



RESEARCH ARTICLE

Widespread phages of endosymbionts: Phage WO genomics and the proposed taxonomic classification of Symbioviridae

Sarah R. Bordenstein ^{1,2*}, Seth R. Bordenstein ^{1,2,3,4}

1 Department of Biological Sciences, Vanderbilt University, Nashville, Tennessee, United States of America, **2** Vanderbilt Microbiome Innovation Center, Vanderbilt University, Nashville, Tennessee, United States of America, **3** Department of Pathology, Microbiology and Immunology, Vanderbilt University, Nashville, Tennessee, United States of America, **4** Vanderbilt Institute of Infection, Immunology, and Inflammation, Vanderbilt University, Nashville, Tennessee, United States of America

* sarah.bordenstein@psu.edu



OPEN ACCESS

Citation: Bordenstein SR, Bordenstein SR (2022) Widespread phages of endosymbionts: Phage WO genomics and the proposed taxonomic classification of Symbioviridae. *PLoS Genet* 18(6): e1010227. <https://doi.org/10.1371/journal.pgen.1010227>

Editor: Ivan Matic, Institut Cochin, FRANCE

Received: November 18, 2021

Accepted: April 29, 2022

Published: June 6, 2022

Copyright: © 2022 Bordenstein, Bordenstein. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its [Supporting Information](#) files.

Funding: This work was supported by National Institutes of Health (NIH, <https://www.nih.gov>) Awards R01 AI132581 and R01 AI143725 to S.R.B. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Abstract

Wolbachia are the most common obligate, intracellular bacteria in animals. They exist worldwide in arthropod and nematode hosts in which they commonly act as reproductive parasites or mutualists, respectively. Bacteriophage WO, the largest of *Wolbachia*'s mobile elements, includes reproductive parasitism genes, serves as a hotspot for genetic divergence and genomic rearrangement of the bacterial chromosome, and uniquely encodes a Eukaryotic Association Module with eukaryotic-like genes and an ensemble of putative host interaction genes. Despite WO's relevance to genome evolution, selfish genetics, and symbiotic applications, relatively little is known about its origin, host range, diversification, and taxonomic classification. Here we analyze the most comprehensive set of 150 *Wolbachia* and phage WO assemblies to provide a framework for discretely organizing and naming integrated phage WO genomes. We demonstrate that WO is principally in arthropod *Wolbachia* with relatives in diverse endosymbionts and metagenomes, organized into four variants related by gene synteny, often oriented opposite the putative origin of replication in the *Wolbachia* chromosome, and the large serine recombinase is an ideal typing tool to distinguish the four variants. We identify a novel, putative lytic cassette and WO's association with a conserved eleven gene island, termed Undecim Cluster, that is enriched with virulence-like genes. Finally, we evaluate WO-like Islands in the *Wolbachia* genome and discuss a new model in which Octomom, a notable WO-like Island, arose from a split with WO. Together, these findings establish the first comprehensive Linnaean taxonomic classification of endosymbiont phages, including non-*Wolbachia* phages from aquatic environments, that includes a new family and two new genera to capture the collective relatedness of these viruses.

Author summary

Despite reduced genome sizes and an obligate intracellular lifestyle, some bacterial endosymbionts contain an ensemble of mobile genetic elements that can influence both

bacterial and eukaryotic host biology. One such exemplar is the genus *Wolbachia* that live within the cells of about half of all arthropod species, making them the most widespread endosymbionts in animals. They are primarily transmitted from mother to offspring and establish various forms of symbiosis ranging from mutualism to reproductive parasitism. Like a matryoshka doll, *Wolbachia*, in turn, harbor various mobile genetic elements, and here we explore the largest of these genetic elements, temperate phage WO, using 150 *Wolbachia* genome sequences. We report patterns of host distribution within and beyond *Wolbachia* that newly extends the distribution of this endosymbiotic phage, identify four variants by their gene synteny and recombinase gene sequence, and present putative functions of highly conserved gene clusters involved in cell lysis, virulence, and rearrangement of the bacterial chromosome. We show how an intensively studied virulence island, Octomom, may have arisen from the splitting of an ancestral phage WO variant. Finally, we propose an established Linnaean classification system within a new taxonomic family Symbioviridae that also includes two new genera.

Introduction

Intracellular, endosymbiotic bacteria comprise some of the most intimate and enduring host-microbe interactions. While reductive evolutionary forces are often presumed to lead to streamlined, tiny genomes, many endosymbionts that host switch contain notable levels of active or relic mobile DNA [1]. An exemplar is the genus *Wolbachia* which harbor transposons [2], temperate phages [3,4], and putative plasmids [5,6]. *Wolbachia* are members of the Anaplasmataceae family [7] that also includes the intracellular genera *Anaplasma*, *Ehrlichia*, *Neorickettsia*, *Aegyptianella*, and several newly classified bacteria. *Wolbachia* occur in a vast number of invertebrates spanning some nematodes and roughly half of all arthropod species, thus making them the most widespread endosymbionts in animals [8]; but unlike its sister genera, it does not naturally occur in mammalian hosts [9]. Transmission routes are predominantly vertical through the germline, and horizontal transmission of *Wolbachia* in arthropods is frequent on an evolutionary timescale [10,11], leading to coinfections and subsequent bacteriophage exchanges in the same host [12–16]. Integrated within the bacterial chromosome, these bacteriophages (collectively termed phage WO) are hot spots of genetic divergence between *Wolbachia* strains [6,17–20].

Many arthropod-associated *Wolbachia* cause various forms of reproductive parasitism including feminization, parthenogenesis, male killing, and cytoplasmic incompatibility (CI). These selfish modifications hijack sex determination, sex ratios, gametogenesis, and/or embryonic viability to enhance the spread of *Wolbachia* through the transmitting matriline [21,22]. Nematode-associated *Wolbachia*, however, generally lack phage WO and more often act as mutualists within their animal host [23,24]. Thus, phage WO was originally hypothesized to contribute to these reproductive manipulations in arthropods through horizontal acquisition and differential expression of parasitism genes that are not part of the core *Wolbachia* genome [20,23,25–28]. Indeed, transgenic expression of two genes from phage WO or WO-like Islands (genomic islands that are associated with and/or derived from phage WO) demonstrated cytoplasmic incompatibility factors *cifA* and *cifB* as the primary cause of *Wolbachia*-induced CI and rescue [29–32]. In addition, transgenic expression of the WO-mediated killing gene *wmk* recapitulates male-specific embryo lethality and is a candidate for male killing [33]. Conversely, lytic activity of phage WO associates with reduced *Wolbachia* densities and CI levels [34].

First observed in 1978 as “virus-like bodies” within the gonads of *Culex pipiens* mosquitoes [35], phage WO is a temperate phage that is integrated in the bacterial genome, termed a prophage, until an event triggers particle production and subsequent lysis of the cell [4,34,36–38]. Unlike phages of free-living bacteria, however, the phage particles of intracellular *Wolbachia* contend with a two-fold cell challenge of bacterial and eukaryotic-derived membranes surrounding *Wolbachia* as well as the cytoplasmic and/or extracellular environments of the eukaryotic host. These unique challenges encountered by phage WO presumably selected for the evolution of a novel Eukaryotic Association Module (EAM) that comprises up to 60% of its genome with genes that are eukaryotic-like in function and/or origin [39]. The phage WO genome also features one of the longest genes ever identified in a phage and an abundance of ankyrin repeat domain genes [20,23,34,40,41], though their function has not been clearly elucidated as it has for the Ankyphages of sponge symbionts that aid in the evasion of the eukaryotic immune system [42]. Given the abundance and importance of phage WO in *Wolbachia* and for understanding genomic flux in endosymbioses worldwide, a firm grasp of its biology, including classification, evolution, and functions, will be important for establishing and comparing the rules across systems of endosymbiotic phages.

Here we survey prophage WO from 150 *Wolbachia* genome assemblies currently available in the NCBI database [43]. We report the patterns of distribution, chromosomal location, and functions of WO, and we propose a Linnaean classification system according to consultation with the International Committee and their guidelines on Taxonomy of Viruses (ICTV) [44,45] in which there are two distinguishable genera within a new taxonomic family encompassing prophages of obligate, intracellular bacteria. We show that WO generally occurs in arthropod-associated *Wolbachia*, and prophage insertions are enriched away from the putative origin of replication in the bacterial chromosome. We fully annotate the EAM boundaries of representative WO genomes and highlight the presence of the CI genes, *cifA* and *cifB*, and a conserved set of eleven genes, defined here as the *Undecim Cluster*. We also establish a new model suggesting Octomom is derived from the EAM of prophage WO, with implications for Octomom-based pathogenicity, and we determine that all intact prophage WO genomes have a putatively novel patatin-based lytic cassette immediately upstream from the tail module. Finally, we report for the first time, to our knowledge, that prophage WO-like variants occur in diverse bacterial endosymbionts as well as metagenomes of putative symbionts from aquatic environments, providing a deeper understanding of WO origins, evolution, and ecology within and between endosymbiotic bacteria.

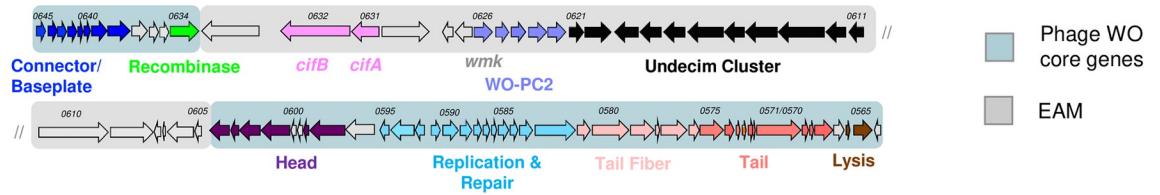
Results

Comprehensive survey of *Wolbachia*'s prophage WO and WO-like islands

Prophage WO elements generally occur in arthropod-associated *Wolbachia*. *Wolbachia* occur in many protosome animal species of the superphylum Ecdysozoa, while prophage WO has previously been described as restricted to arthropod-associated strains. Because WO molecular surveys typically use single gene markers [15,16], we comprehensively explored the NCBI database for prevalence of prophage WO, as determined by presence of one or more core phage WO genes (Fig 1A), throughout all sequenced *Wolbachia* genomes. All *Wolbachia* strains are indicated by a lower-case *w* followed by descriptor of host species, and prophage WO genomes are indicated by a WO prefix followed by the same host descriptor (listed in S1 Table).

Out of 150 assemblies across nematode and arthropod *Wolbachia*, phage WO occurs in arthropod *Wolbachia* with one exception from the mixed host supergroup of F *Wolbachia* (Fig 1B and S1 Table). All arthropod-associated strains contain evidence of intact or relic phage

a Genomic map of prophage WOMelB



b Prophage WO elements across Wolbachia supergroups

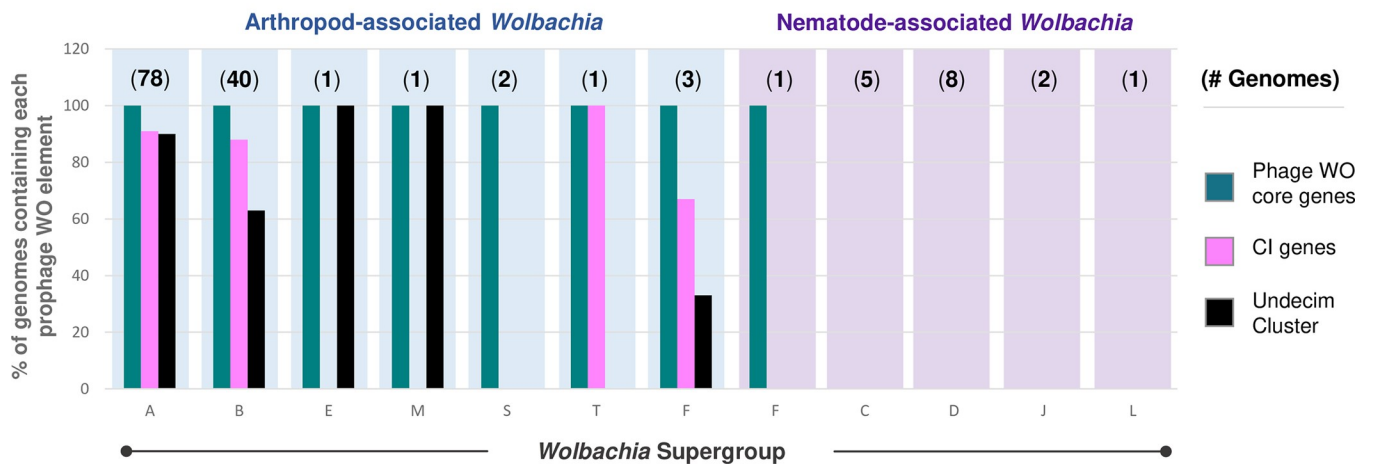


Fig 1. Prophage WO is modular in structure and associates with all arthropod-infecting Wolbachia. (a) A genomic map of prophage WOMelB from the *D. melanogaster* wMel *Wolbachia* strain highlights phage WO core genes in blue and EAM genes in gray. Genes are illustrated as arrows, and direction correlates with forward/reverse strand. The phage WO core consists of recombinase (green), connector/baseplate (royal blue), head (purple), replication and repair (light blue), tail fiber (light pink), tail (salmon), and lysis (brown). The WOMelB EAM encodes *cifA* and *cifB* (pink), WO-PC2 containing HTH_XRE transcriptional regulators (lavender), and a conserved set of genes termed the *Undecim Cluster* (black). (b) At least one phage WO core gene (teal) is associated with all sequenced arthropod-*Wolbachia* Supergroups and Supergroup F, which infects both arthropods (blue) and nematodes (purple). The Undecim Cluster (black) is found in the majority of sequenced Supergroup A, B, E, and M *Wolbachia* genomes, and CI genes (pink) are encoded by the majority of sequenced Supergroup A, B, T, and F genomes. Phage WO elements are absent from all strictly-nematode *Wolbachia* Supergroups. The number of genomes analyzed is listed in parentheses above each Supergroup. Each bar indicates the % of genomes containing each phage WO element. Source data is provided in [S1 Table](#).

<https://doi.org/10.1371/journal.pgen.1010227.g001>

WO, termed WO-like Islands, and the single instance of WO genes in a nematode occurs in strain wMhie from *Madathamugadia hiepei*, a parasite of the insectivorous South African gecko. The wMhie genome encodes four genes that are conserved throughout phage WO’s transcriptional regulation and replication/repair modules (S2 Table) and are not part of the core *Wolbachia* genome. Interestingly, wMhie is a member of Supergroup F that occurs in both arthropods and nematodes. Thus, the presence of phage WO genes in this *Wolbachia* genome may support the horizontal transfer of WO between arthropods and nematodes or indicate an ancestral WO infection that predates the presence of Supergroup F in its nematode host.

In addition to core phage WO genes, we characterized the widespread distribution of two phage WO elements across arthropod *Wolbachia*: (i) the cytoplasmic incompatibility factor genes *cifA* and *cifB* and (ii) Undecim Cluster (Fig 1B). Generally located within phage WO’s Eukaryotic Association Module (EAM [39]; Fig 1A) or in WO-like Islands (genomic islands that are associated with and/or derived from phage WO), *cifA* and *cifB* occur in Supergroups

A, B, F, and T; the latter two are newly reported here. *Wolbachia* strains *wMov* and *wOc* of Supergroup F both encode phylogenetic Type I *cifA* and *cifB* genes, whereas *wChem* of Supergroup T encodes Type II *cifA* and *cifB* genes (S3 Table; See [29,46,47] for a discussion of *cif* Types). Likewise, we identified a highly conserved set of eleven phage WO-associated genes, hereby termed the Undecim Cluster (Fig 1A, discussed below), that is distributed across most arthropod Supergroups but notably absent from all nematode *Wolbachia* genomes.

Characterizing the prophage WO genome

Prophage WO genomes are comprised of conserved structural modules and a Eukaryotic Association Module. Prophage WO genomes have modular organization [18] and thus contain conserved structural gene modules (See discussion in S1 Text) and a Eukaryotic Association Module (EAM) [39]. To date, the EAM is unique to *Wolbachia*'s phage WO and as such is often overlooked by prophage prediction algorithms during the bacterial genome assembly process. Moreover, WO can markedly vary in gene content and synteny, and whether this variation does or does not sort into discrete genomic variants has not been investigated. Thus, we sought to identify conserved and distinguishing genomic features for a comprehensive nomenclature system for the community to classify phage WO major groupings. We mapped and re-annotated prophage WO regions from fully sequenced *Wolbachia* genomes to include the EAM and, more generally, incorporate updated annotations for each module. Together, we propose that all prophage WO genomes comprise a new genus, *Wovirus*, within the class *Caudoviricetes* and new family *Symbioviridae*.

All prophage WO regions were manually curated based on gene content and synteny (Figs 2 and S1–S7) with regards to eight core phage modules (recombinase, replication & repair, head, connector/baseplate, putative tail fiber, tail, putative lysis, and EAM; labeled in Fig 1) and three newly identified and highly conserved gene clusters shown in Fig 2: (i) WO protein cluster 1 (WO-PC1), corresponding to hypothetical proteins WOCauB3_gp2-gp3; (ii) WO protein cluster 2 (WO-PC2), located within the EAM and corresponding to putative HTH_XRE transcriptional regulators, DUF2466 (formerly RadC), and hypothetical proteins WOMelB_WD0622-WD0626; and (iii) the Undecim Cluster, an eleven-gene region located within the EAM and corresponding to WOMelB_WD0611-WD0621.

