Opinion

# Is it time to retire *Wolbachia* multilocus sequence typing (MLST)?

Christoph Bleidorn[1*] & Michael Gerth[3]

[1]Museo Nacional de Ciencias Naturales, Spanish National Research Council (CSIC), Madrid, Spain.

[2]German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, 04103 Leipzig, Germany.

[3]Institute for Integrative Biology, University of Liverpool, Liverpool, UK.

[*]Correspondence to:

Christoph Bleidorn

Museo Nacional de Ciencias Naturales

Spanish National Research Council (CSIC)

Department of Biodiversity and Evolutionary Biology

José Gutiérrez Abascal 2

Madrid 28006

Spain

Email: christoph.bleidorn@gmail.com

Keywords: phylogenetic informativeness, MLST, recombination, strain typing, *Wolbachia*

Abstract

Alphaproteobacteria of the genus *Wolbachia* constitute the most common animal endosymbiont on earth. Analyses of strain diversity and evolutionary relationships of these bacteria often rely on multilocus sequence typing (MLST), where sequence data of five selected housekeeping genes are used for allele designation. Typically, phylogenetic and phylogeographic relationships, as well as questions regarding horizontal transmission are addressed with *Wolbachia* MLST. However, given frequent recombination and limited phylogenetic informativeness of the *Wolbachia* MLST genes we argue that these markers are unsuited to infer phylogenetic relationships of closely related strains, to detect recent horizontal transmissions or to assess movements between populations on an ecological timescale. Instead of relying on *Wolbachia* MLST, we suggest genome-scale approaches for all analyses dealing with phylogenetic approaches, whereas for studies screening the prevalence of *Wolbachia* strains across potential hosts single or few genes (e.g., *ftsz* and *wsp*) are sufficient.

*Wolbachia* is a genus of maternally inherited intracellular Alphaproteobacteria that is found in arthropod and nematode hosts (Werren *et al.* 2008). Meta-analyses suggest that between 40% and 52% of all terrestrial arthropods are infected, making these bacteria the most common animal endosymbiont on earth (Weinert *et al.* 2015; Zug & Hammerstein 2012). Host specificity and type of symbiosis differs between different major lineages of *Wolbachia*, which have been classified into (currently) 16 supergroups named with capital letters from A to Q, consecutively in the order of their description (Gerth 2016; Glowska *et al.* 2015). Supergroups A and B are found in arthropods, representing the vast majority of described *Wolbachia* lineages. Many different types of symbioses, including reproductive parasitism, facultative mutualism and obligate mutualism have been found for these lineages (Zug & Hammerstein 2015). In contrast, supergroups C and D are restricted to filarial nematodes, with which they share a close relationship that can be described as obligate mutualism (Makepeace & Gill 2016). Supergroup F has been found in both nematodes and arthropods and all other supergroups are rather rare, limited to a single or few hosts (Gerth *et al.* 2014).

Several host manipulations have been described for *Wolbachia*, and it is thought that those accelerate their spread in host populations, such as male-killing, feminization, induced parthogenesis, and cytoplasmatic incompatibility (Werren *et al.* 2008). All these manipulations are considered to have a mostly a negative effect on their hosts. However, several positive aspects for hosts have been reported, too. These include provision of the host with amino acids or vitamins, or protection against viruses (Hedges *et al.* 2008; Zug & Hammerstein 2015). It is likely that *Wolbachia* infections are established in their host population due to positive effects on the host fitness (Kriesner *et al.* 2013). Recently, *Wolbachia* received much attention in fighting widespread human pathogens. Field studies demonstrated that mosquito populations can

be artificially infected with fast spreading *Wolbachia* lineages which confer virus resistance to their hosts, thereby suppressing transmission of Dengue fever (Hoffmann *et al.* 2011). However, not all strains of *Wolbachia* are able to confer virus resistance or manipulating their host reproduction (Makepeace & Gill 2016).

Divergent, but highly recombining strains of *Wolbachia* can be found within supergroups, whereas recombination between supergroups is less frequent (Ellegaard *et al.* 2013). Genetic diversity of strains within supergroups has been used to address questions regarding phylogenetic relationships (Bordenstein *et al.* 2009), co-speciation patterns of *Wolbachia* and their hosts (Lefoulon *et al.* 2016), interspecific transmission (Baldo *et al.* 2008; Gerth *et al.* 2013), biogeography (Russell *et al.* 2009), or to assess the ecological timescale at which *Wolbachia* moves between host populations (Bailly-Bechet *et al.* 2017). Even though the number of genome-scale studies is increasing, many studies analysing *Wolbachia* genetic diversity still are based on a single or few genes. Initially *Wolbachia* genetic diversity was characterized using the 16S rRNA gene or the more variable *wsp* gene (Zhou *et al.* 1998). However, in 2006 a multilocus sequence typing (MLST) system was established, and this subsequently became a standard in the community of *Wolbachia* researchers (Baldo *et al.* 2006b).

