Novel Eukaryotic Association Module in Phage WO Genomes from *Wolbachia*

Sarah R. Bordenstein[a] and Seth R. Bordenstein* [a,b]

Affiliations:

[a]Department of Biological Sciences, Vanderbilt University, Nashville, TN 37232, USA.

[b]Department of Pathology, Microbiology, and Immunology, Vanderbilt University, Nashville, TN 37232, USA.

Viruses are trifurcated into eukaryotic, archaeal and bacterial categories. This domain-specific ecology underscores why eukaryotic viruses typically co-opt eukaryotic genes and bacteriophages commonly harbor bacterial genes. However, the presence of bacteriophages in obligate intracellular bacteria of eukaryotes may promote DNA transfers between eukaryotes and bacteriophages. Here we report the first metagenomic analysis of purified bacteriophage WO particles of *Wolbachia* and uncover a novel eukaryotic association module. It encodes domains, such as the black widow latrotoxin-CTD, that are uninterrupted in bacteriophage genomes, enriched with eukaryotic protease cleavage sites, and combined with additional domains to forge one of the largest bacteriophage genes to date (14,256 bp). These domains have never before been reported in packaged bacteriophages, and their phylogeny, distribution and sequence diversity imply lateral transfers between animal and bacteriophage genomes. Finally, the WO genome sequences and identification of attachment sites will potentially advance genetic manipulation of *Wolbachia*.

Viruses are the most abundant and diverse biological entities in the biosphere[1,2]. Infecting organisms across the tree of life, they associate with every ecosystem on the planet. They are generally classified into polythetic groups according to ecological niche and mode of replication[3,4]. While any cellular domain can be infected by a virus, no extant virus is known to traverse more than one domain[5,6]. This domain-specific ecology of viruses underpins the current taxonomic paradigm of trifurcating viruses into eukaryotic, archaeal and bacterial categories, along with recent reappraisals of whether viruses constitute a fourth domain of life[7,8]. As a result of this domain-specific ecology, viruses often integrate host genes via specific highways of lateral gene transfer. Eukaryotic viruses tend to hijack genes directly from their eukaryotic hosts to evade, manipulate and counter-strike anti-viral immune responses[9,10], with the exception of some giant viruses that appear to acquire genes from all domains of life[11]. Bacterial viruses, or bacteriophages (phages), integrate genetic material from their bacterial hosts including toxin[12], photosynthesis[13] and pigment biosynthesis genes[14] that contribute to the fitness of their bacterial host. To date, however, there is no archetypal case of phage particles harboring genomes with eukaryotic DNA.

While all viruses are specific to one of the three domains of life, some bacteriophages target obligate intracellular bacteria of eukaryotic cells. For instance, phage WO infects the obligate intracellular alpha-proteobacteria *Wolbachia*, which in turn infect an estimated 40% of the most speciose group of animals worldwide - arthropods (as well as filarial nematodes). They cause a range of host reproductive pathologies[15,16], primarily infect the cells of host reproductive tissues, exist in Golgi-derived vesicles within the

53 eukaryotic cytoplasm, and are enclosed by a bacterial cell membrane and one or more
54 eukaryotic-derived membranes[17,18]. Nearly all sequenced *Wolbachia* genomes, with the
55 exception of those acting as obligate mutualists, harbor prophage WO[19-21]. They encode
56 conserved structural modules (e.g., head, tail, baseplate) and exhibit *Caudovirales*
57 morphology in electron micrographs of purified phages [20,22-25]. Electron microscopy and
58 quantitative analyses indicate that prophages undergo a lytic phase capable of rupturing
59 bacterial and eukaryotic cell membranes, and phage WO occurs in the extracellular
60 matrix of arthropod gonads[23,26]. Therefore, phage WO appears to uniquely contend with
61 the cellular exit, entry and defense mechanisms of two separate domains of life. WO is
62 also a promising tool for genome editing of *Wolbachia* that has thus far been refractory to
63 genetic modifcation. Here we assemble the first sequenced genomes of phage WO
64 particles, resolve the bacteriophage attachment and bacterial integration sites, report a
65 novel eukaryotic association module in bacteriophages, and discuss lateral gene transfers
66 between eukaryotes and bacteriophages.
67
68 **RESULTS**
69 *Phage WO Genomes Reveal A Novel Eukaryotic Association Module*
70 Here we report the first metagenomic analysis of phage WO particles from *w*VitA-
71 infected *Nasonia girault*i wasps and *w*CauB-infected *Ephestia kuehniella* moths (the *w*-
72 prefix indicates specific *Wolbachia* strain and WO-prefix indicates phage haplotype; see
73 Supplementary Table 1 for a complete list). We identify the phage attachment sites and
74 insertion regions and show from fully sequenced genomes that WO harbors all formerly
75 described phage genetic modules (lysogeny, baseplate, head, replication, virulence, tail

and patatin-like phospholipase[27]) as well as a new group of genes with atypical protein domains indicative of eukaryotic interaction. We collectively group these genes, which include the second largest gene in bacteriophages to date, into a novel "Eukaryotic Association Module" (EAM, white box, Fig. 1). The EAM features genes that (i) encode protein domains and cleavage sites central to eukaryotic functions, (ii) occur in phage and metazoan hosts, (iii) are among the largest genes in phage genomes (up to 14,256 bp) and (iv) are absent from mutualistic, phage-free genomes such as the bedbug-infecting *w*Cle and filarial nematode-infecting *w*Bm and *w*Oo. They occur in all complete prophage WO haplotypes (Supplementary Table 2).

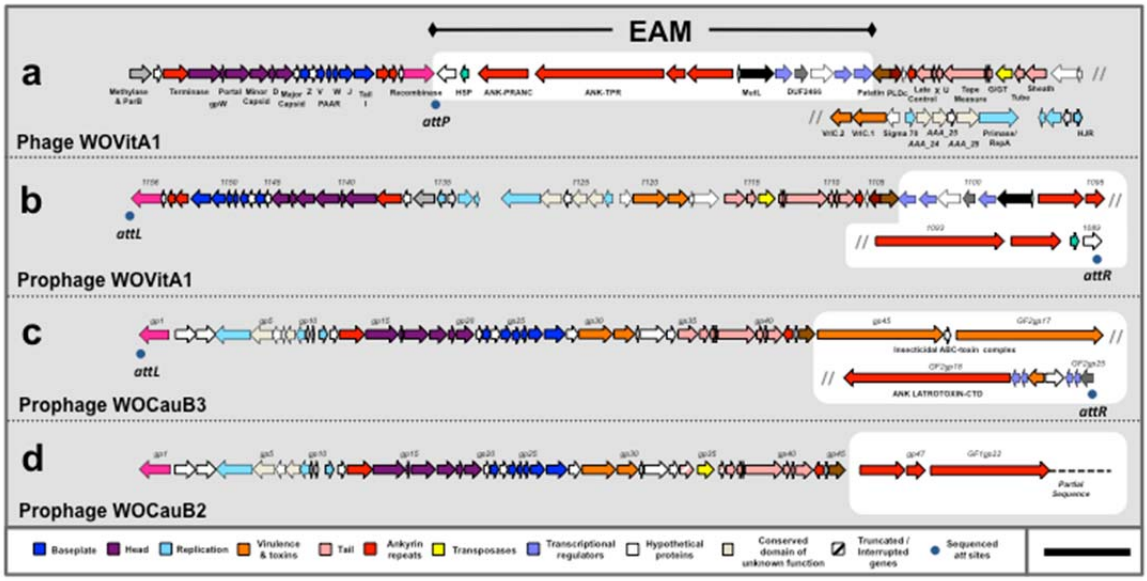


Figure 1 | Phage WO genomes harbor a novel Eukaryotic Association Module (EAM). The complete phage WO genome for (a) WOVitA1 was sequenced directly from purified viral particles using high throughput, metagenomic sequencing. The prophage (b) WOVitA1, (c) WOCauB3 and (d) WOCauB2 genomes were reannotated based on sequencing reads obtained from purified particles; complete genomes of WOCauB3 and WOCauB2 were not obtained. Each genome consists of a bacteriophage-