There are four distinguishable prophage WO variants: sr1WO, sr2WO, sr3WO, and sr4WO. While gene synteny within each core module is generally consistent, the arrangement of modules across prophage genomes is variable and does not correlate with the early organization of *orf7*-based WO clades, WO-A and WO-B [16,48]. To formally update this classification with a more comprehensive classification system, we identified conserved WO loci and modular synteny diagnostic of four WO arrangement groupings. Sequence variation in one gene candidate was consistently associated with similar variation in gene content and synteny: the large serine recombinase [18,49]. Phage-encoded large serine recombinases facilitate integration of the phage genome into specific attachment sites within the bacterial chromosome as well as control the excision, often with the help of an accessory protein, of the prophage genome during the lytic cycle [50]. A BLASTN analysis of the WO serine recombinase gene confirmed that only those associated with comparable WO module arrangement were full-length megablast hits (S4 Table). Phylogenetic analysis of the recombinase peptide sequence also supported four distinct clades of prophage WO (common names sr1WO, sr2WO, sr3WO, and sr4WO; nomenclature proposed in [49] and based on the “serine recombinase”) as well as closely-related recombinases in prophage regions of non-*Wolbachia* endosymbionts, including the *Paramecium* endosymbiont *Holospora obtusa* (Fig 3A). The genomic

WOVIRUS GENOMES

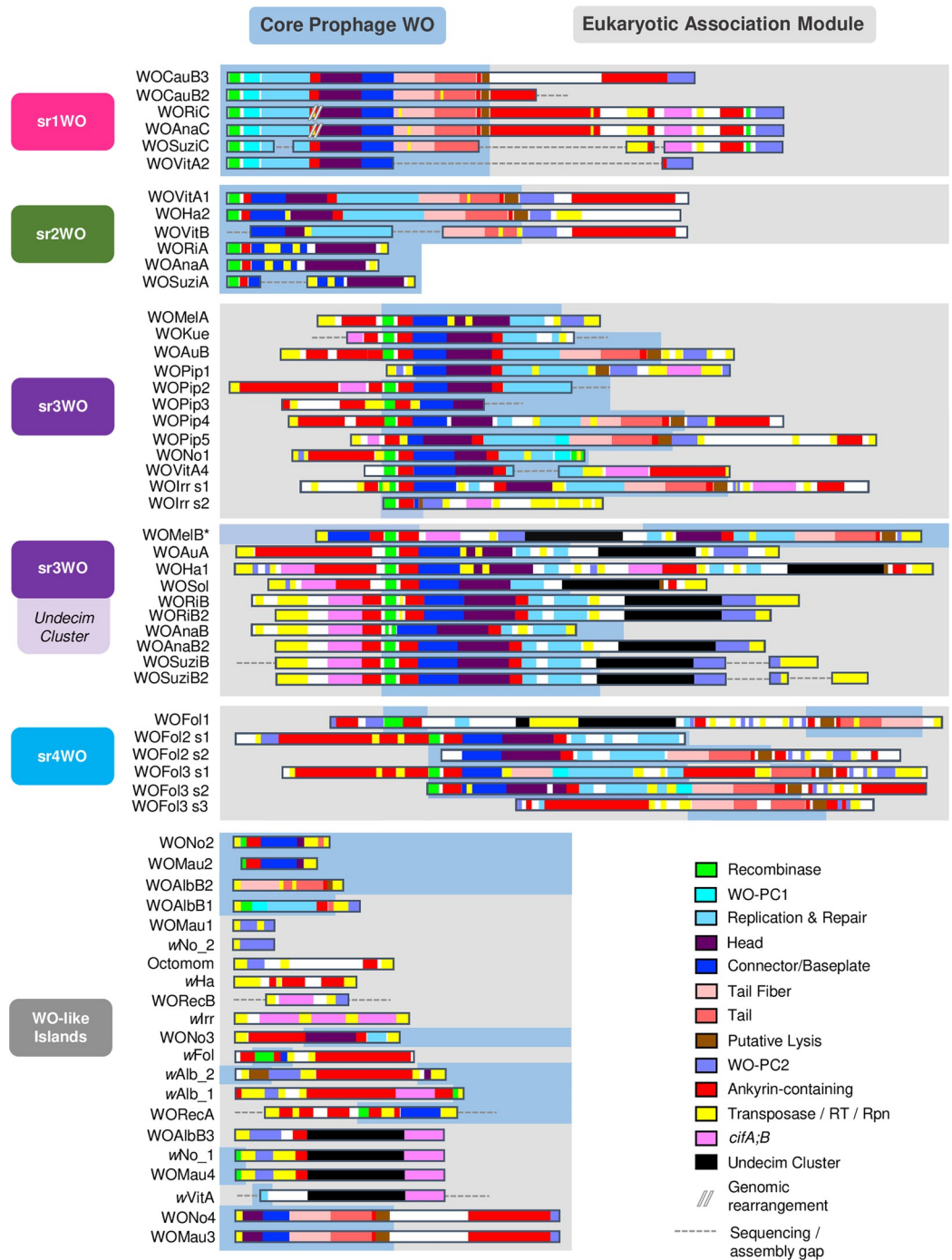


Fig 2. Integrated Wovirus genomes feature distinguishable module synteny. Prophage WO variants are organized by genome content and synteny of their structural modules. Sr1WO and sr2WO feature a 5'-core prophage WO region (blue) and a 3'-EAM (gray). Sr3WO features an internal core prophage WO region that is flanked by EAM genes and mobile elements (yellow). Sr4WO is only present in *wFol* and features three genomic regions with multiple prophage segments. WO-like Islands feature small clusters of prophage WO-like genes; they are comprised of singular structural modules and/or subsets of EAM genes. All modules are color coded: green = recombinase; turquoise = WO-PC1; light blue = replication; purple = head; blue = connector/baseplate; light pink = tail fiber; salmon = tail; brown = putative lysis; lavender = WO-PC2; and black = Undecim Cluster. In addition, ankyrins are shown in red; transposable elements are shown in yellow; and *cifA*/*cifB* are shown in pink. Dotted lines represent breaks in the assembly; module organization is estimated based on closely related variants. Sr1WO is highlighted in hot pink; sr2WO is highlighted in green; sr3WO is highlighted in purple; sr4WO is highlighted in blue; WO-like Islands are highlighted in gray. * The WOMelB genome is rearranged relative to similar variants. Rather than 5'- and 3'-flanking EAM regions, module synteny reflects that of active phage particles whereby the EAM is internally oriented [39].

<https://doi.org/10.1371/journal.pgen.1010227.g002>

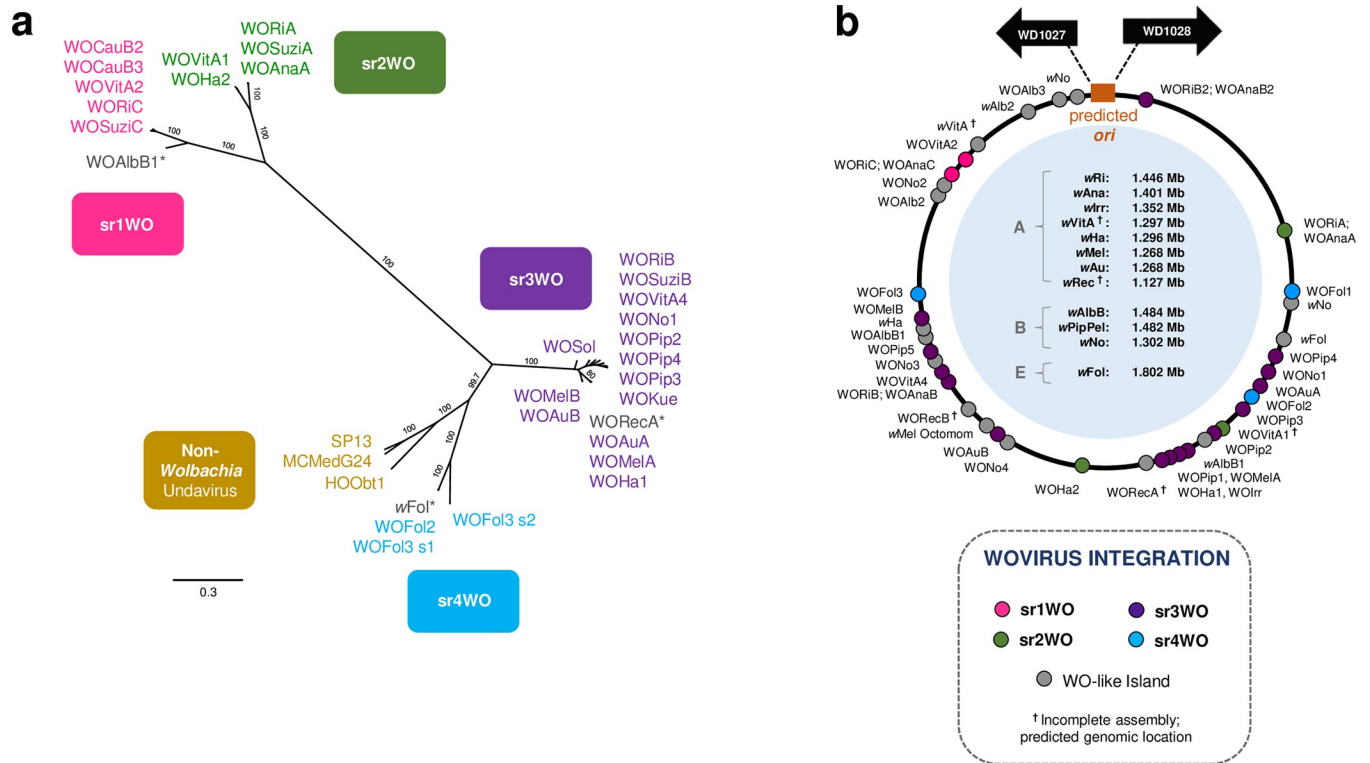


Fig 3. Phylogeny of Wovirus large serine recombinase correlates with module synteny and genomic integration. (a) A phylogenetic tree of the proposed Wovirus recombinase sequence illustrates the utility of this gene as a WO-typing tool to distinguish prophage WO variants. Four distinct clades correlate with sr1WO–sr4WO genome organization shown in Fig 2. Non-*Wolbachia* sequences represent similar prophages (undaviruses, discussed below) from other bacterial hosts, such as the prophage HOObt1 of *Holospira obtusa*, an endonuclear symbiont of *Paramecium*. The tree was generated by Bayesian analysis of 283 amino acids using the JTT-IG model of evolution. Consensus support values are indicated for each branch. (*) indicates that the prophage regions are highly degraded; while they likely originated from the corresponding prophage group, they are now classified as WO-like Islands (S7 Fig). (b) Wovirus integration loci are concentrated opposite the putative origin of replication, *ori*. All *Wolbachia* genomes have been standardized where each dot represents % nucleotide distance calculated by: (nucleotide distance between 5'-WO and *ori* / genome size) * 100. (†) indicates the genome is not closed/circularized; genomic locations are estimated based on alignment of contigs to a reference genome (obtained from authors in [51,52]) and may not reflect true orientation.

<https://doi.org/10.1371/journal.pgen.1010227.g003>

content, organization, and chromosomal integration of each srWO variant are described below.

sr1WO. Most sr1WO recombinases integrate into *Wolbachia*'s magnesium chelatase gene, as we previously reported [39], with portions of the bacterial gene found flanking either side of the prophage region. Two exceptions are in: (i) closely-related *wRi* and *wAna* where the sr1WO prophage has since been rearranged in the *Wolbachia* genome (S1 Fig) with a portion of the magnesium chelatase now associated with each prophage fragment (S8A and S8B Fig); and (ii) *wCauB* which contains at least two sr1WO prophages, and *WOCauB3* has a secondary intergenic attachment site between *sua5* and a hypothetical protein (S8C Fig).

A key characteristic of sr1WO is the single domain HTH_XRE transcriptional regulators of WO-PC2 (S1 Fig, lavender) that are located at the 3'-end of the prophage region. Because the genes are fused in most other WO prophages, they are sometimes annotated as pseudogenes (i.e., *wRi_p006660* and *wRi_p006630* of *WORiC*) in the *Wolbachia* genome; however, conservation across multiple variants suggests they are functional. sr1WOs also lack the methylase/ParB gene that is associated with all other WO prophages. A few genomes (i.e. *WORiC*, *WOAnaC*, *WOSuziC*) harbor *cifA* and *cifB* genes, though the origin of these genes remains inconclusive due to a downstream, highly-pseudogenized sr3WO recombinase (*wRi_p006680*) and

adjacent transposases. Finally, all members of the sr1WO group have a distinct 5'-core-prophage region followed by an ankyrin-rich 3'-EAM (Figs 2 and S1).

sr2WO. sr2WO prophage genes are also organized as 5'-core-prophage followed by 3'-EAM (Figs 2 and S2), yet module synteny is quite distinct from sr1WO: (i) they lack WO-PC1; (ii) the replication, head, and connector/baseplate modules are reversed; and (iii) WO-PC2 is located at the juncture between the core-prophage and EAM regions rather than at the terminal 3'-end of the prophage genome. In Supergroup A *Wolbachia*, the sr2WO recombinase integrates into variable number tandem repeat 105 (VNTR-105) as previously reported [39], a conserved intergenic region used to type closely-related A-*Wolbachia* strains [53]. Similar to the disrupted magnesium chelatase gene flanking sr1WO genomes, disrupted VNTR-105 regions likewise flank the complete sr2WO genome, including the eukaryotic-like *secA* [54] EAM of WOHa2. In newly sequenced B-*Wolbachia* strains, the sr2WO recombinase integrates into specific regions of the bacterial chromosome. Comparative analysis of these regions with a sr2WO-free genome (i.e., *wPip*) can be used to predict prophage and WO-like Island boundaries (listed in S5 Table).

Sr3WO. Unlike the previous groups, sr3WO appears to lack a conserved integration site. Rather, these variants feature a core prophage region that is flanked on either side by EAM regions, are separated from adjacent *Wolbachia* genes by an enrichment of transposase-encoding insertion sequences (Fig 2, yellow and S6 Table), and are concentrated away from the putative origin of replication in the bacterial chromosome (Fig 3B). While their function here is unknown, transposable Mu-like phages replicate via replicative transposition in the bacterial chromosome and, much like phage WO, are associated with severe chromosomal rearrangements and disruptions [55]. Under a similar model, sr3WO transposases could mediate prophage replication and movement throughout the *Wolbachia* genome.

Sr3WO core-prophage module synteny generally resembles that of sr2WO, although a subset of variants also encode an eleven-gene module termed the *Undecim Cluster* (S4 and S5 Figs), discussed in detail below. Most importantly, unlike other prophage WO groups, a majority of the sr3WO variants contain at least one *cifA* and *cifB* gene pair, the locus responsible for *Wolbachia*'s cytoplasmic incompatibility phenotype [29,30,32,46,47].

Sr4WO. The prophage WO group identified strictly in *wFol* of *Folsomia candida* springtails is tentatively labelled sr4WO. Three variants, broken into multiple segments (S6 Fig), loosely resemble the module synteny of sr2WO. WOFol1 is associated with an Undecim Cluster similar to sr3WO, but all variants contain single-domain HTH_XRE genes similar to sr1WO. The sr4WO prophages contain multiple genomic duplications and mobile elements [56]. While they lack *cifA* and *cifB* genes, they are enriched with multiple copies of *ligA* and *resolvase*. More variants of this group are needed to analyze chromosomal integration.

WO-like islands. We identified numerous portions of the prophage WO genome that do not contain enough genetic information to be properly classified. Termed WO-like Islands, they are comprised of single core phage modules, such as a baseplate or tail, and/or genes that are typically associated with the prophage WO genome rather than part of the core *Wolbachia* genome (Figs 2 and S7). Most WO-like Islands are therefore considered “cryptic”, “relic”, or “defective” prophages, and likely originated from an ancestral prophage WO genome where they have since been domesticated by the bacterial host or are in the process of degradation and elimination from the chromosome. Based on studies in other systems, conserved prophage genes or gene modules that are not part of a complete prophage are likely to provide a fitness advantage to their host [57,58] and may interact with, even parasitize, fully intact phages within the same bacterial host [59,60].

Like sr3WO prophages, WO-like Islands are often flanked by at least one insertion sequence (S6 Table) and are commonly associated with CI genes *cifA* and *cifB*. In the unusual

case of the *wIrr* WO-like Island, four CI loci, along with multiple transposases, are arranged in a single genomic cluster that is not associated with conserved WO genes (S7 Fig). We tentatively label the region as a WO-like Island because (i) the *cif* genes and adjacent hypothetical proteins are overwhelmingly associated with prophage WO regions and (ii) there is evidence of a highly disrupted prophage genome about 160kb upstream in the *wIrr* chromosome (illustrated as WO_{Irr} Segment 2 in S4 Fig) that is also enriched with transposases, allowing for the possibility of a prophage WO origin. Such a model for the putative phage WO origin of one highly studied WO-like Island, *wMel*'s Octomom, is discussed in detail below.

Prophage WO is spatially concentrated away from the predicted origin of replication in the *Wolbachia* chromosome. To comprehensively examine the association of each prophage WO variant with its chromosomal location in *Wolbachia*, we mapped integration sites, determined by the recombinase or the most 5'-WO gene, on the chromosome with respect to normalized distance from the putative origin of replication, *ori* [61]. There is a clustering of prophage WO insertion loci, particularly sr3WO, opposite the putative origin of replication (Fig 3B; Chi-square 2-tailed, $p = 0.0035$) that is similar to the localization patterns of temperate phages in *Escherichia*, *Salmonella*, and Negativicutes [62–65]. WO chromosomal location patterns support a model in which prophage insertions and WO-like Islands may not be tolerated in certain regions of the *Wolbachia* chromosome, in this case the region directly surrounding the predicted origin of replication.

