



www.bioinformatics.net  
Volume 19(3)



Research Article

Received March 1, 2023; Revised March 31, 2023; Accepted March 31, 2023, Published March 31, 2023

DOI: 10.6026/97320630019336

**Declaration on Publication Ethics:**

The author's state that they adhere with COPE guidelines on publishing ethics as described elsewhere at <https://publicationethics.org/>. The authors also undertake that they are not associated with any other third party (governmental or non-governmental agencies) linking with any form of unethical issues connecting to this publication. The authors also declare that they are not withholding any information that is misleading to the publisher in regard to this article.

**Declaration on official E-mail:**

The corresponding author declares that lifetime official e-mail from their institution is not available for all authors

**License statement:**

This is an Open Access article which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. This is distributed under the terms of the Creative Commons Attribution License

**Comments from readers:**

Articles published in BIOINFORMATION are open for relevant post publication comments and criticisms, which will be published immediately linking to the original article without open access charges. Comments should be concise, coherent and critical in less than 1000 words.

Edited by P Kanguane

Citation: Sharma & Som, Bioinformatics 19(3): 336-340 (2023)

# Assigning new supergroups V and W to the *Wolbachia* diversity

Amresh Kumar Sharma & Anup Som\*

Centre of Bioinformatics, Institute of Interdisciplinary Studies, University of Allahabad, Prayagraj – 211002, India; \*Corresponding author

**Affiliation URL:**

[www.allduniv.ac.in](http://www.allduniv.ac.in)

**Author contacts:**

AKS: [amresharma1@gmail.com](mailto:amresharma1@gmail.com)

Anup Som: E-mail: [som.anup@gmail.com](mailto:som.anup@gmail.com) & [asom@allduniv.ac.in](mailto:asom@allduniv.ac.in)

**Abstract:**

*Wolbachia* are endosymbiotic and alphaproteobacteria that belong to the order Rickettsiales. They are known to infect half of the insect population and cause host manipulation, and have been categorized into 19 monophyletic lineages called supergroups. Recently, two strains, wCfeJ and wCfeT were isolated from cat fleas (*Ctenocephalides felis*), but their supergroup relationships were not assigned. In this article, we have attempted to classify these two novel strains and establish their evolutionary lineage (i.e., supergroup designation). For this we performed 16S rRNA similarity analysis and reconstructed 16S rRNA phylogeny of 52 *Wolbachia* strains (including two novel strains) belong to 19 supergroups. We also performed average nucleotide identity (ANI) and digital DNA-DNA hybridization (dDDH)

studies to measure genomic similarity between the two novel genomes. The results revealed that 16S rRNA similarity between the two novel strains is 97.94%, which is below the threshold value of 98.6% and phylogeny shows that they are placed at the two different positions (i.e., showing distinct evolutionary lineages). Further, genomic similarity analysis revealed that the novel genomes have ANI and dDDH values 79% and 22.4% respectively, which were below the threshold value of ANI (95%) and dDDH (70%). These results suggested that the novel strains neither shared a species boundary between them nor with any other previously identified supergroups, which designate them as two new supergroups, namely supergroup V (strain wCfeJ) and supergroup W (strain wCfeT).

**Keywords:** *Wolbachia*; Supergroups; 16S rRNA, Phylogeny; Average nucleotide identity (ANI); digital DNA-DNA hybridization (dDDH).

### Background:

*Wolbachia* are alpha-proteobacteria that follows an endosymbiotic life and infect a wide range of arthropods and nematodes [1-2]. These bacteria are gram-negative, obligate and intracellular, and belong to the order Rickettsiales [3]. The genomes of *Wolbachia* have been analyzed to determine the type and nature of symbiotic relationships with their host [4-8]. Their nature of relationships in the hosts is reproductive parasite in arthropods, nutritional mutualists in bed bugs, and obligates mutualism in filarial nematodes [9-10]. *Wolbachia* mediated all the reproductive manipulation in the host (i.e., mostly arthropods and some nematodes), by means of parthenogenesis, feminization, male-killing, by inducing cytoplasmic incompatibility, and nutritional supplement [11-15]. It was estimated that *Wolbachia* infection is up to 40-76% of insects [16-18]. *Wolbachia* have been classified into distinct monophyletic lineages called supergroups, which first came into the appearance in 1998 [19]. Later, Lo et al popularized this concept [20]. Till now, *Wolbachia* have been divided into 19 supergroups namely A-F, H-Q, S-U [21-24]. The 16S rRNA gene is widely used in species determination and identification of new species or strains [25]. Comparison of 16S rRNA gene sequences allows differentiation/delineation of organisms at the genus, species and subspecies level. In the *Wolbachia* classification, the 16S rRNA gene played an important role in identifying and characterizing the new strains. In this paper, we focused on two recently published and