**What is multilocus sequence typing (MLST) and what is it used for?**

The MLST approach has been developed to provide a reproducible and portable method for the molecular characterization of bacterial pathogens. Initially developed to monitor local and global *Neisseria meningitides* outbreaks (Maiden *et al.* 1998), MLST schemes have since been published for many other bacterial species (Maiden 2006). For MLST, five to ten loci (usually conserved housekeeping that are present in all strains) from different regions of the genome are sequenced and each unique allele is assigned with a

number in the order of discovery. For strain typing, a universal nomenclature based on a code of numbers referring to the sequenced loci is assembled. MLST genes are selected under the assumption that they underlie purifying selection, resulting in sequence variation that is mostly neutral. In the absence of recombination, accumulation of substitutions should be approximately linearly with time (Francisco *et al.* 2009). In this case genetic distances between strains would be proportional to divergence time. MLST data is usually provided in a curated form in a freely accessible database (Jolley *et al.* 2004). Based on MLST profiles relationships (or diversity) of typed strains can be either analysed using the designated numbers from coding the alleles (i.e., MLST profiles), or by analysing the allelic nucleotide sequence data directly. For *Wolbachia*, fragments of five ubiquitous housekeeping genes (*gatB*, *coxA*, *hcpA*, *fbpA*, and *ftsZ*) are sequenced, and primers that amplify across the major *Wolbachia* supergroups in infecting arthropods are available (Baldo *et al.* 2006b). Some more, but less frequently used loci have been proposed on the way of developing the current MLST system (Paraskevopoulos *et al.* 2006). Currently, the *Wolbachia* MLST database comprises 3260 sequences and 470 MLST profiles (https://pubmlst.org/wolbachia/, last accessed 24$^{th}$ of March 2017). According to the high number of citations for the original publication (Baldo *et al.* 2006b) as given by ISI Web of Science (329 citations, last accessed April 21st 2017) the approach is well-established and frequently used in the community of *Wolbachia* researchers. Judging from the abstracts and keywords of the citing articles, major questions that are commonly addressed with *Wolbachia* MLST are phylogeny & phylogeography (102 articles with corresponding terms), and horizontal transmission of strains (58 articles). In conclusion, MLST data are routinely used to infer phylogenetic relationships among *Wolbachia* strains. We here argue that there are no convincing reasons to continue this practise.

## The case against using MLST for phylogenetic analyses of *Wolbachia* strains

The level of resolution given by MLST analyses depends on the number and type of genes used, length of the sequences and the genetic diversity of chosen loci (Cooper & Feil 2004). The limits of MLST schemes have been discussed for genetically monomorphic bacteria (Achtman 2008) such as *Mycobacterium tuberculosis* or *Bacillus anthracis* and genome-scale analyses were needed for population genetic analyses (Achtman 2012). Similarly, the missing resolution of MLST-data for closely related *Wolbachia* strains has been criticized (Riegler *et al.* 2012). The genetic diversity within *Wolbachia* supergroups is far from being monomorphic, as evident from the large number of available MLST profiles in the database (see above). However, the actual pace of evolution of *Wolbachia* genes and genomes was an open question. Recently, based on a time-calibrated phylogenomic analyses it was hypothesized that *Wolbachia* lineages are much older than previously assumed – and likewise genetic change due to substitutions or recombination accumulates slower than expected (Gerth & Bleidorn 2016). Not surprisingly, the indistinguishability of closely related strains by using MLST loci is a well-known problem (Riegler *et al.* 2012).

The level of resolution across time for a given gene in a phylogenetic analysis can be estimated by its phylogenetic informativeness (PI) (Townsend 2007). Based on an ultrametric tree (and an alignment) the relative amount of phylogenetic signal to noise across time can be estimated. Analysing the PI of all MLST genes for a set of *Wolbachia* strains covering supergroup A and B reveals that all of them (separately or combined) show the highest phylogenetic resolution on the supergroup level (Fig. 1). According to Gerth and Bleidorn (2016) the supergroups A and B have diverged more than 200 million years ago. MLST genes however provide only little phylogenetic

information for strains that diverged more recently (Fig. 1). Consequently, the MLST approach is unsuited to infer phylogenetic relationships of closely related strains, to detect recent horizontal transmissions or to assess ecological timescales of *Wolbachia* movements between populations.