Transposable elements may facilitate transposition and domestication of prophage WO regions. In addition to specific chromosomal integration patterns, we next surveyed the relationship between WO and its associated mobile elements. With the exception of WOCauB3, all fully sequenced prophage WO genomes and WO-like Islands contained at least one transposable element beyond the phage recombinase. The diversity of the WO-associated transposable elements by prophage variant is listed in S6 Table and includes (i) transposases of insertion sequence families IS3, IS4, IS5, IS6, IS66, IS110, IS256, IS481, IS630, IS982; (ii) recombination-promotion nuclease (Rpn), which encodes a PD-(D/E)XK nuclease family transposase; and (iii) reverse transcriptase of group II intron origin (RT). WO's transposable elements are associated with the genomic rearrangement (e.g., WORiC), degradation or domestication (e.g., WORiA), and copy number variation (e.g., WORiB) of various prophage genomes. As discussed above, flanking transposases of sr3WO variants may also play a role in replicative transposition similar to phage Mu.

We observed that reverse transcriptases of group II intron origin (RT) are associated with chromosomal rearrangements, insertions, and/or duplications of multiple sr3WO and sr4WO prophages (illustrated in S9 Fig). In the case of WORiB (S9B Fig), the entire prophage region is duplicated in the *Wolbachia* chromosome, whereas other observations involve RT-associated rearrangements within a single integrated prophage region. Likewise, we identified numerous associations of *cifA*;B gene pairs with RTs of sr3WO variants (including WOPip1, WOVitA4, WO_{Irr}, WOHa1, WORiB, WOAnaB, WOSuziB) and the *wIrr* WO-like Island. Therefore, the association of CI loci with transposable elements—both within and beyond prophage regions—could be indicative of post-integration genomic rearrangement and/or domestication of the genes, as previously discussed [6]. Below we propose a detailed model and evidence for the most intriguing RT-associated genomic rearrangement, the origin of *wMel*'s Octomom from prophage WOMe1A to generate a WO-like Island.

Unique characteristics of prophage WO

The WO-Octomom model posits that Octomom is derived from the EAM; *Wolbachia* proliferation may be dependent upon a 1:1 ratio of Octomom: prophage WO. Octomom

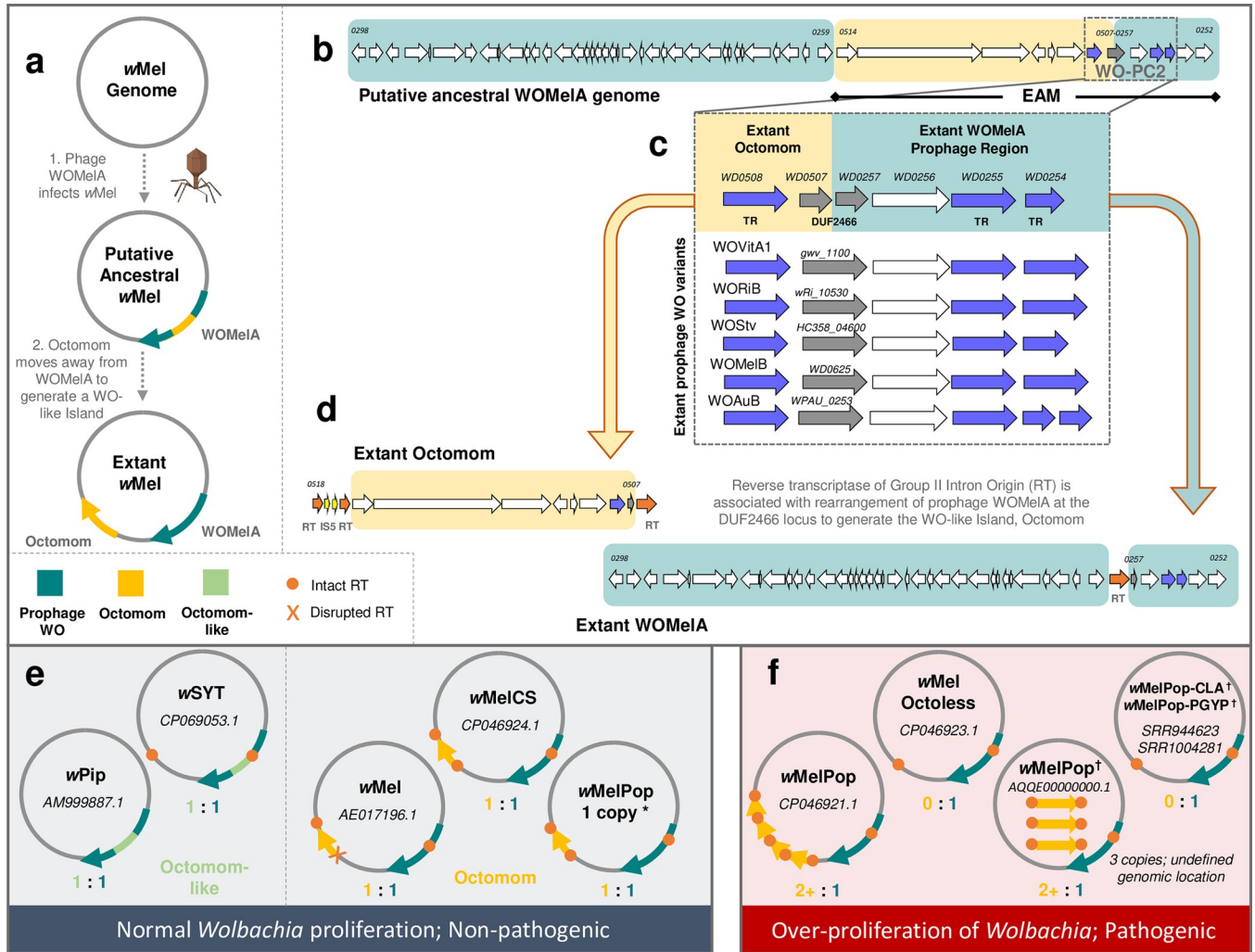


Fig 4. Comparative genomics supports a WO:Octomom origin model for *Wolbachia* proliferation in *wMelPop*. (a) A new model for Octomom origin predicts the initial infection of *wMel* with a WOMeA phage. After integration, Octomom splits from the WOMeA core prophage region to form a WO-like Island. (b) A genome map of the putative, intact, ancestral WOMeA where Octomom is highlighted in yellow and the extant WOMeA genome in teal illustrates placement of Octomom in the WO EAM. (c-d) An alignment of the WO-PC2 region with closely related prophages shows that half of the conserved module (WD0507-WD0508) is now associated with Octomom and the other half (WD0257-WD0254) remained with WOMeA prophage region. DUF2466 is split across the genomic regions and, when concatenated, shares homology to intact DUF2466 genes of WO-PC2 (see S11 Fig). An IS5 insertion (d) is associated with single-copy Octomom stability in the *wMel* chromosome. In *wMelCS*-like genomes, where the flanking RTs are intact (see S10 Fig), Octomom varies in copy number. (e) When Octomom (orange-yellow) and Octomom-like (green, defined by homology to WD0512, WD0513 and WO-PC2; illustrated in S10 Fig) regions exist in a single copy, either within or outside the corresponding prophage region, *Wolbachia* proliferation is normal, and it is non-pathogenic. (f) If the WO-like Island occurs in multiple copies or is absent from the genome, *Wolbachia* over-proliferate and are pathogenic. (*) Restoring the 1:1 (WO:Octomom) ratio returns the *wMelPop* phenotype back to normal levels. The association of Octomom with pathogenicity (i.e., correlation vs. causation) is still to be determined [66–68]. NCBI accession numbers are listed for each genome; (†) indicates circular genomes are unavailable and genomic locations are putative.

<https://doi.org/10.1371/journal.pgen.1010227.g004>

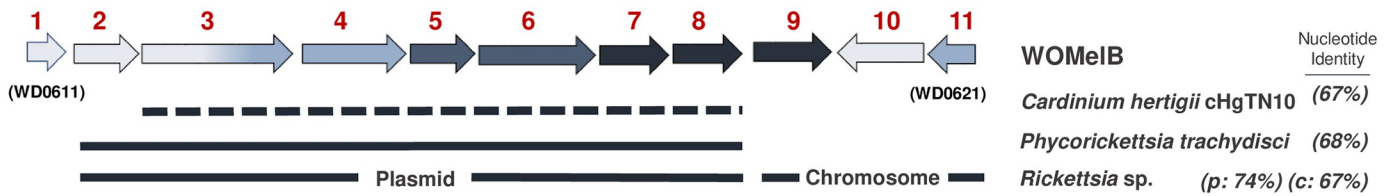
is a cluster of eight genes in the *D. melanogaster wMel Wolbachia* genome that has been described for its resemblance to a bacterial pathogenicity island (see S10 Fig for genome schematic) [69]. Increasing the environmental temperature of flies either containing multiple copies or completely lacking this region results in *Wolbachia* over-proliferation and pathogenicity [67,68]. Based on our observations of RT-associated genomic rearrangement, we present a new WO-Octomom Model (Fig 4A) with genomic evidence (Fig 4B–4D), in which Octomom putatively originated from the EAM of ancestral WOMeA (sr3WO). First, an ancestral phage

WOMeLA with core phage genes as well as an Octomom-encoding EAM infects *wMel* and integrates into the bacterial chromosome. Second, Octomom splits from the prophage EAM region, possibly mediated by RTs, to form an independent WO-like Island about 38kb from the extant WOMeLA (Fig 4A). This is supported by gene synteny of the WO-PC2 variant that is split between Octomom and WOMeLA at the DUF2466 gene (also annotated as *radC*). Notably, by concatenating the two regions at Octomom's WD0507 (5'-DUF2466) and WOMeLA's WD0257 (3'-DUF2466), both the DUF2466 gene sequence (S11 Fig) and module synteny (Fig 4C) form a complete WO-PC2 that closely resembles that of related sr3WO prophages.

Furthermore, Octomom homologs of the two-domain HTH_XRE transcriptional regulator (WD0508) are characteristic of sr2WO and sr3WO prophages, and the *mutL* paralog (WD0509) from Octomom is a phage WO-specific allele [70] that is distinct from the chromosomal *mutL* (WD1306). This supports an ancestral WOMeLA prophage genome comprised of core structural modules and an Octomom-containing EAM with intact WO-PC2 (Fig 4B). An alternative explanation could be that genes WD0512-WD0514 existed as a pathogenicity island in the *Wolbachia* chromosome prior to WOMeLA infection and later acquired adjacent EAM genes from the prophage to form a complete Octomom Island. In this case, we would expect to find at least one other instance of WD0512-WD0514 occurring independent of prophage regions in other *Wolbachia* strains. Instead, the only *Wolbachia* homologs, to date, are associated with the EAMs of WOPip5 and the wSYT (*Wolbachia* of *Drosophila santomea*, *D. yakuba*, and *D. teissieri*, respectively) prophages [6,19,71] (S10 Fig). Likewise, Octomom may have arisen from WOMeLB or another prophage that has since been lost from the genome. However, based on the presence of an intact WO-PC2 in WOMeLB and sequence data confirming that the Octomom DUF2466 pseudogene aligns nearly perfectly with that of WOMeLA, a model of WOMeLA-origin for Octomom is the most parsimonious explanation.

An interesting and robust correlation of this WO-Octomom Model is that one copy relative to prophage WO, either within or outside of the prophage region, is always a distinguishing factor of non-pathogenic *Wolbachia* (Fig 4E), while absence or multiplication of Octomom are notably associated with *Wolbachia* over-proliferation and pathogenicity (Fig 4F). This has been previously reported in context of the *Wolbachia* chromosome [66,67], and we make the distinction here of a prophage association to enable a more fine-tuned exploration of Octomom biology. For example, the disruption (*wMel*) or absence of one (*wSYT*) or both (*wPip*) flanking RTs correlates with a static 1:1 ratio of the Octomom-like region (i.e., containing WD0512-WD0513 and a transcriptional regulation gene) and its corresponding prophage genome (Fig 4E). Conversely, the region is flanked by identical RTs on either side in all *wMel* clade VI strains, including *wMelCS* and the dynamic *wMelPop* that ranges from 0 to multiple copies of the WO-like Island (Fig 4F; *wMel* phylogeny presented in [66,72]). When the 1:1 ratio in clade VI strains is disrupted, possibly in conjunction with flanking RTs, *Wolbachia* develops a pathogenic relationship with its animal host [66,72]. The possible association of RTs with Octomom copy number is also notable due to the observed dependence of both RT activity [73,74] and *wMelPop* pathology [67,68] on environmental conditions, such as temperature. The direct role of Octomom on host phenotype is a subject of debate [66,67], and understanding the association of prophage WO with this region, if any, could inform the biology of this unique system. The two phage-derived regions, for example, may share a common regulatory mechanism since the proposed ancestral splitting of Octomom from WOMeLA broke a cluster of transcriptional regulators, namely one transcriptional regulator (WD0508) from the other two (WD0254 and WD0255) that would typically form an intact module. Alternatively, a split of Octomom from its associated prophage genome may influence epigenetic modifications via WOMeLA's adenine methylase (WD0267; see [66] for a discussion of epigenetic vs. genetic factors).

a *Undecim Cluster*



b

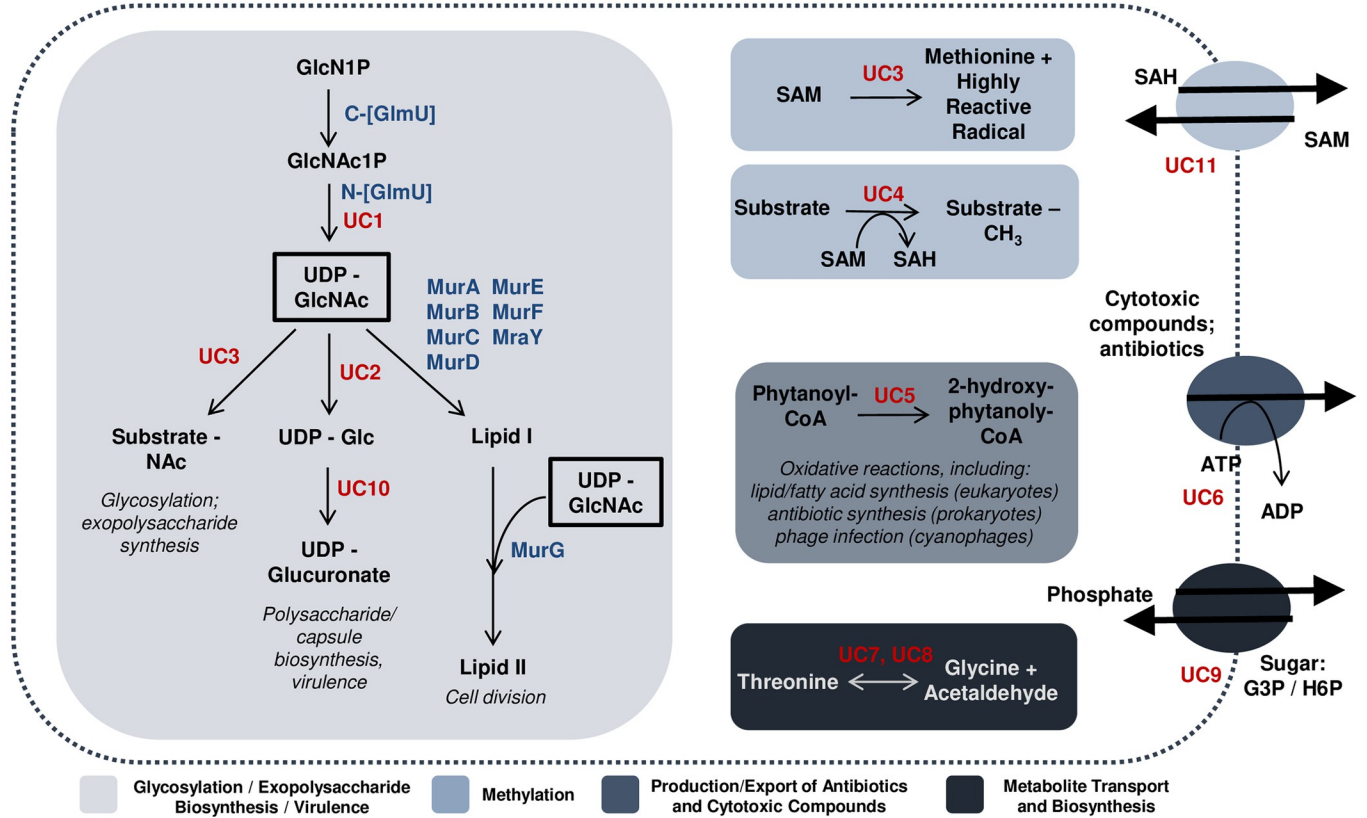


Fig 5. The Undecim Cluster contributes a wide range of cellular processes associated with host-symbiont interactions. (a) A genome map illustrates prophage WO’s Undecim Cluster. Gene labels UC1—UC11 correlate with *wMel* locus tags WD0611–WD0621. Lines under the genes indicate lateral gene transfer events of this region between *Cardinium hertigii* cHgTN10, *Phycorickettsia trachydisci*, and multiple strains of *Rickettsia*, including the *Rickettsia* endosymbiont of *Ixodes scapularis* (REIS) and its plasmid (pREIS2). Nucleotide identity is listed to the right. Dashed lines indicate that the region is not contiguous in the genome. UC1 shares partial homology with a core *Wolbachia* gene, *glmU* (WD0133) and was either not involved in the transfer event or has since been lost from non-*Wolbachia* genomes. (b) A cellular model illustrates the putative functions associated with this region. Cellular reactions are highlighted in boxes and membrane transporters are drawn as ovals. *Wolbachia* genes are labeled in blue; Undecim Cluster genes are labeled in red. UC3 (WD0613) is a fusion protein with an N-terminal glycosyltransferase and C-terminal radical SAM domain; therefore, it is listed twice. Reactions in light gray (UC1, UC2, UC3, and UC10) are likely precursors to multiple pathways in glycosylation, exopolysaccharide biosynthesis, cell division, and/or virulence. Light blue (UC3, UC4, and UC11) is associated with methylation; dark gray (UC5 and UC6) is associated with the production and export of antibiotics and cytotoxic compounds; and dark blue (UC7, UC8, and UC9) is associated with metabolite transport and biosynthesis. The above functions are predicted based on annotation and homology to other systems. Given the contiguous conservation of the Undecim Cluster throughout prophage WO, all functions, including those not captured in this model, are likely interrelated and influence host-symbiont dynamics.