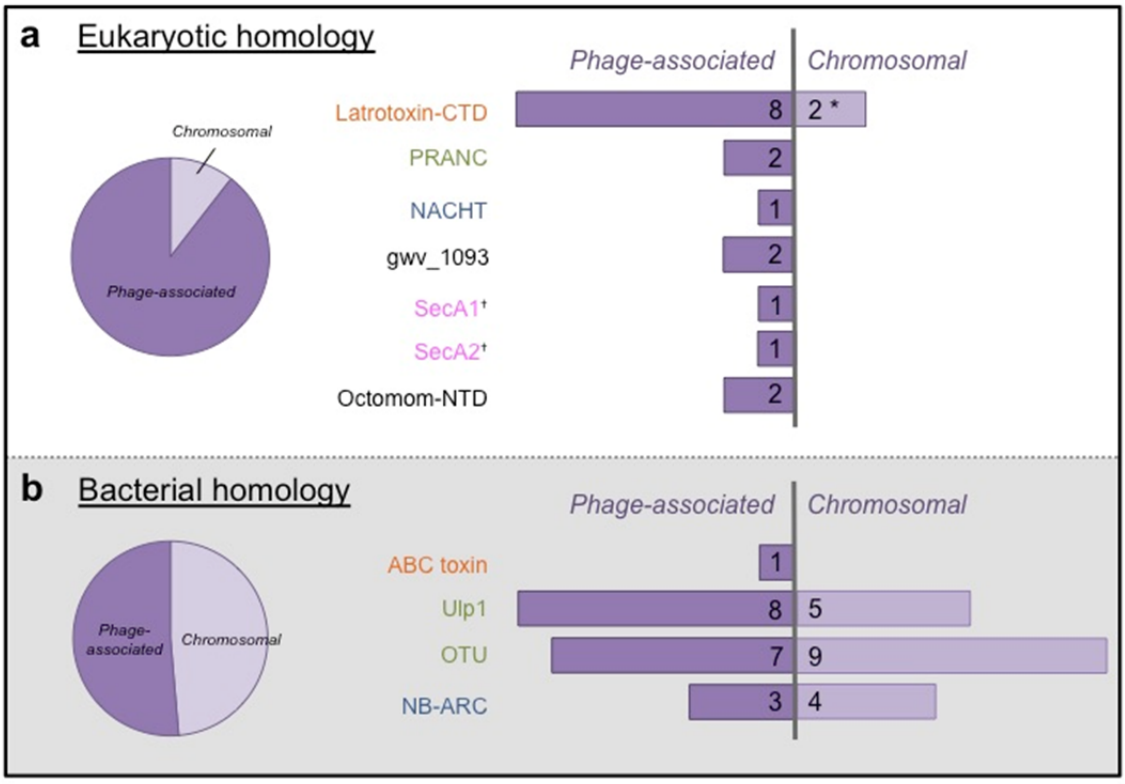
like region (recombinase to patatin) and EAM highlighted in white. Gray slash marks indicate illustrative continuation of the genome. Dark blue dots indicate the discovery of the *attL* and *attR* sites of the prophage, which adjoin in the packaged WO genome to form *attP*. Numbers above the open reading frames indicate locus tags. Scale bar, 5,000 base pairs.

To verify the newly discovered EAM in the phage genome, we identified the terminal prophage WO genes and Sanger sequenced amplicons from an independent sample of phage WOVitA1 (Fig. 1a) across the linear phage *attP* site (hypothetical protein gwv_1089 to recombinase, Supplementary Fig. 1). Next, using the newly identified *attR* and *attL* sites, we extrapolated the bacterial *attB* site in WOVitA1, which is a noncoding, repetitive sequence in *Wolbachia* from *Nasonia* wasps (Supplementary Fig. 1e). The full length of the completely assembled, linear WOVitA1 genome is 65,653 bp, which is 23,531 bp larger than the previous prophage WO annotation. Similarly, we identified the new terminal ends of the WOCauB3 prophage [23,099 bp (51%) larger than original estimate of 45,078 bp], extending the previous observation that the end of the genome is beyond the patatin gene[25], along with internal localization of the EAM genes by Sanger sequencing its *attP* site [Domain of Unknown Function (DUF)2426 to recombinase]. While we were not able to assemble a complete contig for WOCauB2, it is more than 6,854 bp larger than the original estimate of 43,016, includes multiple ankyrin repeat genes homologous to those in WOVitA1, and, like many other prophage haplotypes (e.g., WORiC, WOVitA2, WOSuziC), integrates directly into *Wolbachia*'s magnesium chelatase (*chlI*) gene.

115

116 *The EAM Is Enriched With Eukaryotic-like Domains*

117 We then analyzed each phage WO protein domain for homology and surrounding peptide
118 architecture. Unlike the single domain architecture of phage WO's structural genes, EAM
119 genes are highly polymorphic and encompass fusions of both eukaryotic and bacterial
120 protein domains. By extending the analysis to include homologous prophage regions
121 from all sequenced *Wolbachia* chromosomes, ten types of protein domains with putative
122 eukaryotic functions were uncovered spanning four predicted functions: (i) toxins, (ii)
123 host-microbe interactions, (iii) host cell suicide, and (iv) secretion of proteins through the
124 cell membrane (Fig. 2). Notably, over half of these domain types (6/10; latrotoxin-CTD,
125 PRANC, NACHT, SecA, gwv_1093-NTD, Octomom-NTD) share greater amino acid
126 homology to eukaryotic invertebrates than to bacteria in GenBank. Among this subset
127 with eukaryotic sequence homology, the protein domains are almost exclusively found in
128 the prophage EAM region (N=17) versus the *Wolbachia* chromosome (N=2). In the latter
129 case, the two chromosomal latrotoxin-CTD domains (wNo_10650 and wHa_05390) are
130 flanked by phage-associated genes and transposases, indicating a likely phage WO origin
131 and subsequent genomic rearrangement. This pattern differs from other EAM protein
132 domains with bacterial homology, which are equally dispersed in phage WO (N=19) and
133 the *Wolbachia* chromosome (N=18) (Fig. 2, Fisher's Exact Test, $p = 0.0072$). The
134 difference importantly indicates that the eukaryotic-like protein domains are highly
135 enriched in the EAM, suggesting a near exclusive role in phage WO biology.



136

137 **Figure 2 | Eukaryotic-like EAM genes are enriched in prophage WO regions in the**

138 ***Wolbachia* chromosome.** EAM genes with (a) eukaryotic homology are most likely to be

139 associated with prophage WO while those with (b) bacterial homology are both phage-

140 associated and found scattered throughout the *Wolbachia* chromosome. (*) The two

141 chromosomal latrotoxin-CTD domains (wNo_10650 and wHa_05390) are located within

142 phage-associated genes and transposases, indicating a potential genomic rearrangement.

143 (†) SecA represents one "domain type" but is listed separately because phage WO

144 contains two different homologs (i.e., wHa_3920 and wHa_3930). Putative functional

145 categories are: anti-eukaryotic toxins (orange); host-microbe interactions (green); host

146 cell suicide (blue); secretion of virulence factors (pink); and unknown (black). Octomom

147 refers to WD0513 of the *w*Mel genome.

148

149 *The Black Widow Latrotoxin-CTD*

150 Latrotoxin C-terminal domain (CTD) is the most prevalent eukaryotic domain in
151 prophage WO. Originally described for its major role in the venom of widow spiders
152 (*Latrodectus* species), latrotoxins act extracellularly to cause the formation of ion-
153 permeable membrane pores in their vertebrate or invertebrate victims. The CTD,
154 specifically, is only associated with the latrotoxin precursor molecule (protoxin) and
155 could possibly act intracellularly to facilitate disintegration of the spider's toxin-
156 producing cells[28]. While latrotoxins are generally considered exclusive to spiders, CTD-
157 homologs in *Wolbachia*, *Rickettsiella grylli*[28], and a transcriptome from a *Wolbachia*-
158 infected stink bug[29] have been reported. Here, phylogenetic analysis implies that the
159 latrotoxin-CTD horizontally transferred between widow spiders and phage WO (Fig. 3).
160 Reciprocal search queries using homologous spider and phage CTDs return the same
161 BLASTP hits shown in Fig. 3. Notably, phage WO CTD sequences have the highest
162 amino acid similarity to black widow spider homologs that target invertebrates, which are
163 the primary hosts of *Wolbachia*. While convergent evolution could explain amino acid
164 sequence similarities of the latrotoxin-CTD in black widows and *Wolbachia*, these two
165 taxa occur in overlapping ecological niches (*Wolbachia* are known to infect spiders of the
166 family *Theridiidae*) in which gene transfers are likely to happen[30]. We also confirmed the
167 presence of *Wolbachia* in three independent *Latrodectus geometricus* samples by
168 amplifying *Wolbachia* 16S rDNA and *wsp* membrane protein genes. The transfer event
169 was apparently followed by a relatively more recent transfer from phage WO back to

170 animals in the *Aedes aegypti* genome, where the region is located between genes of
171 mosquito origin [fibrinogen-related protein (AAEL004156) and GalE3 (AAEL004196)].
172
173 *Toxin Activation by Eukaryotic Furin Cleavage*
174 Latrotoxin-CTD is universally located at the 3'-terminal ends of both conserved spider
175 latrotoxin genes[31] and enormous, polymorphic, and eukaryotic-like phage WO genes (up
176 to 14,256 bp). There is a high incidence of eukaryotic furin cleavage sites that
177 immediately precede the latrotoxin-CTD. In spiders, cleavage at these sites by the
178 eukaryotic furin protease in the trans-Golgi network or extracellular matrix is required for
179 latrotoxin activation before the toxin exerts its effects upon the victim. We show that all
180 prophage WO EAMs contain at least one site for eukaryotic furin cleavage
181 (Supplementary Table 3), and the proportion of all EAM genes with predicted furin
182 cleavage sites (25%) is two-fold greater than that of the genes in the core phage genome
183 (11%, Fisher's Exact Test, $p < 0.0001$), defined as the conserved bacteriophage region
184 from recombinase to patatin. In regards to the phage WO latrotoxin-CTD, its preferential
185 localization in prophage WO genomes versus the rest of the *Wolbachia* chromosome,
186 conservation of eukaryotic furin cleaveage sites, large eukaryotic-like length, homology
187 to invertbrate-specific toxins, and reduced divergence relative to the spider venom
188 homologs is consistent with a eukaryotic origin and post-translational processing by furin
189 peptidases.
190