undescribed *Wolbachia* genomes wCfeT and wCfeJ isolated from cat fleas (*Ctenocephalides felis*) found in co-infecting mechanism with the same host with different lifestyle such as strain wCfeJ is parasitic and wCfeT is mutualistic [26]. We used 16S rRNA genes to find out the evolutionary lineages of the novel strains in the *Wolbachia* diversity that currently have 19 supergroups. A total of 50 *Wolbachia* strains from the existing 19 supergroups and two novel strains were used in this study. Further, we did genomic similarity study on the genomes of the two novel strains using average nucleotide identity (ANI) and digital DNA-DNA hybridization (dDDH) test.

### Materials and methods:

#### Data collection:

For finding the supergroup relationships of the two novel *Wolbachia* genomes wCfeT and wCfeJ, first, we took 16S rRNA sequences for phylogenetic analysis because it is highly conserved gene and able to show species delineation. 16S rRNA phylogeny is used to check whether two strains wCfeJ and wCfeT cluster with each other or with any other supergroup(s). For this, we used two novel strains plus other 50 strains from 19 supergroups which consist of a total of 52 *Wolbachia* strains in the study. All the 16S rRNA genes were downloaded from the NCBI database [27]. Details of the 52 *Wolbachia* strains are given in Table 1.

**Table1: Details of 52 *Wolbachia* strains used in this study**

Host Species	Strain	Supergroups	Accession No.
<i>Ctenocephalides felis</i>	wCfeT	Novel strain	*NZ_CP051156.1 (HF197_RS05045)
<i>Ctenocephalides felis</i>	wCfeJ	Novel strain	*NZ_CP051157.1 (HF196_RS00025)
<i>Drosophila melanogaster</i>	wMel	A	*NC_002978.6 (GX67_RS05935)
<i>Nomada panzeri</i>	wNpa	A	*NZ_LYUX01000081.1 (BA050_RS05445)
<i>Nomada flava</i>	wNfla	A	*NZ_LYUW01000079.1 (BA052_RS05590)
<i>Drosophila simulans</i>	wHa	A	*NC_021089.1 (WHA_RS05510)
<i>Drosophila incompta</i>	wInc	A	*CP011148.1 (WG67_RS05300)
<i>Carposina sasaki</i>	wCauA	A	*CP041215.1 (wCauA_RS01020)
<i>Wolbachia sp. (wRi)</i>	wRi	A	*(NC_012416.1 (WRI_RS06005)
<i>Drosophila simulans</i>	wNo	B	*NC_021084.1 (WNO_RS03275)
<i>Culex molestus</i>	wPip Mol	B	*NZ_CTEH01000001.1 (WPM_RS00430)
<i>Nasonia vitripennis</i>	wVitB	B	*NZ_GL883637.1 (WVB_RS0105690)
<i>Aedes albopictus</i>	wAlbB	B	*NZ_RWIK01000001.1 (EJE47_RS00350)
<i>Chrysomya megacephala</i>	wMeg	B	*CP021120.1 (CAI20_RS02875)
<i>Drosophila mauritiana</i>	wMau	B	*CP034335.1 (EJB00_RS01065)
<i>Bemisia tabaci</i>	wBtab	B	*CP016430.1 (BBB02_RS03415)
<i>Onchocerca ochengi</i>	wOo	C	AJ010276.1
<i>Onchocerca volvulus</i>	wOvol	C	AF069069.1
<i>Dirofilaria immitis</i>	wDimm	C	KU255236.1
<i>Brugia malayi</i>	wBm	D	*NC_006833.1 (WBM_RS02885)
<i>Wuchereria bancrofti</i>	wWb	D	*NZ_NJBR02000071.1 (CCY16_RS03430)
<i>Brugia pahangi</i>	wBpah	D	*NZ_CP050521.1 (WBP_RS01130)
<i>Litomosoides sigmodontis</i>	wLsig	D	*CP046577 (GOY07_RS01740)
<i>Folsomia candida</i>	wFol	E	*NZ_CP015510.2 (ASM33_RS07175)
<i>Hypochothionius rufulus</i>	wHruf	E	MN699328.1