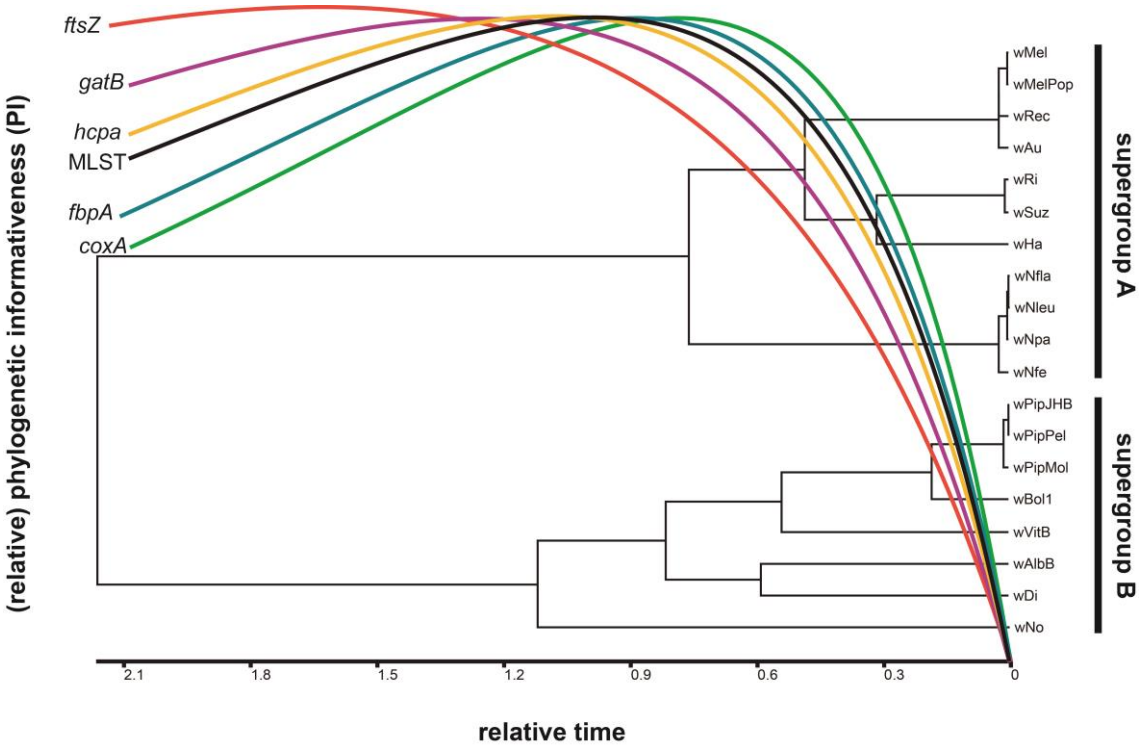
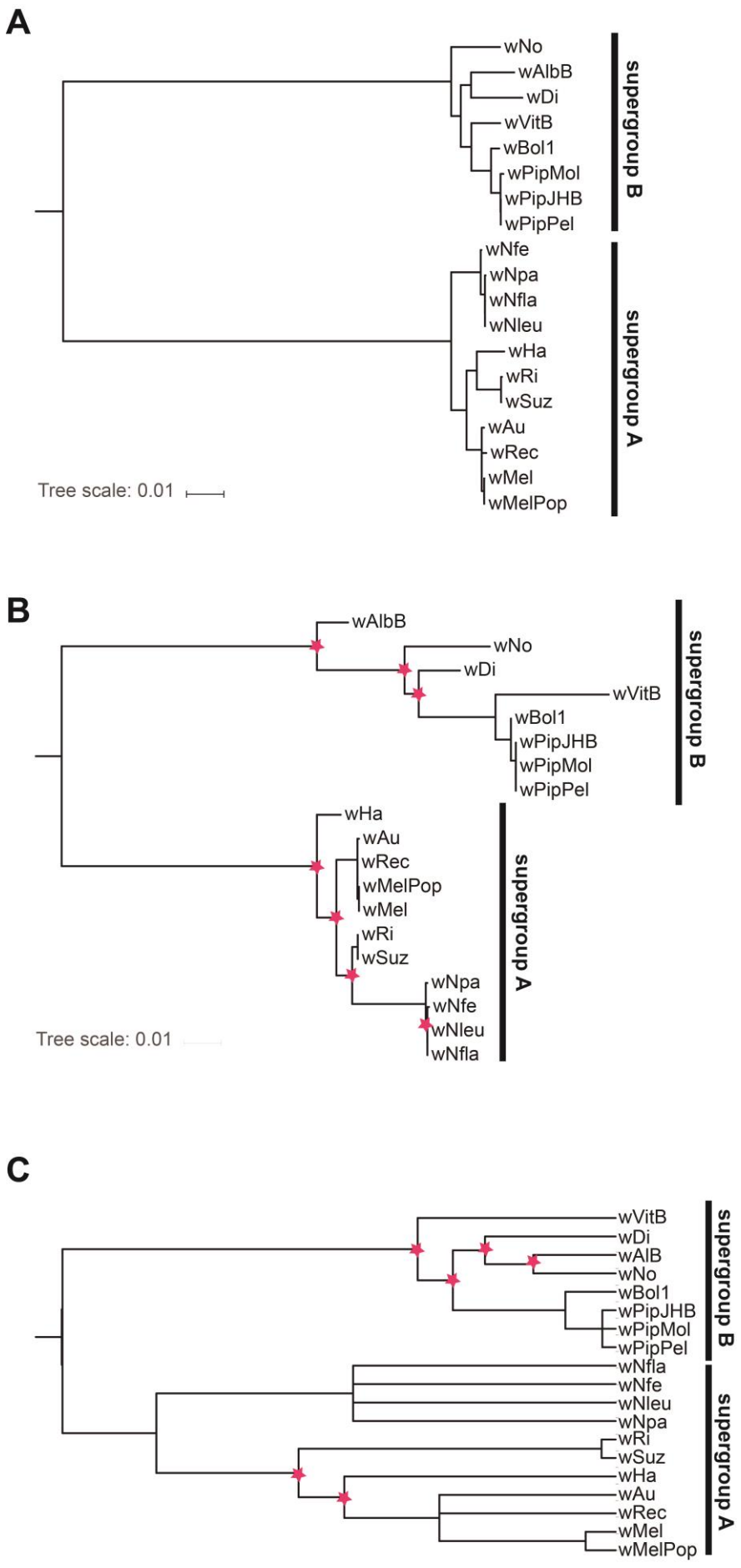