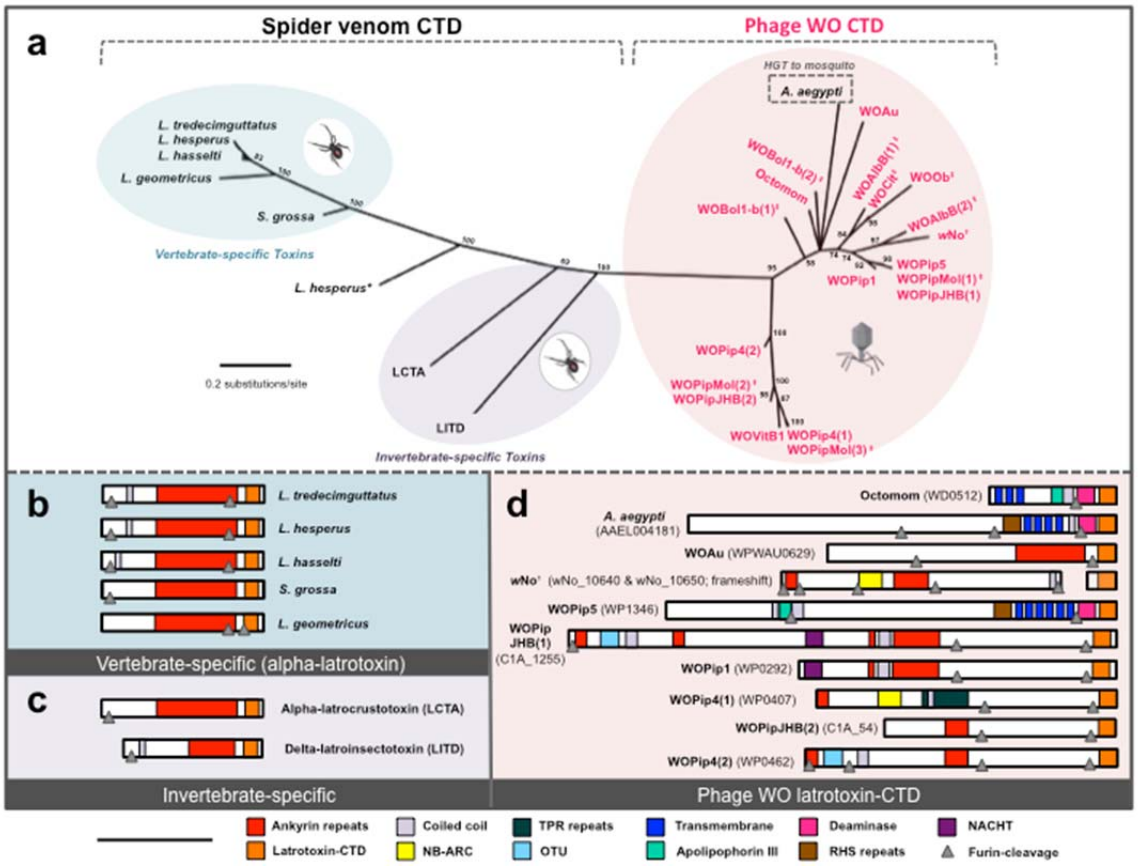


Figure 3 | Latrotoxin-CTD phylogeny and protein architecture support lateral genetic transfers. (a) Phylogeny of phage WO latrotoxin-CTD protein domains and their eukaryotic homologs was constructed by Bayesian analysis of 74 amino acids using the JTT model of evolution. Consensus support values are shown at the nodes. Comparative protein architecture shows that spider venom (b) vertebrate-specific alpha-latrotoxins and (c) invertebrate-specific alpha- and delta-latrotoxins are highly conserved, whereas (d) phage WO are not. Bolded nomenclature in (d) denotes the specific phage WO haplotype (listed as WO). Genome locus tags are listed in parentheses. Predicted furin cleavage sites, listed in Supplementary Table 3, are illustrated with gray triangles. (*) A second *L. hesperus* sequence represents a recently-described downstream paralog with unknown toxin activity[32]. (†) wNo_10650 is located within phage-associated genes and

transposases, indicating a potential genomic rearrangement of a phage region. (‡) Architecture is not shown for sequences on incomplete contigs (WOBol1-b, WOAlbB, WOCit, WOPipMol, WOVitB) because complete peptide information and specific phage association are unknown. Scale bar, 1,000 amino acids.

***P**ox protein **R**epeats of **AN**kyrin **C** terminus (PRANC)*

Domains central to modifying animal proteins are also abundant in the phage WO EAM. The PRANC domain in the WOVitA1 genome (gwv_1092) shares protein sequence homology with corresponding PRANC domains in multiple parasitic wasp hosts (Supplementary Table 4) and their eukaryotic viruses. Reciprocal BLASTP searches retrieve the same best hits and support previous findings that this protein domain horizontally transferred between eukaryotic viruses, animals, and *Proteobacteria*[33]. The discovery here of the eukaryotic-like PRANC domain in phage WO parallels its presence in the *Poxviridae* virus family, in which it functions in evasion of eukaryotic immune responses via modification of host ubiquitination. PRANC is related to amino acid sequences in F-box proteins, which are eukaryotic proteins involved in protein degradation. The PRANC domain also occurs in vaccina virus, ectromelia virus, cowpox virus and Orf virus and can regulate NF-κB signalling pathway to inhibit transcription of inflammatory cytokines[34].

*Conserved Ankyrin and **T**etratrico**p**eptide **R**epeat (TPR) Protein*

Adjacent to the PRANC-encoding gene in WOVitA1's EAM is an ankyrin and TPR-containing gwv_1093. Ankyrin repeats and TPRs mediate a broad range of protein-

protein interactions (apoptosis, cell signaling, inflammatory response, etc.) within eukaryotic cells and are commonly associated with effector proteins of certain intracellular pathogens[35,36]. In *Wolbachia*, ankyrins within the core phage genome have been associated with reproductive manipulation of the insect host[37,38]. While generally rare in viral genomes (Supplementary Fig. 2 and 3, respectively), these repeat regions occur in all prophage WO haplotypes from sequenced *Wolbachia* genomes (N=23). Phylogenetic analysis using reciprocal BLASTP hits (Fig. 4) shows that the N-terminus sequences of the TPR-containing gwv_1093 are embedded within a diverse set of homologs from many athropod lineages (Fig. 4b), with the most recent transfer putatively occurring between phage WO and *Solenopsis invicta* (Fig. 4c). In this species, the gene is located between ant genes bicaudal D and rho guanine nucleotide exchange factor 11. As *S. invicta* can naturally harbor *Wolbachia*[39], either a gene transfer event occurred between these ecologically-associated taxa or the *S. invicta* homolog could be an assembly artifact. This assembly was based on samples from a region rarely infected with *Wolbachia* (Y Wurm, personal communication, April 2016) and there are no other *Wolbachia*/prophage WO homologs in the *S. invicta* genome; therefore, the latter explanation seems unlikely. Moreover, other gwv_1093 homologs are from insect genome sequences of uninfected strains, i.e., *N. vitripennis*, and thus they can not be derived by an assembly artifact. Based on parsimony, the transfer event appears to have occurred from arthopod to phage WO since the arthropod taxa comprise a more diverse set of lineages. However, the reverse is plausible as transfers from *Wolbachia* to their arthropod hosts are common[40-42].

