Eukaryotic association module in phage WO genomes from *Wolbachia*

Sarah R. Bordenstein[1] and Seth R. Bordenstein[1,2]

Affiliations:

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

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

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

Abstract

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 metagenomic analysis of purified bacteriophage WO particles of *Wolbachia* and uncover a eukaryotic association module. It encodes domains, such as the black widow latrotoxin C-terminal domain, 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, to our knowledge, 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*.

Introduction

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). *Wolbachia* cause a range of host reproductive pathologies[15,16],

55 primarily infect the cells of host reproductive tissues, exist in Golgi-derived vesicles
56 within the eukaryotic cytoplasm, and are enclosed by a bacterial cell membrane and one
57 or more eukaryotic-derived membranes[17,18]. Nearly all sequenced *Wolbachia* genomes,
58 with the exception of those acting as obligate mutualists, harbor prophage WO[19-21]. The
59 prophage WO encode conserved structural modules (e.g., head, tail, baseplate) and
60 exhibit *Caudovirales* morphology in electron micrographs of purified phages [20,22-25].
61 Electron microscopy and quantitative analyses indicate that prophages undergo a lytic
62 phase capable of rupturing bacterial and eukaryotic cell membranes, and phage WO
63 occurs in the extracellular matrix of arthropod gonads[23,26]. Therefore, phage WO appears
64 to uniquely contend with the cellular exit, entry and defense mechanisms of two separate
65 domains of life. WO is also a promising tool for genome editing of *Wolbachia* that has
66 thus far been refractory to genetic modifcation.
67
68 Here we assemble the sequenced genomes of phage WO particles, resolve the
69 bacteriophage attachment and bacterial integration sites, report a eukaryotic association
70 module in bacteriophages, and discuss lateral gene transfers between eukaryotes and
71 bacteriophages.
72
73 **Results**
74 **Phage WO genomes reveal a eukaryotic association module.** Here we report the
75 metagenomic analysis of phage WO particles from *w*VitA-infected *Nasonia giraulti*
76 wasps and *w*CauB-infected *Ephestia kuehniella* moths (the *w*-prefix indicates specific
77 *Wolbachia* strain and WO-prefix indicates phage haplotype; see Supplementary Table 1

78 for a complete list). We identify the phage attachment sites and insertion regions and
79 show from fully sequenced genomes that WO harbors all formerly described phage
80 genetic modules (lysogeny, baseplate, head, replication, virulence, tail and patatin-like
81 phospholipase[27]) as well as a new group of genes with atypical protein domains
82 indicative of eukaryotic interaction. We collectively group these genes, which include the
83 second largest gene in bacteriophages to date, into a 'Eukaryotic Association Module'
84 (EAM; Fig. 1, white box). The EAM features genes that (i) encode protein domains and
85 cleavage sites central to eukaryotic functions, (ii) occur in phage and metazoan hosts, (iii)
86 are among the largest genes in phage genomes (up to 14,256 bp) and (iv) are absent from
87 mutualistic, phage-free genomes such as the bedbug-infecting *w*Cle and filarial
88 nematode-infecting *w*Bm and *w*Oo. They occur in all complete prophage WO haplotypes
89 (Supplementary Table 2).
90
91 To verify the newly discovered EAM in the phage genome, we identified the terminal
92 prophage WO genes and Sanger sequenced amplicons from an independent sample of
93 phage WOVitA1 (Fig. 1a) across the linear phage *attP* site (hypothetical protein
94 gwv_1089 to recombinase, Supplementary Fig. 1). Next, using the newly identified *attR*
95 and *attL* sites, we extrapolated the bacterial *attB* site in WOVitA1, which is a noncoding,
96 repetitive sequence in *Wolbachia* from *Nasonia* wasps (Supplementary Fig. 1e). The full
97 length of the completely assembled, linear WOVitA1 genome is 65,653 bp, which is
98 23,531 bp larger than the previous prophage WO annotation. Similarly, we identified the
99 new terminal ends of the WOCauB3 prophage [23,099 bp (51%) larger than original
100 estimate of 45,078 bp], extending the previous observation that the end of the genome is

101 beyond the patatin gene[25], along with internal localization of the EAM genes by Sanger
102 sequencing its *attP* site [Domain of Unknown Function (DUF)2426 to recombinase].
103 While we were not able to assemble a complete contig for WOCauB2, it is more than
104 6,854 bp larger than the original estimate of 43,016 bp, includes multiple ankyrin repeat
105 genes homologous to those in WOVitA1, and, like many other prophage haplotypes (e.g.,
106 WORiC, WOVitA2, WOSuziC), integrates directly into *Wolbachia*'s magnesium
107 chelatase (*chlI*) gene.
108
109 **The EAM is enriched with eukaryotic-like domains.** We then analyzed each phage
110 WO protein domain for homology and surrounding peptide architecture. Unlike the single
111 domain architecture of phage WO's structural genes, EAM genes are highly polymorphic
112 and encompass fusions of both eukaryotic and bacterial protein domains. By extending
113 the analysis to include homologous prophage regions from all sequenced *Wolbachia*
114 chromosomes, ten types of protein domains with putative eukaryotic functions were
115 uncovered spanning four predicted functions: (i) toxins, (ii) host-microbe interactions,
116 (iii) host cell suicide, and (iv) secretion of proteins through the cell membrane (Fig. 2).
117 Notably, over half of these domain types [6/10; latrotoxin C-terminal domain (CTD),
118 PRANC, NACHT, SecA, gwv_1093 N-terminal domain (NTD), Octomom-NTD] share
119 greater amino acid homology to eukaryotic invertebrates than to bacteria in GenBank.
120 Among this subset with eukaryotic sequence homology, the protein domains are almost
121 exclusively found in the prophage EAM region (N=17) versus the *Wolbachia*
122 chromosome (N=2). In the latter case, the two chromosomal latrotoxin-CTD domains
123 (wNo_10650 and wHa_05390) are flanked by phage-associated genes and transposases,

124 indicating a likely phage WO origin and subsequent genomic rearrangement. This pattern
125 differs from other EAM protein domains with bacterial homology, which are equally
126 dispersed in phage WO (N=19) and the *Wolbachia* chromosome (N=18) (Fig. 2, Fisher's
127 Exact Test, p = 0.0072). The difference importantly indicates that the eukaryotic-like
128 protein domains are highly enriched in the EAM, suggesting a near exclusive role in
129 phage WO biology.
130
131 **The black widow latrotoxin-CTD.** Latrotoxin-CTD is the most prevalent eukaryotic
132 domain in prophage WO. Originally described for its major role in the venom of widow
133 spiders (*Latrodectus* species), latrotoxins act extracellularly to cause the formation of ion-
134 permeable membrane pores in their vertebrate or invertebrate victims. The CTD,
135 specifically, is only associated with the latrotoxin precursor molecule (protoxin) and
136 could possibly act intracellularly to facilitate disintegration of the spider's toxin-
137 producing cells[28]. While latrotoxins are generally considered exclusive to spiders, CTD-
138 homologs in *Wolbachia*, *Rickettsiella grylli*[28], and a transcriptome from a *Wolbachia*-
139 infected stink bug[29] have been reported. Here, phylogenetic analysis implies that the
140 latrotoxin-CTD horizontally transferred between widow spiders and phage WO (Fig. 3).
141 Reciprocal search queries using homologous spider and phage CTDs return the same
142 BLASTP hits shown in Fig. 3. Notably, phage WO CTD sequences have the highest
143 amino acid similarity to black widow spider homologs that target invertebrates, which are
144 the primary hosts of *Wolbachia*. While convergent evolution could explain amino acid
145 sequence similarities of the latrotoxin-CTD in black widows and *Wolbachia*, these two
146 taxa occur in overlapping ecological niches (*Wolbachia* are known to infect spiders of the

147 family *Theridiidae*) in which gene transfers are likely to happen[30]. We also confirmed the
148 presence of *Wolbachia* in three independent *Latrodectus geometricus* samples by
149 amplifying *Wolbachia* 16S rDNA and *wsp* membrane protein genes. The transfer event
150 was apparently followed by a relatively more recent transfer from phage WO back to
151 animals in the *Aedes aegypti* genome, where the region is located between genes of
152 mosquito origin [fibrinogen-related protein (AAEL004156) and GalE3 (AAEL004196)].
153
154 **Toxin activation by eukaryotic furin cleavage.** Latrotoxin-CTD is universally located
155 at the 3'-terminal ends of both conserved spider latrotoxin genes[31] and enormous,
156 polymorphic, and eukaryotic-like phage WO genes (up to 14,256 bp). There is a high
157 incidence of eukaryotic furin cleavage sites that immediately precede the latrotoxin-CTD.
158 In spiders, cleavage at these sites by the eukaryotic furin protease in the trans-Golgi
159 network or extracellular matrix is required for latrotoxin activation before the toxin exerts
160 its effects upon the victim. We show that all prophage WO EAMs contain at least one site
161 for eukaryotic furin cleavage (Supplementary Table 3), and the proportion of all EAM
162 genes with predicted furin cleavage sites (25%) is two-fold greater than that of the genes
163 in the core phage genome (11%, Fisher's Exact Test, $p < 0.0001$), defined as the
164 conserved bacteriophage region from recombinase to patatin. In regards to the phage WO
165 latrotoxin-CTD, its preferential localization in prophage WO genomes versus the rest of
166 the *Wolbachia* chromosome, conservation of eukaryotic furin cleaveage sites, large
167 eukaryotic-like length, homology to invertbrate-specific toxins, and reduced divergence
168 relative to the spider venom homologs is consistent with a eukaryotic origin and post-
169 translational processing by furin peptidases.

