

CURRENT OPINION

Wolbachia megadiversity: 99% of these microorganismic manipulators unknown

Matsapume Detcharoen^{*,†}, Wolfgang Arthofer, Birgit C. Schlick-Steiner[†] and Florian M. Steiner[†]

Department of Ecology, Molecular Ecology Group, University of Innsbruck, Technikerstr. 25, 6020 Innsbruck, Austria

*Corresponding author: Department of Ecology, Molecular Ecology Group, University of Innsbruck, Technikerstr. 25, 6020 Innsbruck, Austria. E-mail: mdetcharoen@gmail.com

One sentence summary: We estimate *Wolbachia* diversity at about 100 000 strains and the portion currently known of this diversity is less than 1%, based on quantitative analyzes of the published literature and of 16S DNA sequences of arthropod-hosted *Wolbachia*.

[†]These authors contributed equally to this paper.

Editor: Marcus Horn

[†]Matsapume Detcharoen, <http://orcid.org/0000-0002-4610-4206>

ABSTRACT

Wolbachia (Alphaproteobacteria) are the most widespread endosymbionts of arthropods, manipulating their hosts by various means to maximize the number of host individuals infected. Based on quantitative analyzes of the published literature from Web of Science® and of DNA sequences of arthropod-hosted *Wolbachia* from GenBank, we made plausible that less than 1% of the expected 100 000 strains of *Wolbachia* in arthropods is known. Our findings suggest that more and globally better coordinated efforts in screening arthropods are needed to explore the true *Wolbachia* diversity and to help us understand the ecology and evolution of these host-endosymbiont interactions.

Keywords: endosymbiont; diversity; arthropods

Wolbachia, bacterial (Alphaproteobacteria) endosymbionts of arthropods and nematodes, manipulate their hosts by inducing reproductive barriers among individuals that carry incompatible strains (Werren et al. 2008). Five of six animal species are arthropods (Roskov et al. 2018), and 40%–60% of these are expected to carry *Wolbachia* (Zug and Hammerstein 2012; Weinert et al. 2015; Sazama et al. 2017), making them world's largest pandemics (Werren et al. 2008; Zug and Hammerstein 2012). It is a truism that large fractions of microbial diversity await discovery (Keller and Zengler 2004), but grasping the extent of this lacuna has remained difficult. *Wolbachia* may offer a special opportunity to quantify this lacuna because of their tight relationship to arthropods. Here, we investigate global trends in exploring *Wolbachia* diversity and how much of this diversity is actually known.

We undertook a standardized, quantitative literature study and used articles from the Web of Science® database to analyze trends of *Wolbachia* research starting in 1995, after which year *Wolbachia*-related articles were published annually. We searched all entries of articles published between 1995 and 2016 and found 2806 *Wolbachia* articles and 752 articles about *Wolbachia* in arthropods (Table S3, Supporting Information). We calculated the annual portion of *Wolbachia* articles of all the articles in field of biology and found that *Wolbachia* research, in total, has been growing faster than the whole field of biology (Fig. 1A). Increasing knowledge in molecular techniques (Zhou et al. 1998) and in the availability of complete genomes (Wu et al. 2004) may have caused this increase. In contrast, the rise of publications on *Wolbachia* in arthropods was likewise faster than that of whole biology until 2003 but has been leveling off since.

