Lateral Genetic Transfers Between Eukaryotes and Bacteriophages

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 genes are typically co-opted by eukaryotic viruses and bacterial genes are commonly found in bacteriophages. However, the presence of bacteriophages in symbiotic bacteria that obligately reside in eukaryotes may promote eukayotic DNA transfers to bacteriophages. By sequencing full genomes from purified bacteriophage WO particles of *Wolbachia,* we discover a novel eukaryotic association module with various animal proteins domains, such as the black widow latrotoxin-CTD, that are uninterrupted in intact bacteriophage genomes, enriched with eukaryotic protease cleavage sites, and combined with additional domains to forge some of the largest bacteriophage genes (up to 14,256 bp). These various protein domain families are central to eukaryotic functions and have never before been reported in packaged bacteriophages, and their phylogeny, distribution and sequence diversity implies lateral transfer from animal to bacteriophage genomes. We suggest that the evolution of these eukaryotic protein domains in bacteriophage WO parallels the evolution of eukaryotic genes in canonical eukaryotic viruses, namely those commandeered for viral life cycle adaptations. Analogous selective pressures and evolutionary outcomes may occur in bacteriophage WO as a result of its "two-fold cell challenge" to persist in and traverse cells of obligate intracellular bacteria that strictly reside in animal cells. Finally, the full WO genome sequences and identification of attachment sites will advance eventual genetic manipulation of *Wolbachia* for disease control strategies.

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[3]. They are generally classified into polythetic groups according to ecological niche and mode of replication[4,5]. While any cellular domain can be infected by a virus, no extant virus is known to traverse more than one domain[6,7]. 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[8-10]. 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[11,12], with the exception of some giant viruses that appear to acquire genes from all domains of life[13]. Bacterial viruses, or bacteriophages (phages), only integrate genetic material from their bacterial hosts including toxin[14], photosynthesis[15] and pigment biosynthesis genes[16] 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[17,18], primarily infect the cells of host reproductive tissues, exist in Golgi-derived vesicles within the

58 eukaryotic cytoplasm, and are enclosed by a bacterial cell membrane and one or more
59 eukaryotic-derived membranes[19-21]. Nearly all sequenced *Wolbachia* genomes, with the
60 exception of those acting as obligate mutualists, harbor prophage WO[22-24]. They encode
61 conserved structural modules (e.g., head, tail, baseplate) and exhibit *Caudovirales*
62 morphology in electron micrographs of purified phages[23,25-30]. Electron microscopy and
63 quantitative analyses indicate that prophages undergo a lytic phase capable of rupturing
64 bacterial and eukaryotic cell membranes, and phage WO occurs in the extracellular
65 matrix of arthropod gonads[27,31]. Therefore, phage WO appears to uniquely contend with
66 the cellular exit, entry and defense mechanisms of two separate domains of life. WO is
67 also a promising tool for genome editing of *Wolbachia* that has thus far been refractory to
68 genetic modifcation. Until now, the genomes of bacteriophage WO particles have not
69 been fully sequenced and assembled into circular genomes, and their attachment sites and
70 bacterial integration sites are unresolved.
71
72 **RESULTS**
73 Here we report the first metagenomic analysis of phage WO particles from *w*VitA-
74 infected *Nasonia giraulti* wasps and *w*CauB-infected *Ephestia kuehniella* moths. We
75 identify the phage attachment sites and insertion regions and show from fully sequenced
76 genomes that WO harbor all formerly described phage genetic modules (lysogeny,
77 baseplate, head, replication, virulence, tail and patatin[32]) as well as a new group of genes
78 with atypical protein domains indicative of eukaryotic interaction. We collectively group
79 these genes, which include one of the largest genes in bacteriophages to date, into a novel
80 "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) frequently undergo horizontal transfer between phage and metazoan hosts, (iii) can be much longer (up to 14,256 bp) than those in the bacterial chromosome, (iv) are absent from mutualistic, phage-free genomes such as the bedbug-infecting *w*Cle[33] and filarial nematode-infecting *w*Bm and *w*Oo[34,35]. They occur in all complete phage WO haplotypes (Supplementary Information Table 1).

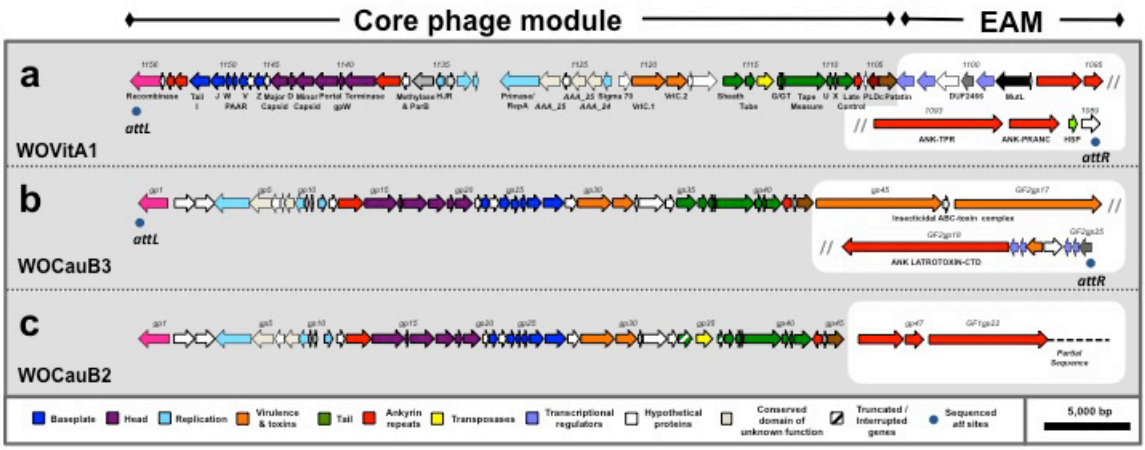
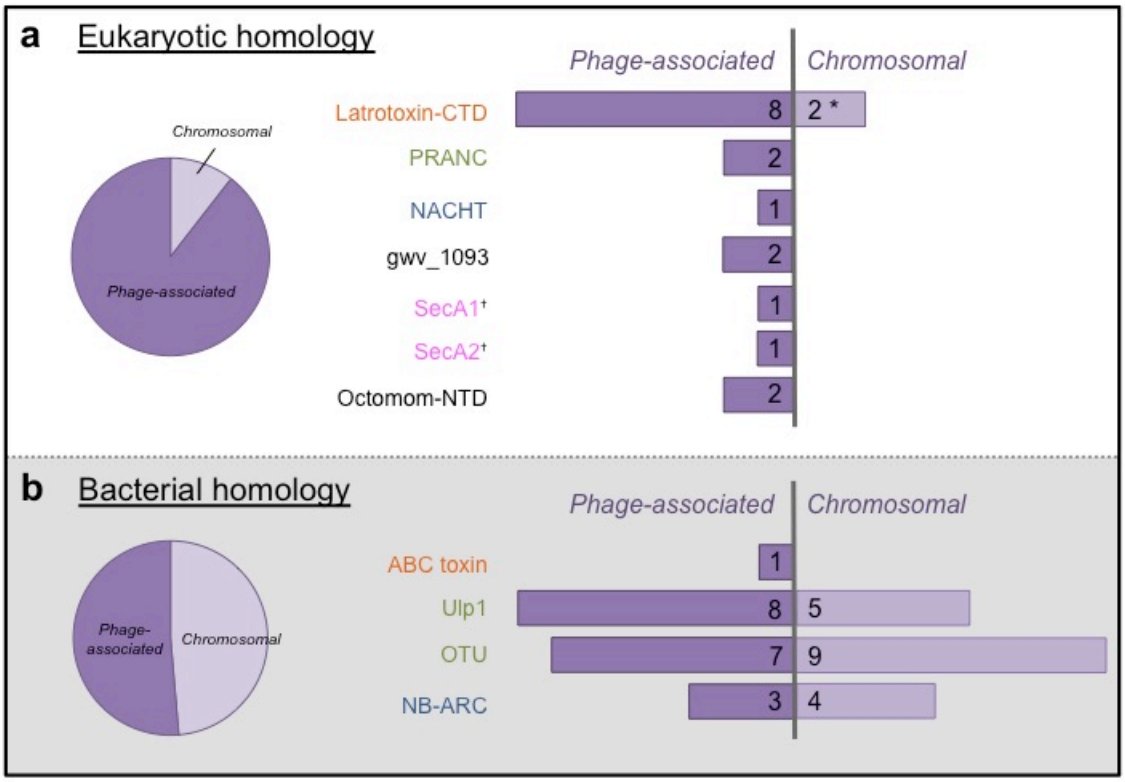


Figure 1 | Complete phage WO genomes harbor a novel Eukaryotic Association Module (EAM).

To verify the newly discovered EAM in the phage genome, we identified the terminal phage WO genes and Sanger sequenced amplicons from an independent sample of phage WOVitA1 (Fig. 1a) across the circularized phage *attP* site (hypothetical protein gwv_1089 to recombinase, Supplementary Information 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

98 Information Fig. 1e). The full length of the completely assembled circular WOVitA1 is
99 66,688 bp, which is 48% larger than any previous prophage WO annotation. Similarly,
100 we identified the new terminal ends of the WOCauB3 phage (23,099 bp (51%) larger
101 than original estimate of 45,078 bp) along with internal localization of the EAM genes by
102 Sanger sequencing its *attP* site [Domain of Unknown Function (DUF)2426 to
103 recombinase]. While we were not able to assemble a complete contig for WOCauB2, it is
104 more than 12,000 bp larger than the original estimate of 43,016, includes multiple
105 ankyrin repeat genes homologous to those in WOVitA1, and, like many other phage
106 haplotypes (e.g., WORiC, WOVitA2, WOSuziC), integrates directly into *Wolbachia*'s
107 magnesium chelatase (*chlI*) gene.
108
109 We then analyzed each phage WO protein domain for homology and surrounding peptide
110 architecture. Unlike the single domain architecture of phage WO's structural genes, EAM
111 genes are highly polymorphic and encompass fusions of both eukaryotic and bacterial
112 protein domains. By extending the analysis to include homologous prophage regions
113 from all sequenced *Wolbachia* chromosomes, ten types of protein domains with putative
114 eukaryotic functions were revealed spanning four predicted functions: (i) toxins, (ii) host-
115 microbe interactions, (iii) host cell suicide, and (iv) secretion of proteins through the cell
116 membrane (Fig. 2). Notably, over half of these domain types (6/10; latrotoxin-CTD,
117 PRANC, NACHT, SecA, gwv_1093-NTD, Octomom-NTD) share greater amino acid
118 homology to eukaryotic invertebrates than to bacteria in GenBank. Among this subset
119 with eukaryotic sequence homology, the protein domains are almost exclusively found in
120 the EAM region (N=17) versus the *Wolbachia* chromosome (N=2). This pattern differs

121 from other EAM protein domains with bacterial homology, which are equally dispersed
122 in phage WO (N=19) and the *Wolbachia* chromosome (N=18) (Fig. 2, Fisher's Exact
123 Test, p = 0.0072). This difference importantly indicates that the eukaryotic-like protein
124 domains are highly enriched in the EAM, suggesting a near exclusive role in phage WO
125 biology.