<https://doi.org/10.1371/journal.pgen.1010227.g005>

Undecim cluster is a unique eleven gene island associated with prophage WO. Another “pathogenicity island” candidate in the *Wolbachia* chromosome is a highly conserved set of genes (WD0611 to WD0621; Fig 5A) defined here as the *Undecim Cluster* (*Undecim* is Latin for “eleven”). We identify it in the majority of WO-containing *Wolbachia* genomes (Fig 1B), particularly in association with *cifA*- and *cifB*-encoding regions of sr3WO (S4 and S5 Figs) and WO-like Islands (S7 Fig). Unlike sr3WO prophages themselves, however, the Undecim Cluster does not occur more than once per *Wolbachia* genome. Its complete absence from both *wPip*

and *wRec* suggests that it is not strictly required for *Wolbachia*'s intracellular survival and/or ability to induce cytoplasmic incompatibility. Rather, it may contribute to variation in host-symbiont interactions [18,48] by encoding a broad spectrum of metabolic functions and transport potential [75,76], including cellular exopolysaccharide and/or lipopolysaccharide (LPS) biosynthesis (WD0611-WD0613; WD0620), methylation (WD0613-WD0614; WD0621), production and export of antibiotics and cytotoxic compounds (WD0615-WD0616) and metabolite transport and biosynthesis (WD0617-WD0619) (Fig 5B). It was identified in phage particle genomes from both *wVitA* and *wCauB* [39], indicating that the region may be transferred between *Wolbachia* strains via the phage. In addition, both RNA-SEQ [77] and mass spectrometry data [75] show that the region is highly expressed. Interestingly, ten of the eleven genes were involved in a lateral gene transfer event between *Wolbachia* and the *Rickettsia* endosymbiont of *Ixodes scapularis* (REIS; [17,76]) with WD0612 to WD0618 sharing 74% nucleotide identity to a region of the Rickettsial plasmid pREIS2 and WD0619 to WD0621 sharing 67% identity to a region of the bacterial chromosome (Fig 5A). We also identified homologs in *Cardinium hertigii* cHgTN10 (CP029619.1; 67% nucleotide identity) and *Phycorickettsia trachydisci* (CP027845.1; 68% nucleotide identity). While not contiguous in *C. hertigii*, adjacent transposases may have facilitated post-integration rearrangement.

Phage WO putatively harbors a novel lytic cassette. The most direct impact on *Wolbachia* cellular biology is the potential for phage WO to induce cell lysis [34,78]. The mechanism of phage-induced cell lysis has been well documented and generally involves a three-component lysis system in gram-negative infecting phages: endolysin, holin, and spanins [79]. This genetic system is noticeably absent from prophage WO genomes, and peptidoglycan, the bacterial target of canonical phage endolysins, has never been detected in *Wolbachia* [80]. We therefore hypothesized that WO phages encode an alternative lytic pathway. The top candidate is a putative and novel patatin-based lytic cassette immediately upstream from the tail module [81].

The cassette contains a patatin-like phospholipase A₂, a small holin-like protein, and an ankyrin-repeat protein. A few prophage WO variants (i.e., WOVitA1, WOAuB, WOPip1, WOPip4, and WOPip5) additionally encode an endonuclease of the phospholipase D family. Patatin-like proteins determine virulence in multiple gram-negative bacteria and specifically facilitate disruption of host cell membranes by *Pseudomonas aeruginosa* and *Rickettsia typhi* [82,83]. They are significantly more common in pathogenic bacteria and symbionts than in non-pathogens, suggesting a role in host-association [84]. Holins are not easily annotated because they do not share conserved domain sequence homology, yet several lines of evidence suggest the small protein adjacent to patatin is a "holin-like" candidate: it (i) encodes a single N-terminal transmembrane domain with no predicted charge; (ii) features a C-terminal coiled coil motif; (iii) is smaller than 150 amino acid residues; and (iv) has a highly charged C-terminal domain (S12A Fig) [79,85,86]. In addition, homologs of this holin-like gene in prophages from bacterial chromosomes other than *Wolbachia* (e.g., a Tara Oceans Prophage SP13 and *Holospora* sp.) are directly adjacent to a GH108 lysozyme, further supporting its holin-like potential (Figs S12B and S12C and 6). The third conserved gene in this module, an ankyrin repeat protein with a C-terminal transmembrane domain, may have the potential to impact membrane stability similar to spanins of the traditional phage lysis model; alternatively, they may play a role in evasion of the arthropod-host immune response similar to those in sponge-associated Ankyphages [42]. Together, this module is fairly conserved across tailed WO phages and is a likely candidate in the exit and/or entry of phage particles through *Wolbachia*'s multiple membranes.

Other prophage genes in the *Wolbachia* chromosome are gene transfer agents (GTAs). In addition to prophage WO, we identified several non-WO prophage genes (S13 Fig) in the majority of *Wolbachia* Supergroups, including those of the filarial nematodes.

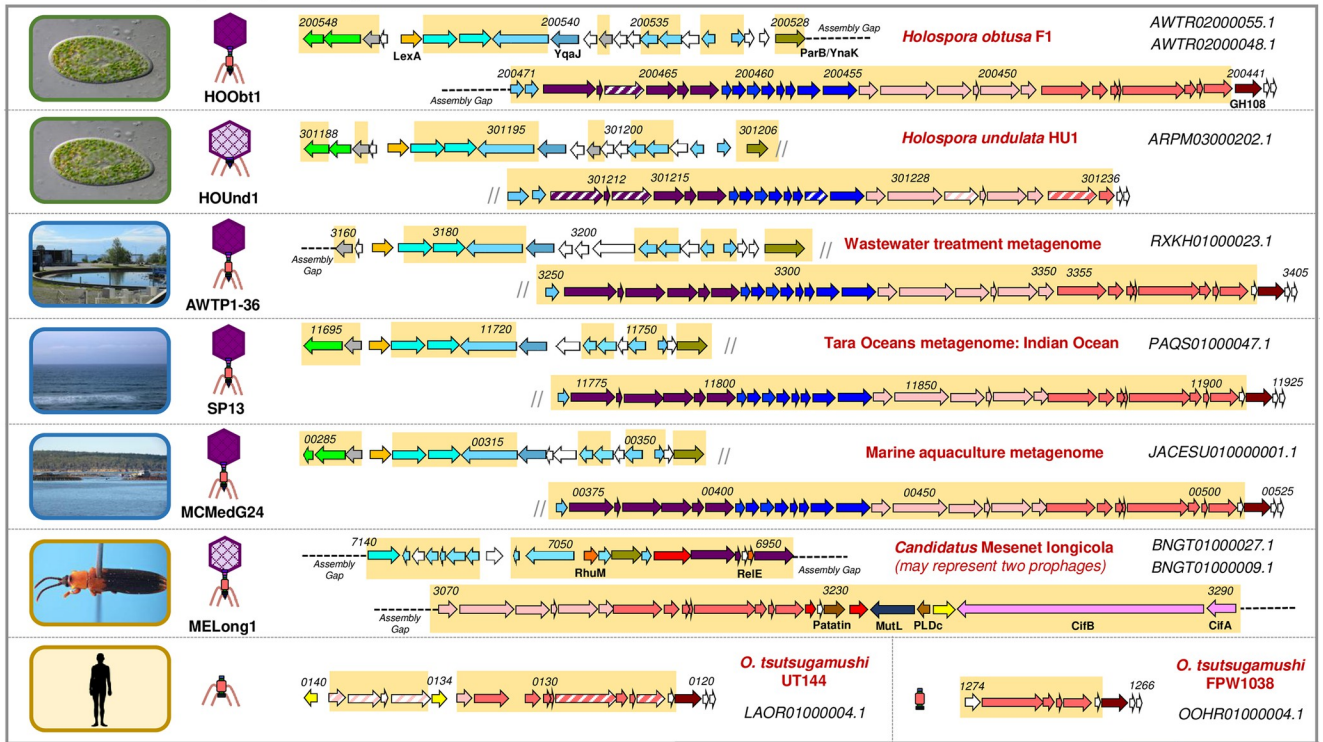
Similar to the well-studied GTA of *Rhodobacter capsulatus* (RcGTA; [87,88]), at least six of these genes encode *E. coli* phage HK97-like conserved domains (S7 Table). We also identified GTA terminase genes associated with the *Wolbachia* chromosome. As previously reported for *Rickettsiales*, the GTA loci are found in multiple locations across the genome rather than organized in an identifiable prophage-like cluster [89]. To investigate the evolutionary relationship of the GTA genes with their *Wolbachia* host, we performed individual nucleotide alignments and recovered two highly conserved genetic groups that demarcate Supergroup A and B *Wolbachia* (S14 Fig), supporting vertical descent with modification across these major Supergroups. While absent from Supergroups J and L of nematodes, they are present across all other *Wolbachia* Supergroups as well as the closely related genera *Candidatus Mesenet*, *Anaplasma*, *Ehrlichia*, and *Rickettsia* (S13B Fig). These results imply that *Wolbachia*'s GTA genes are vertically inherited, codiverge with their bacterial hosts, and likely functional given their intact sequences. They are, however, distinct from phage WO, not indicative of former WO-infections, and may be lost during genome reduction.

Prophage WO beyond *Wolbachia*

Prophage WO-like variants occur in diverse bacterial endosymbionts and metagenomes. We identified multiple prophage WO-like variants beyond the *Wolbachia* genus that have gene synteny and nucleotide identity to prophage WO structural modules in: (i) endonuclear bacterial symbionts of *Paramecium* (*Holospora obtusa*, *H. undulata*, *H. elegans*, and *H. curviuscula*) [90]; (ii) metagenome projects from an advanced water treatment facility [91], the Indian Ocean (*Tara* Oceans circumnavigation expedition [92]), and a marine aquaculture habitat [93]; (iii) *Candidatus Mesenet longicola*, the CI-inducing bacterial endosymbiont of *Brontispa longissima* [94]; and (iv) multiple strains of *Orientia tsutsugamushi* isolated from humans (Fig 6A). While the structural genes closely resembled those of prophage WO, novel genes were identified in the replication/repair and lysis modules (Fig 6A, genes with prophage WO homology are highlighted in yellow). All non-*Wolbachia* variants except *Candidatus Mesenet longicola* lacked signature *Wolbachia* phage WO genes such as patatin, ankyrin repeats, and the EAM that are putatively or definitively involved in phage-by-arthropod interactions.

Relative to the full-length genomes recovered from *Holospora*, *Candidatus Mesenet longicola* and the metagenome projects, *Orientia* prophages appeared to be highly degenerate. These regions featured only tail and lysis genes, but the modules are noticeably intact. Some WO-like Islands, such as WOAlbB2, WONO4, and WOMau3 (Fig 6B), also harbor sole tail and lysis modules. The retention of a complete phage structural module in the bacterial chromosome suggests that it has been domesticated and adapted to benefit the host. For example, several studies report phage-derived bacteriocins that consist of tail and lysis genes and target other strains of the same bacterial species [57]. Similarly, an extracellular contractile injection system (eCIS) comprised of phage tail-like proteins specifically targets eukaryotic cells [95]. Overall, the presence of WO-like variants in non-*Wolbachia* genera continue to support phage WO lateral transfer between unrelated, coinfecting symbionts. This is further evident by the presence of the CI genes, *cifA* and *cifB*, in the *O. tsutsugamushi* genome [96], which may represent a derived variant of phage WO from *Wolbachia* that has since been domesticated by its bacterial host. Alternatively, the association of CI genes in a bacterium harboring WO-like variants could be indicative of two other possible origins—either the last common ancestor of the WO and WO-like phages encoded *cifA* and *cifB*, or the loci may have originated in WO-like phages and transferred to *Wolbachia*. For divergent, horizontally transferred elements, it is often not possible in practice to assign a direction of evolution and origin story.

a WO-like Prophage Regions Beyond *Wolbachia*



b *Wolbachia* WO-like Islands featuring Tail and Lysis genes

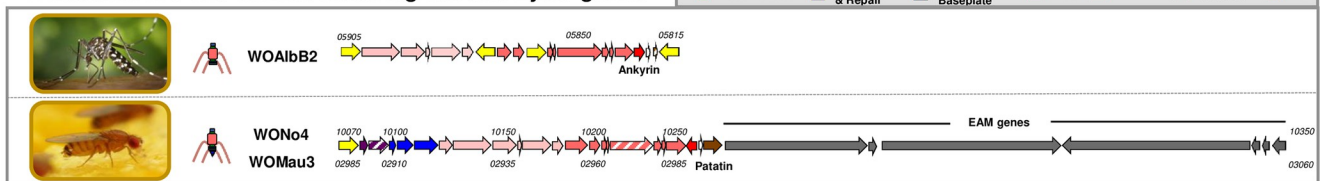


Fig 6. WO-like prophage regions are found in endonuclear *Paramecium* endosymbionts, aquatic environments, and other animal-associated bacteria. (a) Genome maps of non-*Wolbachia* prophage regions illustrate similar gene content and synteny to prophage WO. Locus tags are listed in italics above the genes; NCBI contig accession numbers are shown in the right-hand corner of each genome. Dashed lines represent breaks in the assembly whereas small diagonal lines represent a continuation of the genome onto the next line. Genes with nucleotide homology to prophage WO are highlighted in yellow and genes of similar function are similarly color-coded according to the figure legend. *Candidatus Mesenet longicola* is the only genome to feature EAM genes, including *cifA* and *cifB*. Arrows with diagonal stripes represent genes that may be pseudogenized relative to homologs in other prophage genomes. Genome maps for *H. elegans* and *H. curviuscula* prophages are not shown. (b) WO-like Islands featuring tail and lysis genes share homology with the *Orientia* regions and may represent phage-derived bacteriocins. Predicted physical structures are illustrated to the left of each genome. Images illustrate the isolation source for each prophage: green borders represent protozoa; blue borders represent aquatic environments; gold borders represent animals. All images are available under creative commons or public domain; attribution information is provided in S8 Table.

<https://doi.org/10.1371/journal.pgen.1010227.g006>

Linnaean classification of phage WO. Finally, while phage WO is a model organism to study the tripartite association between viruses, endosymbiotic bacteria, and animal hosts, it is not yet recognized by the International Committee on Taxonomy of Viruses (ICTV). Recently, the ICTV Executive Committee implemented a pipeline for the official classification of viruses from metagenomic datasets [45], including those originating from integrated prophage sequences. Through our comparative analysis of prophage WO sequences here with those that have been sequenced from active particles (i.e., WOVitA1 and WOCauB3), we propose a formal phage WO taxonomy (Fig 7) to align with the ICTV Linnaean-based classification code [44]. The correlation between common name and proposed scientific name for each taxonomic rank is listed in Table 1.

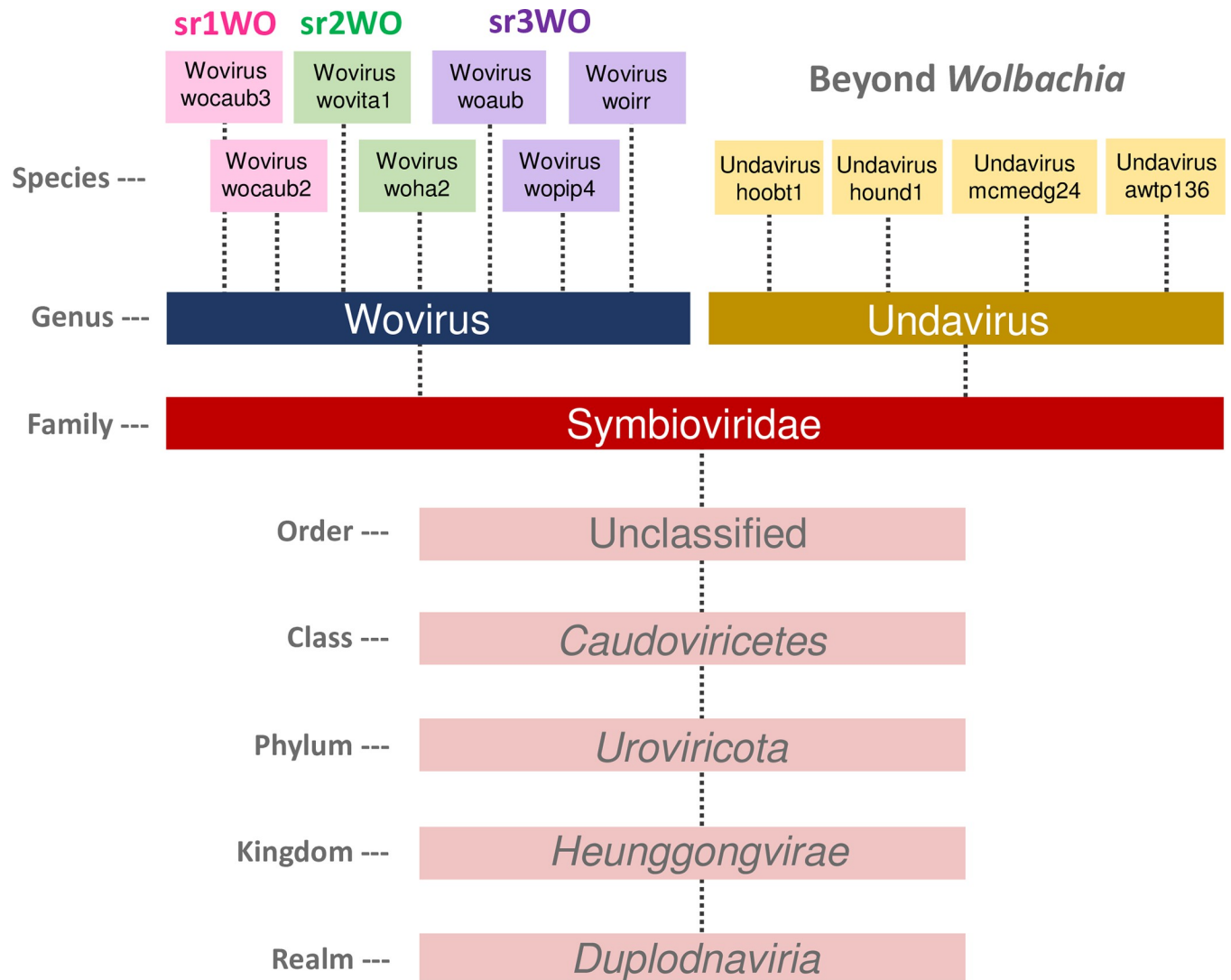


Fig 7. Comparative genomics support a new family-level designation for prophage WO classification. Symbioviridae is proposed as a new taxonomic family of tailed phages within the class *Caudoviricetes*. It contains viruses that primarily infect *Wolbachia* (proposed genus *Wovirus*) and other symbionts including *Holospira* and metagenome-assembled genomes (MAGs) from aquatic environments (proposed genus *Undavirus*).