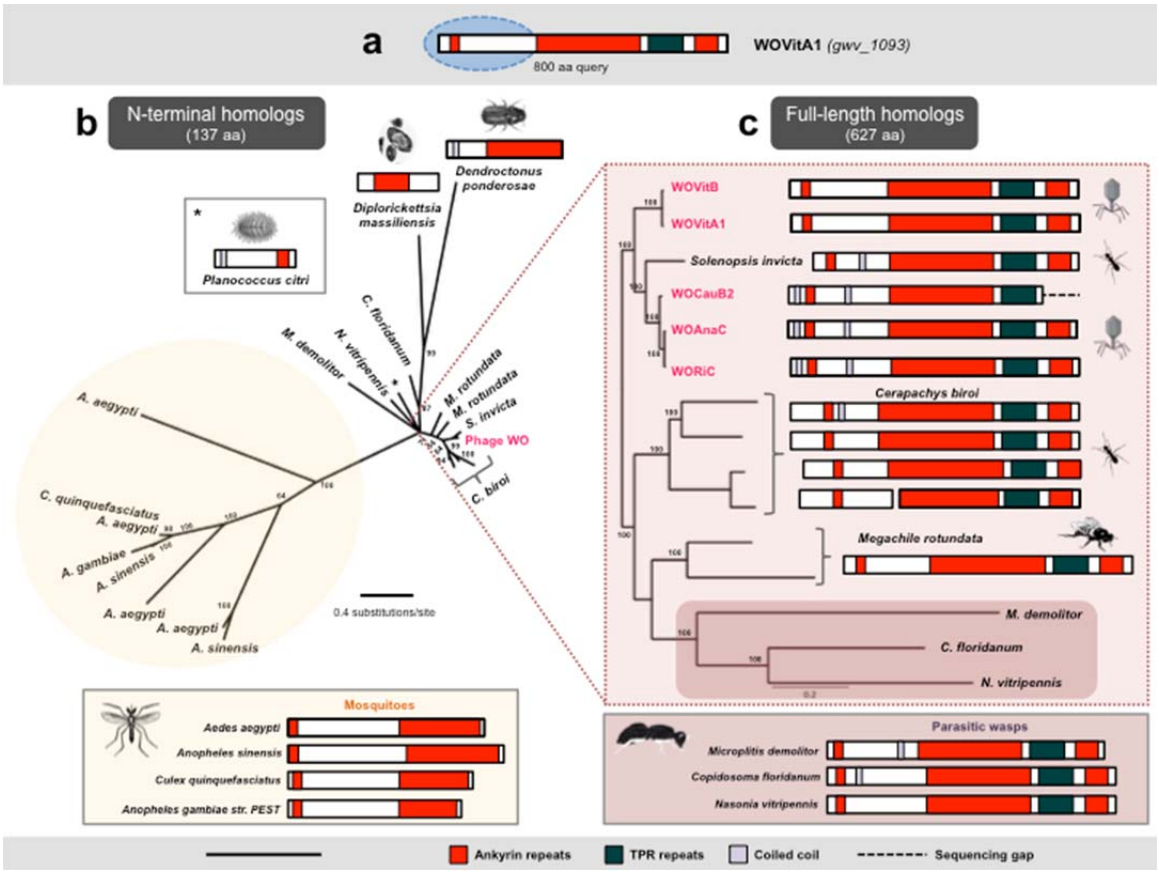


Figure 4 | Phylogeny and protein architecture of a conserved TPR and anyrin-repeat protein support lateral genetic transfers. (a) A BLASTP query of WOVitA1's gwv_1093 N-terminus reveals homologs in moquitoes, ants, beetles, a mealybug, a solitary bee and one obligate intracellular gammaproteobacteria. Bayesian phylogenetic trees were constructed based on (b) a 137-aa alignment of all homologs with E-value less than e^{-40} using the LG+G model of evolution. (c) To resolve taxa closest to phage WO, trees were reconstructed based on a 627-aa alignment of all homologs with an E-value of 0 using the JTT+I+G model of evolution. Isoforms were removed from each alignment. Both trees are unrooted. Consensus support values are shown at the nodes. Chromosomal neighborhood analyses of available animal genome sequences indicate that animal

260 homologs to the phage WO protein are on contigs with other animal genes. Scale bar,
261 1,000 amino acids.
262
263 *NACHT*
264 Another instance of genetic transfer involves the programmed cell death (PCD) domain,
265 NACHT (Fig. 5). Eukaryotic NACHT-containing proteins are typically engaged in PCD
266 by acting as pathogen-sensors and signal transduction molecules of the innate immune
267 system[43]. The polymorphic prophage WO homolog encodes ankyrin repeats and a
268 latrotoxin-CTD directly downstream from the conserved NTPase domain (Fig. 5a).
269 NACHT domains have been identified in animals, fungi and bacteria[44] and phylogenetic
270 patterns indicate multiple instances of horizontal transfer[45]. A NACHT-containing
271 peptide was recently discovered in the *Clostridium difficile*-infecting phage phiCDHM1
272 genome[46] although, in contrast to prophage WO, the phiCDHM1 NACHT domain is
273 bacterial in both amino acid homology and protein architecture. While all BLASTP and
274 reciprocal BLASTP queries of the phiCDHM1 NACHT domain yield only bacterial
275 homologs, BLASTP searches of the prophage WO NACHT domain yield only animals,
276 and reciprocal BLASTP searches of the animal homologs yield only prophage WO and
277 other animals. Similar to the phylogeny of the N-terminus of the TPR-containing
278 gwv_1093, this single NACHT domain sequence in prophage WO is embedded within a
279 more diverse set of homologs in arthropods (Fig. 5b,c). Phylogenetic analyses place the
280 prophage WO variants adjacent to a divergent *Bombyx mori* sequence, though these
281 variants have slightly closer total homology to *Culex quiquefasciatus* mosquitoes that
282 harbor *Wolbachia* with related prophage WO variants.

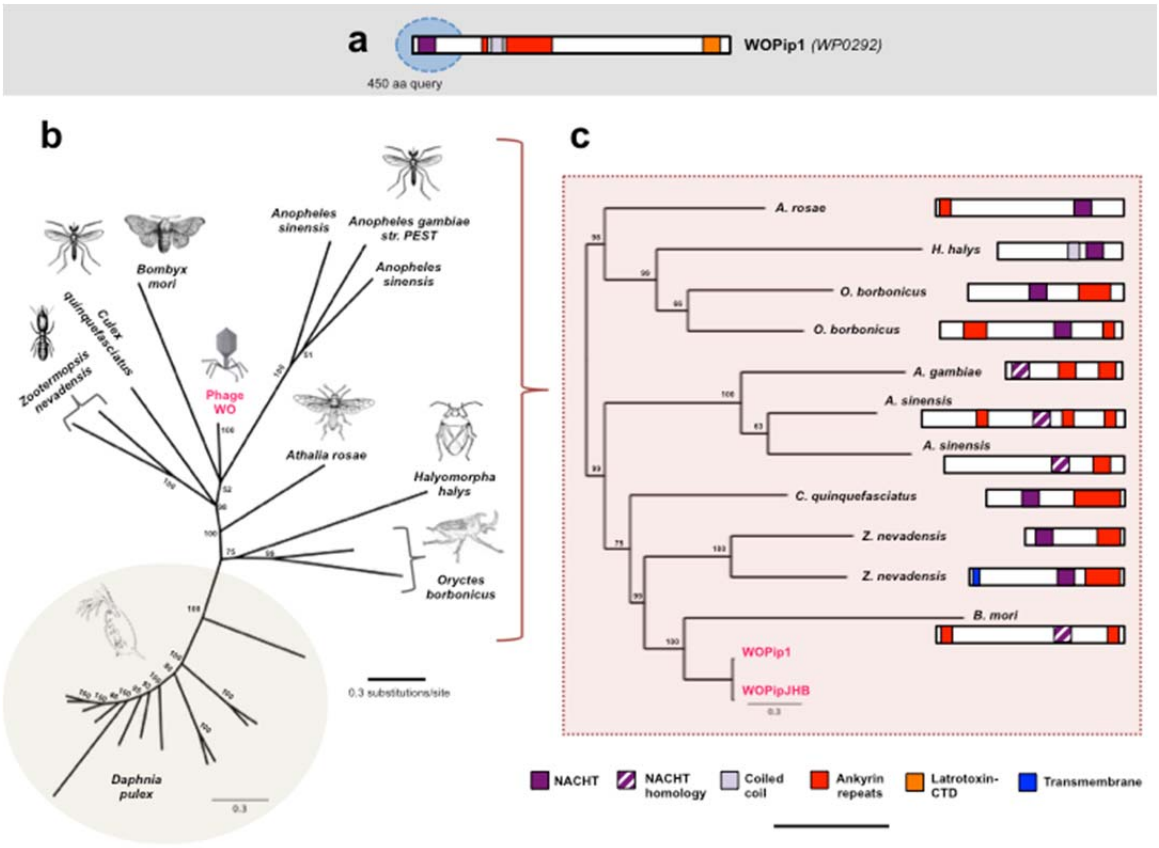


Figure 5 | Phylogeny and protein architecture of the programmed cell death domain, NACHT, support lateral genetic transfers. (a) A BLASTP query of prophage WO's NACHT region reveals homologs throughout arthropods and crustaceans. (b) Bayesian phylogenetic trees were constructed based on a 271-aa alignment of all homologs with E-value less than e^{-15} and coverage greater than 70% using the cpREV+G model of evolution. To resolve taxa closest to prophage WO, all *Daphnia* sequences were removed from the alignment and clusters of highly divergent residues (i.e., 5 or more sequential residues with less than 15% pairwise identity) were trimmed. Trees were reconstructed based on this 262-aa alignment using the LG+G model of evolution. Consensus support values are shown at the nodes. Both trees are unrooted. Chromosomal