126

127 **Figure 2 | Eukaryotic-like EAM genes are enriched in prophage WO regions in the**
128 ***Wolbachia* chromosome.**

129

130 Latrotoxin C-terminal domain (CTD) is the most prevalent eukaryotic domain in phage
131 WO. Originally described for its major role in the venom of widow spiders (*Latrodectus*
132 species), latrotoxins cause the formation of membrane pores in their vertebrate or

133 invertebrate victims. Phylogenetic analysis indicates that the latrotoxin-CTD horizontally
134 transferred between widow spiders and phage WO (Fig. 3). In addition, reciprocal search
135 queries using homologous spider and phage CTDs return the same BLASTp hits shown
136 in Fig. 3. These taxa occur in overlapping ecological niches (*Wolbachia* are known to
137 infect spiders of the family *Theridiidae*) in which gene transfers are more likely to
138 happen[36,37]. We also confirmed the presence of *Wolbachia* in three independent
139 *Latrodectus geometricus* samples by amplifying *Wolbachia* 16S rDNA and *wsp*
140 membrane protein genes. The transfer event was apparently followed by a relatively more
141 recent transfer from phage WO back to animals in the *Aedes aegypti* genome where the
142 region is located between genes of mosquito origin [fibrinogen-related protein
143 (AAEL004156) and GalE3 (AAEL004196)], or *A. aegypti* was the putative donor of the
144 domain to phage WO, followed by a recursive transfer to black widow spiders.
145
146 Latrotoxin-CTD is universally located at the 3'-terminal ends of both conserved spider
147 latrotoxin genes[38] and enormous, polymorphic, and eukaryotic-like phage WO genes (up
148 to 14,256 bp). Notably, phage WO CTD sequences have the highest amino acid similarity
149 to black widow spider homologs that target invertebrates, which are the primary hosts of
150 *Wolbachia*. There is also a high incidence of eukaryotic furin cleavage sites that
151 immediately precede the latrotoxin-CTD. In spiders, cleavage at these sites by the
152 eukaryotic furin protease in the trans-Golgi network or extracellular matrix is required for
153 latrotoxin activation before the toxin exerts its effects upon the victim[39-41]. We show that
154 all phage WO EAMs contain at least one site for eukaryotic furin cleavage
155 (Supplementary Information Table 2), and the proportion of all EAM genes with

predicted furin cleavage sites (25%) is two-fold greater than that of the genes in the core phage genome (11%, Fisher's Exact Test, $p < 0.0001$), defined as the conserved bacteriophage region from recombinase to patatin. In regards to the phage WO latrotoxin-CTD, their packaging in virions, conservation of eukaryotic furin cleaveage sites, large eukaryotic-like length, and reduced CTD divergence relative to the spider venom CTD is consistent with their eukaryotic origin and post-translational processing by furin peptidases.

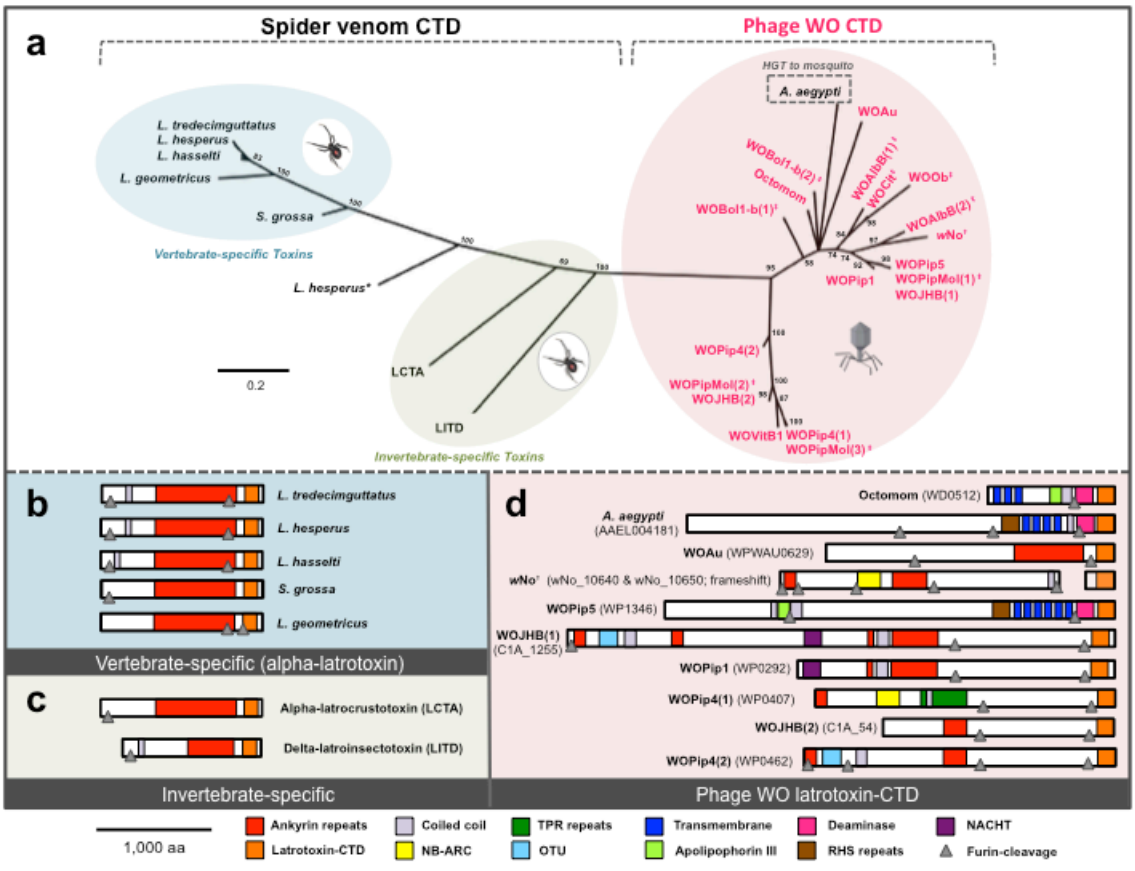


Figure 3 | Latrotoxin-CTD phylogeny and protein architecture reveal lateral genetic transfers between black widow spiders and bacteriophage WO.

168 Domains central to modifying animal proteins are also abundant in the phage EAM. The
169 Pox protein Repeats of ANkyrin C terminus (PRANC) domain in the WOVitA1 genome
170 (gwv_1092) shares protein sequence homology with corresponding PRANC domains in
171 multiple parasitic wasp hosts (Supplementary Information Table 3) and their eukaryotic
172 viruses. Reciprocal BLASTp searches retrieve the same best hits and support previous
173 findings that this protein domain horizontally transferred between eukaryotic viruses,
174 animals, and *Proteobacteria*[42]. The discovery here of the eukaryotic-like PRANC domain
175 in phage WO parallels its presence in the *Poxviridae* virus family, in which it functions in
176 evasion of eukaryotic immune responses via modification of host ubiquitination. PRANC
177 is related to amino acid sequences in F-box proteins, which are eukaryotic proteins
178 involved in protein degradation. The PRANC domain also occurs in vaccina virus,
179 ectromelia virus, cowpox virus and Orf virus and can regulate NF-κB signalling pathway
180 to inhibit transcription of inflammatory cytokines[43].
181
182 Adjacent to the PRANC-encoding gene in WOVitA1 is an ankyrin and tetratricopeptide
183 repeat (TPR)-containing *gwv_1093*. Ankyrin repeats and TPRs mediate a broad range of
184 protein-protein interactions (apoptosis, cell signaling, inflammatory response, etc.) within
185 eukaryotic cells and are commonly associated with effector proteins of certain
186 intracellular pathogens[44-48]. While generally rare in viral genomes (Supplementary
187 Information Fig. 2 and 3, respectively), they occur in all phage WO haplotypes from
188 sequenced *Wolbachia* genomes (N=23). Phylogenetic analysis using reciprocal BLASTp
189 hits (Fig. 4) shows that the N-terminus sequences of the TPR-containing gwv_1093 is
190 embedded within, and likely derived by horizontal transfer from, a deeper and more

diverse set of ancestral lineages in arthropods (Fig. 4b). The event was either followed by a relatively recent recursive transfer from phage WO back to animals in the *Solenopsis invicta* genome (Fig. 4c), where the gene is located between genes of ant origin (bicaudal D and rho guanine nucleotide exchange factor 11), or *Solenopsis invicta* is the putative donor of the region to phage WO.

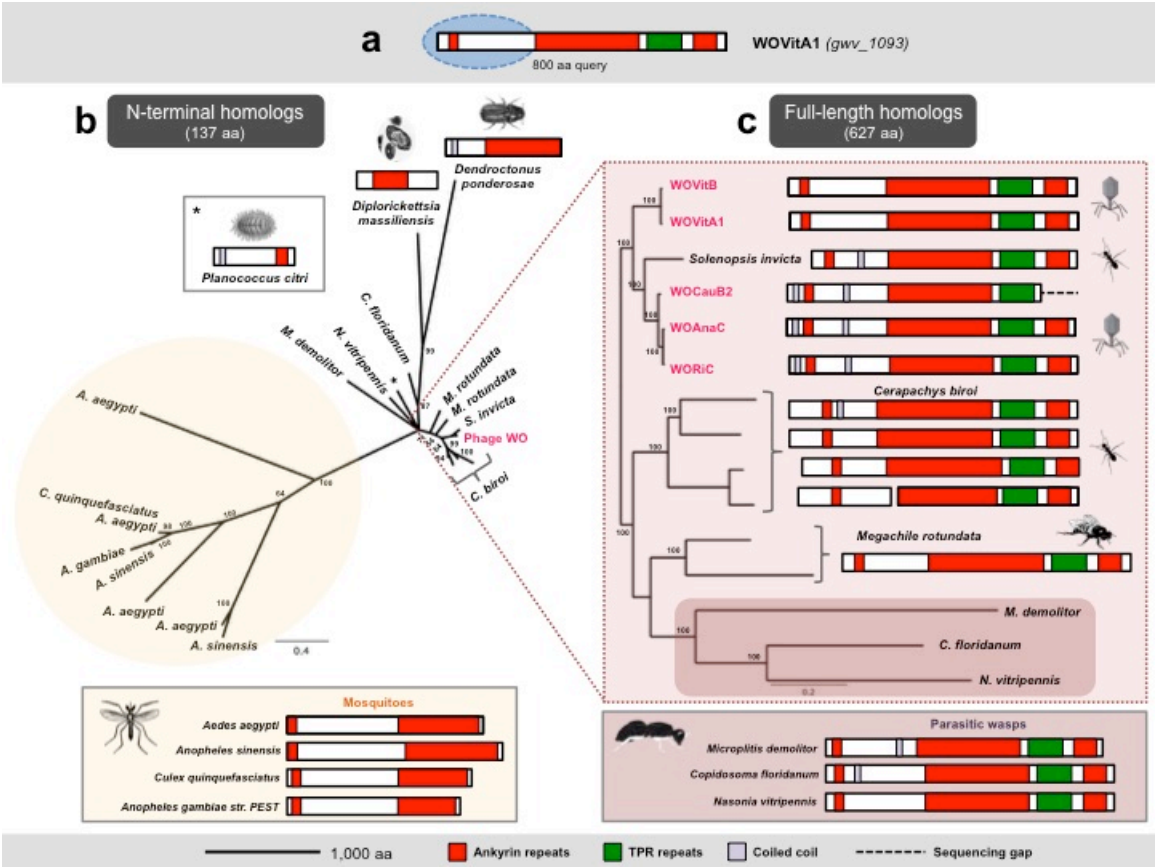


Figure 4 | A conserved TPR and anyrin-repeat protein horizontally transferred from eukaryotes to bacteriophage WO.