<https://doi.org/10.1371/journal.pgen.1010227.g007>

Table 1. The correlation between common name and proposed scientific name is listed for each phage WO exemplar variant and taxonomic rank.

WO Exemplar Variant	Taxonomic Rank	Common Name	Proposed Scientific Name
WOCauB3	Species	WOCauB3	Wovirus wocaub3
	Genus	Phage WO	Wovirus
	Family		Symbioviridae
WOVitA1	Species	WOVitA1	Wovirus wovita1
	Genus	Phage WO	Wovirus
	Family		Symbioviridae
WOAuB	Species	WOAuB	Wovirus woaub
	Genus	Phage WO	Wovirus
	Family		Symbioviridae

<https://doi.org/10.1371/journal.pgen.1010227.t001>

Distinguishing Family Traits		Symbioviridae		
Host: symbiotic bacteria		✓		
Core phage modules: recombinase, replication, head, connector/baseplate, tail fiber, contractile tail, lysis		✓		
Large serine recombinase (typing gene)		✓		
PAAR gene in connector/baseplate module		✓		

Distinguishing Genus Traits		Wovirus	Undavirus
EAM		✓	Absent
Ankyrin-repeat containing proteins		✓	Absent
HTH_XRE transcriptional regulators		✓	Absent
Putative lysis gene		Patatin	GH108

Distinguishing srWO Traits		sr1WO	sr2WO	sr3WO
Chromosomal integration		Magnesium chelataase gene or Sua5-intergenic region	(A) VNTR-105 or (B) intergenic region	Flanked by EAM, mobile elements
Structural module synteny: baseplate -> head -> replication & repair -> tail		✓	----	----
Structural module synteny: replication & repair -> head -> baseplate -> tail		----	✓	✓
Location of EAM in prophage genome		3'-end	3'-end	Flanking both sides
Direction of ankyrin adjacent to late control gene (relative to tail/patatin)		Opposite	Same	Same
WO-PC1		✓	Absent	Some
ParB		Absent	✓	✓
Transcriptional regulators (HTH_XRE)		1-domain	2-domain	2-domain
MutL		Absent	✓	Some

Fig 8. Linnaean classification of prophage WO-like viruses is supported by taxonomic traits at the family and genus level. (a) Proposed family Symbioviridae encompasses viruses that infect symbiotic bacteria, contain a large serine recombinase for integration and a Proline-Alanine-Alanine-arginine repeat (PAAR) gene in the connector/baseplate module, and feature a conserved set of core phage modules. They share nucleotide homology to *Wolbachia*'s prophages. (b) Genera are distinguished by presence (Wovirus) or absence (Undavirus) of an EAM and ankyrin repeat containing proteins. Woviruses may utilize patatin for lysis whereas undaviruses encode a canonical GH108 endolysin. (c) Proposed Wovirus clades are further distinguished by multiple factors including structural module synteny, HTH_XRE domains, and genome composition.

<https://doi.org/10.1371/journal.pgen.1010227.g008>

We propose that all phage WO and WO-like viruses be classified in existing class *Caudoviricetes* (phylum *Uroviricota*; kingdom *Heunggongvirae*; realm *Duplodnaviria*) for tailed phages based on the presence of a tail module and observed tail-like structure in electron microscopy [34,78]. We propose the new family Symbioviridae to recognize the association of these viruses with endosymbionts. Two proposed genera, Wovirus and Undavirus, highlight the first bacterial host identified for the genus (*Wolbachia* endosymbionts of arthropods) and the aquatic environment of protist hosts and metagenomic assemblies (“unda” is Latin for water in motion or wave), respectively. Modules shared across the proposed Symbioviridae family are recombinase, replication, head, connector/baseplate, tail fiber, tail, and a putative lytic cassette (See Fig 8 for a summary of taxonomic traits).

The suggested genus Wovirus encompasses all phage WO and prophage WO variants and is distinguishable by the presence of EAM and eukaryotic-like genes, a patatin-like phospholipase, and multiple ankyrin repeat containing proteins (Fig 8).

Within Woviruses, srWO clades loosely correlate with subgenus-level rankings. Sr1WO core module synteny (replication, head, connector/baseplate) is inverted relative to other

members of the proposed Wovirus genus; the ankyrin located between the tail module and putative lytic cassette is encoded on the opposite strand; and the genome does not contain a methylase/ParB protein (S1 Fig). Current members of the second group, sr2WO, feature discrete integration into the VNTR-105 locus (*A-Wolbachia*) or intergenic regions (*B-Wolbachia*), and the recombinase is adjacent to ankyrin repeats rather than WO-PC1. Finally, sr3WO is the most speciose group of Symbioviridae and, likewise, features the greatest number of degraded prophage regions both within and across diverse *Wolbachia*.

Finally, the WO-like prophages of *Candidatus Mesenet longicola* are likely classified as Wovirus due to nucleotide homology of structural genes and the presence of *cifA:B* containing EAM, but complete sequence information (specifically the recombinase and 5'-region beyond the CI loci) is necessary to definitively classify these phages. Likewise, the *wFol* prophages will remain as “unclassified” until more genomes are sequenced to provide definitive taxonomic characteristics for the sr4WO variants. As more prophage WO genomes are sequenced, we propose using the srWO designation as a “common name” that roughly correlates with subgenus-level demarcation.

The proposed genus Undavirus includes the WO-like prophages from most non-*Wolbachia* metagenomic sequences and is currently comprised of phages from aquatic endosymbionts. They lack an EAM and ankyrin repeat containing proteins, feature a GH108 hydrolase rather than patatin-like phospholipase in the putative lytic cassette, and encode LexA and YqaJ that are generally absent from Wovirus genomes (Fig 6). The first representatives of this genus were identified in *Holospira* spp., endonuclear symbionts of *Paramecium caudatum* and *P. bursaria* [97].

Many Symbioviridae prophage regions do not fulfill all requirements for taxonomic classification. In general, we propose that regions with fewer than two structural modules and/or regions lacking definitive taxonomic traits be considered WO-like Islands. If a region contains all structural modules but lacks the recombinase marker (i.e., WOPip5 or AWTP1-36), putative classifications may be assigned only if gene content, module synteny, and integration patterns distinctly satisfy a taxonomic clade. This will likely result in many prophages being assigned to the proposed genus Wovirus but not a defined srWO group.

In summary, we propose that viruses should be classified as Symbioviridae based on nucleotide homology and shared gene content with the core prophage WO genome. The large serine recombinase can be used as a typing tool (Fig 3A) to distinguish srWO groups of Wovirus and intact genomes for inclusion should include (i) recombinase, (ii) replication and repair, (iii) connector/baseplate, (iv) tail fiber, (v) tail, and (vi) lytic modules. Woviruses are delineated by the presence of a eukaryotic association module (EAM), multiple ankyrin repeats, and a patatin-containing lytic module. Undaviruses are characterized by the absence of an EAM, lack of ankyrin repeats, and a GH108-containing lytic module.

Discussion

The survey of 150 genomes coupled with manual annotations and comparative sequence analyses offers the most comprehensive overview of *Wolbachia* prophage WO genomics, distribution, and classification to date. From these analyses, we propose four major prophage WO variants and support the creation of a new family Symbioviridae (within the *Caudoviricetes*) containing two distinct genera, Wovirus and Undavirus. Results presented above suggest that tailed, intact prophage WO genomes serve as a proxy for estimating prophage autonomy vs. domestication in the *Wolbachia* genome where multiple “degraded” prophages and WO-like Islands are indicative of prophage WO domestication by the bacterial host. WO regions enriched with transposable elements contribute to genome plasticity of the bacterial chromosome and may play a role in the domestication of these prophages. One such region, Octomom,

has a putative WO origin in which a former EAM region is dynamically replicated or eliminated, and is associated with pathogenicity when not in a 1:1 ratio with its ancestral prophage. Finally, while there is currently no transformation system for *Wolbachia*, future applications may take advantage of conserved integration loci associated with each srWO group and utilize the serine recombinase to introduce new genetic material into the bacterial chromosome.

Methods

Prophage WO genome maps and chromosomal integration patterns

Prophage WO regions were manually retrieved from sequenced *Wolbachia* genomes in GenBank via BLASTN searches against each individual *Wolbachia* genome in the Nucleotide (NR/NT) and WGS databases [43]. Genomes from WOCauB3, WOVitA1, WOMelB, WOPip5, and WOFol3 were the primary reference genomes used for each search, and all results were manually inspected and annotated. Because most prophage regions were incomplete and located at the ends of contigs, we selected more complete assemblies for comparative genomics: *wRi*, *wAna*, *wSuzi*, *wVitA*, *wHa*, *wMel*, *wPip*, *wNo*, *wAu*, *wIrr*, *wFol*, *wAlbB*, *wMau*, and the previously described prophage genomes WOKue, WOCauB2, WOCauB3, WOSol, WOREcA, and WOREcB (See S1 Table for accession numbers). All genomes were reannotated in Geneious Prime v2019.2 using the InterProScan [98] plug-in along with information from BLASTP [99], Pfam [100], HHPRED [101], ISFinder [102], and SMART [103] databases. Prophages were then organized into groups based on similar gene content and module organization. Whole genome alignments were performed with the Mauve [104] plug-in in Geneious.

Prophage genomic boundaries for sr1WO and sr2WO were defined by 5' and 3' homology to a known *attP* site (discussed below). Prophage genomic boundaries for sr3WO and sr4WO were identified by translating each prophage gene and “walking out” from the structural modules by using a BLASTP (presence/absence) of each gene product against the core *Wolbachia* genome. If a gene was identified in most *Wolbachia* strains, including those infecting nematodes, as well as in the closely related genera *Ehrlichia* and *Anaplasma*, it was considered a core *Wolbachia* gene and not included in the prophage annotation. If a gene was only present in WO-like regions of other *Wolbachia* genomes or highly divergent taxa (i.e., arthropods, protists, or distant non-Rickettsiales bacteria), it was considered a phage-associated gene. Because the HTH_XRE transcriptional regulators (WO-PC2) were identified in phage purifications from WOCauB3 and WOVitA1, any genes located between the structural modules and WO-PC2 were considered part of the prophage genome. Through this method, we identified flanking 5' and 3' transposases that separated phage-associated genes and the bacterial chromosome in sr3WO and sr4WO regions. Because some transposable elements did not fall within the known IS Groups for *Wolbachia* [2], they were comparably annotated to IS Family using ISFinder.

A step-by-step example of the proposed “walk out” method, specifically developed for prophage WO regions, from the WOIrr recombinase (S4 Fig) identified a putative genomic boundary between E0495_01785 and E0495_01780:

- E0495_01810: An ankyrin-repeat containing protein with only phage-associated *Wolbachia* strains in the first 100 BLASTP results. The closest homolog, *wTei* with 89.70% identity, is also adjacent to a prophage region. Using a revised query that eliminated all *Wolbachia* results, the top homologs were phage WO (E-value 0), arthropods (2e-139), and divergent bacteria not inclusive of the closely related *Ehrlichia* and *Anaplasma*. This gene was therefore predicted to be a mobile element rather than part of the core *Wolbachia* genome.
- E0495_01805: IS630 family transposase

- E0495_01800: IS630 family transposase
- E0495_01795: The top 100 BLASTP results were all arthropods (E-value 0). If this assembly is correct, the gene is predicted to be a mobile element rather than part of the core *Wolbachia* genome.
- E0495_01790: IS5 family transposase
- E0495_01785: TPR containing protein from arthropod associated *Wolbachia*. The only nematode associated *Wolbachia* in the top 100 results was WO-containing *wMhie* from *Madathamugadia hiepei*. Additional results were *Symbiodinium* dinoflagellates (7e-132), *Pelagomonas* heterokont algae (2e-109), *Aureococcus* heterokont algae (2e-100), and *Mycobacterium* (7e-79). *Ehrlichia* and *Anaplasma* were absent from default BLASTP results. Together, these results suggest that the gene is not part of the core *Wolbachia* genome.
- E0495_01780: 2-oxoglutarate dehydrogenase complex dihydrolipoyllysine-residue succinyltransferase. The top 100 hits included phage-free nematode *Wolbachia* such as *wBm* and *wWb* (E-value 0), *Ehrlichia* (E-value 0), and *Candidatus* Neoehrlichia (1e-170). This gene was predicted to be part of the core *Wolbachia* genome and was not included in the prophage WO region.
- E0495_01775: hydroxymethylbilane synthase. The top 100 hits included phage-free nematode *Wolbachia* (E-value 0), *Candidatus* Neoehrlichia (1e-112), *Ehrlichia* (1e-106), and *Anaplasma* (1e-104). This gene was predicted to be part of the core *Wolbachia* genome and was not included in the prophage WO region.

To assess the correlation between large serine recombinase and gene synteny, a BLAST search (Megablast Nucleotide) was performed for each prophage WO grouping using WOCauB3 (sr1WO), WOVitA1 (sr2WO), WOMelB (sr3WO) and WOFol2 (sr4WO) as query sequences under default parameters. The top 100 sequences featuring >50% query coverage were manually inspected for gene synteny. sr1WO regions were generally defined by the following orientation: recombinase >> WO-PC1 >> replication and repair >> head >> connector/baseplate >> tail fiber >> tail >> putative lysis >> EAM. sr2WO and sr3WO regions were generally defined by the following orientation: recombinase >> ankyrins >> connector/baseplate >> head >> replication and repair >> tail fiber >> tail >> putative lysis >> EAM. sr2WO was further distinguished by flanking *Wolbachia* genes at the 3'-end of the recombinase whereas sr3WO was further distinguished by flanking EAM and transposase genes at the 3'-end of the recombinase. Sr4WO recombinase gene synteny was highly variable and did not feature any hits beyond *wFol*. BLAST hits with lower % identity values that likely correlated with degraded and/or domesticated portions of the prophage region were listed as "partial."

Chromosomal integration patterns were analyzed by similarly aligning all circular genomes based on the putative origin of replication, *ori* [61]: WD1027 (CBS domain-containing)-like genes were oriented in the reverse direction and WD1028 (*hemE*)-like genes were oriented in the forward direction. The nt-distance from *ori* to the prophage recombinase, or 5'-gene, was divided by the length of the total *Wolbachia* genome and multiplied by 100 for a % distance from *ori*. The *wVitA* and *wRec* genome arrangements may not be exact as they contain multiple scaffold breaks and genome orientation was estimated based on homology to closely related genomes.

Recombinase homology and phylogenetics

Large serine recombinase genes from each reference genome were translated and aligned using the MUSCLE [105] plugin in Geneious. The best model of evolution, according to

corrected Akaike information criteria, was determined by ProtTest [106,107] and the phylogenetic tree was constructed using default parameters of the MrBayes [108] plugin in Geneious with Rate Matrix = jones and Rate Variation = invgamma. A Consensus Tree was built with a support threshold of 50% and burn-in of 10%.

Phage WO *att* sites

The *attP* sites for WOVitA1 and WOCauB3 were previously identified by sequencing active phage particles and confirmed with PCR and Sanger sequencing [39]. Each *attP* sequence was submitted as a BLASTN query against *Wolbachia* genomes harboring similar prophage haplotypes to identify specific *attL* and *attR* sites by manually inspecting alignments for regions of ~100% identity. The *attB* sites were predicted by concatenating chromosomal sequences adjacent to *attL* and *attR*. The predicted *attB* sites were then used as reciprocal queries in a BLASTN search against *Wolbachia* genomes to confirm that similar sequences exist, uninterrupted, in chromosomes lacking these prophage variants.