Fig. 1: Phylogenetic informativeness (PI) of five MLST-gene alignments from supergroup A and B *Wolbachia* strains analysed separately and combined within a supermatrix using PhyDesign (López-Giráldez & Townsend 2011). Gene alignments are based on *Wolbachia* PubMLST data and are deposited at dataDryad (www.datadryad.org). Ultrametric tree (based on a phylogenomic of 252 loci) taken from (Gerth & Bleidorn 2016). PI profiles for all genes and the combined matrix are given in different colours as indicated in the figure. Note that informativeness of genes/combined genes are only shown to demonstrate relative resolution and cannot be compared regarding the PI of each of them. See supporting information for scaled PI profiles for all partitions.

In summary, the MLST approach is too conserved on the nucleotide level to distinguish recently diverged strains. Moreover, as already mentioned in the original MLST publication (Baldo *et al.* 2006b), four of the five genes show some level of intragenic recombination. Only for *ftsz* no signal for recombination had been detected (Baldo *et al.* 2006a). However, the number of available MLST-alleles for each marker increased dramatically since 2006. We downloaded all available MLST alleles (*gatB*: 260 alleles; *coxA*: 249 alleles; *hcpA*: 290 alleles; *ftsZ*: 228 alleles; *fbpA*: 432: alleles) from *Wolbachia* PubMLST (https://pubmlst.org/wolbachia/, accessed 24th of March 2017) and generated alignments using MAFFT (Katoh & Standley 2013). By using the software RDP4 (Martin *et al.* 2015) and applying several different approaches (RDP, GENECONV, Chimaera, MaxChi), signals of recombination could be detected for all five markers. These results were also supported by highly significant p-values using the PHI test (Bruen *et al.* 2006). The presence of horizontal genetic exchange makes the interpretation of phylogenetic analyses of MLST genes challenging, as the resulting tree may not reflect the evolutionary relationships of the analysed strains (Holmes *et al.* 1999; Jiggins *et al.* 2001). When comparing the phylogenetic reconstruction for the dataset in Fig. 1 using a supermatrix of the five combined MLST-fragments with the original analyses based on 252 orthologs which were selected to show no signs of recombination, several differences in the topology can be found (Fig. 2A, B). In comparison, seven internal nodes are reconstructed differently in the MLST-based analysis (Fig. 2B). Moreover, remarkable differences in the branch lengths within supergroups are obvious, when comparing both analyses. All these differences are likely due to the misleading signal from recombination events. ClonalFrame is a Bayesian phylogenetic framework that was developed especially for MLST datasets and capable of inferring relationships despite the presence of recombination (Didelot & Falush

2007). Nevertheless, analysing our dataset with ClonalFrame led to a similarly high number of conflicting nodes (six) in comparison to the phylogenomic dataset, and multiple polytomies (i.e., unresolved nodes, Fig 2C). This shows that the usage of recombination-aware phylogenetic methods cannot circumvent the inherent problems of *Wolbachia* MLST genes as phylogenetic markers. In summary, phylogenetic analyses of MLST data works well on the level of supergroups, but not within supergroups.