Another instance of genetic transfer between insects and bacteriophages involves the programmed cell death (PCD) domain, NACHT (Fig. 5). Eukaryotic NACHT-containing proteins are typically engaged in PCD by acting as pathogen-sensors and signal

203 transduction molecules of the innate immune system[49]. The polymorphic phage WO
204 homolog encodes ankyrin repeats and a latrotoxin-CTD directly downstream from the
205 conserved NTPase domain (Fig. 5a). NACHT domains have been identified in animals,
206 fungi and bacteria[50] and phylogenetic patterns indicate multiple instances of horizontal
207 transfer[51]. A NACHT-containing peptide was recently discovered in the *Clostridium*
208 *difficile*-infecting phage phiCDHM1 genome[52] although, in contrast to phage WO, the
209 phiCDHM1 NACHT domain is bacterial in both amino acid homology and protein
210 architecture. Similar to the phylogeny of the N-terminus of the TPR-containing
211 gwv_1093, the NACHT domain sequence in phage WO is embedded within, and likely
212 derived by horizontal transfer from, a deeper and more diverse set of ancestral variants in
213 arthropods (Fig. 5b,c).
214

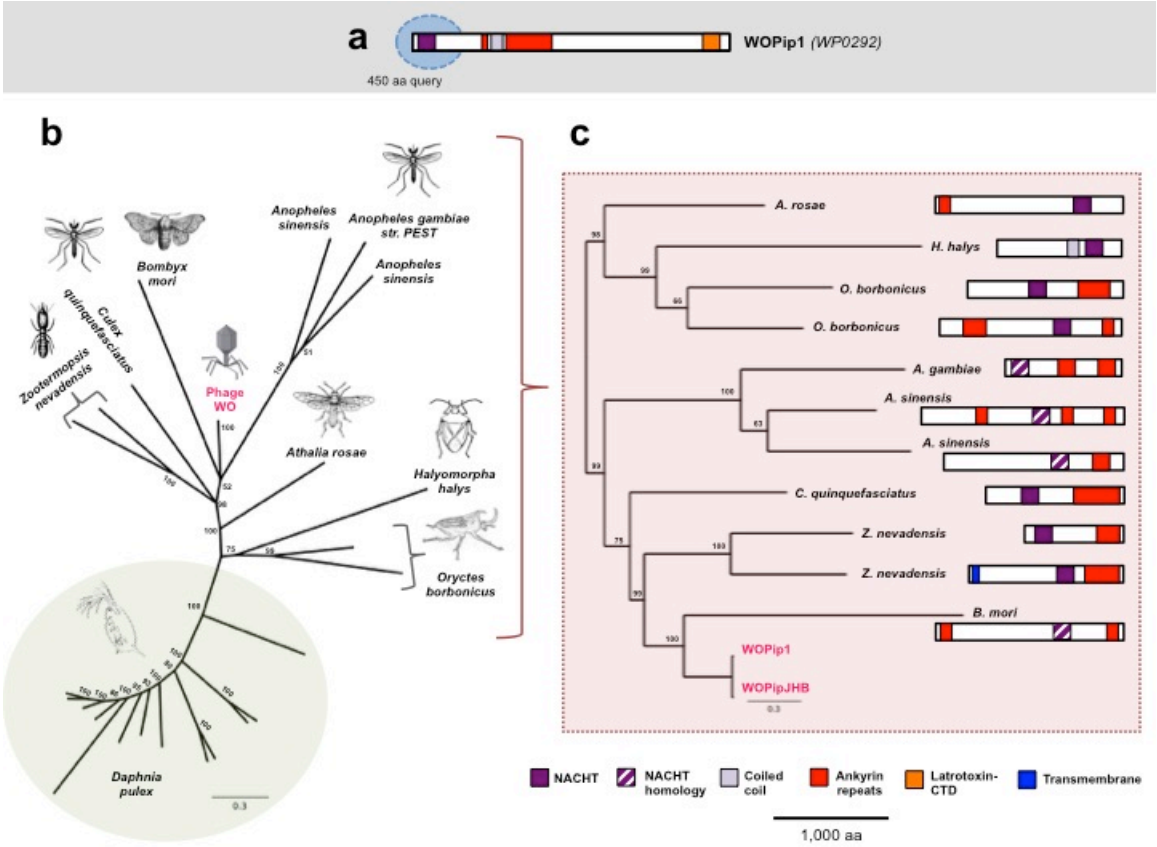


Figure 5 | The programmed cell death domain, NACHT, horizontally transferred from eukaryotes to bacteriophage WO.

DISCUSSION

This inaugural set of completely sequenced phage WO particle genomes, coupled with reciprocal BLAST analyses, phylogenies, annotations of the conserved domains, evolutionary distances, gene lengths, and enrichment of eukaryotic furin cleavage sites in the phage EAM, reveals evidence for lateral genetic transfers from metazoans to bacteriophage. The presence of eukaryotic protein domains in bacteriophage genomes is of special note as they curiously mirror eukaryotic genes in large eukaryotic viruses that aid in viral mimicry and manipulation of host processes[53-55]. Similarly in phage WO,

227 these animal protein domains are central to anti-eukaryotic functions including the black
228 widow latrotoxin, programmed cell death (NACHT), immune evasion (PRANC), and
229 protein-protein interactions. They have never before been reported in bacteriophage
230 genomes because phages have naturally been overlooked as recipients of eukaryotic
231 DNA.
232
233 Bacteriophage WO frequently transfer between *Wolbachia* coinfections in the same
234 animal host[56,57] and to the host genome as part of large transfers of the *Wolbachia*
235 chromosome[58,59]. We previously reported that they were also capable of transferring
236 adjacent flanking non-phage genes in the process of transfer between coinfections[60]. For
237 two of these flanking genes, sequence evidence indicated that *Wolbachia* genomes may
238 be able to receive eukaryotic DNA[61-63]. However, the nature of these lateral genetic
239 transfers remained to be validated and elucidated as these regions were not previously
240 known to be part of the packaged phage genome until now. Based on this work, we
241 suspect that systematic surveys of phage genomes in intimate host-associated bacteria
242 may uncover a broad range of eukaryotic protein domains involved in phage lifecycle
243 adaptations and phage-eukaryote interactions.
244
245 The mechanisms by which eukaryotic protein domains integrate into phage WO are
246 unknown, and could follow at least two models. First, animal genetic material could
247 directly transfer to WO genomes during phage particle packaging in the cytoplasm of
248 animal cells (Fig. 6a) or inside *Wolbachia* cells that are lysing and exposed to the
249 eukaryotic cytoplasmic environment. Packaging of eukaryotic host RNAs, for instance,

occur in the virions of herpesvirus[64,65] and cytomegalovirus[66]. Second, genes may trasnfer from animal genomes to the *Wolbachia* chromosome and then to prophage WO. However, for this scenario to be plausible, animal genetic material transferred presumably in random locations in the *Wolbachia* genome would have to be preferentially lost in non-phage associated domains from the *Wolbachia* chromosome (Fig. 6b) because domains with eukaryotic homology are extremely enriched in the phage/prophage WO EAM versus the rest of the chromosome (Fig. 2).