To determine WORiC *attP*, the *wMel* magnesium chelatase gene (*attB*) was BLASTed against the *wRi* genome (CP001391.1) and each alignment was inspected for regions of shared identity. Of the 45-nt sequence listed in S8 Fig, for example, nucleotides 22–45 share 100% identity to the 3'-end of WORiC (*attR*); nucleotides 1–23 share 100% identity to the 5'-end of WORiC (*attL*). Similarly, the *wPip* intergenic region between WP0133 and WP0134 was used as the *attB* query to confirm WOCauB3's *att* sites.

Phage WO beyond *Wolbachia*

Contigs containing WO-like prophage regions in *Holospira*, *Orientia*, *Candidatus Mesenet*, and multiple metagenome-associated taxa were identified during the prophage WO manual curation and annotation process. If a non-*Wolbachia* hit appeared in the BLASTP result, the nucleotide sequence for each homolog (usually a contig in the WGS database) was manually inspected for WO-like regions. If detected, the boundaries of each prophage region were determined using the similar “walk out” BLASTP approach described above, looking for homology (presence/absence) to other phage or bacterial genes. All non-Anaplasmataceae prophage genomes had concise boundaries (recombinase and lysis module) that did not include an EAM.

Identification of gene transfer agents

The genome annotations used for comparative genomics were manually inspected for keywords *phage*, *capsid*, and *tail*. Any gene not within an annotated prophage WO region was translated and a BLASTP was performed against the NCBI database. Based on top hits, genes were binned into “WO-like” indicating homology to phage WO and “GTA” indicating homology to HK97 phage. The NCBI Conserved Domain E-values from the GTA BLASTP analysis are listed in S7 Table.

Taxonomic classification

The proposed taxonomic classification of phage WO was drafted in accordance with ICTV guidelines for genome-based taxonomy [109] and will be formally reviewed by the Committee in the next cycle. Specifically, it is recommended that phages should be assigned the same species if their genomes are more than 95% identical; assigned the same genus if genomes share 70% nucleotide identity across the genome length and form monophyletic groups based on a phylogenetic tree of signature gene(s); and assigned the same family if they share orthologous genes and form a cohesive and monophyletic group in a proteome-based clustering tool.

Prophage WO taxonomic classification satisfied all demarcation criteria except for genus designation. At the genus level, due to the high variability of the EAM, we applied alternative criteria: genomes should (i) share >70% nucleotide homology across the core prophage WO genome, exclusive of the EAM.

Supporting information

S1 Fig. Sr1WO genome maps. Genome maps of sr1WO prophage regions where genes are drawn to scale in forward and reverse directions. Predicted physical structures are illustrated to the left of each genome. All genomes contain tail modules with the exception of the partial WOVitA2 sequence. Prophage WO Core Genes are shaded in blue and predicted EAM genes are shaded in gray. Genes of similar function are similarly color-coded according to the figure legend. Locus tags, if available, are listed in italics above the genes. The large, black diagonal lines between the recombinase and transposase in WORiC and WOSuziC represent post-integration rearrangement of the prophage region in the *Wolbachia* chromosome. Dashed lines represent breaks in the assembly whereas small diagonal lines represent a continuation of the genome onto the next line. Arrows with diagonal stripes represent genes that may be pseudo-genized relative to homologs in other prophage WO genomes. The putative function for each structural gene is discussed in [S1 Text](#). * Indicates a partial sequence and/or highly degraded genome that may be considered a WO-like Island; gene content, module synteny, and recombinase typing support a putative sr1WO-origin.
(TIF)

S2 Fig. Sr2WO genome maps. Genome maps of sr2WO prophage regions where genes are drawn to scale in forward and reverse directions. Predicted physical structures are illustrated to the left of each genome. WOVitA1-like prophage genomes encode all structural modules (shaded in blue) and an EAM (shaded in gray) whereas WORiA-like prophage genomes encode an intact head module, recombinase, lysozyme, AAA16, and disrupted connector. They lack most other modules. Genes of similar function are similarly color-coded according to the figure legend. Locus tags, if available, are listed in italics above the genes. Dashed lines represent breaks in the assembly whereas small diagonal lines represent a continuation of the genome onto the next line. Arrows with diagonal stripes represent genes that may be pseudo-genized relative to homologs in other prophage WO genomes. The putative function for each structural gene is discussed in [S1 Text](#). * Indicates a partial sequence or highly degraded genome that may be considered a WO-like Island; gene content, module synteny, and recombinase typing support a putative sr2WO-origin.
(TIF)

S3 Fig. Sr3WO genome maps. Genome maps of sr3WO prophage regions where genes are drawn to scale in forward and reverse directions. Three *wPip* prophages exist as one contiguous prophage region in the *Wolbachia* genome and are illustrated here as WOPip1, WOPip2, and WOPip3 (based on [\[110\]](#)). Predicted physical structures are illustrated to the left of each genome. Prophage WO Core Genes are shaded in blue and predicted EAM genes are shaded in gray. Genes of similar function are similarly color-coded according to the figure legend. sr3WO is comprised of highly variable genomes that are often flanked by mobile elements (transposases are shown in yellow). They generally contain a recombinase, connector/base-plate, head, and EAM with only a few genomes encoding a complete tail. Prophages in this group often contain *cifA;B* (pink). Locus tags are listed in italics above the genes. Dashed lines represent breaks in the assembly whereas small diagonal lines represent a continuation of the genome onto the next line. Arrows with diagonal stripes represent genes that may be

pseudogenized relative to homologs in other prophage WO genomes. The putative function for each structural gene is discussed in [S1 Text](#). * Indicates a partial sequence or highly degraded genome that may be considered a WO-like Island; gene content, module synteny, and recombinase typing support a putative sr3WO-origin. † The WONO1 region is a chimera between a 5'-sr3WO and 3'-sr1WO. Definitive boundaries are unknown.

(TIF)

S4 Fig. Sr3WO and sr3WO-Undecim Cluster genome maps. Genome maps of sr3WO prophage regions where genes are drawn to scale in forward and reverse directions. WOIr is one contiguous prophage region in the *Wolbachia* genome that is illustrated here as Segment 1 and Segment 2. A subset of sr3WO prophages is further categorized by the presence of a highly conserved WD0611-WD0621 like region, termed the Undecim Cluster (black). Predicted physical structures are illustrated to the left of each genome. Prophage WO Core Genes are shaded in blue and predicted EAM genes are shaded in gray. Genes of similar function are similarly color-coded according to the figure legend. sr3WO is comprised of highly variable genomes that are often flanked by mobile elements (transposases are shown in yellow). Prophages in this group often contain *cifA;B* (pink). Locus tags are listed in italics above the genes. Dashed lines represent breaks in the assembly whereas small diagonal lines represent a continuation of the genome onto the next line. Arrows with diagonal stripes represent genes that may be pseudogenized relative to homologs in other prophage WO genomes. The putative function for each structural gene is discussed in [S1 Text](#). * Indicates a partial sequence or highly degraded genome that may be considered a WO-like Island; gene content, module synteny, and recombinase typing support a putative sr3WO-origin.

(TIF)

S5 Fig. Sr3WO-Undecim Cluster genome maps. Genome maps of sr3WO prophage regions where genes are drawn to scale in forward and reverse directions. This subset of sr3WO prophages is further categorized by the presence of a highly conserved WD0611-WD0621 like region, termed the Undecim Cluster (black). Predicted physical structures are illustrated to the left of each genome. Prophage WO Core Genes are shaded in blue and predicted EAM genes are shaded in gray. Genes of similar function are similarly color-coded according to the figure legend. sr3WO is comprised of highly variable genomes that are often flanked by mobile elements (transposases are shown in yellow). Prophages in this group often contain *cifA;B* (pink). Locus tags are listed in italics above the genes. Dashed lines represent breaks in the assembly whereas small diagonal lines represent a continuation of the genome onto the next line. Arrows with diagonal stripes represent genes that may be pseudogenized relative to homologs in other prophage WO genomes. The putative function for each structural gene is discussed in [S1 Text](#).

(TIF)

S6 Fig. Sr4WO genome maps. Genome maps of sr4WO prophage regions where genes are drawn to scale in forward and reverse directions. To date, sr4WO prophages have only been identified in the parthenogenic strain of *Folsomia candida*, *wFol*. WOFol2 is one contiguous prophage region in the *Wolbachia* genome that is illustrated here as Segment 1 and Segment 2. Likewise, the WOFol3 prophage region is illustrated as three segments. Predicted physical structures are illustrated to the left of each genome. Prophage WO Core Genes are shaded in blue and predicted EAM genes are shaded in gray. Genes of similar function are similarly color-coded according to the figure legend. Locus tags are listed in italics above the genes. Small diagonal lines represent a continuation of the genome onto the next line. Arrows with diagonal stripes represent genes that may be pseudogenized relative to homologs in other

prophage WO genomes. The putative function for each structural gene is discussed in [S1 Text](#). (TIF)

S7 Fig. WO-like Island genome maps. Genome maps of WO-like Islands where genes are drawn to scale in forward and reverse directions. These regions contain only one structural module and/or group of WO-related genes. Regions flanked by assembly breaks (i.e., WORecB, WORecA, and wVitA) are tentatively classified as WO-like Islands due to lack of a full-length prophage in the genome assembly. Names are based on the original author's description. If it was identified as a prophage in the genome announcement, the reported WO name is listed here. Otherwise, the name simply refers to the encoding *Wolbachia* genome. Many WO-like Islands contain *cifA;B*; some Islands (i.e., wNo, wVitA, WOMau4, and WOAlbB3) contain both Type III *cifA;B* (pink) and the Undecim Cluster (black). Predicted physical structures are illustrated to the left of each genome. Prophage WO Core Genes are shaded in blue and predicted EAM genes are shaded in gray. Genes of similar function are similarly color-coded according to the figure legend. Locus tags are listed in italics above the genes. Dashed lines represent breaks in the assembly. Arrows with diagonal stripes represent genes that may be pseudogenized relative to homologs in other prophage WO genomes. The putative function for each structural gene is discussed in [S1 Text](#). (TIF)

S8 Fig. *In silico* predictions of phage WO attachment (*att*) sites. An integrated prophage sequence contains left and right attachment sites (*attL* and *attR*, respectively) at the points of chromosomal integration. Half of the *att* site is phage-derived (green); the other half is bacterial derived (black). If the DNA sequence of the bacterial attachment site (*attB*, black) is known, a nucleotide alignment of the intact sequence with the integrated prophage genome will correlate with 5'- (*attL*) and 3'- (*attR*) prophage boundaries. (a) WORiC, a member of sr1WO, integrates into wRi's magnesium chelatase gene. By aligning an intact copy of this gene (WD0721) from closely related wMel that does not harbor sr1WO, (b) the juncture points of the disrupted magnesium chelatase indicate the *attL* and *attR* sites for the WORiC prophage region within the wRi genome. (b) The phage attachment site (*attP*, green) is predicted *in silico* by concatenating the non-*Wolbachia* portions of the *attL* and *attR* sites. (c) Likewise, this method can also be applied when the bacterial integration locus is intergenic. The homologous intergenic region of closely related, sr1WO-free wPip can be used to predict *att* sites for WOCauB3. Nucleotides in orange represent a common region, O, that is shared by all four *att* sites. This method was adapted from [39] where the *attP* site was used to predict the *attB* site of WOVitA1. (TIF)

S9 Fig. RT is associated with duplication, inversion, and recombination of the prophage WO genome. (a) The WOMelB prophage genomes of wMel2_a and wMel2_b are duplicated relative to the wMel reference genome [72]. (b) The entire WORiB prophage region is duplicated in wRi [19]. (c) WOHa1 encodes a second, pseudogenized *cifA;B*-containing region relative to closely related WOAuA, WORiB, WOSuziB, and WOSol prophages. (d) A ligase-containing region is duplicated in wFol's WOFol1 and WOFol2 [56]. (e) Based on homology to other prophage regions (Fig 2), the connector/baseplate should be adjacent to a head module and the WOPC-2 and replication genes should be oriented in the opposite direction; this indicates a likely insertion and/or recombination in the WOFol3 prophage region. (f) The WOIrr head module is inverted relative to other sr3WOs. Genes are illustrated as arrows; putative gene annotations are labeled in [S1–S7 Figs](#). In each example, the regions of chromosomal

rearrangement are highlighted in light orange and flanked by at least one RT. (TIF)

S10 Fig. Comparative genomics of Octomom-like variants across diverse *Wolbachia*. Octomom (orange) and Octomom-like (green) regions are illustrated for *wMelCS*, *wMel*, *wSYT* clade, and *wPip*. Characteristics of each region are listed next to the genome schematic. Notably, the *wMelCS* genome, representative of the dynamic *wMelPop*, is distinguished from other variants by intact, flanking reverse transcriptases of group II intron origin (RT) on both sides. *wPip*, the only *Wolbachia* Supergroup B variant, is the most divergent and not associated with an RT, MutL or ankyrin repeat. Rather it is adjacent to WP1349, another gene that has been horizontally transferred between phage and arthropod [71]. (TIF)

S11 Fig. DUF2466 nucleotide alignment supports a WOMeA origin of Octomom. (a) A nucleotide alignment of concatenated WD0507 (Octomom) and WD0257 (WOMeA) illustrates homology with intact DUF2466 genes of similar WO-PC2 modules, except for a 30-bp insertion at the 3'-end of WD0507 (highlighted in red). WD0507 is shaded in gold; WD0257 is shaded in blue. Disagreements relative to consensus (excluding ambiguous disagreements) are shaded in gray. (b) A distance matrix of the alignment confirms that the putative ancestral DUF2466 shares 94% and 68% nucleotide identity with homologous WO-PC2 modules in *wStv* (HC358_04600) and *wAu* (WPAU_0253), respectively. (TIF)

S12 Fig. Prophage WO encodes a putative lytic cassette. Adjacent to the tail module of most prophage WO variants are three phage lysis candidates: ankyrin repeat containing protein (not shown), holin-like, and patatin-like phospholipase. (a) Similar to canonical holins, the prophage WO gene product encodes a single N-terminal transmembrane domain with no predicted charge. It is smaller than 150 amino acid residues, features a C-terminal coiled coil motif, and has a highly charged C-terminal domain. Unlike canonical holins, however, it is adjacent to a patatin-like gene rather than a characterized endolysin. (b) The prophage WO holin-like peptide shares 41.1% amino acid identity to a homolog in the non-*Wolbachia* prophage from the Tara Oceans Project that is directly adjacent to a GH108 lysozyme (complete genome illustrated in Fig 6). (c) A Mauve alignment of these genomic regions (core phage modules only; EAM not included) indicates 50.3% nucleotide identity across the majority of the sequence, including the holin-like gene (marked with a gold star). The similarity of these prophages suggest that prophage WO may utilize a similar holin-like gene with a different lytic enzyme (i.e., patatin rather than lysozyme) to lyse the bacterial cell. (TIF)

S13 Fig. *Wolbachia* contains both prophage regions and GTA-like genes scattered throughout the chromosome. (a) Circular *wMel* contains three prophage WO-like regions (teal) and multiple genes with homology to GTAs (orange) scattered throughout the genome, illustrated relative to the putative origin of replication (*ori*, gray). The Undecim Cluster is highlighted in black, *cifA;B* are highlighted in pink, and *wmk* is highlighted in purple. (b) GTAs are present in at least one strain of each *Wolbachia* Supergroup except Supergroups J and L. They are also present in closely related Anaplasmataceae genera. (TIF)

S14 Fig. Distance matrices of GTA nucleotide homology indicate evolution with the *Wolbachia* chromosome. Nucleotide alignments of GTA genes (a) portal, (b) BRO599, (c) TIM barrel, (d) major capsid, (e) head-tail connector, and (f) terminase indicate strict delineation

based on *Wolbachia* supergroup. This supports evolution with the *Wolbachia* chromosome rather than independent evolution of a phage genome.

(TIF)

S1 Table. Prophage WO genes are associated with arthropod-infecting *Wolbachia*. *Wolbachia* genomes are listed according to (A) host phylum; (B) *Wolbachia* supergroup; (C) *Wolbachia* name (D) host species and (E) host strain/lineage, if applicable; (F) NCBI accession number; (G) genome assembly status; (H) identification of prophage WO core genes; (I) identification of CI genes; and (J) identification of the Undecim Cluster. *Wolbachia* strains that did not include official names in the assembly reports are listed here using a capital letter for host genus and two to three lowercase letters for host species. “Highly pseudogenized” in column H indicates that the prophage genome is highly pseudogenized and encodes very few Core WO genes. (*) indicates that the genome lacks a complete Undecim Cluster but encodes WD0616 and/or WD0621 homologs. (**) indicates that the genome was not included as Source Data for Fig 1B due to incomplete genome information.

(XLSX)

S2 Table. *wMhie* encodes prophage WO genes. *wMhie*, a *Wolbachia* endosymbiont from the nematode *Madathamugadia hiepei*, encodes four genes that are conserved throughout phage WO’s transcriptional regulation and replication/repair modules. Each gene is listed by locus tag, annotation, and nucleotide homology to prophage WOVitA1.

(XLSX)

S3 Table. *cifA* and *cifB* genes are associated with *Wolbachia* Supergroups F and T. *cifA* and *cifB* are identified in Supergroups F and T. NCBI accession numbers and genomic coordinates (or locus tags) are provided for each locus.

(XLSX)

S4 Table. Homologous large serine recombinase sequences are associated with similar gene prophage WO gene synteny. Megablast hits using default parameters and greater than >50% identity cutoff across the recombinase sequence are shown for WOCauB3 (sr1WO), WOVitA1 (sr2WO), WOMelB (sr3WO), and WOFol2 (sr4WO) in columns A-E. Column F lists the synteny of adjacent genes. Partial indicates that it is not a complete prophage, but genes support the associated classification. N/A indicates that the recombinase homolog lacks adjacent prophage genes and/or there is not enough genomic information to make a confident assessment.