Received: 28 June 2019; Accepted: 26 September 2019

© FEMS 2019. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

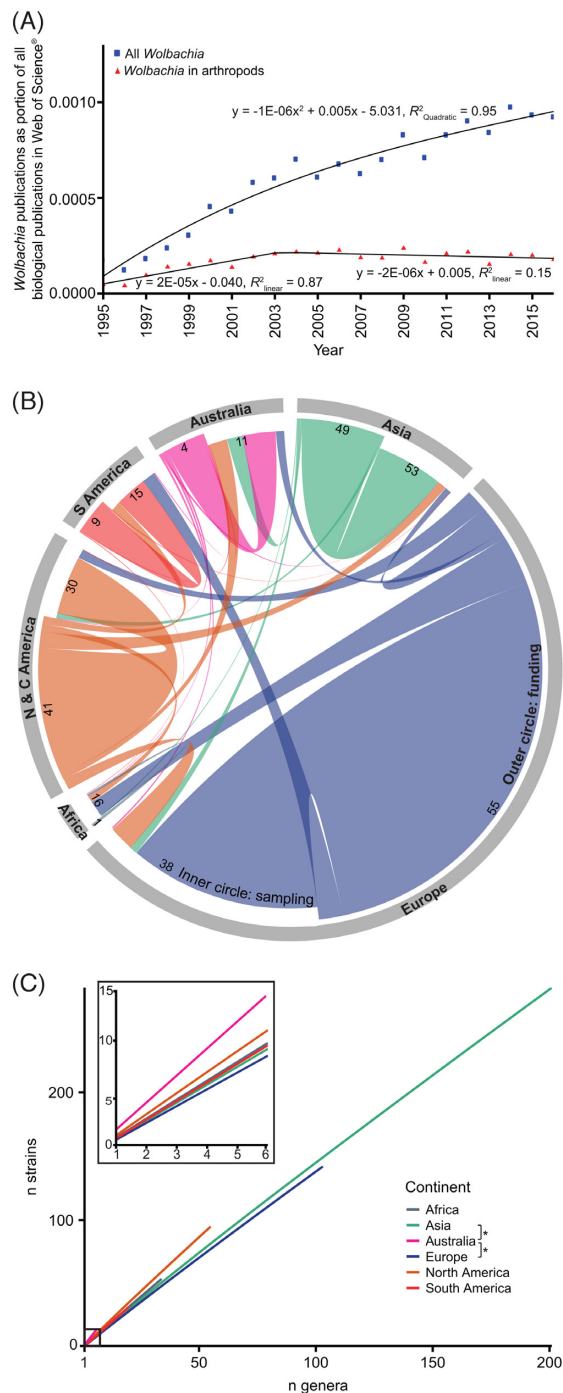


Figure 1. Growth rate of the *Wolbachia* literature, geographic relationship between sampling and funding, and *Wolbachia* strain discovery. (A), Portions of all *Wolbachia* articles (blue) and of articles on *Wolbachia* in arthropods (red) of all biological articles in Web of Science 1995–2016. The results of regression analyses including R^2 are shown. (B), Funding (outer ring) and sampling (inner ring) of articles on *Wolbachia* in arthropods at the continent level; widths standardized by the size of continents as a measure of biodiversity (absolute numbers of articles in numerals). (C), Rarefaction curves illustrating the discovery of *Wolbachia* strains in arthropods depending on the number of host genera analyzed in each continent; significant differences in the discovery rates across continents are marked with asterisks.

We then considered global trends at the continent level in *Wolbachia* research in arthropods using all years together. First, we focused on the geography of funding and sampling. We manually curated the data set by retaining only articles that reported on primary research on *Wolbachia* in arthropods and provided information on both locations of funding agencies and sampling sites. This reduced the data by 752 to 149 articles (Table S5, Supporting Information). Europe played by far the most important role in funding, whereas Africa's role in funding was negligible (Fig. 1B). In terms of sampling activity, Asia scored first and Australia last (Fig. 1B). We also found that most samples were collected in the same continent where funding came from (Fig. 1B). Because of the unequal contribution of continents to global biodiversity (Stork 2018), we then considered the size of a continent as a proxy for its biodiversity and standardized the number of publications by continent size for both funding and sampling. Thereby, Europe's role in both activities became even more prominent, but Africa turned out neglected also in terms of sampling (Fig. 1B, Table S10, Supporting Information). This is in-line with the little overall investment in research and development in Africa compared with the rest of the world (UNESCO Institute for Statistics 2018). Inequalities in sampling and funding among continents might lead to a bias in revealing the true diversity of *Wolbachia* in nature.