A
wNo
wAlbB
wDi
wVitB
wBol1
wPipMol
wPipJHB
wPipPel
supergroup B
wNfe
wNpa
wNfla
wNleu
wHa
wRi
wSuz
wAu
wRec
wMel
wMelPop
supergroup A
Tree scale: 0.01
B
wAlbB
wNo
wDi
wVitB
wBol1
wPipJHB
wPipMol
wPipPel
supergroup B
wHa
wAu
wRec
wMelPop
wMel
wRi
wSuz
wNpa
wNfe
wNleu
wNfla
supergroup A
Tree scale: 0.01
C
wVitB
wDi
wAlB
wNo
wBol1
wPipJHB
wPipMol
wPipPel
supergroup B
wNfla
wNfe
wNleu
wNpa
wRi
wSuz
wHa
wAu
wRec
wMel
wMelPop
supergroup A

188

Fig. 2: Phylogenetic reconstruction of a dataset of supergroup A and B strains as published by Gerth and Bleidorn (2016). A. ML-reconstruction using optimal partitions and models as selected by IQ-TREE (Nguyen *et al.* 2015) for a dataset containing nucleotide data of 252 orthologs which do not show evidence of recombination. B. ML-reconstruction of the same taxon sampling using optimal partitions and models as selected by IQ-TREE for a dataset containing the five MLST-gene fragments. Seven conflicting nodes (red asterisks) compared to the phylogenomic analyses are found. C. ClonalFrame (Didelot & Falush 2007) analysis based on four independent runs (500,000 Burnin; 500,000 posterior sample, otherwise default) always converged on the same topology which shows six conflicting nodes (red asterisks) and some polytomies compared to the phylogenomic analysis.

Homologous recombination is widespread among Bacteria (Didelot & Maiden 2010). Therefore MLST data can be alternatively analysed based on allele designations to estimate relationships among strains. A simple method for this is to cluster strains according to their similarity, which can be visualized as a dendrogram. However, strain similarity does not necessarily reflect common ancestry. A popular and more sophisticated method to analyse allele-based strain data is eBURST (Feil *et al.* 2004). This software incorporates a model of bacterial evolution in which strains increasing in frequency diversify, thereby forming clusters of similar genotypes. For MLST data, so-called clonal complexes are defined as groups that share a predefined number of alleles (e.g., three of five allele designations are identical) with at least one other strain type. After searching for these clonal complexes, the likely founding strain type is inferred, as well as evolutionary relationships within this clonal complex. Simulation studies have shown that when recombination is absent or present in low to moderate levels, the

inferred relationships of clonal complexes are very similar to the (known) true ancestry (Turner *et al.* 2007). However, increasing rates of the frequency of recombination to mutations led to a strong decrease in the reliability of eBURST analyses. In *Wolbachia* the overall ratio of recombination to mutation events to explain the generation of a substitution is ranging from 2.3 to 8.2, depending on the analysed genome (Ellegaard *et al.* 2013). Therefore, the rather high recombination rates in *Wolbachia* genomes make allele-based analyses unreliable.

In addition to problems with recombination, there are also theoretical arguments against using 'eBURST'-like clustering algorithms with *Wolbachia* MLST profiles. Because the only criterion for assigning a novel allele number is at least one nucleotide difference compared to all described alleles, any number of substitutions in one allele is weighted equally. For example, 10 different *Wolbachia* strains may be differentiated by only 9 nucleotide differences in total, or by 50, and could potentially be characterised by identical MLST profiles. This makes comparing these profiles across systems challenging. When sampling is dense and therefore the majority of the allele diversity is known, this will likely not be problematic. However, this is rarely ever the case for *Wolbachia.* Given the large number of infected species, it is essentially impossible to know the true diversity of *Wolbachia* in any ecosystem. Furthermore, because horizontal transmissions are common (Ahmed *et al.* 2016; Baldo *et al.* 2008; Gerth *et al.* 2013; Zug *et al.* 2012), and exact pathways of these transmissions are still discussed (Huigens *et al.* 2004; Le Clec'h *et al.* 2013; Li *et al.* 2017), it does not make sense to define "founding" and "descending" *Wolbachia* genotypes in most cases.

**Alternatives to MLST**

MLST was developed as a replacement for an earlier strain typing approach called multilocus enzyme electrophoresis (MLEE), which measured genetic variation by the resolution of electrophoretic variants (electromorphs) of metabolic enzymes (Maiden 2006). One problem of this method was that experiments were difficult to reproduce across labs. With the availability of affordable and faster Sanger sequencers it was possible to directly use sequence data instead of electromorphs. Nowadays, a wide array of different high-throughput sequencing techniques is available (Bleidorn 2016; Goodwin *et al.* 2016). Due to their rather small size (usually a few mb, less for many symbionts) sequencing of complete bacterial genomes is affordable and routinely carried out using benchtop sequencers in laboratories with standard equipment (Loman & Pallen 2015). Consequently, strain typing methods based on whole genome data have been proposed. In the case of rMLST a set of 53 ribosomal proteins are retrieved for sequence typing from the whole genome data (Jolley *et al.* 2012). Ribosomal proteins were chosen as candidates as they are universally found in bacterial genomes, show a wide distribution across genomes and are expected to underlie stabilizing selection, similar to the above mentioned MLST genes. In the case of *Wolbachia*, ribosomal proteins have been already successfully used for phylogenomic analyses (Nikoh *et al.* 2014). And of course using the complete genome data, as in the case of whole-genome sequence-typing (WGST) has consequently been recommended (Pérez-Losada *et al.* 2013).