<i>Coptotermes acinaciformis</i>	wCaci	F	DQ837197.1
<i>Nasutitermes nigriceps</i>	wNnig	F	DQ837204.1
<i>Cimex lectularius</i>	wCle	F	*NZ_AP013028.1 (WCLE_RS01905)
<i>Madathamugadia hiepei</i>	wMhie	F	*NZ_WQMP01000029.1 (GO685_RS01815)
<i>Zootermopsi angusticollis</i>	wZangu	H	AY764279.1
<i>Zootermopsi nevadensis</i>	wZnev	H	AY764280.1
<i>Orchopeas leucopus</i>	wOleu	I	AY335924.1
<i>Ctenocephalides felis</i>	wCfe	I	AY157512.1
<i>Myodopsylla gentilis</i>	wMgen	I	AY335918.1
<i>Cruorifilaria tubero cauda</i>	wCtub	J	*CP046579 (GOY13_RS01470)
<i>Dipetalonema caudispina</i>	wDcau	J	*CP046580 (GOY14_RS02280)
<i>Bryobia sps</i>	wBry	K	EU499316.1
<i>Radopholus similis</i>	wRsim	L	EU833482.1
<i>Pratylenchus penetrans</i>	wPpe	L	*NZ_MJMG01000007.1 (BIY23_RS03360)
<i>Cinara cedri</i>	wCced	M	JN384079.1
<i>Macrosiphum euphorbiae</i>	wMeu	M	JN109113.1
<i>Pentalonia nigronervosa</i>	wPni	M	KJ786951
<i>Toxoptera auranti</i>	wTaur	N	JN384094.1
<i>Bemisia tabaci</i>	wBt10	O	KF454771.1
<i>Kaburagia rhushicola</i>	wKMKR	O	MT554837.1
<i>Syringophilopsis turdus</i>	wStur	P	KP114103.1
<i>Torotroglia merulae</i>	wTmer	P	KP114102.1
<i>Torotroglia cardueli</i>	wTcard	Q	KP114101.1
<i>Atemnus politus</i>	wApol	S	*NZ_JAAXCS010000017.1 (HET73_RS02035)
<i>Cimex hemipterous</i>	wChem PL13	T	*NZ_CP061738.1 (ID128_RS02485)
<i>Spinturnix mites</i>	wSpin Bat1	U	KP165041
<i>Spinturnix mites</i>	wSpin Bat2	U	KP165042

\*In cases where accession ID for the genes was not available, accession ID of the genome along with locus tag (in the parenthesis) of the gene has been mentioned.

### Sequence similarity measure and Phylogenetic tree reconstruction:

16S rRNA sequence similarity paved a way for species demarcation among all bacterial species. So we firstly performed similarity check on the two novel strains along with other 50 strains using GGDC online server [28]. Further, we did phylogenetic analysis and reconstructed the 16S rRNA phylogeny. We aligned the sequences with CLUSTAL W package [29]. We also performed the model test by using ModelFinder that revealed HKY+F+I+G4 is the best suitable model [30]. Then maximum likelihood (ML) tree was reconstructed by using the IQTREE package with the HKY+F+I+G4 model [31].

### ANI measure and dDDH study:

After 16S rRNA similarity and phylogeny, we also measured the genomic similarity of the novel strains. For that, we performed average nucleotide identity (ANI) test and digital DNA-DNA hybridization (dDDH) study. ANI measures nucleotide-level genomic similarity between the coding regions of two genomes and here we attempted to find the divergence of the genomes to check whether two novel genomes are from the same supergroup or belong to different supergroups. We also carried out dDDH analysis to calculate *in-silico* genome-to-genome comparison using the GGDC tool [28]. The dDDH analysis emerged as an alternative to the tedious wet-lab DNA-DNA hybridization of species delineation. In GGDC tool, we used GBDP (Genome Blast Distance Phylogeny) method to calculate the probability that an inter-genomic distance yielded a dDDH value lower than 70% considered as a novel species-delimitation threshold.