295 neighborhood analyses of available animal genome sequences indicate that animal
296 homologs to the prophage WO protein are on contigs with other animal genes. Scale bar,
297 1,000 amino acids.
298
299 **DISCUSSION**
300 The inaugural metagenomic analysis of the complete genome from phage WO particles
301 reveals all formerly described phage genetic modules (lysogeny, baseplate, head,
302 replication, virulence, tail and patatin-like phospholipase[27]) as well as a new group of
303 genes that we collectively group into a eukaryotic associatoin module (EAM). Some of
304 these genes (i) encode protein domains and cleavage sites central to eukaryotic functions,
305 (ii) occur in both phage and metazoan hosts, (iii) comprise the second largest phage gene
306 to date (14,256 bp) and (iv) are absent from mutualistic, phage-free genomes of
307 *Wolbachia*. Together, these genes increase the phage WO genome size by roughly 50%
308 and include ten types of protein domains with four predicted eukaryotic functions: toxins,
309 host-microbe interactions, host cell suicide, and secretion of proteins through the cell
310 membrane. Notably, over half of these domain types share greater amino acid homology
311 to eukaryotic invertebrates than to bacteria in GenBank. Among this subset with
312 eukaryotic sequence homology, the protein domains are almost exclusively found in the
313 phage EAM. An EAM has never before been reported in bacteriophage genomes,
314 possibly because phages of obligate intracellular bacteria occupy a unique eukaryotic-
315 enclosed niche and are relatively understudied.
316

317 The presence of eukaryotic protein domains in bacteriophage genomes is of special note
318 as they curiously mirror eukaryotic genes in large eukaryotic viruses that aid in viral
319 mimicry and manipulation of host processes[47,48]. In phage WO, these animal protein
320 domains are central to anti-eukaryotic functions including the black widow latrotoxin,
321 programmed cell death (NACHT), immune evasion (PRANC), and protein-protein
322 interactions.
323
324 Bacteriophage WO frequently transfer between *Wolbachia* coinfections in the same
325 animal host[49,50] and to the host genome as part of large transfers of the *Wolbachia*
326 chromosome[40,41]. We previously reported that phage WO in *Wolbachia* of *Nasonia*
327 *vitripennis* were also capable of transferring adjacent flanking non-phage genes in the
328 process of exchange between coinfections[51]. For two of these flanking genes, sequence
329 evidence indicated that *Wolbachia* genomes may be able to receive eukaryotic
330 DNA[42,52,53]. However, the nature of these lateral genetic transfers remained to be
331 elucidated as these regions were not previously known to be part of the packaged phage
332 genome until now. Here, we demonstrate that genes with eukaryotic homology are
333 constituents of phage WO and its EAM, and they either retain conservation of eukaryotic
334 furin cleaveage sites and a large eukaryotic-like length (i.e., latrotoxin-CTD), or they
335 exhibit markedly reduced or no diversity relative to the arthropod homologs as the WO
336 sequences exist as single or a few representatives (NACHT and TPR-containing
337 proteins). Moreover, WO protein domains with eukaryotic homology are highly enriched
338 in the EAM over WO protein domains with bacterial homology. Based on this work, we
339 suspect that systematic surveys of phage genomes in intimate host-associated bacteria

may uncover a broad range of eukaryotic-like protein domains involved in phage lifecycle adaptations and phage-eukaryote interactions. Of particular note is the reported association between phage WO genes, specifically ankyrins, transcriptional regulators and the Ulp1 operon, and *Wolbachia*'s ability to manipulate host reproduction[37,38,54-56].

The mechanisms by which eukaryotic protein domains are exchanged with phage WO are unknown and could follow at least three models (Fig. 6). First, animal genetic material could directly transfer to and from WO genomes during phage particle propagation in the cytoplasm of animal cells (Fig. 6b) or during packaging inside *Wolbachia* cells that are lysing and exposed to the eukaryotic cytoplasmic environment. Packaging of eukaryotic host RNAs, for instance, occur in the virions of herpesvirus[57] and cytomegalovirus[58]. Second, genes may transfer between animal genomes and the *Wolbachia* chromosome and then to prophage WO. For this scenario to be plausible, animal genetic material transferred in random locations in the *Wolbachia* genome would have to be preferentially lost in non-phage associated locations from the *Wolbachia* chromosome (Fig. 6c) because domains with eukaryotic homology are highly enriched in the phage/prophage WO EAM versus the rest of the chromosome (Fig. 2). Third, DNA may transfer first between animal genomes and intermediary entities, such as eukaryotic viruses or other obligate intracellular bacteria, and then to phage WO and/or *Wolbachia* (Fib. 6d). In fact, the PRANC-domain (described in gwv_1092) was named for its discovery in and association with eukaryotic Pox viruses. Finally, once DNA is incorporated into a prophage genome, it is susceptible to recombination with other phage WO haplotypes located in the same *Wolbachia* chromosome and can transfer from one haplotype to another.

363

364 Alternatively, these protein domains could originate in the phage and be particularly
365 prone to transfer, maintenance, and spread in their recipient arthropod genomes (Fig. 6b).
366 For this scenario to be plausible, it would have to imply that phage genetic material
367 independently and repeatedly transfers to athropods and spreads through the host
368 population, which would subsequently be followed by loss of these phage genes or
369 recombination with other non-transferred phage genetic material so that the eukaryotic
370 sequence varation clusters seperately from the phage WO sequence(s). While each mode
371 of transfer is possible, the eukaryotic length of these genes, presence of furin protease
372 domains, and enrichment in the phage WO EAM provides evidence for their eukaryotic
373 origin.

374

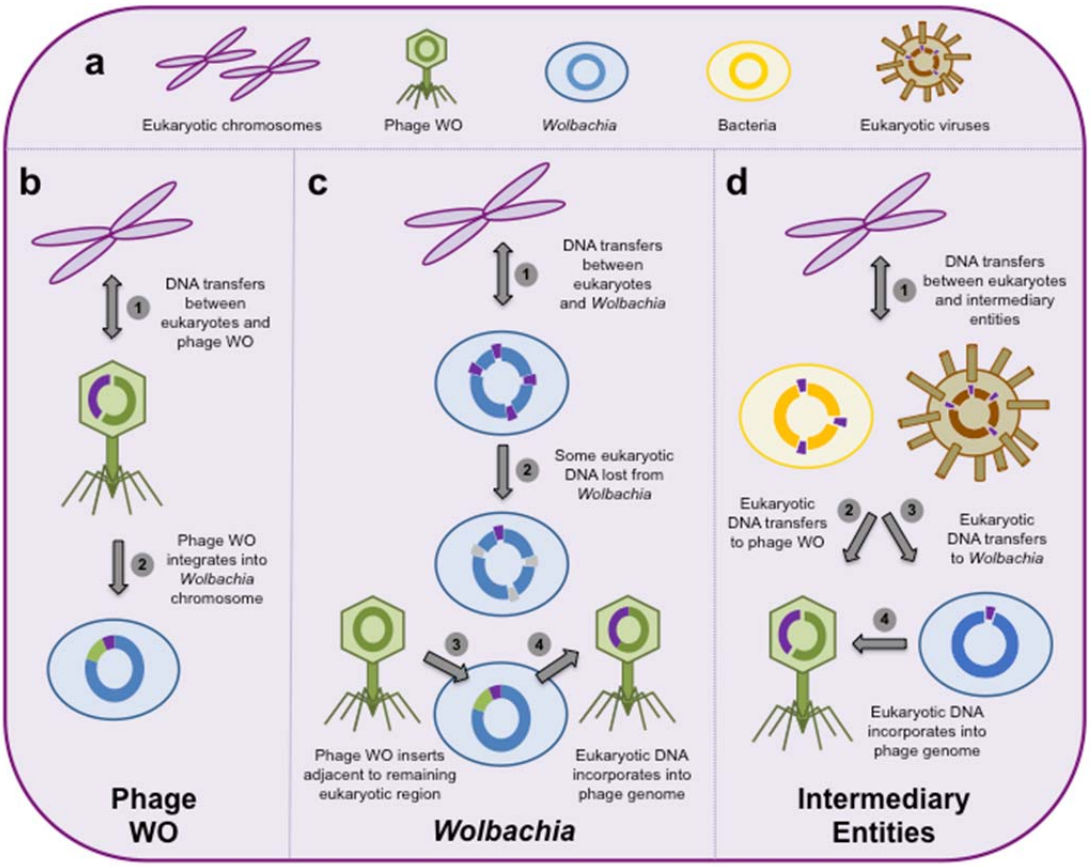


Figure 6 | Models of lateral DNA transfer between eukaryotes and bacteriophages. (a) The eukaryotic cell can harbor multiple microbes capable of horizontal gene transfer. Genetic transfers between eukaryotes and bacteriophages, in particular, can in theory occur (b) directly via incorporation into the phage genome followed by subsequent inclusion in the chromosomal prophage region; (c) indirectly via the transfer of eukaryotic DNA to the *Wolbachia* chromosome or vice versa; and (d) indirectly via the transfer of DNA between eukaryotic chromosomes and intermediary entities, such as eukaryotic viruses and other intracellular bacteria, followed by the transfer to *Wolbachia* and/or phage WO. Since phage EAM genes carrying protein domains central to eukaryotic functions primarily occur in phage/prophage genomes (see Fig. 2), transferred DNA from eukaryotes to non-phage regions in *Wolbachia* is likely eliminated from the

bacterial genome under this model (c). Prophage genomes adjacent to these EAM genes then incorporate the DNA into their packaged genomes and pass it on to new copies of the phage.