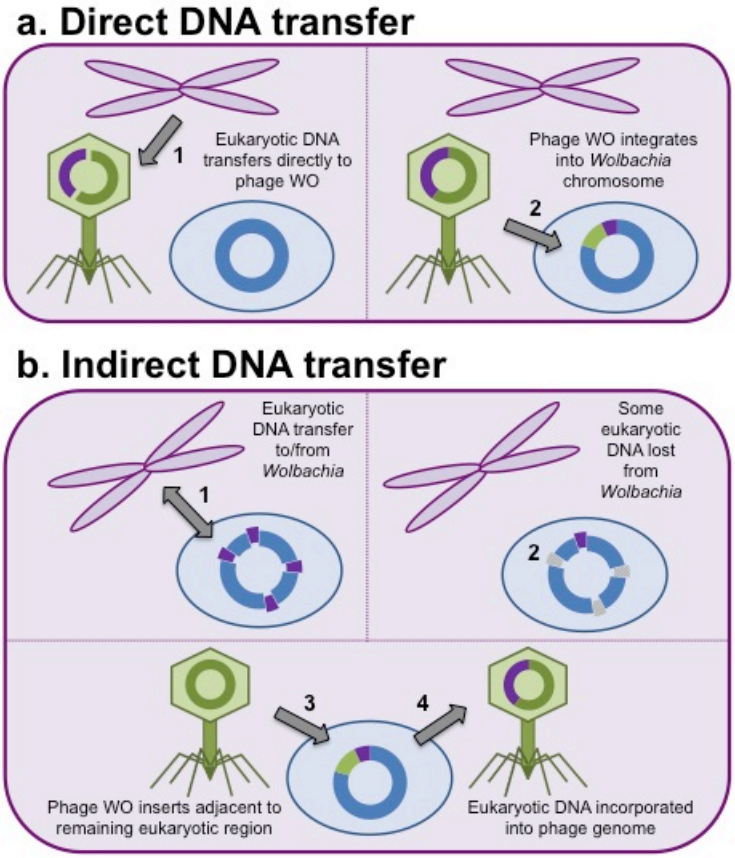


Figure 6 | Models of DNA transfer between eukaryotes and bacteriophages

262 Why are these protein domains present in the EAM of bacteriophage WO? Phages of
263 obligate intracellular bacteria are contained within both bacterial and eukaryotic
264 membranes and can possess an enigmatic "two-fold cell challenge". They may not only
265 have to breach peptidoglycan and permeabilize bacterial membranes, but they may also
266 have to exit (and enter) across the eukaryotic membrane(s) that directly encapsulates the
267 bacteria. Functional studies of homologous domains (i.e., PRANC and NACHT) suggest
268 that these proteins could have eukaryotic viral-like properties that are deployed in
269 processes such as the lysis of eukaryotic cells and post-translational modification of host
270 proteins[67,68]. Phage WO can dwell in the eukaryotic cytoplasm and extracellular matrix
271 that they encounter upon bacterial lysis[31], raising the possibility of direct interaction with
272 the host's biology.
273
274 *Chlamydiomicroviridae* infect obligate intracellular bacteria, yet still do not directly
275 contend with the eukaryotic membrane. Rather, they attach to dormant chlamydial cells
276 (i.e., reticulate bodies) and enter via phagocytosis or endocytosis of the bacteria[69]. The
277 phages then alter development of their bacterial host, which leads to disintegration of the
278 chlamydial inclusion and subsequent lysis of the eukaryotic host cell[70,71]. The nature of
279 phage WO's lifestyle, on the other hand, may require a distinct interaction with multiple
280 membranes and immune responses because lytic activity of phage WO has been
281 associated with typical bacterial cell defects including degraded bacterial DNA, a
282 detached inner membrane, and exit of the phage particles from inside *Wolbachia* and its
283 host cell into the extracellular matrix of the reproductive tissues[31]. Bacteriophages of
284 free-living bacteria also regularly colonize eukaryotic environments, particularly those

285 associated with mucosal surfaces[72]. They, however, do not infect or traverse the
286 eukaryotic membrane and are still within the genomic boundaries of the bacterial
287 virosphere.
288
289 Temperate dsDNA phages also occur in facultative symbionts of aphids[73] and tsetse
290 flies[74]. While *Wolbachia* has never successfully been cultured outside of host cells[75],
291 these facultative symbionts can replicate both intra- and extracellularly (JW Brandt,
292 personal communication, July 2015;[76]) suggesting that their phages are not constrained
293 by the same two-fold cell challenge. In addition, their phages encode a traditional lytic
294 cassette (holin and lysozyme) that correlates with the need to deal only with bacterial
295 membranes. In some cases, the phages harbor bacterial-derived toxins that target
296 eukaryotic cells[77], and these function mutualistically in aphids by arresting parasitoid
297 wasp larvae[73]. Furthermore, unlike phage WO, these phages are readily lost in the
298 absence of parasitoids during laboratory rearing, presumably due to the cost of their
299 toxins[78].
300
301 In addition to providing new insights into the evolution of bacteriophages and showing
302 phage WO to be far more complex than previously described, the findings here reveal
303 that phage evolution in *Wolbachia* leads to a novel example of phage-metazoan genomic
304 chimerism. Acquistion and retooling of intact eukaryotic domains in phage WO appears
305 to be analgous to the commandeering of host genes by eukaryotic viruses. Whether this
306 newly discovered highway of lateral genetic transfer is common in the symbiotic
307 virosphere remains to be determined.
308

309 **METHODS**

310 *Insect and Bacterial Strains*

311 The transfected line of the Mediterranean flour moth *Ephestia kuehniella* harboring
312 *Wolbachia* strain *w*CauB was obtained from Takema Fukatsu and Tetsuhiko Sasaki[26].
313 Moths were maintained at 24˚C and 70% humidity on a diet consisting of wheat bran,
314 glycerol and dried yeast (20:2:1 w/w). The introgressed line of the parasitoid wasp
315 *Nasonia giraulti* harboring *Wolbachia* strain *w*VitA, termed IntG12.1, was previously
316 derived by repeatedly backcrossing *N. vitripennis* (strain 12.1) females to *N. giraulti*
317 males for nine generations[79]. The strain was incubated at 25˚C using the flesh fly
318 *Sarcophaga bullata* as host.

319

320 *Phage Particle Purification*

321 Phage particles were isolated according to Fujii et al[26] with modifications. Approximately
322 4 g of adult insects were homogenized in 29.6 ml cold SM buffer (50mM Tris-HCl, pH
323 7.5, 0.1 M NaCl, 10mM $MgSO_4 \cdot 7H_20$, and 0.1% (w/v) gelatin). NaCl and RNase A
324 were added to a final concentration of 1M and 1ug/ml, respectively. The homogenate was
325 incubated on a shaker at 4˚C for 1 h and then centrifuged at 13,000*g* for 10 min at 4˚C.
326 Polyethylene glycol (PEG) 6000 was added to a final concentration of 10% to precipitate
327 phage particles, incubated at 4˚C for 1 hr with gentle shaking and centrifuged at 13,000g
328 for 10 min. The pellet was resuspended in 5 ml TM buffer (50 mM Tris-HCl, pH 7.5, 10
329 mM $MgCl_2 \cdot 6H_2O$) and mixed with an equal volume chloroform. The suspension was
330 centrifuged at 3,000*g* to remove PEG and the aqueous phase was filtered through a 0.22

331 um filter to remove bacterial cells. The suspension was centrifuged at 60,000*g* for 1 h at
332 4C to collect phage particles. The pellet was suspended in 10 ul TM buffer.
333
334 *Phage DNA Extraction & Metagenomic Sequencing*
335 The phage suspension was treated with RQ1 RNase-Free DNase (Promega) for 30 min at
336 37˚C, followed by heat inactivation for 10 min at 65˚C, to remove host DNA
337 contamination. Phage DNA was extracted from the suspension using the QIAamp
338 MinElute Virus Spin Kit (Qiagen) and amplified using the REPLI-g Mini Kit (Qiagen).
339 Following amplification, paired-end DNA libraries were prepared according to
340 manufacturer's (Illumina) instructions and samples were sequenced with an Illumina
341 HiSeq 2000 (2×100-nt read length).
342
343 *Bioinformatics & Statistics*
344 Metagenomic sequences (reads) were trimmed, paired and assembled into contigs using
345 the CLC Assembler (CLC bio) with bubble size = 50, insertion and deletion cost = 3,
346 mismatch cost = 2, length fraction = 0.6, minimum contig size = 130, similarity = 0.5,
347 minimum distance = 90 and maximum distance = 200. Contigs were compared to the
348 GenBank non-redundant database using NCBI's BLASTN
349 (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and those with similarity to phage WO and/or
350 *Wolbachia* (E-value $<10^{-10}$) were manually annotated using Geneious (Biomatters Ltd.).
351 Individual reads were mapped to reference sequences using Geneious. Open reading
352 frame (ORF) homology searches were performed to determine putative function using
353 NCBI's BLASTP (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and Wellcome Trust Sanger

354 Institute's pfam database (http://pfam.sanger.ac.uk). Coiled coil domains were predicted
355 with EMBL's Simple Modular Architecture Research Tool (SMART, http://smart.embl-
356 heidelberg.de). Furin cleavage sites were identified using PiTou
357 (http://www.nuolan.net/reference.html). The number of genes with and without furin
358 cleavage sites was analyzed with respect to phage-region using Fisher's Exact Test
359 (GraphPad Software). Phylogenetic trees were built using the Bayes plugin in Geneious
360 and model selection for each Bayes analysis was estimated using ProtTest [80].
361
362 *Confirmation of Phage WO Terminal Genes*
363 Genomic DNA was extracted from *w*VitA-infected *N. vitripennis* (strain 12.1) and
364 *w*CauB-infected *E. kuehniella* individuals using the Gentra Puregene Tissue Kit (Qiagen).
365 Primers were designed for both WOVitA1 and WOCauB3 *att* sites, respectively:
366 VitA1_attF (5'- CGA AGA ACC AGC ACA GGG TGG-3'), VitA1_attR (5'- GCT GGA
367 AGA GGG CAT CTG CAT C-3'), CauB3_attF (5'- TCG TGA CTG CCC TAT TGC
368 TGC T – 3') and CauB3_attR (5'- ATG CGG CCA AAG CTG GGT GT – 3').
369 Amplification was performed in a Veriti thermal cycler (Applied Biosystems) using
370 GoTaq green master mix (Promega) under the following conditions: 94C for 2 min; 35
371 cycles of 94C for 30 s, 53C for 30 s, 72C for 1 min; and a final elongation cycle of 72C
372 for 10 min. PCR products were sequenced via Sanger sequencing (Genewiz, Inc).
373

374 **REFERENCES**

375 1. Edwards, R.A. & Rohwer, F. Viral metagenomics. *Nat Rev Microbiol* **3**, 504-10
376 (2005).
377 2. Hendrix, R.W., Smith, M.C., Burns, R.N., Ford, M.E. & Hatfull, G.F. Evolutionary
378 relationships among diverse bacteriophages and prophages: all the world's a
379 phage. *Proc Natl Acad Sci U S A* **96**, 2192-7 (1999).
380 3. Suttle, C.A. Viruses in the sea. *Nature* **437**, 356-61 (2005).
381 4. Brussow, H. The not so universal tree of life or the place of viruses in the living
382 world. *Philos Trans R Soc Lond B Biol Sci* **364**, 2263-74 (2009).
383 5. King, A.M.Q., Adams, M.J., Lefkowitz, E.J. & Carstens, E.B. *Virus taxonomy:*
384 *classification and nomenclature of viruses: Ninth Report of the International*
385 *Committee on Taxonomy of Viruses.*, 1327 (Elsevier, San Diego, 2012).
386 6. Nasir, A., Forterre, P., Kim, K.M. & Caetano-Anolles, G. The distribution and
387 impact of viral lineages in domains of life. *Front Microbiol* **5**, 194 (2014).
388 7. Prangishvili, D., Forterre, P. & Garrett, R.A. Viruses of the Archaea: a unifying
389 view. *Nat Rev Microbiol* **4**, 837-48 (2006).
390 8. Forterre, P. Giant viruses: conflicts in revisiting the virus concept. *Intervirology*
391 **53**, 362-78 (2010).
392 9. Raoult, D. TRUC or the need for a new microbial classification. *Intervirology* **56**,
393 349-53 (2013).
394 10. Nasir, A. & Caetano-Anolles, G. A phylogenomic data-driven exploration of viral
395 origins and evolution. *Science Advances* **1**(2015).
396 11. Elde, N.C. & Malik, H.S. The evolutionary conundrum of pathogen mimicry. *Nat*
397 *Rev Microbiol* **7**, 787-97 (2009).
398 12. Rappoport, N. & Linial, M. Viral proteins acquired from a host converge to
399 simplified domain architectures. *PLoS Comput Biol* **8**, e1002364 (2012).
400 13. Colson, P. & Raoult, D. Gene repertoire of amoeba-associated giant viruses.
401 *Intervirology* **53**, 330-43 (2010).
402 14. Canchaya, C., Fournous, G. & Brussow, H. The impact of prophages on bacterial
403 chromosomes. *Mol Microbiol* **53**, 9-18 (2004).
404 15. Lindell, D. *et al.* Transfer of photosynthesis genes to and from Prochlorococcus
405 viruses. *Proc Natl Acad Sci U S A* **101**, 11013-8 (2004).
406 16. Dammeyer, T., Bagby, S.C., Sullivan, M.B., Chisholm, S.W. & Frankenberg-Dinkel,
407 N. Efficient phage-mediated pigment biosynthesis in oceanic cyanobacteria. *Curr*
408 *Biol* **18**, 442-8 (2008).
409 17. Werren, J.H., Baldo, L. & Clark, M.E. Wolbachia: master manipulators of
410 invertebrate biology. *Nat Rev Microbiol* **6**, 741-51 (2008).
411 18. Zug, R. & Hammerstein, P. Still a host of hosts for Wolbachia: analysis of recent
412 data suggests that 40% of terrestrial arthropod species are infected. *PLoS One* **7**,
413 e38544 (2012).
414 19. Cho, K.O., Kim, G.W. & Lee, O.K. Wolbachia bacteria reside in host Golgi-related
415 vesicles whose position is regulated by polarity proteins. *PLoS One* **6**, e22703
416 (2011).