### Results and Discussion:

#### Sequence similarity and phylogenetic analysis of 16S rRNA gene:

At first, we compared 16S rRNA sequence similarity of the two novel strains and found that similarity is 97.94 %, which is below

the previously described threshold for species demarcation 98.6% [32, 33]. This result indicates that the novel strains were not from the same supergroup. Furthermore, we also analysed the sequence similarity of the novel strains with respect to the 50 other *Wolbachia* strains. Similarity results of the novel strains with other supergroups' strains showed that similarity score of the novel strains is lower than the threshold value of 98.6% (Table 2). Overall, 16S rRNA gene similarity result indicates that the novel strains neither belong to the same supergroup nor belong to any other supergroups.

**Table 2:** 16S rRNA sequence similarity of the novel strains (wCfeT and wCfeJ) with respect to the 50 *Wolbachia* strains belong to 19 supergroups.

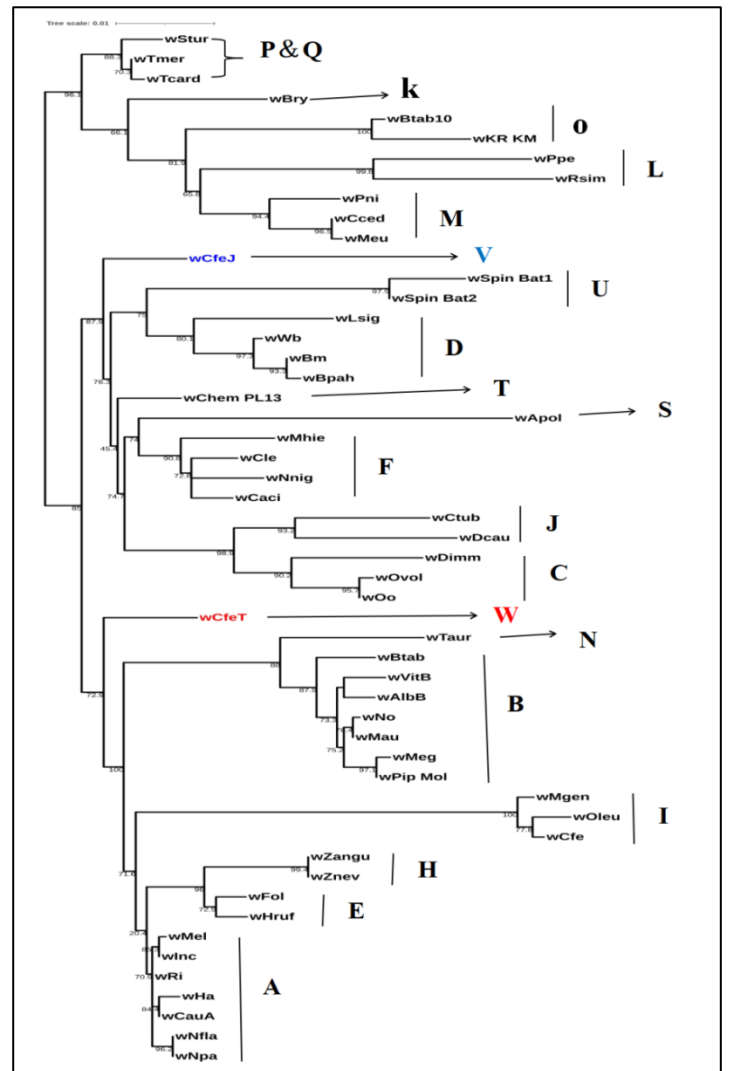
Wolbachia strain	Supergroup	16S rRNA similarity (in %)	
		wCfeT	wCfeJ
wRi	A	98.37	98.47
wInc	A	98.1	98.4
wCauA	A	98.2	98.4
wMel	A	98.34	98.34
wNfla	A	98.47	98.27
wNpa	A	98.47	98.27
wHa	A	98.4	98.2
wPip Mol	B	97.07	97.14
wMau	B	97.07	97.01
wNo	B	97.01	96.94
wAlbB	B	97.27	97.2
wMeg	B	96.94	97.01
wBtab	B	96.87	96.94
wVitB	B	96.74	96.67
wOo	C	96.47	96.47
wOvol	C	96.34	96.47
wDimm	C	96.34	97.07
wBm	D	97.79	97.8
wWb	D	97.6	97.94
wBpah	D	97.34	97.8
wLsig	D	97.27	97.54
wHruf	E	98.38	97.94
wFol	E	98.2	98.01
wMhie	F	97.54	97.87
wCle	F	97.94	98.2