170

171 **Pox protein Repeats of ANkyrin C terminus (PRANC).** Domains central to modifying
172 animal proteins are also abundant in the phage WO EAM. The PRANC domain in the
173 WOVitA1 genome (gwv_1092) shares protein sequence homology with corresponding
174 PRANC domains in multiple parasitic wasp hosts (Supplementary Table 4) and their
175 eukaryotic viruses. Reciprocal BLASTP searches retrieve the same best hits and support
176 previous findings that this protein domain horizontally transferred between eukaryotic
177 viruses, animals, and *Proteobacteria*[33]. The discovery here of the eukaryotic-like PRANC
178 domain in phage WO parallels its presence in the *Poxviridae* virus family, in which it
179 functions in evasion of eukaryotic immune responses via modification of host
180 ubiquitination. PRANC is related to amino acid sequences in F-box proteins, which are
181 eukaryotic proteins involved in protein degradation. The PRANC domain also occurs in
182 vaccina virus, ectromelia virus, cowpox virus and Orf virus and can regulate NF-κB
183 signalling pathway to inhibit transcription of inflammatory cytokines[34].
184

185 **Conserved ankyrin and TetratricoPeptide Repeat (TPR) protein.** Adjacent to the
186 PRANC-encoding gene in WOVitA1's EAM is an ankyrin and TPR-containing
187 gwv_1093. Ankyrin repeats and TPRs mediate a broad range of protein-protein
188 interactions (apoptosis, cell signaling, inflammatory response, etc.) within eukaryotic
189 cells and are commonly associated with effector proteins of certain intracellular
190 pathogens[35,36]**.** In *Wolbachia*, ankyrins within the core phage genome have been
191 associated with reproductive manipulation of the insect host[37,38]. While generally rare in
192 viral genomes (Supplementary Fig. 2 and 3), these repeat regions occur in all prophage

193 WO haplotypes from sequenced *Wolbachia* genomes (N=23). Phylogenetic analysis
194 using reciprocal BLASTP hits (Fig. 4) shows that the N-terminus sequences of the TPR-
195 containing gwv_1093 are embedded within a diverse set of homologs from many
196 athropod lineages (Fig. 4b), with the most recent transfer putatively occurring between
197 phage WO and *Solenopsis invicta* (Fig. 4c). In this species, the gene is located between
198 ant genes bicaudal D and rho guanine nucleotide exchange factor 11. As *S. invicta* can
199 naturally harbor *Wolbachia*[39], either a gene transfer event occurred between these
200 ecologically-associated taxa or the *S. invicta* homolog could be an assembly artifact. This
201 assembly was based on samples from a region rarely infected with *Wolbachia* (Y Wurm,
202 personal communication, April 2016) and there are no other *Wolbachia*/prophage WO
203 homologs in the *S. invicta* genome; therefore, the latter explanation seems unlikely.
204 Moreover, other gwv_1093 homologs are from insect genome sequences of uninfected
205 strains, i.e., *N. vitripennis*, and thus they can not be derived by an assembly artifact.
206 Based on parsimony, the transfer event appears to have occurred from arthopod to phage
207 WO since the arthropod taxa comprise a more diverse set of lineages. However, the
208 reverse is plausible as transfers from *Wolbachia* to their arthropod hosts are common[40-42].
209
210 **NACHT.** Another instance of genetic transfer involves the programmed cell death (PCD)
211 domain, NACHT (Fig. 5). Eukaryotic NACHT-containing proteins are typically engaged
212 in PCD by acting as pathogen-sensors and signal transduction molecules of the innate
213 immune system[43]. The polymorphic prophage WO homolog encodes ankyrin repeats and
214 a latrotoxin-CTD directly downstream from the conserved NTPase domain (Fig. 5a).
215 NACHT domains have been identified in animals, fungi and bacteria[44] and phylogenetic

216 patterns indicate multiple instances of horizontal transfer[45]. A NACHT-containing
217 peptide was recently discovered in the *Clostridium difficile*-infecting phage
218 phiCDHM1[46]. In contrast to prophage WO, it is bacterial in both amino acid homology
219 and protein architecture. While all BLASTP and reciprocal BLASTP queries of the
220 phiCDHM1 NACHT domain yield only bacterial homologs, BLASTP searches of the
221 prophage WO NACHT domain yield only animal homologs, and reciprocal BLASTP
222 searches of these yield only hits to prophage WO and other animals. Similar to the
223 phylogeny of the N-terminus of the TPR-containing gwv_1093, this single NACHT
224 domain sequence in prophage WO is embedded within a more diverse set of homologs in
225 arthropods (Fig. 5b,c). Phylogenetic analyses place the prophage WO variants adjacent to
226 a divergent *Bombyx mori* sequence, though these variants have slightly closer total
227 homology to *Culex quiquefasciatus* mosquitoes that harbor *Wolbachia* with related
228 prophage WO variants.
229
230 **Discussion**
231 Metagenomic analysis of the complete genome from phage WO particles reveals all
232 formerly described phage genetic modules (lysogeny, baseplate, head, replication,
233 virulence, tail and patatin-like phospholipase[27]) as well as a new group of genes that we
234 collectively group into a eukaryotic associatoin module (EAM). Some of these genes (i)
235 encode protein domains and cleavage sites central to eukaryotic functions, (ii) occur in
236 both phage and metazoan hosts, (iii) comprise the second largest phage gene to date
237 (14,256 bp) and (iv) are absent from mutualistic, phage-free genomes of *Wolbachia*.
238 Together, these genes increase the phage WO genome size by roughly 50% and include

239 ten types of protein domains with four predicted eukaryotic functions: toxins, host-
240 microbe interactions, host cell suicide, and secretion of proteins through the cell
241 membrane. Notably, over half of these domain types share greater amino acid homology
242 to eukaryotic invertebrates than to bacteria in GenBank. Among this subset with
243 eukaryotic sequence homology, the protein domains are almost exclusively found in the
244 phage EAM. An EAM has never before been reported in bacteriophage genomes, to our
245 knowledge, possibly because phages of obligate intracellular bacteria occupy a unique
246 eukaryotic-enclosed niche and are relatively understudied.
247
248 The presence of eukaryotic protein domains in bacteriophage genomes is of special note
249 as they curiously mirror eukaryotic genes in large eukaryotic viruses that aid in viral
250 mimicry and manipulation of host processes[47,48]. In phage WO, these animal protein
251 domains are central to anti-eukaryotic functions including the black widow latrotoxin,
252 programmed cell death (NACHT), immune evasion (PRANC), and protein-protein
253 interactions.
254
255 Bacteriophage WO frequently transfer between *Wolbachia* coinfections in the same
256 animal host[49,50] and to the host genome as part of large transfers of the *Wolbachia*
257 chromosome[40,41]. We previously reported that phage WO in *Wolbachia* of *Nasonia*
258 *vitripennis* were also capable of transferring adjacent, flanking, non-phage genes in the
259 process of exchange between coinfections[51]. For two of these flanking genes, sequence
260 evidence indicated that *Wolbachia* genomes may be able to receive eukaryotic
261 DNA[42,52,53]. However, the nature of these lateral genetic transfers remained to be

262 elucidated as these regions were not previously known to be part of the packaged phage
263 genome until now. Here, we demonstrate that genes with eukaryotic homology are
264 constituents of phage WO and its EAM, and they either retain conservation of eukaryotic
265 furin cleaveage sites and a large eukaryotic-like length (i.e., latrotoxin-CTD), or they
266 exhibit markedly reduced or no diversity relative to the arthropod homologs as the WO
267 sequences exist as single or a few representatives (NACHT and TPR-containing
268 proteins). Moreover, WO protein domains with eukaryotic homology are highly enriched
269 in the EAM over WO protein domains with bacterial homology. Based on this work, we
270 suspect that systematic surveys of phage genomes in intimate host-associated bacteria
271 may uncover a broad range of eukaryotic-like protein domains involved in phage
272 lifecycle adaptations and phage-eukaryote interactions. Of particular note is the reported
273 association between phage WO genes, specifically ankyrins, transcriptional regulators
274 and the Ulp1 operon, and *Wolbachia*'s ability to manipulate host reproduction[37,38,54-56].
275
276 The mechanisms by which eukaryotic protein domains are exchanged with phage WO are
277 unknown and could follow at least three models (Fig. 6). First, animal genetic material
278 could directly transfer to and from WO genomes during phage particle propagation in the
279 cytoplasm of animal cells (Fig. 6b) or during packaging inside *Wolbachia* cells that are
280 lysing and exposed to the eukaryotic cytoplasmic environment. Packaging of eukaryotic
281 host RNAs, for instance, occur in the virions of herpesvirus[57] and cytomegalovirus[58].
282 Second, genes may transfer between animal genomes and the *Wolbachia* chromosome
283 and then to prophage WO. For this scenario to be plausible, animal genetic material
284 transferred in random locations in the *Wolbachia* genome would have to be preferentially

285 lost in non-phage associated locations from the *Wolbachia* chromosome (Fig. 6c) because
286 domains with eukaryotic homology are highly enriched in the phage/prophage WO EAM
287 versus the rest of the chromosome (Fig. 2). Third, DNA may transfer first between
288 animal genomes and intermediary entities, such as eukaryotic viruses or other obligate
289 intracellular bacteria, and then to phage WO and/or *Wolbachia* (Fib. 6d). In fact, the
290 PRANC-domain (described in gwv_1092) was named for its discovery in and association
291 with eukaryotic Pox viruses. Finally, once DNA is incorporated into a prophage genome,
292 it is susceptible to recombination with other phage WO haplotypes located in the same
293 *Wolbachia* chromosome and can transfer from one haplotype to another.
294
295 Alternatively, these protein domains could originate in the phage and be particularly
296 prone to transfer, maintenance, and spread in their recipient arthropod genomes (Fig. 6b).
297 For this scenario to be plausible, it would have to imply that phage genetic material
298 independently and repeatedly transfers to athropods and spreads through the host
299 population, which would subsequently be followed by loss of these phage genes or
300 recombination with other non-transferred phage genetic material so that the eukaryotic
301 sequence varation clusters seperately from the phage WO sequence(s). While each mode
302 of transfer is possible, the eukaryotic length of these genes, presence of furin protease
303 domains, and enrichment in the phage WO EAM provides evidence for their eukaryotic
304 origin.
305
306 Why are these protein domains present in the EAM of bacteriophage WO? Some phages
307 of obligate intracellular bacteria may have to overcome two major challenges not