Our second focus in analyzing global trends was to consider *Wolbachia* strain discovery at the continent level. We used a rarefaction approach on all primary research studies that reported defined *Wolbachia* strains in arthropods ($n = 366$). In our study, the names of *Wolbachia* strains were used as defined by the authors of the publications included. We note that *Wolbachia* strain identities reported by authors may not reflect the true diversity as *Wolbachia* strain identification methods are still under debate; the use of genomic characterization may change strain identification in the future (Bleidorn and Gerth 2018). When only the supergroup identity was given, we combined the supergroup name with the host genus name and used this combination as a proxy to strain identity; we note that this approach may have underestimated *Wolbachia* diversity in that, in fact, more than a single, host-specific strain per supergroup may exist in a host genus and that it may have overestimated diversity, in that it does not account for infection of multiple genera by the same strain. In all continents, the analysis of new host genera brought to light new endosymbiont strains, that is, the rarefaction curves were nearly linear suggesting an early stage of *Wolbachia* diversity exploration (Fig. 1C). Then, we addressed whether the discovery rate was the same across continents. We found that Australia, although sampled the least in absolute numbers, had a higher discovery rate than the two most sampled continents, Europe and Asia, and probably harbors the highest *Wolbachia* diversity per genus.

Combining our data from all the continents, we saw 513 *Wolbachia* strains in 342 arthropod genera, that is, 1.5 strains per genus. Two issues have to be considered in the context of this, the age of host genera and the number of species per host genus. Considering the first issue in more detail, we had to use genera, but difference in age across genera entails that different taxa are differently represented in our dataset with regard to *Wolbachia* diversity (i.e. a taxon with young genera is over-represented and vice versa). However, first, the same problem applies to species. Second, it is difficult to evaluate the effect of different ages of arthropod genera because time until infection has been acquired has been estimated at about 9 million years (Baillly-Bechet et al. 2017) but the average age of arthropod genera is not known. If

Table 1. Estimates of infected arthropod genera and of percentage of *Wolbachia* strains known. (A), Simulated number (n) of arthropod genera infected (Table S8, Supporting Information) according to three assumed species infection rates (40% (Zug and Hammerstein 2012), 52% (Weinert et al. 2015) and 60% (Sazama et al. 2017)). (B), Expected number of *Wolbachia* strains when assuming the ratio of strains to arthropod genera as found in the analysis using Web of Science® presented in this study. (C), Percentage of known *Wolbachia* strains of expected strains as calculated from minimum (c1) and maximum cutoffs (c2) for operational taxonomic units estimated using sequences from GenBank (Table S7, Supporting Information) and as calculated based on the results of searching Web of Science® (c3).

Assumed percentage of arthropod species infected	Scenarios		
	40%	52%	60%
a. n infected genera (simulation)	66 008	73 891	78 425
b. n <i>Wolbachia</i> strains (1.5 strains per genus, Web of Science)	99 012	110 837	117 638
c1. % <i>Wolbachia</i> strains known (n = 21, GenBank, 7% cut-off)	0.02	0.02	0.02
c2. % <i>Wolbachia</i> strains known (n = 164, GenBank, 1% cut-off)	0.17	0.15	0.14
c3. % <i>Wolbachia</i> strains known (n = 513, Web of Science)	0.52	0.46	0.44

the acquisition time of *Wolbachia* infection and the age of arthropod genera are of comparable dimensions, the effect of different ages across arthropod genera is relevant; if, in contrast, genera get older than the acquisition time of infection, the effect is not relevant. Considering the second issue in more detail, it is clear that species-rich arthropod genera can host more than 1.5 *Wolbachia* strains per genus (for example, see Gerth et al. 2015). In interpreting the relevance of our finding of 1.5 strains per genus on average, obviously also the number of species per genus would have to be taken into account. Unfortunately, we do not know how many species were analyzed per genus because many authors did not identify the host taxa to the species level, but if the average number of species per genus analyzed was lower than the global average number of about 10 species per arthropod genus (1007 839 arthropod species from 96 191 genera), this would entail an underestimation of the actual *Wolbachia* diversity, and vice versa. Using published infection rates (Zug and Hammerstein 2012; Weinert et al. 2015; Sazama et al. 2017), we estimated 66 008–78 425 genera to be infected and assuming that the ratio of 1.5 of *Wolbachia* strains per arthropod genus in our sample is representative for all arthropods, just 0.44%–0.52% of the *Wolbachia* diversity in arthropods—99 012–117 638 strains—is currently known (Table 1).