wNnig	F	97.81	97.81
wCaci	F	98.1	98.1
wZangu	H	98.16	98.28
wZnev	H	98.16	98.28
wCfe	I	95.76	95.76
wMgen	I	95.7	95.77
wOleu	I	95.46	95.69
wDcau	J	95.94	96.54
wCtub	J	96.54	97.07
wBry	K	97.47	97.26
wRsim	L	95.97	95.38
wPpe	L	95.93	95.41
wMeu	M	97.78	97.59
wPni	M	97.07	96.94
wCced	M	97.72	97.8
wTaur	N	96.63	96.44
wKR KM	O	96.96	96.77
wBtab10	O	96.65	96.37
wTmer	P	98.23	98.23
wStur	P	97.88	97.88
wTcard	Q	97.92	98
wApol	S	95.5	95.91
wChem PL13	T	98.07	98.47
wSpin Bat1	U	97.69	98.27
wSpin Bat2	U	97.41	97.98

Further, 16S rRNA ML phylogenetic tree was reconstructed using the HKY+F+I+G4 model of nucleotide evolution given in Figure 1. In this tree, two novel strains found in the different evolutionary lineages (i.e., not having a common ancestry) indicating that they did not belong to any other previously described supergroups. Here, we found that strain wCfeJ, shows parasitic nature with its hosts, is placed as an outgroup with supergroups C, D, F, J, S, T, and U. The bootstrap value at the node (87.9%) showing the reliability of the node. And strain wCfeT, shows mutualistic nature with its hosts, is placed as an outgroup with supergroups A, B, E, H, I, and N. The bootstrap value at this node is 72.9%, showing reliability of this node. In summary, 16S rRNA gene phylogeny confirms that strains wCfeJ and wCfeT having different lifestyle and has distinct evolutionary lineages.

#### Genome comparison of novel strains:

The supergroups are sub-species level and their genomes are close enough to each other, so genomic divergence analysis is required for supergroup identification. Accordingly, we performed ANI analysis and dDDH study. We found that ANI and dDDH value between strains wCfeT and wCfeJ were found to be 79% and 22.4% respectively. The threshold value for ANI and dDDH are >95% and >70% respectively when the species belong to the same supergroup [32]. This result clearly indicates that the two novel strains are different. Further, the genomic contents of wCfeJ (NZ\_CP051157.1) were 1.50Mb genome length; 1228 proteins; 35.2 GC%; 3 rRNA; 34 tRNA; 4 other RNAs; 1,463 genes and 194 pseudogenes. For the wCfeT (NZ\_CP051156.1) were 1.20Mb genome length; 1070 proteins; 35.6 GC%; 3 rRNA; 34 tRNA 4 other RNAs; 1,155 genes and 44 pseudogenes. These genomic features show that the novel strains have genomic variations. The genomic analysis results also indicate that the novel strains do not share a species boundary.



**Figure 1:** 16S rRNA gene phylogeny of 52 *Wolbachia* strains with 19 supergroups showing position of two novel strains wCfeT (W) and wCfeJ (V). Maximum likelihood tree was reconstructed by using IQTREE with model HKY+F+I+G4. Supergroups are indicated against the strains.

#### Conclusions:

The results of 16S rRNA based similarity analysis and phylogenetic study, and furthermore genomic ANI and dDDH analyses suggested that the novel strains neither shared a species boundary between them nor with any other previously identified supergroups, which designate them as two new supergroups, namely supergroup V (strain wCfeJ) and supergroup W (strain wCfeT). Therefore, our results aid new insights into the *Wolbachia* diversity and dynamics that will be useful in future comparative studies.