308 encountered by the well-studied phages of free-living bacteria. First, they are contained
309 within both bacterial and eukaryotic membranes, posing an enigmatic "two-fold cell
310 challenge". They may not only have to breach peptidoglycan and permeabilize bacterial
311 membranes, but they may also have to exit (and enter) across the eukaryotic membrane(s)
312 that directly encapsulates the bacteria. Second, like their bacterial hosts, they must
313 survive the internal cellular environment of the animal host, including the innate immune
314 response and autophagy, while searching for phage-susceptible bacteria. Phage WO can
315 dwell in the eukaryotic cytoplasm and extracellular matrix that they encounter upon
316 bacterial lysis[26], raising the likelihood of direct interaction with host membranes and
317 intracellular biology. In this context, EAM protein domains are prime candidates to aid in
318 functions including cell lysis (latrotoxin-CTD), manipulation of programmed cell death
319 (NACHT and NB-ARC), host ubiquitination (OTU and Ulp1), insecticidal toxicity (ABC
320 toxin) and interaction with host proteins (ankryin repeats and TPRs). Rather than simply
321 act as virulence factors to benefit their bacterial host, their massive proportion of genomic
322 real estate (up to 60% of the prophage genome, Supplementary Fig. 4) implies that they
323 may be necessary to phage biology and likely have a direct impact on phage propagation.
324 The concept of phage-mediated ecosystem modification as an alternative to bacterial
325 virulence is not new[59] but, much like the biology of phage WO, is relatively
326 understudied.
327
328 Phage WO is not the only virus described within obligate intracellular bacteria.
329 *Chlamydiomicroviridae* infect obligate intracellular bacteria, yet still do not directly
330 contend with the eukaryotic membrane. Rather, they attach to dormant chlamydial cells

331 (i.e., reticulate bodies) and enter via phagocytosis or endocytosis of the bacteria[60]. The
332 phages then alter development of their bacterial host, which leads to disintegration of the
333 chlamydial inclusion and subsequent lysis of the eukaryotic host cell[61,62]. The nature of
334 phage WO's lifestyle, on the other hand, may require a distinct interaction with multiple
335 membranes and immune responses because lytic activity of phage WO has been
336 associated with typical bacterial cell defects including degraded bacterial DNA, a
337 detached inner membrane, and exit of the phage particles from inside *Wolbachia* and its
338 host cell into the extracellular matrix of the reproductive tissues[26]. Bacteriophages of
339 free-living bacteria also regularly colonize eukaryotic environments, particularly those
340 associated with mucosal surfaces[63]. They, however, do not infect or traverse the
341 eukaryotic membrane and are still within the genomic boundaries of the bacterial
342 virosphere.
343
344 Temperate dsDNA phages also occur in facultative symbionts of aphids[64] and tsetse
345 flies[65]. While *Wolbachia* has never successfully been cultured outside of host cells[66],
346 these facultative symbionts can replicate both intra- and extracellularly (JW Brandt,
347 personal communication, July 2015) suggesting that their phages are not constrained by
348 the same two-fold cell challenge. In addition, their phages encode a traditional lytic
349 cassette (holin and lysozyme) that correlates with the need to deal only with bacterial
350 membranes. In some cases, the phages harbor bacterial-derived toxins that target
351 eukaryotic cells[67], and these function mutualistically in aphids by arresting development
352 of parasitoid wasp larvae[64]. Furthermore, unlike phage WO that is stably maintained in

353 the lab, these phages are readily lost in the absence of parasitoids during laboratory
354 rearing, presumably due to the cost of their toxins[68].
355
356 In addition to providing new insights into the evolution of bacteriophages and showing
357 phage WO genomes to be far more complex than previously described, the findings here
358 reveal evidence for gene sharing between metazoan hosts and phages of obligate
359 intracellular bacteria. We suggest that the putative acquistion and retooling of intact
360 eukaryotic domains in phage WO is analgous to the commandeering of host genes by
361 eukaryotic viruses. Whether lateral genetic transfers between metazoans and
362 bacteriophages are common in the symbiotic virosphere remains to be determined.
363

364 **Methods**

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

373

374 **Phage particle purification.** Phage particles were isolated according to Fujii et al[22] with
375 modifications. Approximately 4 g of adult insects were homogenized in 29.6 ml cold SM
376 buffer (50mM Tris-HCl, pH 7.5, 0.1 M NaCl, 10mM $MgSO_4 \cdot 7H_20$, and 0.1% (w/v)
377 gelatin). NaCl and RNase A were added to a final concentration of 1M and 1μg/ml,
378 respectively. The homogenate was incubated on a shaker at 4˚C for 1 h and then
379 centrifuged at 13,000*g* for 10 min at 4˚C. Polyethylene glycol (PEG) 6000 was added to a
380 final concentration of 10% to precipitate phage particles, incubated at 4˚C for 1 hr with
381 gentle shaking and centrifuged at 13,000g for 10 min. The pellet was resuspended in 5 ml
382 TM buffer (50 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2 \cdot 6H_2O$) and mixed with an equal
383 volume chloroform. The suspension was centrifuged at 3,000*g* to remove PEG and the
384 aqueous phase was filtered through a 0.22 μm filter to remove bacterial cells. The
385 suspension was centrifuged at 60,000*g* for 1 h at 4˚C to collect phage particles. The pellet
386 was suspended in 10 μl TM buffer.

387

388 **Phage DNA extraction and metagenomic sequencing.** The phage suspension was
389 treated with RQ1 RNase-Free DNase (Promega) for 30 min at 37˚C, followed by heat
390 inactivation for 10 min at 65˚C, to remove host DNA contamination. Phage DNA was
391 extracted from the suspension using the QIAamp MinElute Virus Spin Kit (Qiagen) and
392 amplified using the REPLI-g Mini Kit (Qiagen). Following amplification, paired-end
393 DNA libraries were prepared according to manufacturer's (Illumina) instructions and
394 samples were sequenced with an Illumina HiSeq 2000 (2×100-nt read length).
395

396 **Bioinformatics and statistics.** Metagenomic sequences (reads) were trimmed, paired and
397 assembled into contigs using the CLC Assembler (CLC bio) with bubble size = 50,
398 insertion and deletion cost = 3, mismatch cost = 2, length fraction = 0.6, minimum contig
399 size = 130, similarity = 0.5, minimum distance = 90 and maximum distance = 200.
400 Contigs were compared to the GenBank non-redundant database using NCBI's BLASTN
401 (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and those with similarity to phage WO and/or
402 *Wolbachia* (E-value $<10^{-10}$) were manually annotated using Geneious (Biomatters Ltd.).
403 Individual reads were mapped to reference sequences using Geneious. Open reading
404 frame (ORF) homology searches were performed to determine putative function using
405 NCBI's BLASTP (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and Wellcome Trust Sanger
406 Institute's pfam database (http://pfam.sanger.ac.uk). Coiled coil domains were predicted
407 with EMBL's Simple Modular Architecture Research Tool (SMART, http://smart.embl-
408 heidelberg.de). Furin cleavage sites were identified using PiTou
409 (http://www.nuolan.net/reference.html). The number of genes with and without furin

410 cleavage sites was analyzed with respect to phage-region using Fisher's Exact Test
411 (GraphPad Software). Phylogenetic trees were built using the Bayes plugin in Geneious
412 and model selection for each Bayes analysis was estimated using ProtTest[70].
413
414 **Confirmation of phage WO terminal genes.** Genomic DNA was extracted from *w*VitA-
415 infected *N. vitripennis* (strain 12.1) and *w*CauB-infected *E. kuehniella* individuals using
416 the Gentra Puregene Tissue Kit (Qiagen). Primers were designed for both WOVitA1 and
417 WOCauB3 *attP* sites, respectively: VitA1_attF (5'- CGA AGA ACC AGC ACA GGG
418 TGG-3'), VitA1_attR (5'- GCT GGA AGA GGG CAT CTG CAT C-3'), CauB3_attF
419 (5'- TCG TGA CTG CCC TAT TGC TGC T – 3') and CauB3_attR (5'- ATG CGG CCA
420 AAG CTG GGT GT – 3'). Amplification was performed in a Veriti thermal cycler
421 (Applied Biosystems) using GoTaq green master mix (Promega) under the following
422 conditions: 94˚C for 2 min; 35 cycles of 94˚C for 30 s, 53˚C for 30 s, 72˚C for 1 min; and
423 a final elongation cycle of 72˚C for 10 min. PCR products were sequenced via Sanger
424 sequencing (Genewiz, Inc).
425
426 **Data availability.** The phage WOVitA1 genome assembly reported in this paper has
427 been deposited in NCBI under accession number KX522565. The *N. vitripennis* viral
428 metagenome sequences have been deposited in the SRA under accession number
429 SRR3560636 and BioProject PRJNA321548. The *w*CauB-infected *E. kuehniella* viral
430 metagenome sequences have been deposited in the SRA under accession number
431 SRR3536639 and BioProject PRJNA321549.
432

433 Data referenced in this study are available in NCBI with accession codes AE017196

434 (*w*Mel), AM999887 (*w*Pip), CTEH00000000 (*w*PipMol), ABZA00000000 (*w*PipJHB)

435 CP001391 (*w*Ri), CAOU00000000 (*w*Suzi), AMZJ00000000 (*w*Di), AAGB01000001

436 (*w*Ana), CAGB00000000 (*w*AlbB), CAOH00000000 (*w*Bol1-b), JYPC00000000 (*w*Ob),