417 20. Henrichfreise, B. *et al.* Functional conservation of the lipid II biosynthesis
418 pathway in the cell wall-less bacteria Chlamydia and Wolbachia: why is lipid II
419 needed? *Mol Microbiol* **73**, 913-23 (2009).
420 21. Louis, C. & Nigro, L. Ultrastructual evidence of Wolbachia Rickettsiales in
421 Drosophila simulans and their relationships with unidirectional cross-
422 incompatibility. *Journal of Invertebrate Pathology* **54**, 39-44 (1989).
423 22. Gavotte, L. *et al.* A Survey of the bacteriophage WO in the endosymbiotic
424 bacteria Wolbachia. *Mol Biol Evol* **24**, 427-35 (2007).
425 23. Kent, B.N. & Bordenstein, S.R. Phage WO of Wolbachia: lambda of the
426 endosymbiont world. *Trends Microbiol* **18**, 173-81 (2010).
427 24. Metcalf, J.A. & Bordenstein, S.R. The complexity of virus systems: the case of
428 endosymbionts. *Curr Opin Microbiol* **15**, 546-52 (2012).
429 25. Chauvatcharin, N., Ahantarig, A., Baimai, V. & Kittayapong, P. Bacteriophage WO-
430 B and Wolbachia in natural mosquito hosts: infection incidence, transmission
431 mode and relative density. *Mol Ecol* **15**, 2451-61 (2006).
432 26. Fujii, Y., Kubo, T., Ishikawa, H. & Sasaki, T. Isolation and characterization of the
433 bacteriophage WO from Wolbachia, an arthropod endosymbiont. *Biochem*
434 *Biophys Res Commun* **317**, 1183-8 (2004).
435 27. Masui, S. *et al.* Bacteriophage WO and virus-like particles in Wolbachia, an
436 endosymbiont of arthropods. *Biochem Biophys Res Commun* **283**, 1099-104
437 (2001).
438 28. Sanogo, Y.O. & Dobson, S.L. WO bacteriophage transcription in Wolbachia-
439 infected Culex pipiens. *Insect Biochem Mol Biol* **36**, 80-5 (2006).
440 29. Tanaka, K., Furukawa, S., Nikoh, N., Sasaki, T. & Fukatsu, T. Complete WO phage
441 sequences reveal their dynamic evolutionary trajectories and putative functional
442 elements required for integration into the Wolbachia genome. *Appl Environ*
443 *Microbiol* **75**, 5676-86 (2009).
444 30. Wright, J.D., Sjostrand, F.S., Portaro, J.K. & Barr, A.R. The ultrastructure of the
445 rickettsia-like microorganism Wolbachia pipientis and associated virus-like
446 bodies in the mosquito Culex pipiens. *J Ultrastruct Res* **63**, 79-85 (1978).
447 31. Bordenstein, S.R., Marshall, M.L., Fry, A.J., Kim, U. & Wernegreen, J.J. The
448 tripartite associations between bacteriophage, Wolbachia, and arthropods. *PLoS*
449 *Pathog* **2**, e43 (2006).
450 32. Kent, B.N., Funkhouser, L.J., Setia, S. & Bordenstein, S.R. Evolutionary genomics
451 of a temperate bacteriophage in an obligate intracellular bacteria (Wolbachia).
452 *PLoS One* **6**, e24984 (2011).
453 33. Hosokawa, T., Koga, R., Kikuchi, Y., Meng, X.Y. & Fukatsu, T. Wolbachia as a
454 bacteriocyte-associated nutritional mutualist. *Proc Natl Acad Sci U S A* **107**, 769-
455 74 (2010).
456 34. Darby, A.C. *et al.* Analysis of gene expression from the Wolbachia genome of a
457 filarial nematode supports both metabolic and defensive roles within the
458 symbiosis. *Genome Res* **22**, 2467-77 (2012).
459 35. Foster, J. *et al.* The Wolbachia genome of Brugia malayi: endosymbiont evolution
460 within a human pathogenic nematode. *PLoS Biol* **3**, e121 (2005).

461 36. Goodacre, S.L., Martin, O.Y., Thomas, C.F. & Hewitt, G.M. Wolbachia and other
462 endosymbiont infections in spiders. *Mol Ecol* **15**, 517-27 (2006).
463 37. Vanthournout, B., Swaegers, J. & Hendrickx, F. Spiders do not escape
464 reproductive manipulations by Wolbachia. *BMC Evol Biol* **11**, 15 (2011).
465 38. Garb, J.E. & Hayashi, C.Y. Molecular evolution of alpha-latrotoxin, the
466 exceptionally potent vertebrate neurotoxin in black widow spider venom. *Mol*
467 *Biol Evol* **30**, 999-1014 (2013).
468 39. Gordon, V.M. & Leppla, S.H. Proteolytic activation of bacterial toxins: role of
469 bacterial and host cell proteases. *Infect Immun* **62**, 333-40 (1994).
470 40. Remacle, A.G. *et al.* Selective and potent furin inhibitors protect cells from
471 anthrax without significant toxicity. *Int J Biochem Cell Biol* **42**, 987-95 (2010).
472 41. Tsuneoka, M. *et al.* Evidence for involvement of furin in cleavage and activation
473 of diphtheria toxin. *J Biol Chem* **268**, 26461-5 (1993).
474 42. Werren, J.H. *et al.* Functional and evolutionary insights from the genomes of
475 three parasitoid Nasonia species. *Science* **327**, 343-8 (2010).
476 43. Chang, S.J. *et al.* Poxvirus host range protein CP77 contains an F-box-like domain
477 that is necessary to suppress NF-kappaB activation by tumor necrosis factor
478 alpha but is independent of its host range function. *J Virol* **83**, 4140-52 (2009).
479 44. Cerveny, L. *et al.* Tetratricopeptide repeat motifs in the world of bacterial
480 pathogens: role in virulence mechanisms. *Infect Immun* **81**, 629-35 (2013).
481 45. Jernigan, K.K. & Bordenstein, S.R. Ankyrin domains across the Tree of Life. *PeerJ*
482 **2**, e264 (2014).
483 46. Jernigan, K.K. & Bordenstein, S.R. Tandem-repeat protein domains across the
484 tree of life. *PeerJ* **3**, e732 (2015).
485 47. Li, J., Mahajan, A. & Tsai, M.D. Ankyrin repeat: a unique motif mediating protein-
486 protein interactions. *Biochemistry* **45**, 15168-78 (2006).
487 48. Pan, X., Luhrmann, A., Satoh, A., Laskowski-Arce, M.A. & Roy, C.R. Ankyrin repeat
488 proteins comprise a diverse family of bacterial type IV effectors. *Science* **320**,
489 1651-4 (2008).
490 49. Koonin, E.V. & Aravind, L. Origin and evolution of eukaryotic apoptosis: the
491 bacterial connection. *Cell Death Differ* **9**, 394-404 (2002).
492 50. Koonin, E.V. & Aravind, L. The NACHT family - a new group of predicted NTPases
493 implicated in apoptosis and MHC transcription activation. *Trends Biochem Sci* **25**,
494 223-4 (2000).
495 51. Leipe, D.D., Koonin, E.V. & Aravind, L. STAND, a class of P-loop NTPases including
496 animal and plant regulators of programmed cell death: multiple, complex
497 domain architectures, unusual phyletic patterns, and evolution by horizontal
498 gene transfer. *J Mol Biol* **343**, 1-28 (2004).
499 52. Hargreaves, K.R., Kropinski, A.M. & Clokie, M.R. What does the talking?: quorum
500 sensing signalling genes discovered in a bacteriophage genome. *PLoS One* **9**,
501 e85131 (2014).
502 53. Alcami, A. & Koszinowski, U.H. Viral mechanisms of immune evasion. *Immunol*
503 *Today* **21**, 447-55 (2000).

