site (pp-hC1A_5) in Wolbachia pipiens (wPip). Molecular Ecology Resources 2011, 11(4):735-744.

41. Wu M, Sun LV, Vamathevan J, Riegler M, Deboy R, Brownlie JC, McGraw EA, Martin W, Esser C, Ahmadinejad N et al. Phylogenomics of the reproductive parasite Wolbachia wMel: a streamlined genome overrun by mobile genetic elements. Public Library of Science Biology 2004, 2(3):237-241.

42. Mosavi LK, Cammett TJ, Desrosiers DC, Peng ZY. DNA repeats lead to the accelerated loss of gene order in Wolbachia pipientis genomes in multiple Drosophila species. Systematic Biology 2007, 56(1):143-1468.

43. Pan X, Lührmann A, Satoh A, Laskowski-Arce MA, Roy CR. Cross-species expansion of a family of bacterial repeat induced by double-strand break repair. Molecular and Cellular Biology 1998, 18(4):2045-2054.

44. Merçot H, Poinsot D. Infection by Wolbachia: from passengers to residents. Comptes Rendus - Biologist 2009, 332(23):264-297.

45. Hoffmann AA, Turelli M, Simmons GM. Wolbachia and motif variability of ankyrin domain genes in the filarial worm W. bancrofti. Molecular Biology and Evolution 2002, 19(11):2425-2434.

46. Sironi M, Bandi C, Sacchi L, DiSacco B, Damiani G, Genchi C. The genome of the heartwater agent Ehrlia ruminantium contains multiple tandem repeats of actively variable copy number. Proceedings of the National Academy of Sciences of the United States of America 2005, 102(3):838-843.

47. Weitzman MN, Woodward KJ, Usdin K. DNA secondary structures and the evolution of hypervariable tandem arrays. Journal of Biological Chemistry 1997, 272(14):9517-9523.

48. Dover G. Molecular drive: A cohesive mode of species evolution. Nature 1982, 299:579-581.

49. Amos W. A comparative approach to the study of microsatellite evolution. In: Microsatellites Evolution and Applications. Oxford: Oxford University Press; Goldstein DB, Schlötterer C 1999:66-79.

50. Reynolds KT, Hoffmann AA. High frequencies of short frameshifts in poly-C/ATG tandem repeats borne by bacteriophage M13 in Escherichia coli K-12. Nucleic Acids Research 1987, 15(13):5323-5338.

51. Faber FE, Van Heerden H, Josemans A, Van Kleef M et al. The genome of the Wolbachia strain Pip from the fruit fly Drosophila melanogaster. Genome Research 2002, 12(11):178-180.

52. Zabalou S, Riegler M, Theodorakopoulos M, Staeuffer M, Savakis C, Bourtzis K. Wolbachia-induced cytoplasmic incompatibility as a means for insect pest population control. Proceedings of the National Academy of Sciences of the United States of America 2004, 101(15):5042-5045.

53. Dyson EA, Hurst GD. Persistence of an extreme sex-ratio bias in natural populations of the fruit fly Drosophila pseudoobscura. Proceedings of the National Academy of Sciences USA 2004, 101(20):7161-7164.

54. Amos W. A comparative approach to the study of microsatellite evolution. In: Microsatellites Evolution and Applications. Oxford: Oxford University Press; Goldstein DB, Schlötterer C 1999:66-79.

55. Zabalou S, Charlat S, Nirgianaki A, Lachaise D, Merçot H, Bourtzis K. Sironi M. Bidirectional incompatibility between conspecific Wolbachia strains in Drosophila simulans: investigating the outcomes of host-symbiont coevolution. Applied and Environmental Microbiology 2004, 70(1):273-279.

56. Letunic I, Copley RR, Schmidt S, Ciccarelli FD, Doerks T, Schultz J, Ponting CP, Bork P. SMART 4.0: towards genomic data integration. Nucleic Acids Research 2004, 32(Database issue):D142-D144.

57. Riegler M, Pip from the fruit fly Drosophila melanogaster. Genome Research 2002, 12(11):178-180.

58. Letunic I, Copley RR, Schmidt S, Ciccarelli FD, Doerks T, Schultz J, Ponting CP, Bork P. SMART 4.0: towards genomic data integration. Nucleic Acids Research 2004, 32(Database issue):D142-D144.

59. Riegler M, Pip from the fruit fly Drosophila melanogaster. Genome Research 2002, 12(11):178-180.

60. Reynolds KT, Hoffmann AA. High frequencies of short frameshifts in poly-C/ATG tandem repeats borne by bacteriophage M13 in Escherichia coli K-12. Nucleic Acids Research 1987, 15(13):5323-5338.
79. Merçot H, Llorente B, Jacques M, Atlan A, Montchampmoreau C. Variability within the Seychelles cytoplasmic incompatibility system in Drosophila simulans. Genetics 1995, 141(3):1015-1023.

80. Giordano R, O'Neill SL, Robertson HM. Wolbachia infections and the expression of cytoplasmic incompatibility in Drosophila sechellia and D. mauritiana. Genetics 1995, 140(4):1307-1317.

81. Hornett EA, Duplouy AMR, Davies N, Rodenick GK, Wedell N, Hurst GDD, Charlat S. You can't keep a good parasite down: evolution of a male-killer suppressor uncovers cytoplasmic incompatibility. Evolution 2008, 62(5):1258-1263.

82. Yamada R, Iturbe-Ormaetxe I, Brownlie JC, O'Neill SL. Functional test of the influence of Wolbachia genes on cytoplasmic incompatibility expression in Drosophila melanogaster. Insect Molecular Biology 2011, 20(1):75-85.

doi:10.1186/1471-2180-12-S1-S12

Cite this article as: Riegler et al. Tandem repeat markers as novel diagnostic tools for high resolution fingerprinting of Wolbachia. BMC Microbiology 2012 12(Suppl 1):S12.