437 CP003884 (*w*Ha), CP003883 (*w*No), LK055284 (*w*Au), AP013028 (*w*Cle), HE660029

438 (*w*Oo), PRJNA213627 (*w*VitA), AB478515 (WOCauB2), AB478516 (WOCauB3),

439 KC955252 (WOSol), HQ906665 and HQ906666 (WOVitB).

440

441 **References**

442 1. Edwards, R.A. & Rohwer, F. Viral metagenomics. *Nat Rev Microbiol* **3**, 504-10
443 (2005).
444 2. Hendrix, R.W., Smith, M.C., Burns, R.N., Ford, M.E. & Hatfull, G.F.
445 Evolutionary relationships among diverse bacteriophages and prophages: all the
446 world's a phage. *Proc Natl Acad Sci U S A* **96**, 2192-7 (1999).
447 3. Brussow, H. The not so universal tree of life or the place of viruses in the living
448 world. *Philos Trans R Soc Lond B Biol Sci* **364**, 2263-74 (2009).
449 4. King, A.M.Q., Adams, M.J., Lefkowitz, E.J. & Carstens, E.B. *Virus taxonomy:*
450 *classification and nomenclature of viruses: Ninth Report of the International*
451 *Committee on Taxonomy of Viruses.*, 1327 (Elsevier, San Diego, 2012).
452 5. Nasir, A., Forterre, P., Kim, K.M. & Caetano-Anolles, G. The distribution and
453 impact of viral lineages in domains of life. *Front Microbiol* **5**, 194 (2014).
454 6. Prangishvili, D., Forterre, P. & Garrett, R.A. Viruses of the Archaea: a unifying
455 view. *Nat Rev Microbiol* **4**, 837-48 (2006).
456 7. Forterre, P. Giant viruses: conflicts in revisiting the virus concept. *Intervirology*
457 **53**, 362-78 (2010).
458 8. Raoult, D. TRUC or the need for a new microbial classification. *Intervirology* **56**,
459 349-53 (2013).
460 9. Elde, N.C. & Malik, H.S. The evolutionary conundrum of pathogen mimicry. *Nat*
461 *Rev Microbiol* **7**, 787-97 (2009).
462 10. Rappoport, N. & Linial, M. Viral proteins acquired from a host converge to
463 simplified domain architectures. *PLoS Comput Biol* **8**, e1002364 (2012).
464 11. Colson, P. & Raoult, D. Gene repertoire of amoeba-associated giant viruses.
465 *Intervirology* **53**, 330-43 (2010).
466 12. Canchaya, C., Fournous, G. & Brussow, H. The impact of prophages on bacterial
467 chromosomes. *Mol Microbiol* **53**, 9-18 (2004).
468 13. Lindell, D. *et al.* Transfer of photosynthesis genes to and from Prochlorococcus
469 viruses. *Proc Natl Acad Sci U S A* **101**, 11013-8 (2004).

470 14. Dammeyer, T., Bagby, S.C., Sullivan, M.B., Chisholm, S.W. & Frankenberg-
471 Dinkel, N. Efficient phage-mediated pigment biosynthesis in oceanic
472 cyanobacteria. *Curr Biol* **18**, 442-8 (2008).
473 15. Werren, J.H., Baldo, L. & Clark, M.E. Wolbachia: master manipulators of
474 invertebrate biology. *Nat Rev Microbiol* **6**, 741-51 (2008).
475 16. Zug, R. & Hammerstein, P. Still a host of hosts for Wolbachia: analysis of recent
476 data suggests that 40% of terrestrial arthropod species are infected. *PLoS One* **7**,
477 e38544 (2012).
478 17. Cho, K.O., Kim, G.W. & Lee, O.K. Wolbachia bacteria reside in host Golgi-
479 related vesicles whose position is regulated by polarity proteins. *PLoS One* **6**,
480 e22703 (2011).
481 18. Louis, C. & Nigro, L. Ultrastructual evidence of Wolbachia Rickettsiales in
482 Drosophila simulans and their relationships with unidirectional cross-
483 incompatibility. *Journal of Invertebrate Pathology* **54**, 39-44 (1989).
484 19. Gavotte, L. *et al.* A Survey of the bacteriophage WO in the endosymbiotic
485 bacteria Wolbachia. *Mol Biol Evol* **24**, 427-35 (2007).
486 20. Kent, B.N. & Bordenstein, S.R. Phage WO of Wolbachia: lambda of the
487 endosymbiont world. *Trends Microbiol* **18**, 173-81 (2010).
488 21. Metcalf, J.A. & Bordenstein, S.R. The complexity of virus systems: the case of
489 endosymbionts. *Curr Opin Microbiol* **15**, 546-52 (2012).
490 22. Fujii, Y., Kubo, T., Ishikawa, H. & Sasaki, T. Isolation and characterization of the
491 bacteriophage WO from Wolbachia, an arthropod endosymbiont. *Biochem*
492 *Biophys Res Commun* **317**, 1183-8 (2004).
493 23. Masui, S. *et al.* Bacteriophage WO and virus-like particles in Wolbachia, an
494 endosymbiont of arthropods. *Biochem Biophys Res Commun* **283**, 1099-104
495 (2001).
496 24. Sanogo, Y.O. & Dobson, S.L. WO bacteriophage transcription in Wolbachia-
497 infected Culex pipiens. *Insect Biochem Mol Biol* **36**, 80-5 (2006).
498 25. Tanaka, K., Furukawa, S., Nikoh, N., Sasaki, T. & Fukatsu, T. Complete WO
499 phage sequences reveal their dynamic evolutionary trajectories and putative
500 functional elements required for integration into the Wolbachia genome. *Appl*
501 *Environ Microbiol* **75**, 5676-86 (2009).
502 26. Bordenstein, S.R., Marshall, M.L., Fry, A.J., Kim, U. & Wernegreen, J.J. The
503 tripartite associations between bacteriophage, Wolbachia, and arthropods. *PLoS*
504 *Pathog* **2**, e43 (2006).
505 27. Kent, B.N., Funkhouser, L.J., Setia, S. & Bordenstein, S.R. Evolutionary
506 genomics of a temperate bacteriophage in an obligate intracellular bacteria
507 (Wolbachia). *PLoS One* **6**, e24984 (2011).
508 28. Zhang, D., de Souza, R.F., Anantharaman, V., Iyer, L.M. & Aravind, L.
509 Polymorphic toxin systems: Comprehensive characterization of trafficking modes,
510 processing, mechanisms of action, immunity and ecology using comparative
511 genomics. *Biol Direct* **7**, 18 (2012).
512 29. Ioannidis, P. *et al.* Rapid transcriptome sequencing of an invasive pest, the brown
513 marmorated stink bug Halyomorpha halys. *BMC Genomics* **15**, 738 (2014).
514 30. Goodacre, S.L., Martin, O.Y., Thomas, C.F. & Hewitt, G.M. Wolbachia and other
515 endosymbiont infections in spiders. *Mol Ecol* **15**, 517-27 (2006).

516 31. Garb, J.E. & Hayashi, C.Y. Molecular evolution of alpha-latrotoxin, the
517 exceptionally potent vertebrate neurotoxin in black widow spider venom. *Mol*
518 *Biol Evol* **30**, 999-1014 (2013).
519 32. Bhere, K.V., Haney, R.A., Ayoub, N.A. & Garb, J.E. Gene structure, regulatory
520 control, and evolution of black widow venom latrotoxins. *FEBS Lett* **588**, 3891-7
521 (2014).
522 33. Werren, J.H. *et al.* Functional and evolutionary insights from the genomes of
523 three parasitoid Nasonia species. *Science* **327**, 343-8 (2010).
524 34. Chang, S.J. *et al.* Poxvirus host range protein CP77 contains an F-box-like
525 domain that is necessary to suppress NF-kappaB activation by tumor necrosis
526 factor alpha but is independent of its host range function. *J Virol* **83**, 4140-52
527 (2009).
528 35. Cerveny, L. *et al.* Tetratricopeptide repeat motifs in the world of bacterial
529 pathogens: role in virulence mechanisms. *Infect Immun* **81**, 629-35 (2013).
530 36. Pan, X., Luhrmann, A., Satoh, A., Laskowski-Arce, M.A. & Roy, C.R. Ankyrin
531 repeat proteins comprise a diverse family of bacterial type IV effectors. *Science*
532 **320**, 1651-4 (2008).
533 37. Pichon, S. *et al.* The expression of one ankyrin pk2 allele of the WO prophage is
534 correlated with the Wolbachia feminizing effect in isopods. *BMC Microbiol* **12**,
535 55 (2012).
536 38. Walker, T. *et al.* Ankyrin repeat domain-encoding genes in the wPip strain of
537 Wolbachia from the Culex pipiens group. *BMC Biol* **5**, 39 (2007).
538 39. Shoemaker, D.D., Ross, K.G., Keller, L., Vargo, E.L. & Werren, J.H. Wolbachia
539 infections in native and introduced populations of fire ants (Solenopsis spp.).
540 *Insect Mol Biol* **9**, 661-73 (2000).
541 40. Dunning Hotopp, J.C. *et al.* Widespread lateral gene transfer from intracellular
542 bacteria to multicellular eukaryotes. *Science* **317**, 1753-6 (2007).
543 41. Funkhouser-Jones, L.J. *et al.* Wolbachia co-infection in a hybrid zone: discovery
544 of horizontal gene transfers from two Wolbachia supergroups into an animal
545 genome. *PeerJ* **3**, e1479 (2015).
546 42. Klasson, L., Kambris, Z., Cook, P.E., Walker, T. & Sinkins, S.P. Horizontal gene
547 transfer between Wolbachia and the mosquito Aedes aegypti. *BMC Genomics* **10**,
548 33 (2009).
549 43. Koonin, E.V. & Aravind, L. Origin and evolution of eukaryotic apoptosis: the
550 bacterial connection. *Cell Death Differ* **9**, 394-404 (2002).
551 44. Koonin, E.V. & Aravind, L. The NACHT family - a new group of predicted
552 NTPases implicated in apoptosis and MHC transcription activation. *Trends*
553 *Biochem Sci* **25**, 223-4 (2000).
554 45. Leipe, D.D., Koonin, E.V. & Aravind, L. STAND, a class of P-loop NTPases
555 including animal and plant regulators of programmed cell death: multiple,
556 complex domain architectures, unusual phyletic patterns, and evolution by
557 horizontal gene transfer. *J Mol Biol* **343**, 1-28 (2004).
558 46. Hargreaves, K.R., Kropinski, A.M. & Clokie, M.R. What does the talking?:
559 quorum sensing signalling genes discovered in a bacteriophage genome. *PLoS*
560 *One* **9**, e85131 (2014).