504 54. Piekna-Przybylska, D., DiChiacchio, L., Mathews, D.H. & Bambara, R.A. A
505 sequence similar to tRNA 3 Lys gene is embedded in HIV-1 U3-R and promotes
506 minus-strand transfer. *Nat Struct Mol Biol* **17**, 83-9 (2010).
507 55. Seet, B.T. *et al.* Poxviruses and immune evasion. *Annu Rev Immunol* **21**, 377-423
508 (2003).
509 56. Bordenstein, S.R. & Wernegreen, J.J. Bacteriophage flux in endosymbionts
510 (Wolbachia): infection frequency, lateral transfer, and recombination rates. *Mol*
511 *Biol Evol* **21**, 1981-91 (2004).
512 57. Masui, S., Kamoda, S., Sasaki, T. & Ishikawa, H. Distribution and evolution of
513 bacteriophage WO in Wolbachia, the endosymbiont causing sexual alterations in
514 arthropods. *J Mol Evol* **51**, 491-7 (2000).
515 58. Dunning Hotopp, J.C. *et al.* Widespread lateral gene transfer from intracellular
516 bacteria to multicellular eukaryotes. *Science* **317**, 1753-6 (2007).
517 59. Funkhouser-Jones, L.J. *et al.* Wolbachia co-infection in a hybrid zone: discovery
518 of horizontal gene transfers from two Wolbachia supergroups into an animal
519 genome. *PeerJ* **3**, e1479 (2015).
520 60. Kent, B.N. *et al.* Complete bacteriophage transfer in a bacterial endosymbiont
521 (Wolbachia) determined by targeted genome capture. *Genome Biol Evol* **3**, 209-
522 18 (2011).
523 61. Duplouy, A. *et al.* Draft genome sequence of the male-killing Wolbachia strain
524 wBol1 reveals recent horizontal gene transfers from diverse sources. *BMC*
525 *Genomics* **14**, 20 (2013).
526 62. Klasson, L., Kambris, Z., Cook, P.E., Walker, T. & Sinkins, S.P. Horizontal gene
527 transfer between Wolbachia and the mosquito Aedes aegypti. *BMC Genomics*
528 **10**, 33 (2009).
529 63. Woolfit, M., Iturbe-Ormaetxe, I., McGraw, E.A. & O'Neill, S.L. An ancient
530 horizontal gene transfer between mosquito and the endosymbiotic bacterium
531 Wolbachia pipientis. *Mol Biol Evol* **26**, 367-74 (2009).
532 64. Amen, M.A. & Griffiths, A. Identification and expression analysis of herpes B
533 virus-encoded small RNAs. *J Virol* **85**, 7296-311 (2011).
534 65. Amen, M.A. & Griffiths, A. Packaging of Non-Coding RNAs into Herpesvirus
535 Virions: Comparisons to Coding RNAs. *Front Genet* **2**, 81 (2011).
536 66. Terhune, S.S., Schroer, J. & Shenk, T. RNAs are packaged into human
537 cytomegalovirus virions in proportion to their intracellular concentration. *J Virol*
538 **78**, 10390-8 (2004).
539 67. Bergsbaken, T., Fink, S.L. & Cookson, B.T. Pyroptosis: host cell death and
540 inflammation. *Nat Rev Microbiol* **7**, 99-109 (2009).
541 68. Zhang, L., Villa, N.Y. & McFadden, G. Interplay between poxviruses and the
542 cellular ubiquitin/ubiquitin-like pathways. *FEBS Lett* **583**, 607-14 (2009).
543 69. Sliwa-Dominiak, J., Suszynska, E., Pawlikowska, M. & Deptula, W. Chlamydia
544 bacteriophages. *Arch Microbiol* **195**, 765-71 (2013).
545 70. Hsia, R., Ohayon, H., Gounon, P., Dautry-Varsat, A. & Bavoil, P.M. Phage infection
546 of the obligate intracellular bacterium, Chlamydia psittaci strain guinea pig
547 inclusion conjunctivitis. *Microbes Infect* **2**, 761-72 (2000).

548 71. Salim, O., Skilton, R.J., Lambden, P.R., Fane, B.A. & Clarke, I.N. Behind the
549 chlamydial cloak: the replication cycle of chlamydiaphage Chp2, revealed.
550 *Virology* **377**, 440-5 (2008).
551 72. Barr, J.J. *et al.* Bacteriophage adhering to mucus provide a non-host-derived
552 immunity. *Proc Natl Acad Sci U S A* **110**, 10771-6 (2013).
553 73. Moran, N.A., Degnan, P.H., Santos, S.R., Dunbar, H.E. & Ochman, H. The players
554 in a mutualistic symbiosis: insects, bacteria, viruses, and virulence genes. *Proc*
555 *Natl Acad Sci U S A* **102**, 16919-26 (2005).
556 74. Belda, E., Moya, A., Bentley, S. & Silva, F.J. Mobile genetic element proliferation
557 and gene inactivation impact over the genome structure and metabolic
558 capabilities of Sodalis glossinidius, the secondary endosymbiont of tsetse flies.
559 *BMC Genomics* **11**, 449 (2010).
560 75. Rasgon, J.L., Gamston, C.E. & Ren, X. Survival of Wolbachia pipientis in cell-free
561 medium. *Appl Environ Microbiol* **72**, 6934-7 (2006).
562 76. Weiss, B.L., Wu, Y., Schwank, J.J., Tolwinski, N.S. & Aksoy, S. An insect symbiosis
563 is influenced by bacterium-specific polymorphisms in outer-membrane protein
564 A. *Proc Natl Acad Sci U S A* **105**, 15088-93 (2008).
565 77. Degnan, P.H. & Moran, N.A. Diverse phage-encoded toxins in a protective insect
566 endosymbiont. *Appl Environ Microbiol* **74**, 6782-91 (2008).
567 78. Oliver, K.M., Degnan, P.H., Hunter, M.S. & Moran, N.A. Bacteriophages encode
568 factors required for protection in a symbiotic mutualism. *Science* **325**, 992-4
569 (2009).
570 79. Chafee, M.E. *et al.* Decoupling of host-symbiont-phage coadaptations following
571 transfer between insect species. *Genetics* **187**, 203-15 (2011).
572 80. Abascal, F., Zardoya, R. & Posada, D. ProtTest: selection of best-fit models of
573 protein evolution. *Bioinformatics* **21**, 2104-5 (2005).
574 81. Bhere, K.V., Haney, R.A., Ayoub, N.A. & Garb, J.E. Gene structure, regulatory
575 control, and evolution of black widow venom latrotoxins. *FEBS Lett* **588**, 3891-7
576 (2014).
577

578

579 **FIGURE LEGENDS**
580

581 **Figure 1 | Complete phage WO genomes harbor a novel Eukaryotic Association**
582 **Module (EAM).** The complete phage WO genomes for (a) WOVitA1, (b) WOCauB3
583 and (c) WOCauB2 were sequenced directly from purified phage particles using high
584 throughput, metagenomic sequencing. Each genome consists of a bacteriophage-like
585 region (recombinase to patatin) and an EAM highlighted in white. Gray slash marks
586 indicate illustrative continuation of the genome. Dark blue dots indicate the discovery of
587 the *attL* and *attR* sites of the prophage, which adjoin in the complete circular WO genome
588 to form *attP*.
589

590 **Figure 2 | Eukaryotic-like EAM genes are enriched in prophage WO regions in the**
591 ***Wolbachia* chromosome.** EAM genes with (a) eukaryotic homology are most likely to be
592 associated with prophage WO while those with (b) bacterial homology are both phage-
593 associated and found scattered throughout the *Wolbachia* chromosome. (*) The two
594 chromosomal latrotoxin-CTD domains (wNo_10650 and wHa_05390) are located within
595 phage-associated genes and transposases, indicating a potential genomic rearrangement.
596 ([†]) SecA represents one "domain type" but is listed separately because phage WO
597 contains two different homologs (i.e., wHa_3920 and wHa_3930). Putative functional
598 categories are: anti-eukaryotic toxins (orange); host-microbe interactions (green); host
599 cell suicide (blue); secretion of virulence factors (pink); and unknown (black). Octomom
600 refers to WD0513 of the *w*Mel genome.
601

602 **Figure 3 | Latrotoxin-CTD phylogeny and protein architecture reveal lateral genetic**
603 **transfers between black widow spiders and bacteriophage WO.**
604 (a) Phylogeny of phage WO latrotoxin-CTD protein domains and their eukaryotic
605 homologs was constructed by Bayesian analysis of 74 amino acids using the JTT model
606 of evolution. Consensus support values are shown at the nodes. Comparative protein
607 architecture shows that spider venom (b) vertebrate-specific alpha-latrotoxins and (c)
608 invertebrate-specific alpha- and delta-latrotoxins are highly conserved, whereas (d) phage
609 WO are not. Bolded nomenclature in (d) denotes the specific phage WO haplotype (listed
610 as WO). Genome locus tags are listed in parentheses. Predicted furin cleavage sites, listed
611 in Supplementary Information Table 2, are illustrated with gray triangles. (*) A second *L.*
612 *hesperus* sequence represents a recently-described downstream paralog with unknown
613 toxin activity[81]. (†) wNo_10650 is located within phage-associated genes and
614 transposases, indicating a potential genomic rearrangement of a phage region.
615 (‡) Architecture is not shown for sequences on incomplete contigs (WOBol1-b,
616 WOAlbB, WOCit, WOPipMol, WOVitB) because complete peptide information and
617 specific phage association are unknown.
618
619 **Figure 4 | A conserved TPR and anyrin-repeat protein horizontally transferred**
620 **from eukaryotes to bacteriophage WO.** (a) An 800-aa BLASTp query of WOVitA1's
621 gwv_1093 N-terminus reveals homologs throughout moquitoes, ants, beetles, a mealybug
622 and one obligate intracellular gammaproteobacteria. Bayesian phylogenetic trees were
623 constructed based on (b) a 137-aa alignment of all homologs with E-value less than e^{-40}
624 using the LG +G model of evolution. (c) To resolve taxa closest to phage WO, trees were

625 reconstructed based on a 627-aa alignment of all homologs with an E-value of 0 using the
626 JTT +I+G model of evolution. Isoforms were removed from each alignment. Both trees
627 are unrooted. Consensus support values are shown at the nodes. Chromosomal
628 neighborhood analyses of available animal genome sequences indicate that animal
629 homologs to the phage WO protein are on contigs with other animal genes.
630
631 **Figure 5 | The programmed cell death domain, NACHT, horizontally transferred**
632 **from eukaryotes to bacteriophage WO.** (a) A 450-aa BLASTp query of phage WO's
633 NACHT region reveals homologs throughout arthropods and crustaceans. (b) Bayesian
634 phylogenetic trees were constructed based on a 271-aa alignment of all homologs with E-
635 value less than e^{-15} and coverage greater than 70% using the cpREV+G model of
636 evolution. To resolve taxa closest to phage WO, all *Daphnia* sequences were removed
637 from the alignment and clusters of highly divergent residues (i.e., 5 or more sequential
638 residues with less than 15% pairwise identity) were trimmed. Trees were reconstructed
639 based on this 262-aa alignment using the LG +G model of evolution. Consensus support
640 values are shown at the nodes. Both trees are unrooted. Chromosomal neighborhood
641 analyses of available animal genome sequences indicate that animal homologs to the
642 phage WO protein are on contigs with other animal genes.
643
644 **Figure 6 | Models of DNA transfer between eukaryotes and bacteriophages.** Genetic
645 transfers between eukaryotes and bacteriophages can occur either (a) directly via
646 incorporation into the phage genome followed by subsequent inclusion in the
647 chromosomal prophage region or (b) indirectly via the transfer of eukaryotic DNA to the

648 bacterial chromosome or vice versa. Since phage EAM genes carrying protein domains
649 central to eukaryotic functions primarily occur in phage/prophage genomes (see Fig. 2),
650 transferred DNA from eukaryotes to non-phage regions in *Wolbachia* is likely eliminated
651 from the bacterial genome. Prophage genomes adjacent to these EAM genes then
652 incorporate the DNA into their packaged genomes and pass it on to new copies of the
653 phage.
654
655

656 **SUPPLEMENTARY INFORMATION**

657

658 **Supplementary Information Figure 1** | Sequencing reveals the phage, prophage and

659 bacterial *att* sites for WOVitA1.