Why are these protein domains present in the EAM of bacteriophage WO? Some phages of obligate intracellular bacteria may have to overcome two major challenges not encountered by the well-studied phages of free-living bacteria. First, they are contained within both bacterial and eukaryotic membranes, posing an enigmatic "two-fold cell challenge". They may not only have to breach peptidoglycan and permeabilize bacterial membranes, but they may also have to exit (and enter) across the eukaryotic membrane(s) that directly encapsulates the bacteria. Second, like their bacterial hosts, they must survive the internal cellular environment of the animal host, including the innate immune response and autophagy, while searching for phage-susceptible bacteria. Phage WO can dwell in the eukaryotic cytoplasm and extracellular matrix that they encounter upon bacterial lysis[26], raising the likelihood of direct interaction with host membranes and intracellular biology. In this context, EAM protein domains are prime candidates to aid in functions including cell lysis (latrotoxin-CTD), manipulation of programmed cell death (NACHT and NB-ARC), host ubiquitination (OTU and Ulp1), insecticidal toxicity (ABC toxin) and interaction with host proteins (ankryin repeats and TPRs). Rather than simply act as virulence factors to benefit their bacterial host, their massive proportion of genomic real estate (up to 60% of the prophage genome, Supplementary Fig. 4) implies that they may be necessary to phage biology and likely have a direct impact on phage propagation. The concept of phage-mediated ecosystem modification as an alternative to bacterial

410 virulence is not new[59] but, much like the biology of phage WO, is relatively
411 understudied.
412
413 Phage WO is not the only virus described within obligate intracellular bacteria.
414 *Chlamydiomicroviridae* infect obligate intracellular bacteria, yet still do not directly
415 contend with the eukaryotic membrane. Rather, they attach to dormant chlamydial cells
416 (i.e., reticulate bodies) and enter via phagocytosis or endocytosis of the bacteria[60]. The
417 phages then alter development of their bacterial host, which leads to disintegration of the
418 chlamydial inclusion and subsequent lysis of the eukaryotic host cell[61,62]. The nature of
419 phage WO's lifestyle, on the other hand, may require a distinct interaction with multiple
420 membranes and immune responses because lytic activity of phage WO has been
421 associated with typical bacterial cell defects including degraded bacterial DNA, a
422 detached inner membrane, and exit of the phage particles from inside *Wolbachia* and its
423 host cell into the extracellular matrix of the reproductive tissues[26]. Bacteriophages of
424 free-living bacteria also regularly colonize eukaryotic environments, particularly those
425 associated with mucosal surfaces[63]. They, however, do not infect or traverse the
426 eukaryotic membrane and are still within the genomic boundaries of the bacterial
427 virosphere.
428
429 Temperate dsDNA phages also occur in facultative symbionts of aphids[64] and tsetse
430 flies[65]. While *Wolbachia* has never successfully been cultured outside of host cells[66],
431 these facultative symbionts can replicate both intra- and extracellularly (JW Brandt,
432 personal communication, July 2015) suggesting that their phages are not constrained by

433 the same two-fold cell challenge. In addition, their phages encode a traditional lytic
434 cassette (holin and lysozyme) that correlates with the need to deal only with bacterial
435 membranes. In some cases, the phages harbor bacterial-derived toxins that target
436 eukaryotic cells[67], and these function mutualistically in aphids by arresting development
437 of parasitoid wasp larvae[64]. Furthermore, unlike phage WO that is stably maintained in
438 the lab, these phages are readily lost in the absence of parasitoids during laboratory
439 rearing, presumably due to the cost of their toxins[68].
440
441 In addition to providing new insights into the evolution of bacteriophages and showing
442 phage WO genomes to be far more complex than previously described, the findings here
443 reveal evidence for gene sharing between metazoan hosts and phages of obligate
444 intracellular bacteria. We suggest that the putative acquistion and retooling of intact
445 eukaryotic domains in phage WO is analgous to the commandeering of host genes by
446 eukaryotic viruses. Whether lateral genetic transfers between metazoans and
447 bacteriophages are common in the symbiotic virosphere remains to be determined.
448

METHODS

Insect and Bacterial Strains

The transfected line of the Mediterranean flour moth *Ephestia kuehniella* harboring *Wolbachia* strain *w*CauB was obtained with the help of Takema Fukatsu and Tetsuhiko Sasaki[22]. Moths were maintained at 24˚C and 70% humidity on a diet consisting of wheat bran, glycerol and dried yeast (20:2:1 w/w). The introgressed line of the parasitoid wasp *Nasonia girault*i harboring *Wolbachia* strain *w*VitA, termed IntG12.1, was previously derived by repeatedly backcrossing *N. vitripennis* (strain 12.1) females to *N. girault*i males for nine generations[69]. The strain was incubated at 25˚C using the flesh fly *Sarcophaga bullata* as host.

Phage Particle Purification

Phage particles were isolated according to Fujii et al[22] with modifications. Approximately 4 g of adult insects were homogenized in 29.6 ml cold SM buffer (50mM Tris-HCl, pH 7.5, 0.1 M NaCl, 10mM $MgSO_4 \cdot 7H_20$, and 0.1% (w/v) gelatin). NaCl and RNase A were added to a final concentration of 1M and 1μg/ml, respectively. The homogenate was incubated on a shaker at 4˚C for 1 h and then centrifuged at 13,000*g* for 10 min at 4˚C. Polyethylene glycol (PEG) 6000 was added to a final concentration of 10% to precipitate phage particles, incubated at 4˚C for 1 hr with gentle shaking and centrifuged at 13,000g for 10 min. The pellet was resuspended in 5 ml TM buffer (50 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2 \cdot 6H_2O$) and mixed with an equal volume chloroform. The suspension was centrifuged at 3,000*g* to remove PEG and the aqueous phase was filtered through a 0.22

μm filter to remove bacterial cells. The suspension was centrifuged at 60,000*g* for 1 h at 4˚C to collect phage particles. The pellet was suspended in 10 μl TM buffer.

Phage DNA Extraction & Metagenomic Sequencing

The phage suspension was treated with RQ1 RNase-Free DNase (Promega) for 30 min at 37˚C, followed by heat inactivation for 10 min at 65˚C, to remove host DNA contamination. Phage DNA was extracted from the suspension using the QIAamp MinElute Virus Spin Kit (Qiagen) and amplified using the REPLI-g Mini Kit (Qiagen). Following amplification, paired-end DNA libraries were prepared according to manufacturer's (Illumina) instructions and samples were sequenced with an Illumina HiSeq 2000 (2×100-nt read length).

Bioinformatics & Statistics

Metagenomic sequences (reads) were trimmed, paired and assembled into contigs using the CLC Assembler (CLC bio) with bubble size = 50, insertion and deletion cost = 3, mismatch cost = 2, length fraction = 0.6, minimum contig size = 130, similarity = 0.5, minimum distance = 90 and maximum distance = 200. Contigs were compared to the GenBank non-redundant database using NCBI's BLASTN (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and those with similarity to phage WO and/or *Wolbachia* (E-value $<10^{-10}$) were manually annotated using Geneious (Biomatters Ltd.). Individual reads were mapped to reference sequences using Geneious. Open reading frame (ORF) homology searches were performed to determine putative function using NCBI's BLASTP (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and Wellcome Trust Sanger

494 Institute's pfam database (http://pfam.sanger.ac.uk). Coiled coil domains were predicted
495 with EMBL's Simple Modular Architecture Research Tool (SMART, http://smart.embl-
496 heidelberg.de). Furin cleavage sites were identified using PiTou
497 (http://www.nuolan.net/reference.html). The number of genes with and without furin
498 cleavage sites was analyzed with respect to phage-region using Fisher's Exact Test
499 (GraphPad Software). Phylogenetic trees were built using the Bayes plugin in Geneious
500 and model selection for each Bayes analysis was estimated using ProtTest[70].
501
502 *Confirmation of Phage WO Terminal Genes*
503 Genomic DNA was extracted from *w*VitA-infected *N. vitripennis* (strain 12.1) and
504 *w*CauB-infected *E. kuehniella* individuals using the Gentra Puregene Tissue Kit (Qiagen).
505 Primers were designed for both WOVitA1 and WOCauB3 *attP* sites, respectively:
506 VitA1_attF (5'- CGA AGA ACC AGC ACA GGG TGG-3'), VitA1_attR (5'- GCT GGA
507 AGA GGG CAT CTG CAT C-3'), CauB3_attF (5'- TCG TGA CTG CCC TAT TGC
508 TGC T – 3') and CauB3_attR (5'- ATG CGG CCA AAG CTG GGT GT – 3').
509 Amplification was performed in a Veriti thermal cycler (Applied Biosystems) using
510 GoTaq green master mix (Promega) under the following conditions: 94˚C for 2 min; 35
511 cycles of 94˚C for 30 s, 53˚C for 30 s, 72˚C for 1 min; and a final elongation cycle of
512 72˚C for 10 min. PCR products were sequenced via Sanger sequencing (Genewiz, Inc).
513