561 47. Alcami, A. & Koszinowski, U.H. Viral mechanisms of immune evasion. *Immunol*
562 *Today* **21**, 447-55 (2000).
563 48. Seet, B.T. *et al.* Poxviruses and immune evasion. *Annu Rev Immunol* **21**, 377-423
564 (2003).
565 49. Bordenstein, S.R. & Wernegreen, J.J. Bacteriophage flux in endosymbionts
566 (Wolbachia): infection frequency, lateral transfer, and recombination rates. *Mol*
567 *Biol Evol* **21**, 1981-91 (2004).
568 50. Masui, S., Kamoda, S., Sasaki, T. & Ishikawa, H. Distribution and evolution of
569 bacteriophage WO in Wolbachia, the endosymbiont causing sexual alterations in
570 arthropods. *J Mol Evol* **51**, 491-7 (2000).
571 51. Kent, B.N. *et al.* Complete bacteriophage transfer in a bacterial endosymbiont
572 (Wolbachia) determined by targeted genome capture. *Genome Biol Evol* **3**, 209-18
573 (2011).
574 52. Duplouy, A. *et al.* Draft genome sequence of the male-killing Wolbachia strain
575 wBol1 reveals recent horizontal gene transfers from diverse sources. *BMC*
576 *Genomics* **14**, 20 (2013).
577 53. Woolfit, M., Iturbe-Ormaetxe, I., McGraw, E.A. & O'Neill, S.L. An ancient
578 horizontal gene transfer between mosquito and the endosymbiotic bacterium
579 Wolbachia pipientis. *Mol Biol Evol* **26**, 367-74 (2009).
580 54. Beckmann, J.F. & Fallon, A.M. Detection of the Wolbachia protein WPIP0282 in
581 mosquito spermathecae: implications for cytoplasmic incompatibility. *Insect*
582 *Biochem Mol Biol* **43**, 867-78 (2013).
583 55. Pinto, S.B. *et al.* Transcriptional regulation of Culex pipiens mosquitoes by
584 Wolbachia influences cytoplasmic incompatibility. *PLoS Pathog* **9**, e1003647
585 (2013).
586 56. Sutton, E.R., Harris, S.R., Parkhill, J. & Sinkins, S.P. Comparative genome
587 analysis of Wolbachia strain wAu. *BMC Genomics* **15**, 928 (2014).
588 57. Amen, M.A. & Griffiths, A. Packaging of Non-Coding RNAs into Herpesvirus
589 Virions: Comparisons to Coding RNAs. *Front Genet* **2**, 81 (2011).
590 58. Terhune, S.S., Schroer, J. & Shenk, T. RNAs are packaged into human
591 cytomegalovirus virions in proportion to their intracellular concentration. *J Virol*
592 **78**, 10390-8 (2004).
593 59. Abedon, S.T. & Lejeune, J.T. Why bacteriophage encode exotoxins and other
594 virulence factors. *Evol Bioinform Online* **1**, 97-110 (2005).
595 60. Sliwa-Dominiak, J., Suszynska, E., Pawlikowska, M. & Deptula, W. Chlamydia
596 bacteriophages. *Arch Microbiol* **195**, 765-71 (2013).
597 61. Hsia, R., Ohayon, H., Gounon, P., Dautry-Varsat, A. & Bavoil, P.M. Phage
598 infection of the obligate intracellular bacterium, Chlamydia psittaci strain guinea
599 pig inclusion conjunctivitis. *Microbes Infect* **2**, 761-72 (2000).
600 62. Salim, O., Skilton, R.J., Lambden, P.R., Fane, B.A. & Clarke, I.N. Behind the
601 chlamydial cloak: the replication cycle of chlamydiaphage Chp2, revealed.
602 *Virology* **377**, 440-5 (2008).
603 63. Barr, J.J. *et al.* Bacteriophage adhering to mucus provide a non-host-derived
604 immunity. *Proc Natl Acad Sci U S A* **110**, 10771-6 (2013).

605 64. Moran, N.A., Degnan, P.H., Santos, S.R., Dunbar, H.E. & Ochman, H. The
606 players in a mutualistic symbiosis: insects, bacteria, viruses, and virulence genes.
607 *Proc Natl Acad Sci U S A* **102**, 16919-26 (2005).
608 65. Belda, E., Moya, A., Bentley, S. & Silva, F.J. Mobile genetic element
609 proliferation and gene inactivation impact over the genome structure and
610 metabolic capabilities of Sodalis glossinidius, the secondary endosymbiont of
611 tsetse flies. *BMC Genomics* **11**, 449 (2010).
612 66. Rasgon, J.L., Gamston, C.E. & Ren, X. Survival of Wolbachia pipientis in cell-
613 free medium. *Appl Environ Microbiol* **72**, 6934-7 (2006).
614 67. Degnan, P.H. & Moran, N.A. Diverse phage-encoded toxins in a protective insect
615 endosymbiont. *Appl Environ Microbiol* **74**, 6782-91 (2008).
616 68. Oliver, K.M., Degnan, P.H., Hunter, M.S. & Moran, N.A. Bacteriophages encode
617 factors required for protection in a symbiotic mutualism. *Science* **325**, 992-4
618 (2009).
619 69. Chafee, M.E. *et al.* Decoupling of host-symbiont-phage coadaptations following
620 transfer between insect species. *Genetics* **187**, 203-15 (2011).
621 70. Abascal, F., Zardoya, R. & Posada, D. ProtTest: selection of best-fit models of
622 protein evolution. *Bioinformatics* **21**, 2104-5 (2005).
623

624

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 contributions

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.

Figure legends

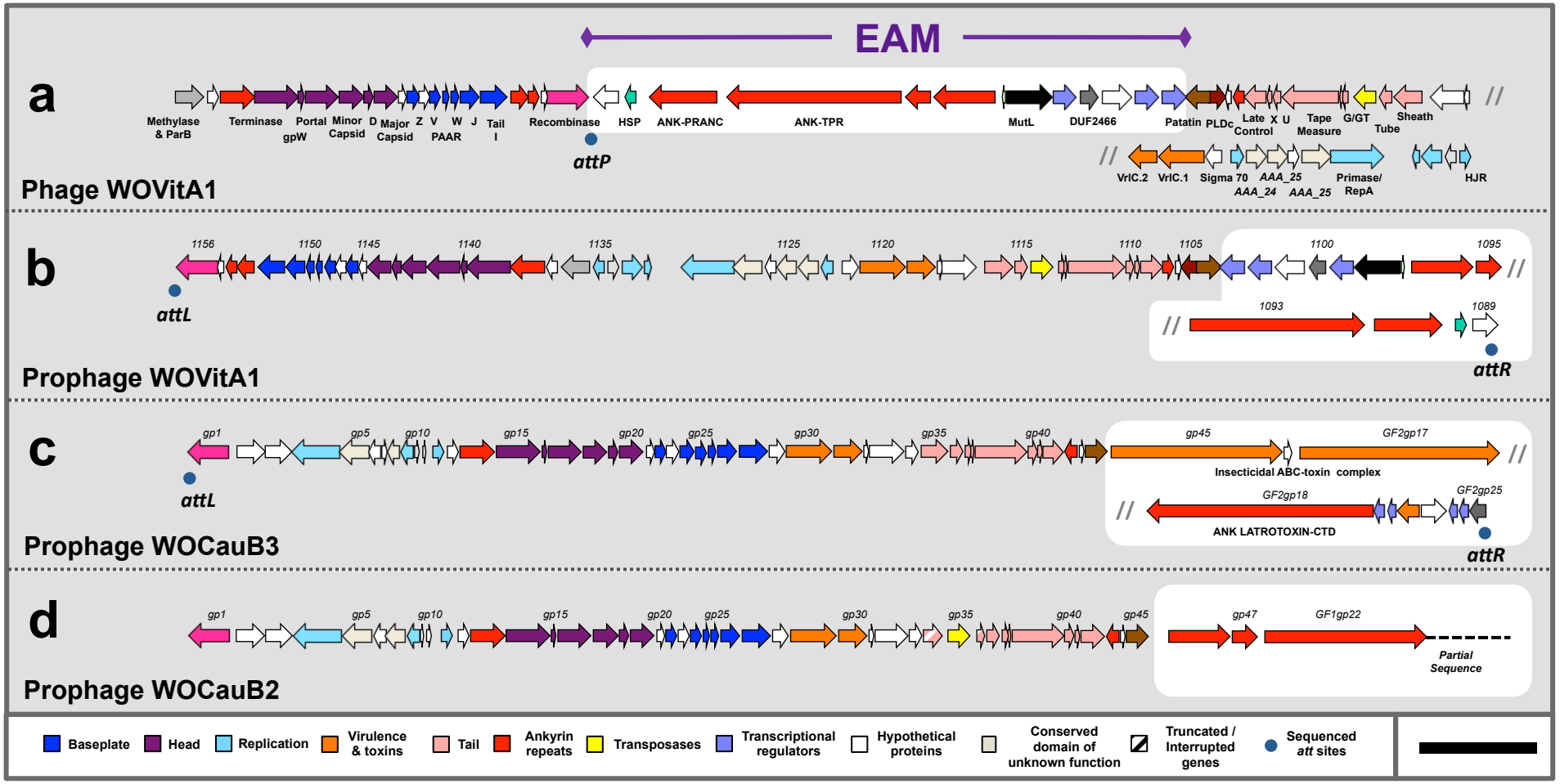
Figure 1 | Phage WO genomes harbor a Eukaryotic Association Module (EAM)