(XLSX)

S5 Table. Flanking bacterial genes can predict sr2WO-like regions in B-*Wolbachia*. Sr2WO recombinase sequences are adjacent to core bacterial genes in the *Wolbachia* chromosome (defined as the 5’-end of the prophage region). The adjacent bacterial genes in the *wPip* genome, a B-*Wolbachia* that lacks sr2WO prophages, correlate with 3’- boundaries of the predicted prophage or WO-like Island region. *prsA* refers to ribose-phosphate diphosphokinase; *tkt* refers to transketolase; *tpiA* refers to triose-phosphate isomerase; *murD* refers to UDP-N-acetylmuramoylalanine: D-glutamate ligase.

(XLSX)

S6 Table. Diversity of prophage WO mobile elements. All mobile elements, both flanking and internal, are listed for each prophage WO genome according to original genome annotations and ISFinder [102]. The sr1WO group and full-length prophages of the sr2WO group do not feature transposases on the 5’- and 3’- flanking regions. The WORiA-like prophages of the sr2WO group are associated with 3’- transposases; these correlate with putative truncations of

the prophage regions. Most genomes within the sr3WO group feature mobile elements on both 5'- and 3'- ends. IS refers to Insertion Sequence Family; RT refers to reverse transcriptase of group II intron origin; Rpn refers to recombination promoting nuclease. (*) indicates a sequencing gap or artificial join in the *Wolbachia* genome. Complete sequence information is unknown. (**) indicates that these prophage sequences were obtained from contigs and may be segmented in the *Wolbachia* chromosome; the exact 5' and 3' ends are uncertain. Genomic locations for each mobile element are illustrated in [S1–S7 Figs.](#)

(XLSX)

S7 Table. *Wolbachia* GTA genes. The annotation of *Wolbachia*'s distributed GTA genes is based on a BLASTP against NCBI Conserved Domains; E-values are listed in column B.

(XLSX)

S8 Table. [Fig 6](#) image attribution. Attribution information is listed for each thumbnail image in [Fig 6](#).

(XLSX)

S1 Text. Phage WO Structural Modules. Phage WO structural genes are organized into head, connector/baseplate, tail, and tail fiber modules. The predicted function of each gene is discussed based on conserved protein domains and homology to other model systems, including lambda, T4, P2, and Mu phages.

(DOCX)

Acknowledgments

We thank Dylan Shropshire, Anne Duploux, Rupinder Kaur, Irene Newton, Julien Martinez, and Sean Oladipupo for insightful feedback on the proposed phage WO taxonomy; Evelien Adriaenssens and Elliot Lefkowitz for ICTV guidance; and Brittany Leigh for countless conversations about phage WO biology.

Author Contributions

Conceptualization: Sarah R. Bordenstein, Seth R. Bordenstein.

Data curation: Sarah R. Bordenstein.

Formal analysis: Sarah R. Bordenstein, Seth R. Bordenstein.

Funding acquisition: Seth R. Bordenstein.

Investigation: Sarah R. Bordenstein.

Methodology: Sarah R. Bordenstein, Seth R. Bordenstein.

Project administration: Sarah R. Bordenstein.

Resources: Sarah R. Bordenstein, Seth R. Bordenstein.

Software: Sarah R. Bordenstein.

Supervision: Sarah R. Bordenstein, Seth R. Bordenstein.

Validation: Sarah R. Bordenstein, Seth R. Bordenstein.

Visualization: Sarah R. Bordenstein.

Writing – original draft: Sarah R. Bordenstein.

Writing – review & editing: Sarah R. Bordenstein, Seth R. Bordenstein.

References

1. Newton IL, Bordenstein SR. Correlations between bacterial ecology and mobile DNA. *Curr Microbiol*. 2011; 62(1):198–208. <https://doi.org/10.1007/s00284-010-9693-3> PMID: 20577742
2. Cerveau N, Leclercq S, Leroy E, Bouchon D, Cordaux R. Short- and long-term evolutionary dynamics of bacterial insertion sequences: insights from *Wolbachia* endosymbionts. *Genome Biol Evol*. 2011; 3:1175–86. <https://doi.org/10.1093/gbe/evr096> PMID: 21940637
3. Masui S, Kamoda S, Sasaki T, Ishikawa H. Distribution and evolution of bacteriophage WO in *Wolbachia*, the endosymbiont causing sexual alterations in arthropods. *J Mol Evol*. 2000; 51(5):491–7. <https://doi.org/10.1007/s002390010112> PMID: 11080372
4. Kent BN, Bordenstein SR. Phage WO of *Wolbachia*: lambda of the endosymbiont world. *Trends Microbiol*. 2010; 18(4):173–81. <https://doi.org/10.1016/j.tim.2009.12.011> PMID: 20083406
5. Reveillaud J, Bordenstein SR, Cruaud C, Shaiber A, Esen OC, Weill M, et al. The *Wolbachia* mobilome in *Culex pipiens* includes a putative plasmid. *Nat Commun*. 2019; 10(1):1051. <https://doi.org/10.1038/s41467-019-08973-w> PMID: 30837458
6. Baiao GC, Janice J, Galinou M, Klasson L. Comparative genomics reveals factors associated with phenotypic expression of *Wolbachia*. *Genome Biol Evol*. 2021. <https://doi.org/10.1093/gbe/evab111> PMID: 34003269
7. Kaur R, Shropshire JD, Cross KL, Leigh B, Mansueto AJ, Stewart V, et al. Living in the endosymbiotic world of *Wolbachia*: A centennial review. *Cell Host Microbe*. 2021; 29(6):879–93. <https://doi.org/10.1016/j.chom.2021.03.006> PMID: 33945798
8. Weinert LA, Araujo-Jnr EV, Ahmed MZ, Welch JJ. The incidence of bacterial endosymbionts in terrestrial arthropods. *Proc Biol Sci*. 2015; 282(1807):20150249. <https://doi.org/10.1098/rspb.2015.0249> PMID: 25904667
9. Pruneau L, Moumene A, Meyer DF, Marcelino I, Lefrancois T, Vachieri N. Understanding *Anaplasmataceae* pathogenesis using "Omics" approaches. *Front Cell Infect Microbiol*. 2014; 4:86. <https://doi.org/10.3389/fcimb.2014.00086> PMID: 25072029
10. Vavre F, Fleury F, Lepetit D, Fouillet P, Bouletreau M. Phylogenetic evidence for horizontal transmission of *Wolbachia* in host-parasitoid associations. *Mol Biol Evol*. 1999; 16(12):1711–23. <https://doi.org/10.1093/oxfordjournals.molbev.a026084> PMID: 10605113
11. Boyle L, O'Neill SL, Robertson HM, Karr TL. Interspecific and intraspecific horizontal transfer of *Wolbachia* in *Drosophila*. *Science*. 1993; 260(5115):1796–9. <https://doi.org/10.1126/science.8511587> PMID: 8511587
12. Chafee ME, Funk DJ, Harrison RG, Bordenstein SR. Lateral phage transfer in obligate intracellular bacteria (*Wolbachia*): verification from natural populations. *Mol Biol Evol*. 2010; 27(3):501–5. <https://doi.org/10.1093/molbev/msp275> PMID: 19906794
13. Wang N, Jia S, Xu H, Liu Y, Huang D. Multiple Horizontal Transfers of Bacteriophage WO and Host *Wolbachia* in Fig Wasps in a Closed Community. *Front Microbiol*. 2016; 7:136. <https://doi.org/10.3389/fmicb.2016.00136> PMID: 26913026
14. Kent BN, Salichos L, Gibbons JG, Rokas A, Newton IL, Clark ME, et al. Complete bacteriophage transfer in a bacterial endosymbiont (*Wolbachia*) determined by targeted genome capture. *Genome Biol Evol*. 2011; 3:209–18. <https://doi.org/10.1093/gbe/evr007> PMID: 21292630
15. Gavotte L, Henri H, Stouthamer R, Charif D, Charlat S, Bouletreau M, et al. A Survey of the bacteriophage WO in the endosymbiotic bacteria *Wolbachia*. *Mol Biol Evol*. 2007; 24(2):427–35. <https://doi.org/10.1093/molbev/msl171> PMID: 17095536
16. Bordenstein SR, Wernegreen JJ. Bacteriophage flux in endosymbionts (*Wolbachia*): infection frequency, lateral transfer, and recombination rates. *Mol Biol Evol*. 2004; 21(10):1981–91. <https://doi.org/10.1093/molbev/msh211> PMID: 15254259
17. Ishmael N, Dunning Hotopp JC, Ioannidis P, Biber S, Sakamoto J, Siozios S, et al. Extensive genomic diversity of closely related *Wolbachia* strains. *Microbiology*. 2009; 155(Pt 7):2211–22. <https://doi.org/10.1099/mic.0.027581-0> PMID: 19389774
18. Kent BN, Funkhouser LJ, Setia S, Bordenstein SR. Evolutionary genomics of a temperate bacteriophage in an obligate intracellular bacteria (*Wolbachia*). *PLoS One*. 2011; 6(9):e24984. <https://doi.org/10.1371/journal.pone.0024984> PMID: 21949820
19. Klasson L, Westberg J, Sapountzis P, Naslund K, Lutnaes Y, Darby AC, et al. The mosaic genome structure of the *Wolbachia* wRi strain infecting *Drosophila simulans*. *Proc Natl Acad Sci U S A*. 2009; 106(14):5725–30. <https://doi.org/10.1073/pnas.0810753106> PMID: 19307581
20. Sanogo YO, Dobson SL. WO bacteriophage transcription in *Wolbachia*-infected *Culex pipiens*. *Insect Biochem Mol Biol*. 2006; 36(1):80–5. <https://doi.org/10.1016/j.ibmb.2005.11.001> PMID: 16360953

21. Werren JH, Baldo L, Clark ME. *Wolbachia*: master manipulators of invertebrate biology. *Nat Rev Microbiol*. 2008; 6(10):741–51. <https://doi.org/10.1038/nrmicro1969> PMID: 18794912
22. Charlat S, Hurst GD, Mercot H. Evolutionary consequences of *Wolbachia* infections. *Trends Genet*. 2003; 19(4):217–23. [https://doi.org/10.1016/S0168-9525\(03\)00024-6](https://doi.org/10.1016/S0168-9525(03)00024-6) PMID: 12683975
23. Fenn K, Blaxter M. *Wolbachia* genomes: revealing the biology of parasitism and mutualism. *Trends Parasitol*. 2006; 22(2):60–5. <https://doi.org/10.1016/j.pt.2005.12.012> PMID: 16406333
24. Slatko BE, Taylor MJ, Foster JM. The *Wolbachia* endosymbiont as an anti-filarial nematode target. *Symbiosis*. 2010; 51(1):55–65. <https://doi.org/10.1007/s13199-010-0067-1> PMID: 20730111
25. 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 Mol Biol*. 2011; 20(1):75–85. <https://doi.org/10.1111/j.1365-2583.2010.01042.x> PMID: 20854481
26. Sinkins SP, Walker T, Lynd AR, Steven AR, Makepeace BL, Godfray HC, et al. *Wolbachia* variability and host effects on crossing type in *Culex* mosquitoes. *Nature*. 2005; 436(7048):257–60. <https://doi.org/10.1038/nature03629> PMID: 16015330
27. Duron O, Fort P, Weill M. Hypervariable prophage WO sequences describe an unexpected high number of *Wolbachia* variants in the mosquito *Culex pipiens*. *Proc Biol Sci*. 2006; 273(1585):495–502. <https://doi.org/10.1098/rspb.2005.3336> PMID: 16615218
28. Duron O, Bernard C, Unal S, Berthomieu A, Berticat C, Weill M. Tracking factors modulating cytoplasmic incompatibilities in the mosquito *Culex pipiens*. *Mol Ecol*. 2006; 15(10):3061–71. <https://doi.org/10.1111/j.1365-294X.2006.02996.x> PMID: 16911221
29. LePage DP, Metcalf JA, Bordenstein SR, On J, Perlmutter JI, Shropshire JD, et al. Prophage WO genes recapitulate and enhance *Wolbachia*-induced cytoplasmic incompatibility. *Nature*. 2017; 543(7644):243–7. <https://doi.org/10.1038/nature21391> PMID: 28241146
30. Shropshire JD, On J, Layton EM, Zhou H, Bordenstein SR. One prophage WO gene rescues cytoplasmic incompatibility in *Drosophila melanogaster*. *Proc Natl Acad Sci U S A*. 2018; 115(19):4987–91. <https://doi.org/10.1073/pnas.1800650115> PMID: 29686091
31. Chen H, Ronau JA, Beckmann JF, Hochstrasser M. A *Wolbachia* nuclease and its binding partner provide a distinct mechanism for cytoplasmic incompatibility. *Proc Natl Acad Sci U S A*. 2019; 116(44):22314–21. <https://doi.org/10.1073/pnas.1914571116> PMID: 31615889
32. Beckmann JF, Ronau JA, Hochstrasser M. A *Wolbachia* deubiquitylating enzyme induces cytoplasmic incompatibility. *Nat Microbiol*. 2017; 2:17007. <https://doi.org/10.1038/nmicrobiol.2017.7> PMID: 28248294
33. Perlmutter JI, Bordenstein SR, Unckless RL, LePage DP, Metcalf JA, Hill T, et al. The phage gene *wmk* is a candidate for male killing by a bacterial endosymbiont. *PLoS Pathog*. 2019; 15(9):e1007936. <https://doi.org/10.1371/journal.ppat.1007936> PMID: 31504075
34. Bordenstein SR, Marshall ML, Fry AJ, Kim U, Wernegreen JJ. The tripartite associations between bacteriophage, *Wolbachia*, and arthropods. *PLoS Pathog*. 2006; 2(5):e43. <https://doi.org/10.1371/journal.ppat.0020043> PMID: 16710453
35. Wright JD, Sjostrand FS, Portaro JK, Barr AR. The ultrastructure of the *Rickettsia*-like microorganism *Wolbachia pipientis* and associated virus-like bodies in the mosquito *Culex pipiens*. *J Ultrastruct Res*. 1978; 63(1):79–85. [https://doi.org/10.1016/s0022-5320\(78\)80046-x](https://doi.org/10.1016/s0022-5320(78)80046-x) PMID: 671578
36. Biliske JA, Batista PD, Grant CL, Harris HL. The bacteriophage WORiC is the active phage element in *wRi* of *Drosophila simulans* and represents a conserved class of WO phages. *BMC Microbiol*. 2011; 11:251. <https://doi.org/10.1186/1471-2180-11-251> PMID: 22085419
37. Fujii Y, Kubo T, Ishikawa H, Sasaki T. Isolation and characterization of the bacteriophage WO from *Wolbachia*, an arthropod endosymbiont. *Biochem Biophys Res Commun*. 2004; 317(4):1183–8. <https://doi.org/10.1016/j.bbrc.2004.03.164> PMID: 15094394
38. Tanaka K, Furukawa S, Nikoh N, Sasaki T, Fukatsu T. Complete WO phage sequences reveal their dynamic evolutionary trajectories and putative functional elements required for integration into the *Wolbachia* genome. *Appl Environ Microbiol*. 2009; 75(17):5676–86. <https://doi.org/10.1128/AEM.01172-09> PMID: 19592535
39. Bordenstein SR, Bordenstein SR. Eukaryotic association module in phage WO genomes from *Wolbachia*. *Nat Commun*. 2016; 7:13155. <https://doi.org/10.1038/ncomms13155> PMID: 27727237
40. Iturbe-Ormaetxe I, Burke GR, Riegler M, O'Neill SL. Distribution, expression, and motif variability of ankyrin domain genes in *Wolbachia pipientis*. *J Bacteriol*. 2005; 187(15):5136–45. <https://doi.org/10.1128/JB.187.15.5136-5145.2005> PMID: 16030207
41. Siozios S, Ioannidis P, Klasson L, Andersson SG, Braig HR, Bourtzis K. The diversity and evolution of *Wolbachia* ankyrin repeat domain genes. *PLoS One*. 2013; 8(2):e55390. <https://doi.org/10.1371/journal.pone.0055390> PMID: 23390535