Even though *Wolbachia* harbour small genomes (1 to 1.5 mbp in size) (Makepeace & Gill 2016), sequencing and assembly is more difficult than for most other Bacteria. First of all, *Wolbachia* are strictly intracellular endosymbionts and keeping them in cell cultures is possible (Dobson *et al.* 2002), but laborious and often not applicable for strains from non-model hosts. Alternatively, a metagenome containing host and

*Wolbachia* DNA is sequenced, and *Wolbachia* sequence data is retrieved using BLAST-searches and read mapping (Gerth *et al.* 2014). However, in this case much more sequencing depth per genome is needed, as typically only 1% to 5% of the reads will be of *Wolbachia* origin. For more efficient sequencing of *Wolbachia* genomes, a target enrichment protocol (Lemmon & Lemmon 2013) is available (Dunning Hotopp *et al.* 2017; Geniez *et al.* 2012). Also the assembly itself usually leads to highly discontinuous genome sequences, as *Wolbachia* genomes are rich in mobile genetic elements with repetitive sequence motives (Wu *et al.* 2004). However, for analyses focussing on sequence data of selected loci and not on synteny, incompletely assembled *Wolbachia* draft genomes are sufficient. Working with complete (or draft) genomes has the advantage that comparative analyses can be used to retrieve large sets of orthologous and recombination-free loci (Comandatore *et al.* 2013). Such datasets circumvent problems of allele-based analyses, as all analyses can be performed sequence-based. Moreover, complete genomes further allow to search for hypervariable regions such as tandem repeat markers (Riegler *et al.* 2012) or ankyrin repeat domains (Siozios *et al.* 2013). However, in the case where genomic approaches might be too cost- and time intensive, approaches relying on few genetic markers are needed. In this case we suggest that a combination of conservative (and rarely recombining) *ftsz* gene and the more variable (but also often recombining) *wsp* gene are sufficient for supergroup designation and strain typing.

## Conclusion

Since their initial publication MLST analyses are widely used in the community of *Wolbachia* researchers and a large database for comparative studies is available. New approaches may always lead to the problem that available resources cannot be used for

comparisons. The two main problems of the MLST markers are their conservative nature and the presence of recombination. The lack of resolution of recent strains means that mainly supergroup affinities can be firmly recognized with these markers, which could be also achieved by relying on single MLST marker. Whereas for studies screening the prevalence of *Wolbachia* strains across potential hosts single genes are sufficient, we suggest genome-scale approaches when discerning closely related strains or analysing the evolutionary history of strains. However, new and more efficient approaches for *Wolbachia* genome sequencing are desirable. In summary, we do not find reasons to pursue MLST of *Wolbachia* strains for phylogenetic purposes in the future.

**Acknowledgments**

C.B. is a 'Ramon y Cajal' fellow supported by the Spanish Ministry of Science and Education (MEC) (RYC-2014-15615). M.G. is funded by the European Molecular Biology Organization (ALTF 48-2015) and co-funded by Marie-Curie Actions of the European Commission (LTFCOFUND2013, GA-2013-609409). This work was supported by the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig.

References

Achtman M (2008) Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. *Annual Review of Microbiology* **62**, 53-70.

Achtman M (2012) Insights from genomic comparisons of genetically monomorphic bacterial pathogens. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**, 860-867.

Ahmed MZ, Breinholt JW, Kawahara AY (2016) Evidence for common horizontal transmission of *Wolbachia* among butterflies and moths. *Bmc Evolutionary Biology* **16**, 118.

Bailly-Bechet M, Martins- Simões P, Szöllősi G, *et al.* (2017) How long does *Wolbachia* remain on board? *Molecular Biology and Evolution* **in press**.

Baldo L, Ayoub NA, Hayashi CY, *et al.* (2008) 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* **17**, 557-569.

Baldo L, Bordenstein S, Wernegreen JJ, Werren JH (2006a) Widespread recombination throughout *Wolbachia* genomes. *Molecular Biology and Evolution* **23**, 437-449.

Baldo L, Dunning Hotopp JC, Jolley KA, *et al.* (2006b) Multilocus sequence typing system for the endosymbiont *Wolbachia pipientis*. *Applied and Environmental Microbiology* **72**, 7098-7110.

Bleidorn C (2016) Third generation sequencing: technology and its potential impact on evolutionary biodiversity research. *Systematics and Biodiversity* **14**, 1-8.