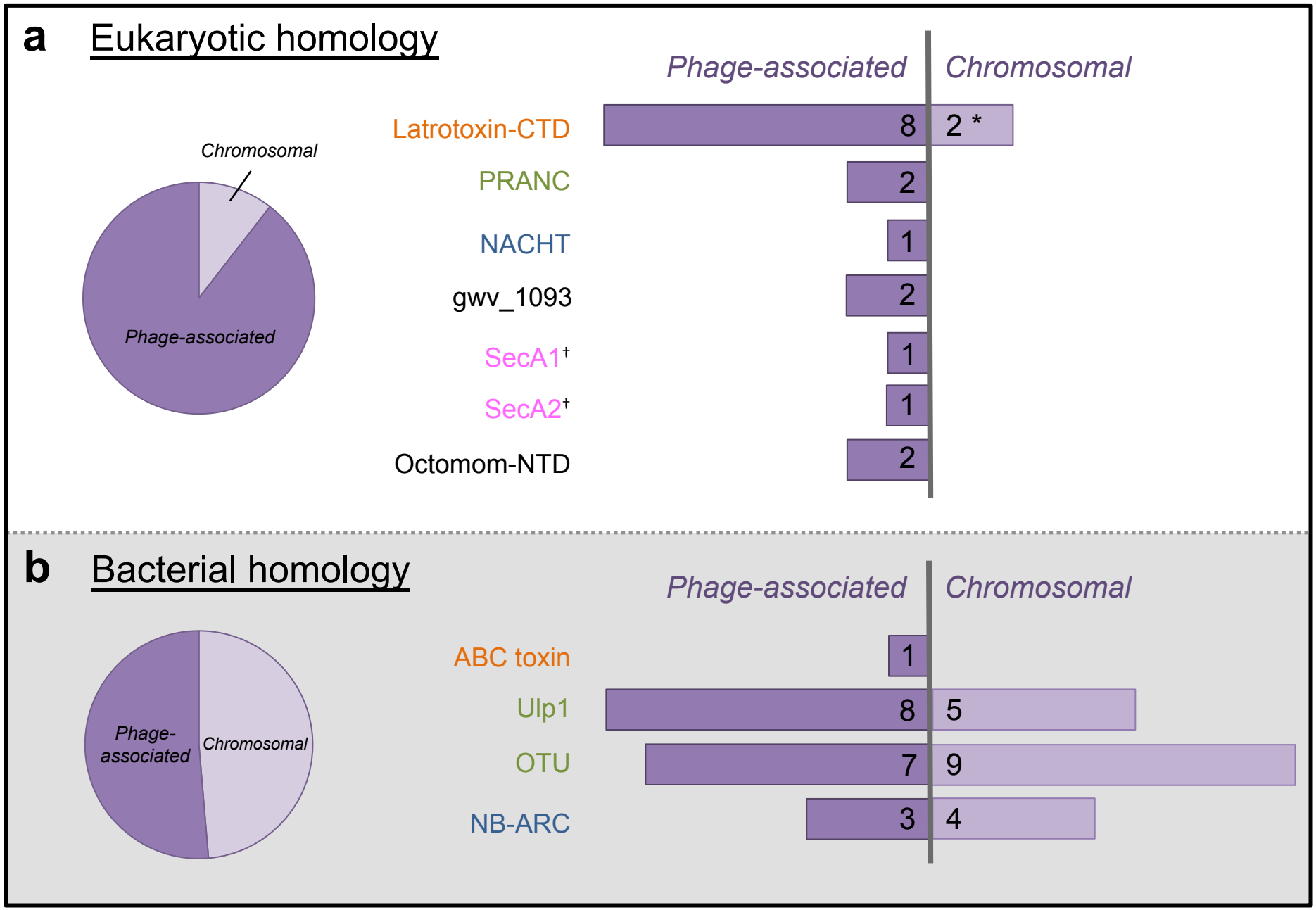
The complete phage WO genome for (a) WOVitA1 was sequenced directly from purified viral particles using high throughput, metagenomic sequencing. The prophage (b)

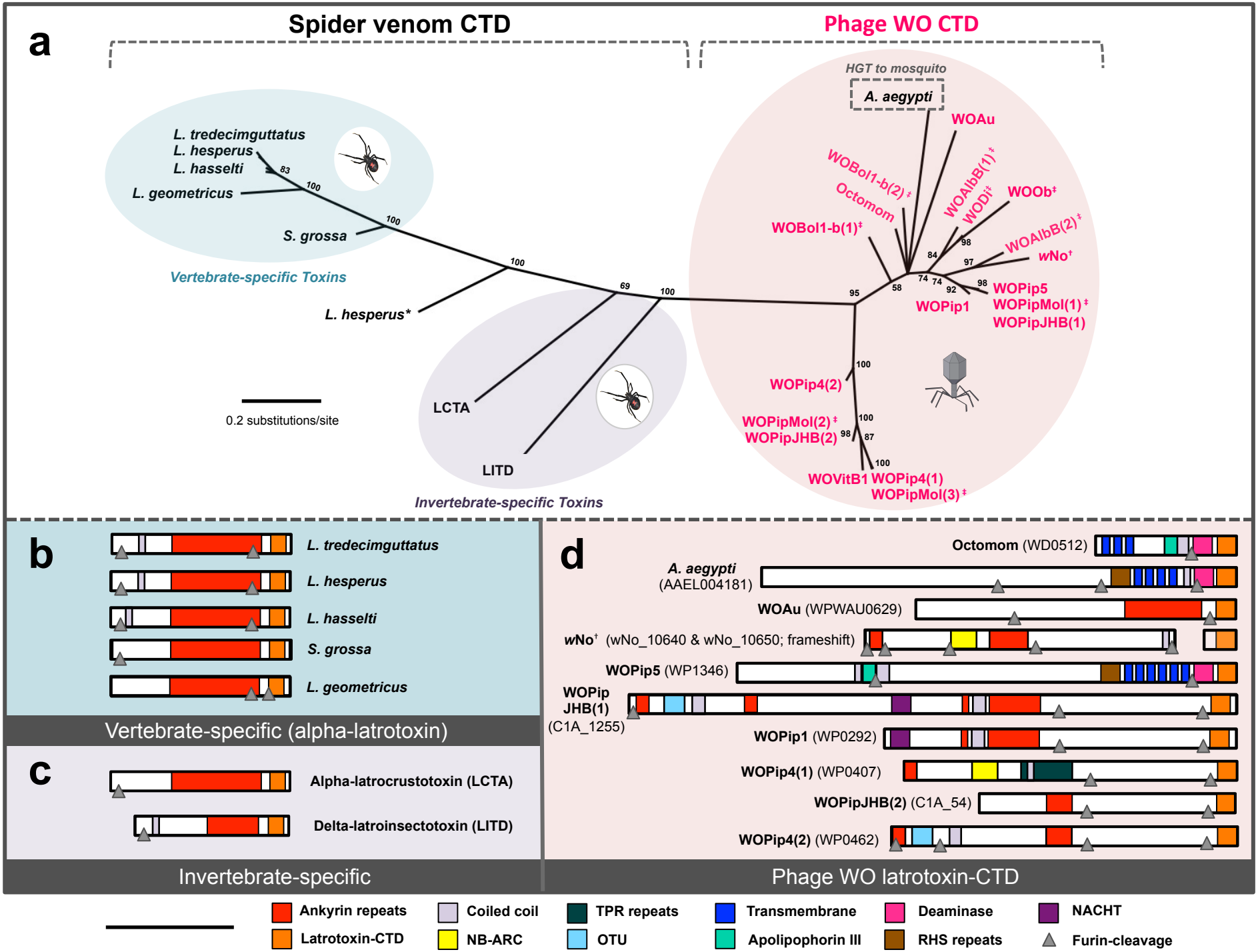
649 WOVitA1, (c) WOCauB3 and (d) WOCauB2 genomes were reannotated based on
650 sequencing reads obtained from purified particles; complete genomes of WOCauB3 and
651 WOCauB2 were not obtained. Each genome consists of a bacteriophage-like region
652 (recombinase to patatin) and EAM highlighted in white. Gray slash marks indicate
653 illustrative continuation of the genome. Dark blue dots indicate the discovery of the *attL*
654 and *attR* sites of the prophage, which adjoin in the packaged WO genome to form *attP*.
655 Numbers above the open reading frames indicate locus tags. Scale bar, 5,000 base pairs.
656
657 **Figure 2 | Eukaryotic-like EAM genes are enriched in prophage WO regions**
658 EAM genes with (a) eukaryotic homology are most likely to be associated with prophage
659 WO while those with (b) bacterial homology are both phage-associated and found
660 scattered throughout the *Wolbachia* chromosome. (*) The two chromosomal latrotoxin-
661 CTD domains (wNo_10650 and wHa_05390) are located within phage-associated genes
662 and transposases, indicating a potential genomic rearrangement. (†) SecA represents one
663 'domain type' but is listed separately because phage WO contains two different homologs
664 (i.e., wHa_3920 and wHa_3930). Putative functional categories are: anti-eukaryotic
665 toxins (orange); host-microbe interactions (green); host cell suicide (blue); secretion of
666 virulence factors (pink); and unknown (black). Octomom refers to WD0513 of the *w*Mel
667 genome.
668
669 **Figure 3 | Latrotoxin-CTD phylogeny supports lateral genetic transfers**
670 (a) Phylogeny of phage WO latrotoxin-CTD protein domains and their eukaryotic
671 homologs was constructed by Bayesian analysis of 74 amino acids using the JTT model