To validate this literature-based estimate using a complementary approach, we extracted from GenBank all sequences from *Wolbachia* in arthropods of the 16S gene, a gene frequently used for diversity assessments of bacteria, to calculate operational taxonomic units (Data set S1–S4). At a mere 0.02%–0.17%, the resulting portion of known from the estimated global *Wolbachia* strain diversity was even lower than in the literature-based approach (Table 1; Table S9, Supporting Information). We stress that here we do not take into account the ongoing discovery of additional arthropod taxa (Troudet et al. 2017; Stork 2018), which are likely to carry additional *Wolbachia* diversity. We consider it therefore as very safe to conclude that less than 1% of global *Wolbachia* diversity is currently known. To the best of our knowledge, these are the first estimates of global *Wolbachia* strain diversity and its portion known so far.

We cannot fathom what knowing the unknown 99% of *Wolbachia* diversity will mean to ecology and evolution in detail but it should help us understand, for example, host-endosymbiont interactions that define phenotypes including mutualistic ones (Engelstädter and Hurst 2009) and help resolve incompatibility problems in genetic rescue programs (Hamm et al. 2014). Increasing knowledge in molecular evolution could gain answers about host diversification (Engelstädter and Hurst 2009), endosymbiont recombination generating new diversity (Baldo et al. 2006),

the evolutionary tempo of *Wolbachia* (Gerth and Bleidorn 2016) and how mutualism evolves into parasitism and vice versa (Martinez et al. 2015; Zug and Hammerstein 2015; Gerth and Bleidorn 2016). A better knowledge on *Wolbachia* diversity should help us understand the systematics of *Wolbachia* and the still unresolved transformation of supergroups into species (Werren et al. 2008; Ellegaard et al. 2013). Different host species respond differently when infected with the same *Wolbachia* strain, and thus, *Wolbachia* strain improvement is necessary, for example, in application such as Dengue fever control (Moreira et al. 2009). Apart from biocontrol, we should become able to compare the genomes of the entire *Wolbachia* diversity with those of pathogenic Rickettsiales to identify potential candidate regions for human and plant pathogenicity. To facilitate all such progress, we need more field research, more international collaborations, and more funding resources. All this together will compensate the biases in terms of sampling and funding identified here and allow us to understand *Wolbachia*'s coevolution with and effects on their hosts in their entirety.

METHODS

Standardized literature search

Articles about *Wolbachia* in arthropods published between 1995 and 2016 were downloaded from Web of Science® using the Advanced Search option with the search string (TS = (*Wolbachia* AND arthropod*)) AND LANGUAGE: (English) AND DOCUMENT TYPES: (Article), Indexes = SCI-EXPANDED, Timespan = 1995–2016. A total of 752 articles were retrieved (Table S3, Supporting Information). Proceedings, computational articles, reviews and book chapters as well as articles not related to *Wolbachia* and arthropods were excluded, resulting in 557 articles (Table S4, Supporting Information).

All *Wolbachia* articles published in the same period were retrieved using the advanced search string TS = (*Wolbachia*), Indexes = SCI-EXPANDED, Timespan = 1995–2016. To retrieve the number of all articles from the whole field of biology, the same keywords were used as defined in Steiner et al. (2015). The results of searching all *Wolbachia* articles and all biological articles were analyzed by year of publication (Table S1 and S2, Supporting Information, respectively).

For the geographic analysis of funding and sampling, articles reporting sampling sites and funding agencies (Table S5, Supporting Information) were selected. Funding and sampling sites were indexed by the continent they belong to. With some articles, samples and/or funding agencies came from multiple continents; in these cases, all n continents contributing were included for the article and their contributions divided by n. The

development from 1995 to 2016 of the annual numbers of articles on *Wolbachia* in arthropods and of all *Wolbachia* articles as portions of the number of all articles from the field of biology as well as the relationships between sampling and funding sites were calculated in R version 3.5.1 (R Core Team 2018) and were visualized using the package shinyCircos (Yu et al. 2018).

Number of *Wolbachia* strains and infected genera

The genus level was used for arthropod hosts because many articles did not report the species level. The identities of the infected arthropod genera and of *Wolbachia* strains and supergroups were collected. Some articles reported the supergroup but not the strain name of the detected *Wolbachia* strains. In these cases, each combination of a supergroup with a particular host genus was counted as one strain. Regression analysis of *Wolbachia* strains against arthropod host genera found in each continent was done in PAST version 3.21 (Hammer et al. 2001). Statistical difference among rarefaction curves of continents was inferred when 95% confidence intervals did not overlap.