Bordenstein SR, Paraskevopoulos C, Dunning Hotopp JC, *et al.* (2009) Parasitism and mutualism in *Wolbachia*: What the phylogenomic trees can and cannot say. *Molecular Biology and Evolution* **26**, 231-241.

Bruen TC, Philippe H, Bryant D (2006) A Simple and Robust Statistical Test for Detecting the Presence of Recombination. *Genetics* **172**, 2665-2681.

Comandatore F, Sassera D, Montagna M, *et al.* (2013) Phylogenomics and analysis of shared genes suggest a single transition to mutualism in *Wolbachia* of nematodes. *Genome Biology and Evolution* **5**, 1668-1674.

Cooper JE, Feil EJ (2004) Multilocus sequence typing – what is resolved? *Trends in Microbiology* **12**, 373-377.

Didelot X, Falush D (2007) Inference of bacterial microevolution using multilocus sequence data. *Genetics* **175**, 1251-1266.

Didelot X, Maiden MCJ (2010) Impact of recombination on bacterial evolution. *Trends in Microbiology* **18**, 315-322.

Dobson SL, Marsland EJ, Veneti Z, Bourtzis K, O'Neill SL (2002) Characterization of *Wolbachia* host cell range via the In vitro establishment of infections. *Applied and Environmental Microbiology* **68**, 656-660.

Dunning Hotopp JC, Slatko BE, Foster JM (2017) Targeted enrichment and sequencing of recent endosymbiont-host lateral gene transfers. *Scientific Reports* **7**, 857.

Ellegaard KM, Klasson L, Näslund K, Bourtzis K, Andersson SGE (2013) Comparative Genomics of *Wolbachia* and the Bacterial Species Concept. *PLoS Genetics* **9**, e1003381.

Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG (2004) eBURST: Inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. *Journal of Bacteriology* **186**, 1518-1530.

Francisco AP, Bugalho M, Ramirez M, Carriço JA (2009) Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. *BMC Bioinformatics* **10**, 152.

Geniez S, Foster JM, Kumar S, *et al.* (2012) Targeted genome enrichment for efficient purification of endosymbiont DNA from host DNA. *Symbiosis* **58**, 201-207.

Gerth M (2016) Classification of *Wolbachia* (Alphaproteobacteria, Rickettsiales): No evidence for a distinct supergroup in cave spiders. *Infection, Genetics and Evolution* **43**, 378-380.

Gerth M, Bleidorn C (2016) Comparative genomics provides a timeframe for *Wolbachia* evolution and exposes a recent biotin synthesis operon transfer. *Nature Microbiology* **2**, 16241.

Gerth M, Gansauge MT, Weigert A, Bleidorn C (2014) Phylogenomic analyses uncover origin and spread of the *Wolbachia* pandemic. *Nature Communications* **5**, 5117.

Gerth M, Röthe J, Bleidorn C (2013) Tracing horizontal *Wolbachia* movements among bees (Anthophila): a combined approach using multilocus sequence typing data and host phylogeny. *Molecular Ecology* **22**, 6149-6162.

Glowska E, Dragun-Damian A, Dabert M, Gerth M (2015) New *Wolbachia* supergroups detected in quill mites (Acari: Syringophilidae). *Infection, Genetics and Evolution* **30**, 140-146.

Goodwin S, McPherson JD, McCombie WR (2016) Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet* **17**, 333-351.

Hedges LM, Brownlie JC, O'Neill SL, Johnson KN (2008) *Wolbachia* and virus protection in insects. *Science* **322**, 702-702.

Hoffmann AA, Montgomery BL, Popovici J, *et al.* (2011) Successful establishment of *Wolbachia* in *Aedes* populations to suppress dengue transmission. *Nature* **476**, 454-457.

Holmes EC, Urwin R, Maiden MC (1999) The influence of recombination on the population structure and evolution of the human pathogen *Neisseria meningitidis*. *Molecular Biology and Evolution* **16**, 741-749.

Huigens ME, de Almeida RP, Boons PAH, Luck RF, Stouthamer R (2004) Natural interspecific and intraspecific horizontal transfer of parthenogenesis–inducing *Wolbachia* in *Trichogramma* wasps. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **271**, 509-515.

Jiggins FM, Schulenburg JHGvd, Hurst GDD, Majerus MEN (2001) Recombination confounds interpretations of *Wolbachia* evolution. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **268**, 1423-1427.

Jolley KA, Bliss CM, Bennett JS, *et al.* (2012) Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. *Microbiology* **158**, 1005-1015.

Jolley KA, Chan M-S, Maiden MC (2004) mlstdbNet – distributed multi-locus sequence typing (MLST) databases. *BMC Bioinformatics* **5**, 86.

Katoh K, Standley DM (2013) MAFFT Multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* **30**, 772-780.