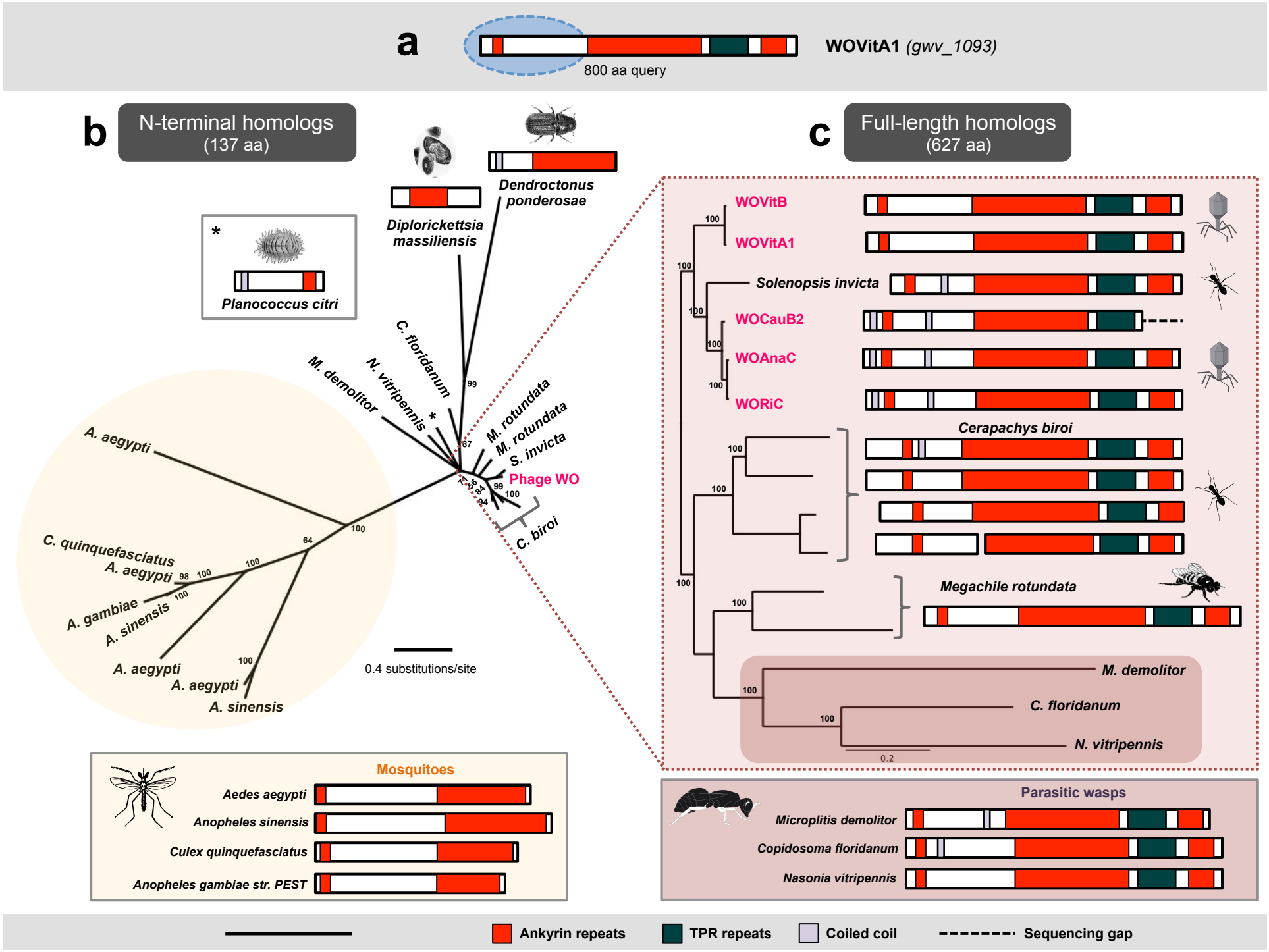
672 of evolution. Consensus support values are shown at the nodes. Comparative protein
673 architecture shows that spider venom (b) vertebrate-specific alpha-latrotoxins and (c)
674 invertebrate-specific alpha- and delta-latrotoxins are highly conserved, whereas (d) phage
675 WO are not. Bolded nomenclature in (d) denotes the specific phage WO haplotype (listed
676 as WO). Genome locus tags are listed in parentheses. Predicted furin cleavage sites, listed
677 in Supplementary Table 3, are illustrated with gray triangles. (*) A second *L. hesperus*
678 sequence represents a recently-described downstream paralog with unknown toxin
679 activity[32]. (†) wNo_10650 is located within phage-associated genes and transposases,
680 indicating a potential genomic rearrangement of a phage region. (‡) Architecture is not
681 shown for sequences on incomplete contigs (WOBol1-b, WOAlbB, WODi, WOPipMol,
682 WOVitB) because complete peptide information and specific phage association are
683 unknown. Scale bar, 1,000 amino acids.
684
685 **Figure 4 | Conserved TPR and anyrin proteins support lateral genetic transfer**
686 (a) A BLASTP query of WOVitA1's gwv_1093 N-terminus reveals homologs in
687 moquitoes, ants, beetles, a mealybug, a solitary bee and one obligate intracellular
688 gammaproteobacteria. Bayesian phylogenetic trees were constructed based on (b) a 137-
689 aa alignment of all homologs with E-value less than e^{-40} using the LG+G model of
690 evolution. (c) To resolve taxa closest to phage WO, trees were reconstructed based on a
691 627-aa alignment of all homologs with an E-value of 0 using the JTT+I+G model of
692 evolution. Isoforms were removed from each alignment. Both trees are unrooted.
693 Consensus support values are shown at the nodes. Chromosomal neighborhood analyses

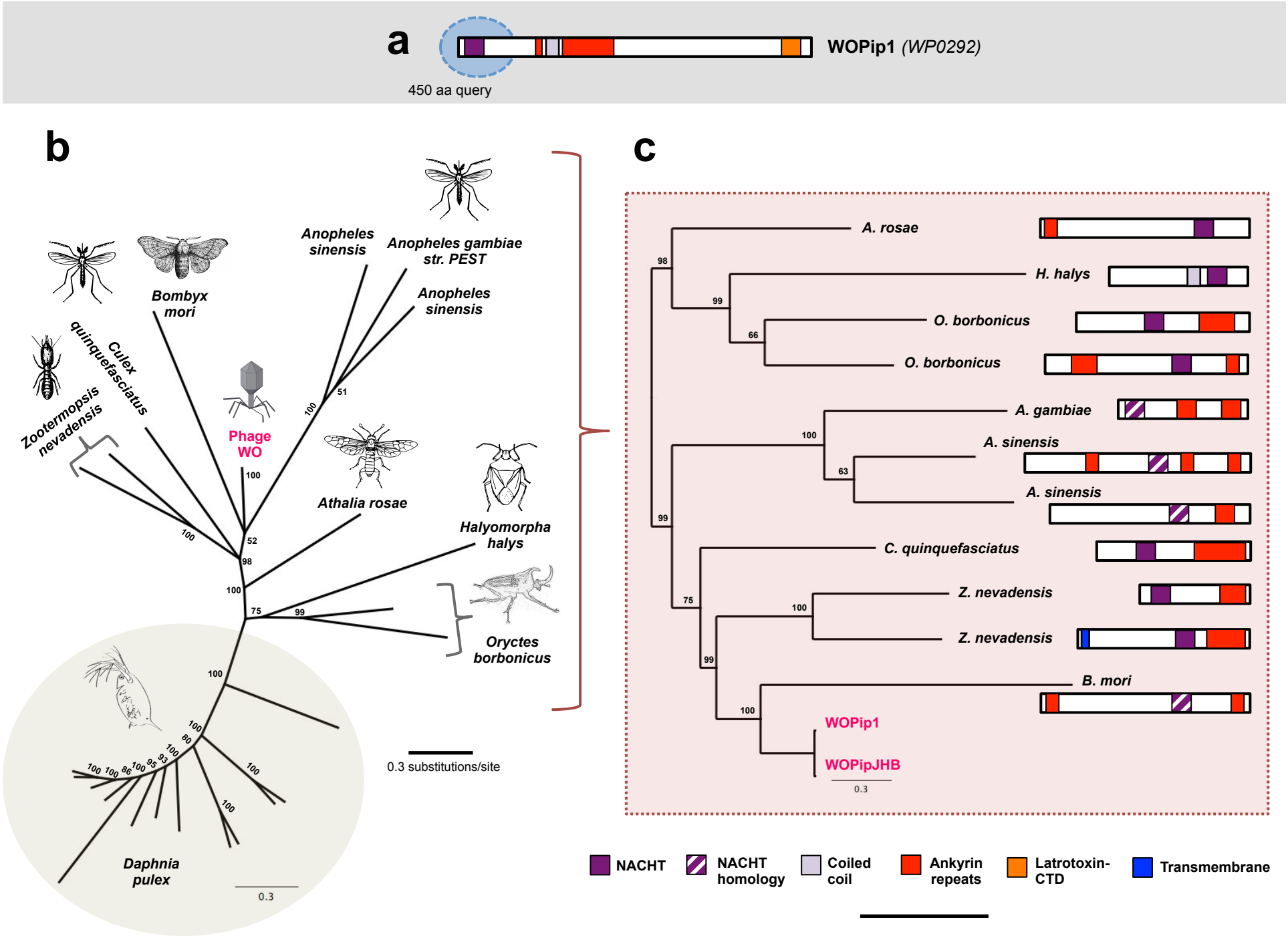
694 of available animal genome sequences indicate that animal homologs to the phage WO
695 protein are on contigs with other animal genes. Scale bar, 1,000 amino acids.
696
697 **Figure 5 | Phylogeny and protein architecture of the cell death domain, NACHT**
698 (a) A BLASTP query of prophage WO's NACHT region reveals homologs throughout
699 arthropods and crustaceans. (b) Bayesian phylogenetic trees were constructed based on a
700 271-aa alignment of all homologs with E-value less than e^{-15} and coverage greater than
701 70% using the cpREV+G model of evolution. To resolve taxa closest to prophage WO,
702 all *Daphnia* sequences were removed from the alignment and clusters of highly divergent
703 residues (i.e., 5 or more sequential residues with less than 15% pairwise identity) were
704 trimmed. Trees were reconstructed based on this 262-aa alignment using the LG+G
705 model of evolution. Consensus support values are shown at the nodes. Both trees are
706 unrooted. Chromosomal neighborhood analyses of available animal genome sequences
707 indicate that animal homologs to the prophage WO protein are on contigs with other
708 animal genes. Scale bar, 1,000 amino acids.
709
710 **Figure 6 | Models of lateral DNA transfer between eukaryotes and bacteriophages**
711 (a) The eukaryotic cell can harbor multiple microbes capable of horizontal gene transfer.
712 Genetic transfers between eukaryotes and bacteriophages can, in theory, occur (b)
713 directly between eukaryotic chromosomes and phage genomes; (c) indirectly between
714 eukaryotic and *Wolbachia* chromosomes; or (d) indirectly between eukaryotic
715 chromosomes and intermediary entities, such as eukaryotic viruses and other intracellular
716 bacteria.