REFERENCES

1. Edwards, R.A. & Rohwer, F. Viral metagenomics. *Nat Rev Microbiol* **3**, 504-10 (2005).
2. Hendrix, R.W., Smith, M.C., Burns, R.N., Ford, M.E. & Hatfull, G.F. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. *Proc Natl Acad Sci U S A* **96**, 2192-7 (1999).
3. Brussow, H. The not so universal tree of life or the place of viruses in the living world. *Philos Trans R Soc Lond B Biol Sci* **364**, 2263-74 (2009).
4. King, A.M.Q., Adams, M.J., Lefkowitz, E.J. & Carstens, E.B. *Virus taxonomy: classification and nomenclature of viruses: Ninth Report of the International Committee on Taxonomy of Viruses.*, 1327 (Elsevier, San Diego, 2012).
5. Nasir, A., Forterre, P., Kim, K.M. & Caetano-Anolles, G. The distribution and impact of viral lineages in domains of life. *Front Microbiol* **5**, 194 (2014).
6. Prangishvili, D., Forterre, P. & Garrett, R.A. Viruses of the Archaea: a unifying view. *Nat Rev Microbiol* **4**, 837-48 (2006).
7. Forterre, P. Giant viruses: conflicts in revisiting the virus concept. *Intervirology* **53**, 362-78 (2010).
8. Raoult, D. TRUC or the need for a new microbial classification. *Intervirology* **56**, 349-53 (2013).
9. Elde, N.C. & Malik, H.S. The evolutionary conundrum of pathogen mimicry. *Nat Rev Microbiol* **7**, 787-97 (2009).
10. Rappoport, N. & Linial, M. Viral proteins acquired from a host converge to simplified domain architectures. *PLoS Comput Biol* **8**, e1002364 (2012).
11. Colson, P. & Raoult, D. Gene repertoire of amoeba-associated giant viruses. *Intervirology* **53**, 330-43 (2010).
12. Canchaya, C., Fournous, G. & Brussow, H. The impact of prophages on bacterial chromosomes. *Mol Microbiol* **53**, 9-18 (2004).
13. Lindell, D. *et al.* Transfer of photosynthesis genes to and from Prochlorococcus viruses. *Proc Natl Acad Sci U S A* **101**, 11013-8 (2004).
14. Dammeyer, T., Bagby, S.C., Sullivan, M.B., Chisholm, S.W. & Frankenberg-Dinkel, N. Efficient phage-mediated pigment biosynthesis in oceanic cyanobacteria. *Curr Biol* **18**, 442-8 (2008).
15. Werren, J.H., Baldo, L. & Clark, M.E. Wolbachia: master manipulators of invertebrate biology. *Nat Rev Microbiol* **6**, 741-51 (2008).
16. Zug, R. & Hammerstein, P. Still a host of hosts for Wolbachia: analysis of recent data suggests that 40% of terrestrial arthropod species are infected. *PLoS One* **7**, e38544 (2012).
17. Cho, K.O., Kim, G.W. & Lee, O.K. Wolbachia bacteria reside in host Golgi-related vesicles whose position is regulated by polarity proteins. *PLoS One* **6**, e22703 (2011).
18. Louis, C. & Nigro, L. Ultrastructual evidence of Wolbachia Rickettsiales in Drosophila simulans and their relationships with unidirectional cross-incompatibility. *Journal of Invertebrate Pathology* **54**, 39-44 (1989).
19. Gavotte, L. *et al.* A Survey of the bacteriophage WO in the endosymbiotic bacteria Wolbachia. *Mol Biol Evol* **24**, 427-35 (2007).

20. Kent, B.N. & Bordenstein, S.R. Phage WO of Wolbachia: lambda of the endosymbiont world. *Trends Microbiol* **18**, 173-81 (2010).
21. Metcalf, J.A. & Bordenstein, S.R. The complexity of virus systems: the case of endosymbionts. *Curr Opin Microbiol* **15**, 546-52 (2012).
22. Fujii, Y., Kubo, T., Ishikawa, H. & Sasaki, T. Isolation and characterization of the bacteriophage WO from Wolbachia, an arthropod endosymbiont. *Biochem Biophys Res Commun* **317**, 1183-8 (2004).
23. Masui, S. *et al.* Bacteriophage WO and virus-like particles in Wolbachia, an endosymbiont of arthropods. *Biochem Biophys Res Commun* **283**, 1099-104 (2001).
24. Sanogo, Y.O. & Dobson, S.L. WO bacteriophage transcription in Wolbachia-infected Culex pipiens. *Insect Biochem Mol Biol* **36**, 80-5 (2006).
25. 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* **75**, 5676-86 (2009).
26. Bordenstein, S.R., Marshall, M.L., Fry, A.J., Kim, U. & Wernegreen, J.J. The tripartite associations between bacteriophage, Wolbachia, and arthropods. *PLoS Pathog* **2**, e43 (2006).
27. Kent, B.N., Funkhouser, L.J., Setia, S. & Bordenstein, S.R. Evolutionary genomics of a temperate bacteriophage in an obligate intracellular bacteria (Wolbachia). *PLoS One* **6**, e24984 (2011).
28. Zhang, D., de Souza, R.F., Anantharaman, V., Iyer, L.M. & Aravind, L. Polymorphic toxin systems: Comprehensive characterization of trafficking modes, processing, mechanisms of action, immunity and ecology using comparative genomics. *Biol Direct* **7**, 18 (2012).
29. Ioannidis, P. *et al.* Rapid transcriptome sequencing of an invasive pest, the brown marmorated stink bug Halyomorpha halys. *BMC Genomics* **15**, 738 (2014).
30. Goodacre, S.L., Martin, O.Y., Thomas, C.F. & Hewitt, G.M. Wolbachia and other endosymbiont infections in spiders. *Mol Ecol* **15**, 517-27 (2006).
31. Garb, J.E. & Hayashi, C.Y. Molecular evolution of alpha-latrotoxin, the exceptionally potent vertebrate neurotoxin in black widow spider venom. *Mol Biol Evol* **30**, 999-1014 (2013).
32. Bhere, K.V., Haney, R.A., Ayoub, N.A. & Garb, J.E. Gene structure, regulatory control, and evolution of black widow venom latrotoxins. *FEBS Lett* **588**, 3891-7 (2014).
33. Werren, J.H. *et al.* Functional and evolutionary insights from the genomes of three parasitoid Nasonia species. *Science* **327**, 343-8 (2010).
34. Chang, S.J. *et al.* Poxvirus host range protein CP77 contains an F-box-like domain that is necessary to suppress NF-kappaB activation by tumor necrosis factor alpha but is independent of its host range function. *J Virol* **83**, 4140-52 (2009).
35. Cerveny, L. *et al.* Tetratricopeptide repeat motifs in the world of bacterial pathogens: role in virulence mechanisms. *Infect Immun* **81**, 629-35 (2013).