Kriesner P, Hoffmann AA, Lee SF, Turelli M, Weeks AR (2013) Rapid sequential spread of two *Wolbachia* variants in *Drosophila simulan*s. *PLOS Pathogens* **9**, e1003607.

Le Clec'h W, Chevalier FD, Genty L, *et al.* (2013) Cannibalism and predation as paths for horizontal passage of *Wolbachia* between terrestrial isopods. *Plos One* **8**, e60232.

Lefoulon E, Bain O, Makepeace B, *et al.* (2016) Breakdown of coevolution between symbiotic bacteria *Wolbachia* and their filarial hosts. *PeerJ* **4**, e1840.

Lemmon EM, Lemmon AR (2013) High-throughput genomic data in systematics and phylogenetics. *Annual Review of Ecology, Evolution, and Systematics* **44**, 99-121.

Li S-J, Ahmed MZ, Lv N, *et al.* (2017) Plantmediated horizontal transmission of *Wolbachia* between whiteflies. *ISME J* **11**, 1019-1028.

Loman NJ, Pallen MJ (2015) Twenty years of bacterial genome sequencing. *Nat Rev Micro* **13**, 787-794.

López-Giráldez F, Townsend JP (2011) PhyDesign: an online application for profiling phylogenetic informativeness. *Bmc Evolutionary Biology* **11**, 152.

Maiden MCJ (2006) Multilocus sequence typing of Bacteria. *Annual Review of Microbiology* **60**, 561-588.

Maiden MCJ, Bygraves JA, Feil E, *et al.* (1998) Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Sciences* **95**, 3140-3145.

Makepeace BL, Gill AC (2016) *Wolbachia*. In: *Rickettsiales: Biology, Molecular Biology, Epidemiology, and Vaccine Development* (ed. Thomas S), pp. 465-512. Springer International Publishing, Cham.

Martin D, Murrell B, Golden M, Khoosal A, Muhire B (2015) RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evolution* **1**, vev003.

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution* **32**, 268-274.

Nikoh N, Hosokawa T, Moriyama M, *et al.* (2014) Evolutionary origin of insect–*Wolbachia* nutritional mutualism. *Proceedings of the National Academy of Sciences* **111**, 10257-10262.

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. *Current Microbiology* **53**, 388-395.

Pérez-Losada M, Cabezas P, Castro-Nallar E, Crandall KA (2013) Pathogen typing in the genomics era: MLST and the future of molecular epidemiology. *Infection, Genetics and Evolution* **16**, 38-53.

Riegler M, Iturbe-Ormaetxe I, Woolfit M, Miller WJ, O'Neill SL (2012) Tandem repeat markers as novel diagnostic tools for high resolution fingerprinting of *Wolbachia*. *BMC Microbiology* **12**, S12.

Russell JA, Goldman-Huertas B, Moreau CS, *et al.* (2009) Specialization and geographic isolationamong *Wolbachia* symbionts from ants and lycaenid butterflies. *Evolution* **63**, 624-640.

Siozios S, Ioannidis P, Klasson L, *et al.* (2013) The diversity and evolution of *Wolbachia* ankyrin repeat domain genes. *Plos One* **8**, e55390.

Townsend JP (2007) Profiling Phylogenetic Informativeness. *Systematic Biology* **56**, 222-231.

Turner KM, Hanage WP, Fraser C, Connor TR, Spratt BG (2007) Assessing the reliability of eBURST using simulated populations with known ancestry. *BMC Microbiology* **7**, 30.

Weinert LA, Araujo-Jnr EV, Ahmed MZ, Welch JJ (2015) The incidence of bacterial endosymbionts in terrestrial arthropods. *Proceedings of the Royal Society B: Biological Sciences* **282**.

Werren JH, Baldo L, Clark ME (2008) *Wolbachia*: master manipulators of invertebrate biology. *Nature Reviews Microbiology* **6**, 741-751.

Wu M, Sun LV, Vamathevan J, *et al.* (2004) Phylogenomics of the reproductive parasite *Wolbachia pipientis w*Mel: a streamlined genome overrun by mobile genetic elements. *PLoS Biol* **2**, e69.

Zhou W, Rousset F, O'Neill S (1998) Phylogeny and PCR–based classification of *Wolbachia* strains using wsp gene sequences. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **265**, 509-515.

Zug R, Hammerstein P (2012) Still a host of hosts for *Wolbachia*: Analysis of recent data suggests that 40% of terrestrial arthropod species are infected. *Plos One* **7**, e38544.

Zug R, Hammerstein P (2015) Bad guys turned nice? A critical assessment of *Wolbachia* mutualisms in arthropod hosts. *Biological Reviews* **90**, 89-111.

Zug R, Koehncke A, Hammerstein P (2012) Epidemiology in evolutionary time: the case of *Wolbachia* horizontal transmission between arthropod host species. *Journal of Evolutionary Biology* **25**, 2149-2160.