Operational taxonomic units from *Wolbachia* 16S sequences

The following string was used to search for *Wolbachia* 16S sequences in GenBank: ((((((‘*Wolbachia*’[Organism] OR *Wolbachia*[All Fields]) AND 16s[All Fields]) NOT (‘uncultured bacterium’[Organism] OR uncultivated bacterium[All Fields])) NOT genome[All Fields]) NOT chromosome[All Fields]) NOT mitochondria[All Fields]) NOT (‘*Spiroplasma*’[Organism] OR *Spiroplasma*[All Fields])). About 2733 sequences were retrieved. Sequences of *Francisella persica* and of other non-arthropods were excluded from the data set. The sequences were aligned with Clustal Omega (Sievers and Higgins 2014) implemented in CLC Main Workbench 8.1.

Based on sequence lengths and their alignment positions, sequences were divided into two groups, thereafter groups A and B. For each group, one alignment including all sequences trimmed to the shortest and another one excluding some sequences for the sake of alignment length were elaborated. This resulted in alignments of 146, 630, 240 and 380 base pairs, defined as short-A, long-A, short-B, and long-B respectively (Data set S1–4).

The numbers of operational taxonomic units were calculated using DOTUR version 1.53 as implemented in mothur version 1.41.1 (Schloss et al. 2009) from distance files generated with the DNAdist program version 3.69 of the Phylip package (Felsenstein 1993). Distances across sequences were calculated using CLC Main Workbench 8.1. Four different cut-off values were applied to all four alignments: 0.01 (the lowest cut-off commonly used in the literature to define operational taxonomic units in bacteria (Needham et al. 2017)), 0.02 (average distance across strains in our data), 0.03 (the maximum distance among sequences of the same strain) and 0.07 (the maximum distance across strains) (Table S7, Supporting Information).

Infected arthropod genera

The names of 1007839 arthropod species, belonging to 96191 genera, were downloaded from the Catalogue of Life database, version October 30, 2018 (Roskov et al. 2018) (Data set S5). A custom Python script was used to randomly declare *n* percent of the

species as infected based on recent estimates of infected arthropod species (Zug and Hammerstein 2012; Weinert et al. 2015; Sazama et al. 2017) and then count the number of genera harboring at least one infected species. For each *n*, 100 repetitions were performed and their mean values used (Table S8, Supporting Information).

SUPPLEMENTARY DATA

Supplementary data are available at FEMSEC online.

ACKNOWLEDGMENTS

We thank David Reiter for helping with data curation and Michael Gerth, Tania Holtzem, Markus Möst, Patrick Krapf and Ruben Sommaruga as well as three anonymous reviewers for critical comments on earlier manuscript versions.

Conflicts of interest. None declared.