660 **Supplementary Information Figure 2** | Ankyrin repeat domain.

661 **Supplementary Information Figure 3** | TPR domain.

662 **Supplementary Information Table 1** | Comparative genomics of phage WO.

663 **Supplementary Information Table 2** | Phage WO EAM furin cleavage sites.

664 **Supplementary Information Table 3** | The phage WO PRANC domain shares amino

665 acid homology with multiple eukaryotic host peptides.

666

667 **ACKNOWLEDGEMENTS**

668 We thank Tetsuhiko Sasaki and Takema Fukatsu for providing the *w*CauB-infected

669 moths; Rini Pauly for computational assistance; Merri Lynn Casem for providing widow

670 spiders; and Kerry Oliver for information on *Hamiltonella*. This research was funded by

671 NIH Awards R01 GM085163 and R21 HD086833 and NSF Awards DEB 1046149 and

672 IOS 1456778 to SRB. The funders had no role in study design, data collection and

673 interpretation or the decision to submit the work for publication.

674

675 **AUTHOR CONTRIBUTION**

676 Sarah Bordenstein designed and performed the experiments, analyzed the data, prepared

677 figures and tables, wrote and reviewed drafts of the paper.

678

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

AUTHOR INFORMATION

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

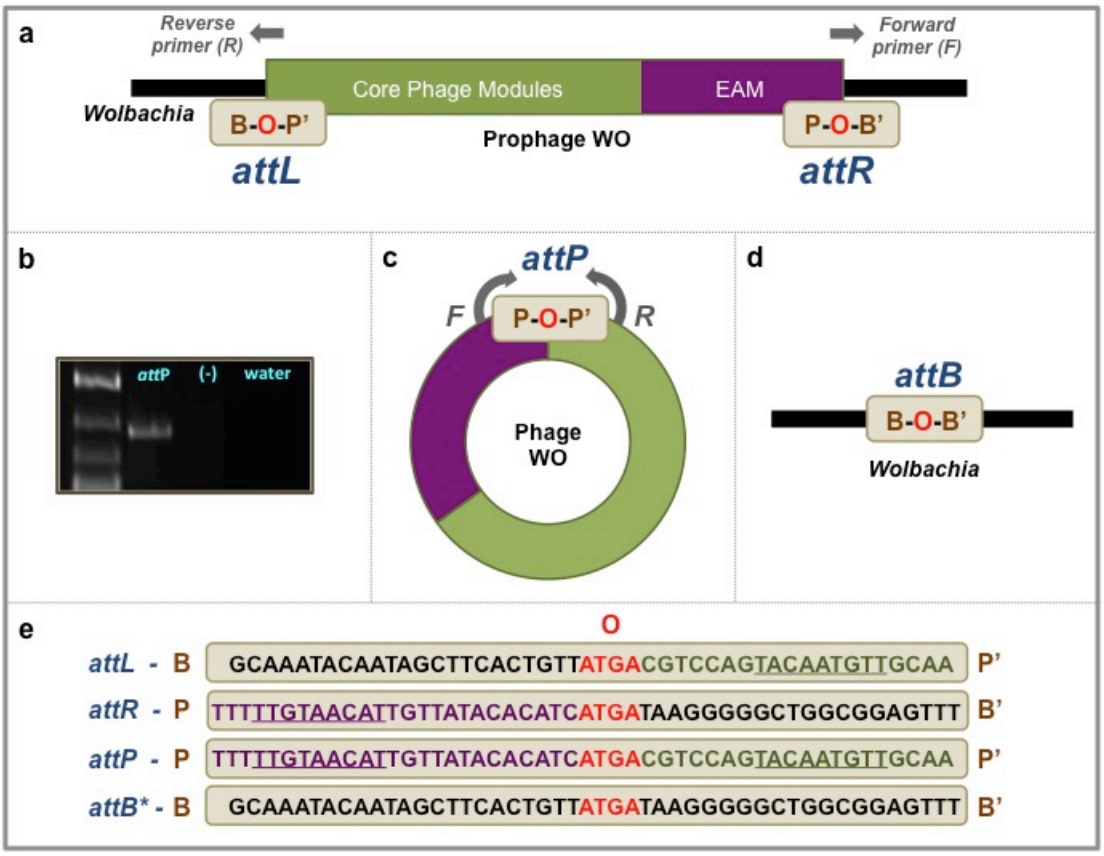
Lateral Genetic Transfers Between Eukaryotes and Bacteriophages

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

Affiliations:

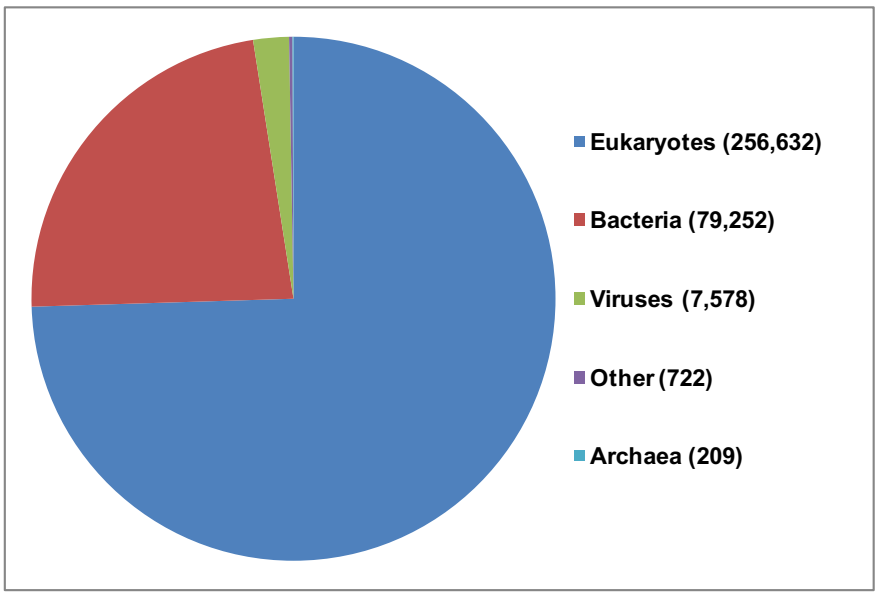
[a]Department of Biological Sciences, Vanderbilt University, Nashville, TN 37232, USA.

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

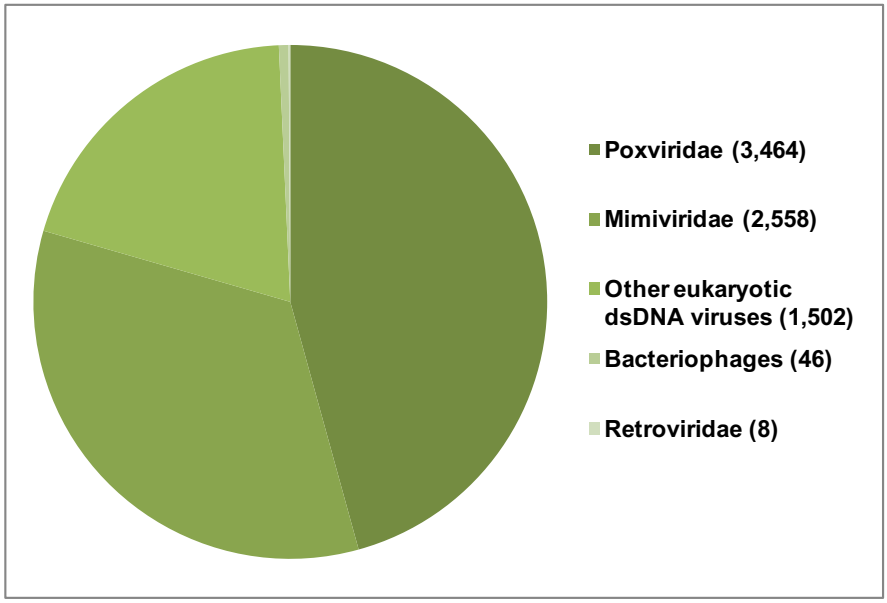


Supplementary Information Figure 1 | Sequencing reveals the phage, prophage and bacterial *att* sites for WOVitA1. (a) The entire prophage WO genome, including all core phage modules (from recombinase to patatin, green) and the EAM (purple), is integrated into the *Wolbachia* chromosome. Attachment sites are designated as BOP' (*attL*) and POB' (*attR*) with B representing 'bacterial' and P representing 'phage'-associated nucleotides. Gray arrows indicate the direction of PCR primers used to amplify the (b) phage attachment (*attP*) site, designated as POP' in the (c) circular WOVitA1 genome. (d) The bacterial attachment site (*attB*) is designated as BOB' on the *Wolbachia* chromosome. (e) All four *att* sites share a common region, O. Underlined nucleotides represent an inverted repeat region. *The *attB* site was predicted based on the *attL* and *attR* sequences. A BLASTN search of this sequence identified the putative *attB* site as a non-coding, repetitive sequence in closely related *Wolbachia* taxa lacking the WOVitA1 infection (e.g., *w*Au, *w*Mel, and *w*Ri).