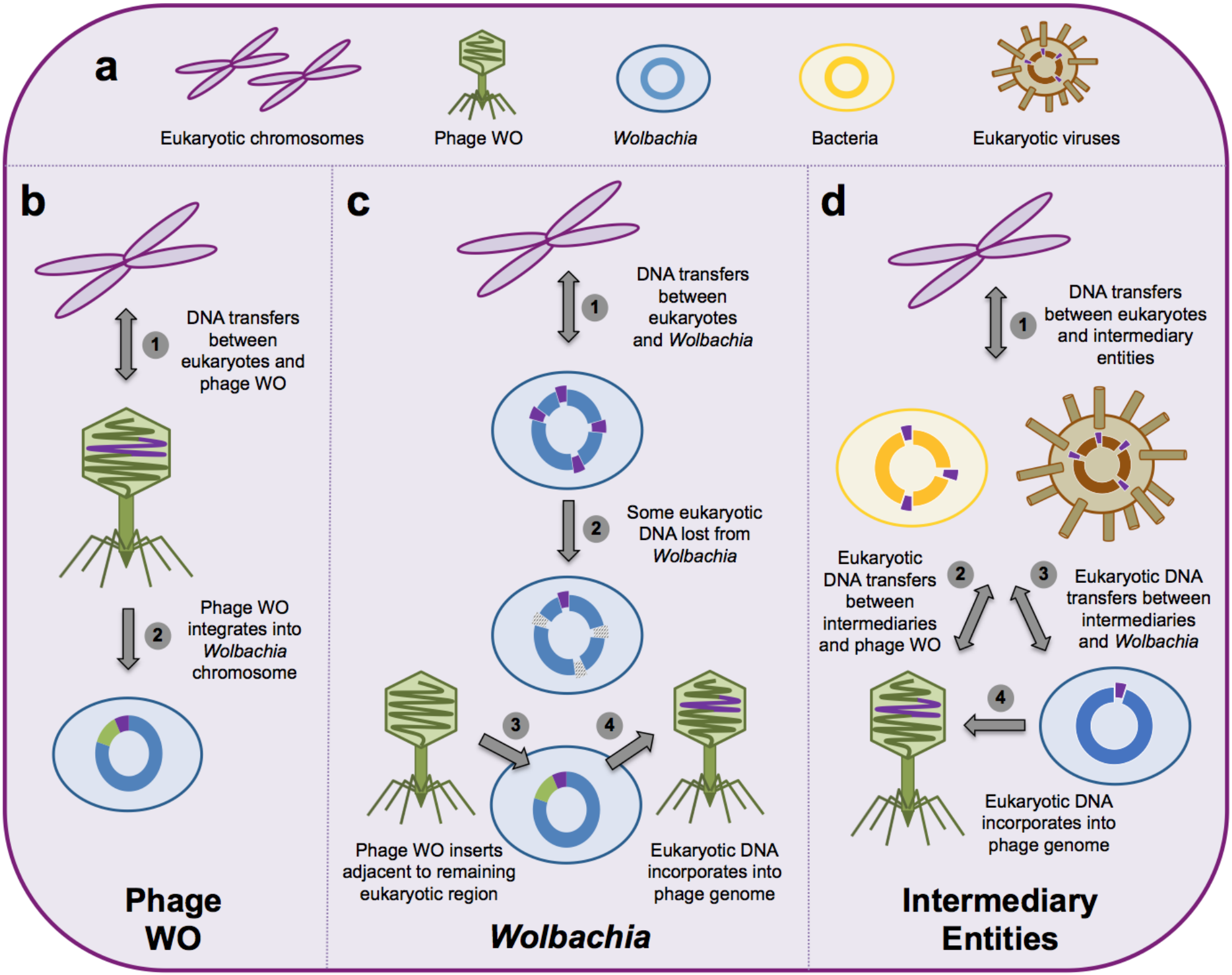

EAM
a
Methylase & ParB
Terminase
Portal
gpW
Minor Capsid
D
Major Capsid
Z
V
W
PAAR
J
Tail
I
Recombinase
attP
HSP
ANK-PRANC
ANK-TPR
MutL
DUF2466
Patatin
PLDc
Late Control
X
U
Tape Measure
G/GT
Tube
Sheath
VrlC.2
VrlC.1
Sigma 70
AAA_24
AAA_25
AAA_25
Primase/ RepA
HJR
Phage WOVitA1
b
1156
1150
1145
1140
1135
1125
1120
1115
1110
1105
1100
1095
1093
1089
attL
attR
Prophage WOVitA1
c
gp1
gp5
gp10
gp15
gp20
gp25
gp30
gp35
gp40
gp45
GF2gp17
Insecticidal ABC-toxin complex
GF2gp18
GF2gp25
ANK LATROTOXIN-CTD
attL
attR
Prophage WOCauB3
d
gp1
gp5
gp10
gp15
gp20
gp25
gp30
gp35
gp40
gp45
gp47
GF1gp22
Partial Sequence
Prophage WOCauB2
Baseplate
Head
Replication
Virulence & toxins
Tail
Ankyrin repeats
Transposases
Transcriptional regulators
Hypothetical proteins
Conserved domain of unknown function
Truncated / Interrupted genes
Sequenced att sites


a
Eukaryotic homology
Chromosomal
Phage-associated
Phage-associated
Chromosomal
Latrotoxin-CTD
8
2 *
PRANC
2
NACHT
1
gwv_1093
2
SecA1†
1
SecA2†
1
Octomom-NTD
2
b
Bacterial homology
Phage-associated
Chromosomal
Phage-associated
Chromosomal
ABC toxin
1
Ulp1
8
5
OTU
7
9
NB-ARC
3
4


a
Spider venom CTD
Phage WO CTD
L. tredecimguttatus
L. hesperus
L. hasselti
L. geometricus
S. grossa
Vertebrate-specific Toxins
L. hesperus*
LCTA
LITD
Invertebrate-specific Toxins
0.2 substitutions/site
HGT to mosquito
A. aegypti
WOAu
WOBol1-b(2)‡
Octomom
WOBol1-b(1)‡
WOAlbB(1)‡
WODi‡
WOOb‡
WOAlbB(2)‡
wNo†
WOPip5
WOPipMol(1)‡
WOPipJHB(1)
WOPip1
WOPip4(2)
WOPipMol(2)‡
WOPipJHB(2)
WOVitB1
WOPip4(1)
WOPipMol(3)‡
b
L. tredecimguttatus
L. hesperus
L. hasselti
S. grossa
L. geometricus
Vertebrate-specific (alpha-latrotoxin)
c
Alpha-latrocrustotoxin (LCTA)
Delta-latroinsectotoxin (LITD)
Invertebrate-specific
d
Octomom (WD0512)
A. aegypti (AAEL004181)
WOAu (WPWAU0629)
wNo† (wNo_10640 & wNo_10650; frameshift)
WOPip5 (WP1346)
WOPip JHB(1) (C1A_1255)
WOPip1 (WP0292)
WOPip4(1) (WP0407)
WOPipJHB(2) (C1A_54)
WOPip4(2) (WP0462)
Phage WO latrotoxin-CTD
Ankyrin repeats
Latrotoxin-CTD
Coiled coil
NB-ARC
TPR repeats
OTU
Transmembrane
Apolipophorin III
Deaminase
RHS repeats
NACHT
Furin-cleavage


a
WOVitA1 (gwv_1093)
800 aa query
b
N-terminal homologs
(137 aa)
Dendroctonus ponderosae
Diplorickettsia massiliensis
Planococcus citri
C. floridanum
N. vitripennis
M. demolitor
M. rotundata
M. rotundata
S. invicta
Phage WO
C. biroi
A. aegypti
C. quinquefasciatus
A. aegypti
A. gambiae
A. sinensis
A. aegypti
A. aegypti
A. sinensis
0.4 substitutions/site
c
Full-length homologs
(627 aa)
WOVitB
WOVitA1
Solenopsis invicta
WOCauB2
WOAnaC
WORiC
Cerapachys biroi
Megachile rotundata
M. demolitor
C. floridanum
N. vitripennis
0.2
Mosquitoes
Aedes aegypti
Anopheles sinensis
Culex quinquefasciatus
Anopheles gambiae str. PEST
Parasitic wasps
Microplitis demolitor
Copidosoma floridanum
Nasonia vitripennis
Ankyrin repeats
TPR repeats
Coiled coil
Sequencing gap


a
WOPip1 (WP0292)
450 aa query
b
Anopheles sinensis
Anopheles gambiae str. PEST
Anopheles sinensis
Bombyx mori
Culex quinquefasciatus
Zootermopsis nevadensis
Phage WO
Athalia rosae
Halyomorpha halys
Oryctes borbonicus
Daphnia pulex
0.3 substitutions/site
0.3
c
A. rosae
H. halys
O. borbonicus
O. borbonicus
A. gambiae
A. sinensis
A. sinensis
C. quinquefasciatus
Z. nevadensis
Z. nevadensis
B. mori
WOPip1
WOPipJHB
0.3
NACHT
NACHT homology
Coiled coil
Ankyrin repeats
Latrotoxin-CTD
Transmembrane


a
Eukaryotic chromosomes
Phage WO
Wolbachia
Bacteria
Eukaryotic viruses
b
1
DNA transfers between eukaryotes and phage WO
2
Phage WO integrates into Wolbachia chromosome
Phage WO
c
1
DNA transfers between eukaryotes and Wolbachia
2
Some eukaryotic DNA lost from Wolbachia
3
4
Phage WO inserts adjacent to remaining eukaryotic region
Eukaryotic DNA incorporates into phage genome
Wolbachia
d
1
DNA transfers between eukaryotes and intermediary entities
Eukaryotic DNA transfers between intermediaries and phage WO
2
3
Eukaryotic DNA transfers between intermediaries and Wolbachia
4
Eukaryotic DNA incorporates into phage genome
Intermediary Entities