603 36. Pan, X., Luhrmann, A., Satoh, A., Laskowski-Arce, M.A. & Roy, C.R. Ankyrin
604 repeat proteins comprise a diverse family of bacterial type IV effectors. *Science*
605 **320**, 1651-4 (2008).
606 37. Pichon, S. *et al.* The expression of one ankyrin pk2 allele of the WO prophage is
607 correlated with the Wolbachia feminizing effect in isopods. *BMC Microbiol* **12**,
608 55 (2012).
609 38. Walker, T. *et al.* Ankyrin repeat domain-encoding genes in the wPip strain of
610 Wolbachia from the Culex pipiens group. *BMC Biol* **5**, 39 (2007).
611 39. Shoemaker, D.D., Ross, K.G., Keller, L., Vargo, E.L. & Werren, J.H. Wolbachia
612 infections in native and introduced populations of fire ants (Solenopsis spp.).
613 *Insect Mol Biol* **9**, 661-73 (2000).
614 40. Dunning Hotopp, J.C. *et al.* Widespread lateral gene transfer from intracellular
615 bacteria to multicellular eukaryotes. *Science* **317**, 1753-6 (2007).
616 41. Funkhouser-Jones, L.J. *et al.* Wolbachia co-infection in a hybrid zone: discovery
617 of horizontal gene transfers from two Wolbachia supergroups into an animal
618 genome. *PeerJ* **3**, e1479 (2015).
619 42. Klasson, L., Kambris, Z., Cook, P.E., Walker, T. & Sinkins, S.P. Horizontal gene
620 transfer between Wolbachia and the mosquito Aedes aegypti. *BMC Genomics* **10**,
621 33 (2009).
622 43. Koonin, E.V. & Aravind, L. Origin and evolution of eukaryotic apoptosis: the
623 bacterial connection. *Cell Death Differ* **9**, 394-404 (2002).
624 44. Koonin, E.V. & Aravind, L. The NACHT family - a new group of predicted
625 NTPases implicated in apoptosis and MHC transcription activation. *Trends*
626 *Biochem Sci* **25**, 223-4 (2000).
627 45. Leipe, D.D., Koonin, E.V. & Aravind, L. STAND, a class of P-loop NTPases
628 including animal and plant regulators of programmed cell death: multiple,
629 complex domain architectures, unusual phyletic patterns, and evolution by
630 horizontal gene transfer. *J Mol Biol* **343**, 1-28 (2004).
631 46. Hargreaves, K.R., Kropinski, A.M. & Clokie, M.R. What does the talking?:
632 quorum sensing signalling genes discovered in a bacteriophage genome. *PLoS*
633 *One* **9**, e85131 (2014).
634 47. Alcami, A. & Koszinowski, U.H. Viral mechanisms of immune evasion. *Immunol*
635 *Today* **21**, 447-55 (2000).
636 48. Seet, B.T. *et al.* Poxviruses and immune evasion. *Annu Rev Immunol* **21**, 377-423
637 (2003).
638 49. Bordenstein, S.R. & Wernegreen, J.J. Bacteriophage flux in endosymbionts
639 (Wolbachia): infection frequency, lateral transfer, and recombination rates. *Mol*
640 *Biol Evol* **21**, 1981-91 (2004).
641 50. Masui, S., Kamoda, S., Sasaki, T. & Ishikawa, H. Distribution and evolution of
642 bacteriophage WO in Wolbachia, the endosymbiont causing sexual alterations in
643 arthropods. *J Mol Evol* **51**, 491-7 (2000).
644 51. Kent, B.N. *et al.* Complete bacteriophage transfer in a bacterial endosymbiont
645 (Wolbachia) determined by targeted genome capture. *Genome Biol Evol* **3**, 209-18
646 (2011).

52. Duplouy, A. *et al.* Draft genome sequence of the male-killing Wolbachia strain wBol1 reveals recent horizontal gene transfers from diverse sources. *BMC Genomics* **14**, 20 (2013).
53. Woolfit, M., Iturbe-Ormaetxe, I., McGraw, E.A. & O'Neill, S.L. An ancient horizontal gene transfer between mosquito and the endosymbiotic bacterium Wolbachia pipientis. *Mol Biol Evol* **26**, 367-74 (2009).
54. Beckmann, J.F. & Fallon, A.M. Detection of the Wolbachia protein WPIP0282 in mosquito spermathecae: implications for cytoplasmic incompatibility. *Insect Biochem Mol Biol* **43**, 867-78 (2013).
55. Pinto, S.B. *et al.* Transcriptional regulation of Culex pipiens mosquitoes by Wolbachia influences cytoplasmic incompatibility. *PLoS Pathog* **9**, e1003647 (2013).
56. Sutton, E.R., Harris, S.R., Parkhill, J. & Sinkins, S.P. Comparative genome analysis of Wolbachia strain wAu. *BMC Genomics* **15**, 928 (2014).
57. Amen, M.A. & Griffiths, A. Packaging of Non-Coding RNAs into Herpesvirus Virions: Comparisons to Coding RNAs. *Front Genet* **2**, 81 (2011).
58. Terhune, S.S., Schroer, J. & Shenk, T. RNAs are packaged into human cytomegalovirus virions in proportion to their intracellular concentration. *J Virol* **78**, 10390-8 (2004).
59. Abedon, S.T. & Lejeune, J.T. Why bacteriophage encode exotoxins and other virulence factors. *Evol Bioinform Online* **1**, 97-110 (2005).
60. Sliwa-Dominiak, J., Suszynska, E., Pawlikowska, M. & Deptula, W. Chlamydia bacteriophages. *Arch Microbiol* **195**, 765-71 (2013).
61. Hsia, R., Ohayon, H., Gounon, P., Dautry-Varsat, A. & Bavoil, P.M. Phage infection of the obligate intracellular bacterium, Chlamydia psittaci strain guinea pig inclusion conjunctivitis. *Microbes Infect* **2**, 761-72 (2000).
62. Salim, O., Skilton, R.J., Lambden, P.R., Fane, B.A. & Clarke, I.N. Behind the chlamydial cloak: the replication cycle of chlamydiaphage Chp2, revealed. *Virology* **377**, 440-5 (2008).
63. Barr, J.J. *et al.* Bacteriophage adhering to mucus provide a non-host-derived immunity. *Proc Natl Acad Sci U S A* **110**, 10771-6 (2013).
64. Moran, N.A., Degnan, P.H., Santos, S.R., Dunbar, H.E. & Ochman, H. The players in a mutualistic symbiosis: insects, bacteria, viruses, and virulence genes. *Proc Natl Acad Sci U S A* **102**, 16919-26 (2005).
65. Belda, E., Moya, A., Bentley, S. & Silva, F.J. Mobile genetic element proliferation and gene inactivation impact over the genome structure and metabolic capabilities of Sodalis glossinidius, the secondary endosymbiont of tsetse flies. *BMC Genomics* **11**, 449 (2010).
66. Rasgon, J.L., Gamston, C.E. & Ren, X. Survival of Wolbachia pipientis in cell-free medium. *Appl Environ Microbiol* **72**, 6934-7 (2006).
67. Degnan, P.H. & Moran, N.A. Diverse phage-encoded toxins in a protective insect endosymbiont. *Appl Environ Microbiol* **74**, 6782-91 (2008).
68. Oliver, K.M., Degnan, P.H., Hunter, M.S. & Moran, N.A. Bacteriophages encode factors required for protection in a symbiotic mutualism. *Science* **325**, 992-4 (2009).

692 69. Chafee, M.E. *et al.* Decoupling of host-symbiont-phage coadaptations following
693 transfer between insect species. *Genetics* **187**, 203-15 (2011).
694 70. Abascal, F., Zardoya, R. & Posada, D. ProtTest: selection of best-fit models of
695 protein evolution. *Bioinformatics* **21**, 2104-5 (2005).
696

697

SUPPLEMENTARY INFORMATION

Supplementary Figure 1 | Sequencing reveals the phage, prophage and bacterial *att* sites for WOVitA1.

Supplementary Figure 2 | Ankyrin repeat domain.

Supplementary Figure 3 | TPR domain.

Supplementary Figure 4 | Prophage WO dedicates nearly half of its genome to eukaryotic host association.

Supplementary Table 1 | *Wolbachia* and phage WO nomenclature.

Supplementary Table 2 | Comparative genomics of prophage WO.

Supplementary Table 3 | Phage WO EAM furin cleavage sites.

Supplementary Table 4 | The phage WO PRANC domain shares amino acid homology with multiple eukaryotic host peptides.

DATA AVAILABILITY

The phage WOVitA1 genome assembly reported in this paper has been deposited in NCBI under accession number KX522565.

The *N. vitripennis* viral metagenome sequences have been deposited in the SRA under accession number SRR3560636 and BioProject PRJNA321548. The *w*CauB-infected *E. kuehniella* viral metagenome sequences have been deposited in the SRA under accession number SRR3536639 and BioProject PRJNA321549.

ACKNOWLEDGEMENTS

We thank Tetsuhiko Sasaki and Takema Fukatsu for assistance with providing the *w*CauB-infected moths; Rini Pauly for computational assistance; Merri Lynn Casem for providing widow spiders; Kerry Oliver for information on *Hamiltonella*; Yannick Wurm for information on *Solenopsis*; and Michael Gerth, Julie Dunning Hotopp, Kevin Kohl, Jason Metcalf for feedback on the manuscript. We apologize to our colleagues for not being able to include all possible references due to citation restrictions. This research was funded by NIH Awards R01 GM085163 and R21 HD086833 and NSF Awards DEB 1046149 and IOS 1456778 to SRB. The funders had no role in study design, data collection and interpretation or the decision to submit the work for publication.

AUTHOR CONTRIBUTION

Sarah Bordenstein designed and performed the experiments, analyzed the data, prepared figures and tables, wrote and reviewed drafts of the paper.

Seth Bordenstein conceived and helped design the experiments, analyzed the data, wrote and reviewed drafts of the paper.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

AUTHOR INFORMATION

Correspondence and requests for materials should be addressed to s.bordenstein@vanderbilt.edu or sarah.bordenstein@vanderbilt.edu.