42. Jahn MT, Arkhipova K, Markert SM, Stigloher C, Lachnit T, Pita L, et al. A Phage Protein Aids Bacterial Symbionts in Eukaryote Immune Evasion. *Cell Host Microbe*. 2019; 26(4):542–50 e5. <https://doi.org/10.1016/j.chom.2019.08.019> PMID: 31561965
43. National Library of Medicine (US) NCfBI. National Center for Biotechnology Information (NCBI) Bethesda (MD)1988. Available from: <https://www.ncbi.nlm.nih.gov/>.
44. International Committee on Taxonomy of Viruses Executive C. The new scope of virus taxonomy: partitioning the virosphere into 15 hierarchical ranks. *Nat Microbiol*. 2020; 5(5):668–74. <https://doi.org/10.1038/s41564-020-0709-x> PMID: 32341570
45. Simmonds P, Adams MJ, Benko M, Breitbart M, Brister JR, Carstens EB, et al. Consensus statement: Virus taxonomy in the age of metagenomics. *Nat Rev Microbiol*. 2017; 15(3):161–8. <https://doi.org/10.1038/nrmicro.2016.177> PMID: 28134265
46. Lindsey ARI, Rice DW, Bordenstein SR, Brooks AW, Bordenstein SR, Newton ILG. Evolutionary Genetics of Cytoplasmic Incompatibility Genes *cifA* and *cifB* in Prophage WO of *Wolbachia*. *Genome Biol Evol*. 2018; 10(2):434–51. <https://doi.org/10.1093/gbe/evy012> PMID: 29351633
47. Martinez J, Klasson L, Welch JJ, Jiggins FM. Life and Death of Selfish Genes: Comparative Genomics Reveals the Dynamic Evolution of Cytoplasmic Incompatibility. *Mol Biol Evol*. 2021; 38(1):2–15. <https://doi.org/10.1093/molbev/msaa209> PMID: 32797213
48. Wu M, Sun LV, Vamathevan J, Riegler M, Deboy R, Brownlie JC, et al. Phylogenomics of the reproductive parasite *Wolbachia pipientis* wMel: a streamlined genome overrun by mobile genetic elements. *PLoS Biol*. 2004; 2(3):E69. <https://doi.org/10.1371/journal.pbio.0020069> PMID: 15024419
49. Crainey JL, Hurst J, Lamberton PHL, Cheke RA, Griffin CE, Wilson MD, et al. The Genomic Architecture of Novel *Simulium damnosum* *Wolbachia* Prophage Sequence Elements and Implications for Onchocerciasis Epidemiology. *Front Microbiol*. 2017; 8:852. <https://doi.org/10.3389/fmicb.2017.00852> PMID: 28611731
50. Smith MCM. Phage-encoded Serine Integrases and Other Large Serine Recombinases. *Microbiol Spectr*. 2015; 3(4). <https://doi.org/10.1128/microbiolspec.MDNA3-0059-2014> PMID: 26350324
51. Metcalf JA, Jo M, Bordenstein SR, Jaenike J, Bordenstein SR. Recent genome reduction of *Wolbachia* in *Drosophila recens* targets phage WO and narrows candidates for reproductive parasitism. *PeerJ*. 2014; 2:e529. <https://doi.org/10.7717/peerj.529> PMID: 25165636
52. Newton IL, Clark ME, Kent BN, Bordenstein SR, Qu J, Richards S, et al. Comparative Genomics of Two Closely Related *Wolbachia* with Different Reproductive Effects on Hosts. *Genome Biol Evol*. 2016; 8(5):1526–42. <https://doi.org/10.1093/gbe/evw096> PMID: 27189996
53. Riegler M, Iturbe-Ormaetxe I, Woolfit M, Miller WJ, O'Neill SL. Tandem repeat markers as novel diagnostic tools for high resolution fingerprinting of *Wolbachia*. *BMC Microbiol*. 2012; 12 Suppl 1:S12. <https://doi.org/10.1186/1471-2180-12-S1-S12> PMID: 22375862
54. Duploux A, Iturbe-Ormaetxe I, Beatson SA, Szubert JM, Brownlie JC, McMeniman CJ, et al. Draft genome sequence of the male-killing *Wolbachia* strain wBol1 reveals recent horizontal gene transfers from diverse sources. *BMC Genomics*. 2013; 14:20. <https://doi.org/10.1186/1471-2164-14-20> PMID: 23324387
55. Harshey RM. Transposable Phage Mu. *Microbiol Spectr*. 2014; 2(5). <https://doi.org/10.1128/microbiolspec.MDNA3-0007-2014> PMID: 26104374
56. Kampfraath AA, Klasson L, Anvar SY, Vossen R, Roelofs D, Kraaijeveld K, et al. Genome expansion of an obligate parthenogenesis-associated *Wolbachia* poses an exception to the symbiont reduction model. *BMC Genomics*. 2019; 20(1):106. <https://doi.org/10.1186/s12864-019-5492-9> PMID: 30727958
57. Bobay LM, Touchon M, Rocha EP. Pervasive domestication of defective prophages by bacteria. *Proc Natl Acad Sci U S A*. 2014; 111(33):12127–32. <https://doi.org/10.1073/pnas.1405336111> PMID: 25092302
58. Ramisetty BCM, Sudhakari PA. Bacterial 'Grounded' Prophages: Hotspots for Genetic Renovation and Innovation. *Front Genet*. 2019; 10:65. <https://doi.org/10.3389/fgene.2019.00065> PMID: 30809245
59. Fillol-Salom A, Martinez-Rubio R, Abdulrahman RF, Chen J, Davies R, Penades JR. Phage-inducible chromosomal islands are ubiquitous within the bacterial universe. *ISME J*. 2018; 12(9):2114–28. <https://doi.org/10.1038/s41396-018-0156-3> PMID: 29875435
60. Penades JR, Christie GE. The Phage-Inducible Chromosomal Islands: A Family of Highly Evolved Molecular Parasites. *Annu Rev Virol*. 2015; 2(1):181–201. <https://doi.org/10.1146/annurev-virology-031413-085446> PMID: 26958912
61. Ioannidis P, Dunning Hotopp JC, Sapountzis P, Siozios S, Tsiamis G, Bordenstein SR, et al. New criteria for selecting the origin of DNA replication in *Wolbachia* and closely related bacteria. *BMC Genomics*. 2007; 8:182. <https://doi.org/10.1186/1471-2164-8-182> PMID: 17584494

62. Rands CM, Brussow H, Zdobnov EM. Comparative genomics groups phages of Negativicutes and classical Firmicutes despite different Gram-staining properties. *Environ Microbiol.* 2019; 21(11):3989–4001. <https://doi.org/10.1111/1462-2920.14746> PMID: 31314945
63. Oliveira PH, Touchon M, Cury J, Rocha EPC. The chromosomal organization of horizontal gene transfer in bacteria. *Nat Commun.* 2017; 8(1):841. <https://doi.org/10.1038/s41467-017-00808-w> PMID: 29018197
64. Canchaya C, Fournous G, Brussow H. The impact of prophages on bacterial chromosomes. *Mol Microbiol.* 2004; 53(1):9–18. <https://doi.org/10.1111/j.1365-2958.2004.04113.x> PMID: 15225299
65. Bobay LM, Rocha EP, Touchon M. The adaptation of temperate bacteriophages to their host genomes. *Mol Biol Evol.* 2013; 30(4):737–51. <https://doi.org/10.1093/molbev/mss279> PMID: 23243039
66. Woolfit M, Iturbe-Ormaetxe I, Brownlie JC, Walker T, Riegler M, Seleznev A, et al. Genomic evolution of the pathogenic *Wolbachia* strain, wMelPop. *Genome Biol Evol.* 2013; 5(11):2189–204. <https://doi.org/10.1093/gbe/evt169> PMID: 24190075
67. Duarte EH, Carvalho A, Lopez-Madrigal S, Costa J, Teixeira L. Forward genetics in *Wolbachia*: Regulation of *Wolbachia* proliferation by the amplification and deletion of an addictive genomic island. *PLoS Genet.* 2021; 17(6):e1009612. <https://doi.org/10.1371/journal.pgen.1009612> PMID: 34143770
68. Rohrscheib CE, Frentiu FD, Horn E, Ritchie FK, van Swinderen B, Weible MW 2nd, et al. Intensity of Mutualism Breakdown Is Determined by Temperature Not Amplification of *Wolbachia* Genes. *PLoS Pathog.* 2016; 12(9):e1005888. <https://doi.org/10.1371/journal.ppat.1005888> PMID: 27661080
69. Chrostek E, Teixeira L. Mutualism breakdown by amplification of *Wolbachia* genes. *PLoS Biol.* 2015; 13(2):e1002065. <https://doi.org/10.1371/journal.pbio.1002065> PMID: 25668031
70. Fallon AM. DNA recombination and repair in *Wolbachia*: RecA and related proteins. *Mol Genet Genomics.* 2021; 296(2):437–56. <https://doi.org/10.1007/s00438-020-01760-z> PMID: 33507381
71. Klasson L, Kambris Z, Cook PE, Walker T, Sinkins SP. Horizontal gene transfer between *Wolbachia* and the mosquito *Aedes aegypti*. *BMC Genomics.* 2009; 10:33. <https://doi.org/10.1186/1471-2164-10-33> PMID: 19154594
72. Chrostek E, Marialva MS, Esteves SS, Weinert LA, Martinez J, Jiggins FM, et al. *Wolbachia* variants induce differential protection to viruses in *Drosophila melanogaster*: a phenotypic and phylogenomic analysis. *PLoS Genet.* 2013; 9(12):e1003896. <https://doi.org/10.1371/journal.pgen.1003896> PMID: 24348259
73. Matsuura M, Saldanha R, Ma H, Wank H, Yang J, Mohr G, et al. A bacterial group II intron encoding reverse transcriptase, maturase, and DNA endonuclease activities: biochemical demonstration of maturase activity and insertion of new genetic information within the intron. *Genes Dev.* 1997; 11(21):2910–24. <https://doi.org/10.1101/gad.11.21.2910> PMID: 9353259
74. Dong X, Qu G, Piazza CL, Belfort M. Group II intron as cold sensor for self-preservation and bacterial conjugation. *Nucleic Acids Res.* 2020; 48(11):6198–209. <https://doi.org/10.1093/nar/gkaa313> PMID: 32379323
75. Baldrige GD, Markowski TW, Witthuhn BA, Higgins L, Baldrige AS, Fallon AM. The *Wolbachia* WO bacteriophage proteome in the *Aedes albopictus* C/wStr1 cell line: evidence for lytic activity? *In Vitro Cell Dev Biol Anim.* 2016; 52(1):77–88. <https://doi.org/10.1007/s11626-015-9949-0> PMID: 26427709
76. Gillespie JJ, Joardar V, Williams KP, Driscoll T, Hostetler JB, Nordberg E, et al. A *Rickettsia* genome overrun by mobile genetic elements provides insight into the acquisition of genes characteristic of an obligate intracellular lifestyle. *J Bacteriol.* 2012; 194(2):376–94. <https://doi.org/10.1128/JB.06244-11> PMID: 22056929
77. Gutzwiller F, Carmo CR, Miller DE, Rice DW, Newton IL, Hawley RS, et al. Dynamics of *Wolbachia pipientis* Gene Expression Across the *Drosophila melanogaster* Life Cycle. *G3 (Bethesda).* 2015; 5(12):2843–56. <https://doi.org/10.1534/g3.115.021931> PMID: 26497146
78. Masui S, Kuroiwa H, Sasaki T, Inui M, Kuroiwa T, Ishikawa H. Bacteriophage WO and virus-like particles in *Wolbachia*, an endosymbiont of arthropods. *Biochem Biophys Res Commun.* 2001; 283(5):1099–104. <https://doi.org/10.1006/bbrc.2001.4906> PMID: 11355885
79. Young R. Bacteriophage lysis: mechanism and regulation. *Microbiol Rev.* 1992; 56(3):430–81. <https://doi.org/10.1128/mr.56.3.430-481.1992> PMID: 1406491
80. Otten C, Brilli M, Vollmer W, Viollier PH, Salje J. Peptidoglycan in obligate intracellular bacteria. *Mol Microbiol.* 2018; 107(2):142–63. <https://doi.org/10.1111/mmi.13880> PMID: 29178391
81. Bordenstein SR, Bordenstein SR. Temperature affects the tripartite interactions between bacteriophage WO, *Wolbachia*, and cytoplasmic incompatibility. *PLoS One.* 2011; 6(12):e29106. <https://doi.org/10.1371/journal.pone.0029106> PMID: 22194999

82. Rahman MS, Ammerman NC, Sears KT, Ceraul SM, Azad AF. Functional characterization of a phospholipase A(2) homolog from *Rickettsia typhi*. *J Bacteriol.* 2010; 192(13):3294–303. <https://doi.org/10.1128/JB.00155-10> PMID: 20435729
83. Sato H, Frank DW. ExoU is a potent intracellular phospholipase. *Mol Microbiol.* 2004; 53(5):1279–90. <https://doi.org/10.1111/j.1365-2958.2004.04194.x> PMID: 15387809
84. Banerji S, Flieger A. Patatin-like proteins: a new family of lipolytic enzymes present in bacteria? *Microbiology.* 2004; 150(Pt 3):522–5. <https://doi.org/10.1099/mic.0.26957-0> PMID: 14993300
85. Kamilla S, Jain V. Mycobacteriophage D29 holin C-terminal region functionally assists in holin aggregation and bacterial cell death. *FEBS J.* 2016; 283(1):173–90. <https://doi.org/10.1111/febs.13565> PMID: 26471254
86. Wang IN, Smith DL, Young R. Holins: the protein clocks of bacteriophage infections. *Annu Rev Microbiol.* 2000; 54:799–825. <https://doi.org/10.1146/annurev.micro.54.1.799> PMID: 11018145
87. Lang AS, Beatty JT. Genetic analysis of a bacterial genetic exchange element: the gene transfer agent of *Rhodobacter capsulatus*. *Proc Natl Acad Sci U S A.* 2000; 97(2):859–64. <https://doi.org/10.1073/pnas.97.2.859> PMID: 10639170
88. Westbye AB, Beatty JT, Lang AS. Guaranteeing a captive audience: coordinated regulation of gene transfer agent (GTA) production and recipient capability by cellular regulators. *Curr Opin Microbiol.* 2017; 38:122–9. <https://doi.org/10.1016/j.mib.2017.05.003> PMID: 28599143
89. Lang AS, Beatty JT. Importance of widespread gene transfer agent genes in alpha-proteobacteria. *Trends Microbiol.* 2007; 15(2):54–62. <https://doi.org/10.1016/j.tim.2006.12.001> PMID: 17184993
90. Dohra H, Tanaka K, Suzuki T, Fujishima M, Suzuki H. Draft genome sequences of three *Holospira* species (*Holospira obtusa*, *Holospira undulata*, and *Holospira elegans*), endonuclear symbiotic bacteria of the ciliate *Paramecium caudatum*. *FEMS Microbiol Lett.* 2014; 359(1):16–8. <https://doi.org/10.1111/1574-6968.12577> PMID: 25115770
91. Kantor RS, Miller SE, Nelson KL. The Water Microbiome Through a Pilot Scale Advanced Treatment Facility for Direct Potable Reuse. *Front Microbiol.* 2019; 10:993. <https://doi.org/10.3389/fmicb.2019.00993> PMID: 31139160
92. Tully BJ, Graham ED, Heidelberg JF. The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. *Sci Data.* 2018; 5:170203. <https://doi.org/10.1038/sdata.2017.203> PMID: 29337314
93. Egan S, Gardiner M. Microbial Dysbiosis: Rethinking Disease in Marine Ecosystems. *Front Microbiol.* 2016; 7:991. <https://doi.org/10.3389/fmicb.2016.00991> PMID: 27446031
94. Takano SI, Gotoh Y, Hayashi T. "Candidatus Mesenet longicola": Novel Endosymbionts of *Brontispa longissima* that Induce Cytoplasmic Incompatibility. *Microb Ecol.* 2021. <https://doi.org/10.1007/s00248-021-01686-y> PMID: 33454808
95. Rocchi I, Ericson CF, Malter KE, Zargar S, Eisenstein F, Pilhofer M, et al. A Bacterial Phage Tail-like Structure Kills Eukaryotic Cells by Injecting a Nuclease Effector. *Cell Rep.* 2019; 28(2):295–301 e4. <https://doi.org/10.1016/j.celrep.2019.06.019> PMID: 31291567
96. Gillespie JJ, Driscoll TP, Verhoeve VI, Rahman MS, Macaluso KR, Azad AF. A Tangled Web: Origins of Reproductive Parasitism. *Genome Biol Evol.* 2018; 10(9):2292–309. <https://doi.org/10.1093/gbe/evy159> PMID: 30060072
97. Garushyants SK, Beliavskaia AY, Malko DB, Logacheva MD, Rautian MS, Gelfand MS. Comparative Genomic Analysis of *Holospira* spp., Intracellular Symbionts of Paramecia. *Front Microbiol.* 2018; 9:738. <https://doi.org/10.3389/fmicb.2018.00738> PMID: 29713316
98. Zdobnov EM, Apweiler R. InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics.* 2001; 17(9):847–8. <https://doi.org/10.1093/bioinformatics/17.9.847> PMID: 11590104
99. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990; 215(3):403–10. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2) PMID: 2231712
100. Finn RD, Bateman A, Clements J, Coghill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein families database. *Nucleic Acids Res.* 2014; 42(Database issue):D222–30. <https://doi.org/10.1093/nar/gkt1223> PMID: 24288371
101. Soding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* 2005; 33(Web Server issue):W244–8. <https://doi.org/10.1093/nar/gki408> PMID: 15980461
102. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res.* 2006; 34(Database issue):D32–6. <https://doi.org/10.1093/nar/gkj014> PMID: 16381877

103. Schultz J, Copley RR, Doerks T, Ponting CP, Bork P. SMART: a web-based tool for the study of genetically mobile domains. *Nucleic Acids Res.* 2000; 28(1):231–4. <https://doi.org/10.1093/nar/28.1.231> PMID: 10592234
104. Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 2004; 14(7):1394–403. <https://doi.org/10.1101/gr.2289704> PMID: 15231754
105. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004; 32(5):1792–7. <https://doi.org/10.1093/nar/gkh340> PMID: 15034147
106. Darriba D, Taboada GL, Doallo R, Posada D. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics.* 2011; 27(8):1164–5. <https://doi.org/10.1093/bioinformatics/btr088> PMID: 21335321
107. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol.* 2003; 52(5):696–704. <https://doi.org/10.1080/10635150390235520> PMID: 14530136
108. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol.* 2012; 61(3):539–42. <https://doi.org/10.1093/sysbio/sys029> PMID: 22357727
109. Turner D, Kropinski AM, Adriaenssens EM. A Roadmap for Genome-Based Phage Taxonomy. *Viruses.* 2021; 13(3):506. <https://doi.org/10.3390/v13030506> PMID: 33803862
110. Klasson L, Walker T, Sebahia M, Sanders MJ, Quail MA, Lord A, et al. Genome evolution of *Wolbachia* strain *wPip* from the *Culex pipiens* group. *Mol Biol Evol.* 2008; 25(9):1877–87. <https://doi.org/10.1093/molbev/msn133> PMID: 18550617