**Acknowledgments:**

The authors thank Aparna Chaturvedi for various useful discussions. AKS thanks the University Grants Commission, India for providing financial support to carry out this research work.

**Conflict of interest:**

The authors declare that they have no conflict of interest.

**References:**

- [1] Sironi *et al. Molecular and Biochemical Parasitology*. 1995 **74**:223–227 [PMID: 8719164].
- [2] Werren *et al. Proc R Soc Lond B*. 1995 **261**:55–63 [PMID: 7644549].
- [3] Harris HL *et al. Symbiosis*. 2010 **51**:37–53 [https://doi.org/10.1007/s13199-010-0065-3].
- [4] Hoerauf *et al. Medical Microbiology and Immunology*. 2003 **192**:211–216 [PMID: 12684759].
- [5] Duron *et al. Heredity*. 2007 **98**:368–374 [PMID: 17519957].
- [6] Hosokawa *et al. Proc Natl Acad Sci USA*. 2010 **107**:769–774 [https://doi.org/10.1073/pnas.0911476107].
- [7] Lindsey *et al. Genes Genomes Genetics*. 2016. **6**:2113–2123 [PMID: 27194801].
- [8] Badawi *et al. Genes*. 2018 **9**:290 [PMID: 29890648].
- [9] Bouchon *et al. Proc R Soc Lond B*. 1998 **265**:1081–1090 [PMID: 9684374].
- [10] Kageyama *et al. Heredity*. 2002 **88**:444–449 [PMID: 12180086].
- [11] Werren *et al. Nat Rev Microbiol*. 2008 **6**:741–751 [PMID: 18794912].
- [12] Cordaux *et al Trends in Genetics*. 2011 **27**:332–341 [PMID: 21663992].
- [13] Miyata *et al. Biol Lett*. 2017 **13**:20170153 [PMID: 28566542].
- [14] Beckmann *et al Nat Microbiol*. 2017 **2**:17007 [PMID: 28248294].
- [15] Perlmutter *et al PLoS Pathog*. 2019 **15**:e1007936 [PMID: 31504075].
- [16] Hilgenboecker *et al. FEMS Microbiology Letters* **281**:215–220 [PMID: 18312577].
- [17] Zug *et al PLoS ONE*. 2012 **7**:e38544 [PMID: 22685581].
- [18] Kajtoch *et al PeerJ*. 2018 **6**:e4471 [PMID: 29568706].
- [19] Zhou *et al Proc Biol Sci*. 1998 **22**:265(1395):509–15 [PMID: 9569669].
- [20] Lo *et al Molecular Biology and Evolution*. 2002 **19**(3):341–6 [PMID: 11861893].
- [21] Lefoulon *et al. BMC Microbiol*. 2020 **20**:188 [PMID: 32605600].
- [22] Laidoudi *et al. IJMS*. 2020 **21**:8064 [PMID: 33138055].
- [23] Konecka *et al. Infection, Genetics and Evolution*. 2021 **91**:104829 [PMID: 33794350].
- [24] Olanratmanee *et al. Southeast Asian J. Trop. Med. Public Health*. 2021 **52**: 48–55 [https://journal.seameotropmednetwork.org/index.php/jtropmed/article/view/396].
- [25] Hassler *et al. Microbiome*. 2022 **10**: 104 [PMID: 35799218].
- [26] Driscoll *et al. PeerJ*. 2020 **8**:e10646 [PMID: 33362982]
- [27] https://www.ncbi.nlm.nih.gov/
- [28] Meier-Kolthoff *et al. BMC Bioinformatics*. 2013 **14**:60 [PMID: 23432962].
- [29] Thompson *et al. Nucleic Acids Res*. 1994 **22**(22):4673–80 [PMID: 7984417].
- [30] Kalyaanamoorthy *et al. Nat Methods*. 2017 **14**:587–589 [PMID: 28481363].
- [31] Trifinopoulos *et al. Nucleic Acids Res*. 2016 **44**:W232–W235 [PMID: 27084950].
- [32] Kim *et al. International Journal of Systematic and Evolutionary Microbiology*. 2014 **64**(2):346–51 [PMID: 24505072].
- [33] Caudill *et al. Microorganisms*. 2022 **10**(3):605 [PMID: 35336180].