REFERENCES

- Bailly-Bechet M, Martins-Simões P, Szöllősi GJ et al. How long does *Wolbachia* remain on board? *Mol Biol Evol* 2017;34:1183–93.
- Baldo L, Bordenstein S, Wernegreen JJ et al. Widespread recombination throughout *Wolbachia* genomes. *Mol Biol Evol* 2006;23:437–49.
- Bleidorn C, Gerth M. A critical re-evaluation of multilocus sequence typing (MLST) efforts in *Wolbachia*. *FEMS Microbiol Ecol* 2018;94.
- Ellegaard KM, Klasson L, Näslund K et al. Comparative genomics of *Wolbachia* and the bacterial species concept. *PLOS Genet* 2013;9:e1003381.
- Engelstädter J, Hurst GDD. The ecology and evolution of microbes that manipulate host reproduction. *Annu Rev Ecol Syst* 2009;40:127–49.
- Felsenstein J. PHYLIP (phylogeny inference package). Dep Genet Univ Washington:Seattle, 1993.
- Gerth M, Bleidorn C. Comparative genomics provides a time-frame for *Wolbachia* evolution and exposes a recent biotin synthesis operon transfer. *Nat Microbiol* 2016;2:16241.
- Gerth M, Saeed A, White JA et al. Extensive screen for bacterial endosymbionts reveals taxon-specific distribution patterns among bees (Hymenoptera, Anthophila). *FEMS Microbiol Ecol* 2015;91:1–12.
- Hamm CA, Handley CA, Pike A et al. *Wolbachia* infection and Lepidoptera of conservation concern. *J Insect Sci* 2014;14:6.
- Hammer Ø, Harper D, Ryan P. Past: Paleontological statistics software package for education and data analysis. *Paleontol Electron* 2001;4:1–9.
- Keller M, Zengler K. Tapping into microbial diversity. *Nat Rev Microbiol* 2004;2:141–50.
- Martinez J, Ok S, Smith S et al. Should symbionts be nice or selfish? Antiviral effects of *Wolbachia* are costly but reproductive parasitism is not. *PLOS Pathog* 2015;11:e1005021.
- Moreira LA, Iturbe-Ormaetxe I, Jeffery JA et al. A *Wolbachia* symbiont in *Aedes aegypti* limits infection with Dengue, Chikungunya, and *Plasmodium*. *Cell* 2009;139:1268–78.
- Needham DM, Sachdeva R, Fuhrman JA. Ecological dynamics and co-occurrence among marine phytoplankton, bacteria and myoviruses shows microdiversity matters. *ISME J* 2017;11:1614–29.

- R Core Team. R: A Language and Environment for Statistical Computing. 2018.
- Roskov Y, Orrell T, Nicolson D et al.(eds). *Species 2000 & ITIS Catalogue of Life*. 30th October 2018, 2018th ed, Leiden, The Netherlands: Species 2000: Naturalis. 2018.
- Sazama EJ, Bosch MJ, Shouldis CS et al. Incidence of *Wolbachia* in aquatic insects. *Ecol Evol* 2017;**7**:1165–9.
- Schloss PD, Westcott SL, Ryabin T et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 2009;**75**:7537–41.
- Sievers F, Higgins DG. Clustal omega, accurate alignment of very large numbers of sequences. *Methods Mol Biol* 2014;**1079**:105–16.
- Steiner FM, Pautasso M, Zettel H et al. A falsification of the citation impediment in the taxonomic literature. *Syst Biol* 2015;**64**:860–8.
- Stork NE. How many species of insects and other terrestrial arthropods are there on Earth? *Annu Rev Entomol* 2018;**63**:31–45.
- Troudet J, Grandcolas P, Blin A et al. Taxonomic bias in biodiversity data and societal preferences. *Sci Rep* 2017;**7**:9132.
- UNESCO Institute for Statistics. R&D Data June 2018 release. 2018. <http://uis.unesco.org/en/news/rd-data-release>
- Weinert LA, Araujo-Jnr E V., Ahmed MZ et al. The incidence of bacterial endosymbionts in terrestrial arthropods. *Proc R Soc B Biol Sci* 2015;**282**:20150249.
- Werren JH, Baldo L, Clark ME. *Wolbachia*: master manipulators of invertebrate biology. *Nat Rev Microbiol* 2008;**6**:741–51.
- Wu M, Sun L V, Vamathevan J et al. Phylogenomics of the reproductive parasite *Wolbachia pipientis* wMel: a streamlined genome overrun by mobile genetic elements. *PLOS Biol* 2004;**2**:e69.
- Yu Y, Ouyang Y, Yao W. ShinyCircos: an R/Shiny application for interactive creation of Circos plot. *Bioinformatics* 2018;**34**:1229–31.
- Zhou WG, Braig HR, O'Neill SL et al. Cloning and characterization of a gene encoding the major surface protein of the bacterial endosymbiont *Wolbachia pipientis*. *J Bacteriol* 1998;**180**:2373–8.
- Zug R, Hammerstein P. Bad guys turned nice? A critical assessment of *Wolbachia* mutualisms in arthropod hosts. *Biol Rev Camb Philos Soc* 2015;**90**:89–111.
- Zug R, Hammerstein P. Still a host of hosts for *Wolbachia*: analysis of recent data suggests that 40% of terrestrial arthropod species are infected. *PLoS One* 2012;**7**:e38544.