a



b

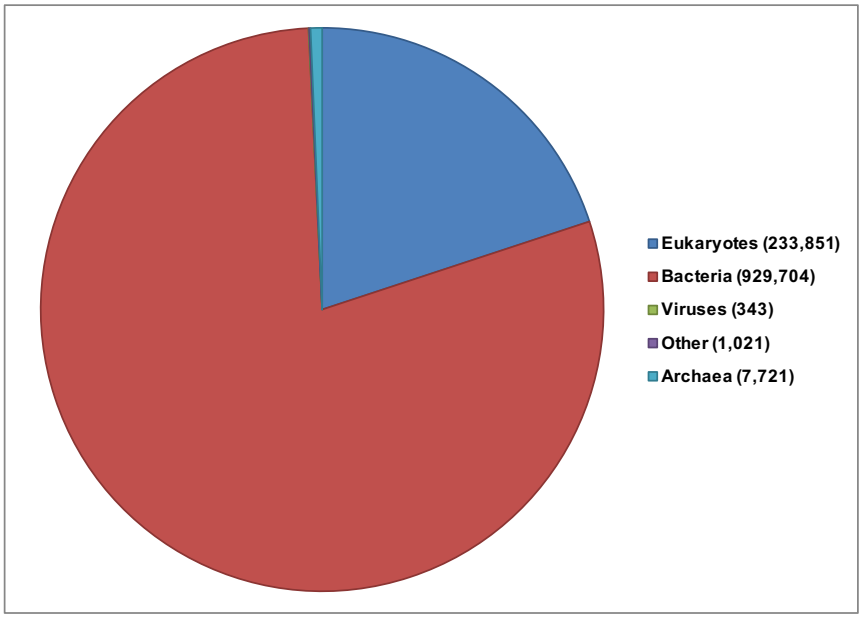


C

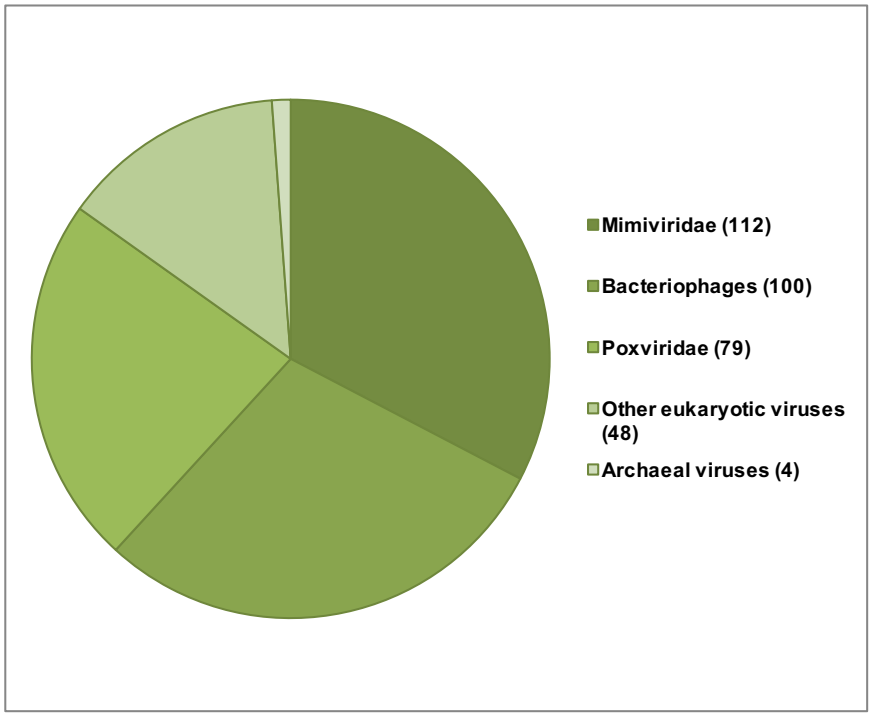
Bacteriophages	# Ankyrin-containing protein records in NCBI	Length (# aa residues)
Phage WO	29	148-2474
Campylobacter phage	4 (identical)	95
Gordonia phage	2 (identical)	102
Lactococcal phage	3 (identical)	169
Leptospira phage	2 (identical)	260
Salmonella phage	3	69-103
Uncultured Mediterranean phage uvMED	3	173-191

Supplementary Information Figure 2 | Ankyrin repeat domain.
(a) The ankyrin domain is most commonly associated with eukaryotes as determined by a taxonomic search of the domain in NCBI's protein database. "Other" refers to synthetically-engineered constructs and metagenomic sequences. (b) Of the viral proteins that contain an ankyrin repeat, the majority are associated with either *Poxviridae* or *Mimiviridae*. (c) Only seven bacteriophages in the NCBI database contain annotated ankyrin domains. Of these, only phage WO ankyrins are eukaryotic in length. This analysis does not include prophage regions within bacterial chromosomes.

a



b



Supplementary Information Figure 3 | TPR domain.
(a) The TPR domain is most commonly associated with bacteria as determined by a taxonomic search of the domain in NCBI's protein database. "Other" refers to synthetically-engineered constructs and metagenomic sequences. (b) The majority of viral proteins are associated with *Mimiviridae*. This analysis does not include prophage regions within bacterial chromosomes.

GENOME	CORE PHAGE REGION	EAM REGION	(WD0611-WD0621)-like
WOVitA1	gwv_1104-gwv_1156	gwv_1089-gwv_1103	na
WOVitA2	gwv_426-gwv_459	gwv_458-gwv_472; gwv_484-gwv_496	gwv_473-gwv_483
WOVitA4	gwv_146-gwv_175	gwv_131-gwv_145(T); gwv_176-gwv_178	na
WOCauB2	gp1-gp45	gp46-*(partial sequence)*	na
WOCauB3	gp1-gp44	gp45-GF2gp25	na
WOSol	So0001-So0022	wSo0003(T)-wSo0014; So0023-So0025; So0026-wSo0028(T)	wSo0015-wSo0026
WOMelA	WD0259-WD0288	WD0289-WD0296(T); WD0253(T)-WD0258	na
WOMelB	WD0634-WD0647(T); WD0563(T) - WD0604	WD0605-WD0610; WD0622-WD0633	WD0611-WD0621
*w*Mel WO-Island	na	WD0507-WD0514	na
WOPip1	WP0236(T)-WP0272	WP0273-WP0293	na
WOPip2	WP0297-WP0322	WP0294-WP0296	na
WOPip3	WP0323-WP0342	WP0343(T)-WP0354	na
WOPip4	WP0411-WP0455	WP0405(T)-WP0410; WP0456-WP0465	na
WOPip5	WP1294(T)-WP1340	WP1289(T)-WP1293; WP1341-WP1352(T)	na
WORiA	wRi_012450-wRi_012680(T)	na	na
WORiB	wRi_005400-wRi_005660	wRi_005310(T)-wRi_005390; wRi_005670-wRi_005720; wRi_005860-wRi_005930(T)	wRi_005730-wRi_005830
WORiB(2)	wRi_010060-wRi_010320	wRi_009980(T)-wRi_010050; wRi_010330-wRi_010380; wRi_010500-wRi_010580(T)	wRi_010390-wRi_010490
WORiC	wRi_006880-wRi_007230(T); wRi_007550-wRi_007660(T)	wRi_006620-wRi_006870	na
WOHa1	wHa_02360-wHa_02620	wHa_02010(T)-wHa_2050; wHa_02170-wHa_02350; wHa_02630-wHa_02760(T)	wHa_02060-wHa_02160
WOHa2	wHa_03390-wHa_03840	wHa_03850-wHa_03930	na
WONo1	wNo_01400-wNo01060	wNo_01000(T)-wNo_01050(T)	na
WONo2	wNo_07170(T)-wNo_07380(T)	na	na
WONo3	wNo_09000-wNo_09120	wNo_09130-wNo_09150(T)	na
WONo4	wNo_10070(T)-wNo_10280	wNo_10290-wNo_10350	na
*w*No WO-Islands	na	wNo_01980-wNo_1990; wNo_2110-wNo_02200; wNo_05070-wNo_05150(T)	wNo_2000-wNo_2100

Supplementary Information Table 1 | Comparative genomics of phage WO. All intact phage WO genomes contain a bacteriophage-like region (termed the "Core Phage Region") either adjacent to or flanked by a EAM region, listed with corresponding locus tags. Locus tags followed by (T) indicate that they are flanking transposases and may or may not be part of the prophage region. The (WD0611-WD0621)-like region represents a highly conserved cluster of bacterial genes with a broad range of metabolic function and transport potential. na, not applicable. NCBI accession numbers for the analyzed phage regions are: WOCauB2 – AB478515; WOCauB3 – AB478516; WOSol – KC955252; *w*Mel - AE017196; wPip - AM999887; wRi - CP001391; wVitA - PRJDB1504; *w*Ha - CP003884; *w*No - CP003883.

GENOME	EAM FURIN CLEAVAGE
WOCauB2	*(partial sequence)*
WOCauB3	gp45, GF2gp18-GF2gp22
WOSol	wSo0011, So0023, So0025
WOMelA	WD0257-WD0258
WOMelB	WD0610, WD0630-WD0632
*w*Mel WO-Island	WD0512-WD0514
WOPip1	WP0280, WP0283, WP0290, WP0292-WP0293
WOPip2	WP0294, WP0319-WP0320
WOPip3	WP0338
WOPip4	WP0407, WP0410, WP0428, WP0433, WP0457, WP0460, WP0462-WP0463
WOPip5	WP1291, WP1341, WP1344, WP1346, WP1349, WP1351
WORiA	(haplotype does not have an EAM)
WORiB	wRi_005330, wRi_005360, wRi_005670, wRi_005720
WORiB(2)	wRi_009990, wRi_010020, wRi_010330, wRi_010380, wRi_010570
WORiC	wRi_006630-wRi_006660, wRi_006740
WOVitA1	gwv_1095
WOVitA2	gwv_464, gwv_484, gwv_489, gwv_491-gwv_495
WOVitA4	gwv_134, gwv_141-gwv_142, gwv_144-gwv_145
WOHa1	wHa_02170, wHa_02200, wHa_02290, wHa_02350, wHa_02690, wHa_02730
WOHa2	wHa_03920
WONo1	wNo_01030, wNo_01060
WONo2	(haplotype does not have an EAM)
WONo3	wNo_09000, wNo_09080, wNo_09140
WONo4	wNo_10290, wNo_10320-wNo_10340
*w*No WO-Islands	wNo_01990, wNo_02030, wNo_02070, wNo_02120, wNo_02130, wNo_05080-wNo_05090, wNo_05110-wNo_05130

Supplementary Information Table 2 | Phage WO EAM furin cleavage sites.
Genes with predicted furin cleavage sites, indicative of potential host-induced protein modification, were identified within every prophage WO EAM. NCBI accession numbers for the analyzed phage regions are: WOCauB2 – AB478515; WOCauB3 – AB478516; WOSol – KC955252; *w*Mel - AE017196; wPip - AM999887; wRi - CP001391; wVitA - PRJDB1504; *w*Ha - CP003884; *w*No - CP003883.

EUKARYOTIC HOMOLOG	ACCESSION	E-VALUE	QUERY COVERAGE	IDENTITY
Microplitis demolitor	XP_014298115.1	8.00E-43	75%	49%
Nasonia vitripennis	XP_003426146.1	5.00E-23	76%	37%
Glypta fumiferanae	AKD28025.1	4.00E-21	71%	39%
Trichogramma pretiosum	XP_014232168.1	2.00E-16	73%	33%
Ceratosolen solmsi marchali	XP_011505281.1	8.00E-15	69%	31%
Copidosoma floridanum	XP_014206311.1	3.00E-12	58%	32%
Diaphorina citri	XP_008470724.1	9.00E-10	49%	31%

Supplementary Information Table 3 | The phage WO PRANC domain shares amino acid homology with multiple eukaryotic host peptides.

The PRANC domain in WOVitA1's gwv_1092 shares homology with multiple insect hosts. The best BLASTp hit for each species is listed above with E-value, query coverage and identity.