Lateral Genetic Transfers Between Eukaryotes and Bacteriophages

Sarah R. Bordenstein\textsuperscript{a} and Seth R. Bordenstein\textsuperscript{a,\textast}}

Affiliations:

\textsuperscript{a}Department of Biological Sciences, Vanderbilt University, Nashville, TN 37232, USA.

\textsuperscript{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 eukaryotic 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\textsuperscript{1,2}. Infecting organisms across the tree of life, they associate with every ecosystem on the planet\textsuperscript{3}. They are generally classified into polythetic groups according to ecological niche and mode of replication\textsuperscript{4,5}. While any cellular domain can be infected by a virus, no extant virus is known to traverse more than one domain\textsuperscript{6,7}. This domain-specific ecology of viruses underpins the current taxonomic paradigm of trifurcating viruses into eukaryotic, archaenal and bacterial categories, along with recent reappraisals of whether viruses constitute a fourth domain of life\textsuperscript{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\textsuperscript{11,12}, with the exception of some giant viruses that appear to acquire genes from all domains of life\textsuperscript{13}. Bacterial viruses, or bacteriophages (phages), only integrate genetic material from their bacterial hosts including toxin\textsuperscript{14}, photosynthesis\textsuperscript{15} and pigment biosynthesis genes\textsuperscript{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 \textit{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\textsuperscript{17,18}, primarily infect the cells of host reproductive tissues, exist in Golgi-derived vesicles within the
eukaryotic cytoplasm, and are enclosed by a bacterial cell membrane and one or more
eukaryotic-derived membranes\(^{19-21}\). Nearly all sequenced *Wolbachia* genomes, with the
exception of those acting as obligate mutualists, harbor prophage WO\(^{22-24}\). They encode
conserved structural modules (e.g., head, tail, baseplate) and exhibit *Caudovirales*
morphology in electron micrographs of purified phages\(^{23,25-30}\). Electron microscopy and
quantitative analyses indicate that prophages undergo a lytic phase capable of rupturing
bacterial and eukaryotic cell membranes, and phage WO occurs in the extracellular
matrix of arthropod gonads\(^{27,31}\). Therefore, phage WO appears to uniquely contend with
the cellular exit, entry and defense mechanisms of two separate domains of life. WO is
also a promising tool for genome editing of *Wolbachia* that has thus far been refractory to
genetic modification. Until now, the genomes of bacteriophage WO particles have not
been fully sequenced and assembled into circular genomes, and their attachment sites and
bacterial integration sites are unresolved.

**RESULTS**
Here we report the first metagenomic analysis of phage WO particles from \(w\)VitA-
infected *Nasonia giraulti* wasps and \(w\)CauB-infected *Ephestia kuehniella* moths. We
identify the phage attachment sites and insertion regions and show from fully sequenced
genomes that WO harbor all formerly described phage genetic modules (lysogeny,
baseplate, head, replication, virulence, tail and patatin\(^{32}\)) as well as a new group of genes
with atypical protein domains indicative of eukaryotic interaction. We collectively group
these genes, which include one of the largest genes in bacteriophages to date, into a novel
"Eukaryotic Association Module" (EAM, white box, Fig. 1). The EAM features genes
that (i) encode protein domains and cleavage sites central to eukaryotic functions, (ii) 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 \( w^{\text{Cle}} \) and filarial nematode-infecting \( w^{\text{Bm}} \) and \( w^{\text{Oo}} \). 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 (hypothesis 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 Information Table 1).
Information Fig. 1e). The full length of the completely assembled circular WOVitA1 is 66,688 bp, which is 48% larger than any previous prophage WO annotation. Similarly, we identified the new terminal ends of the WOCauB3 phage (23,099 bp (51%) larger than original estimate of 45,078 bp) along with internal localization of the EAM genes by Sanger sequencing its attP site [Domain of Unknown Function (DUF)2426 to recombinase]. While we were not able to assemble a complete contig for WOCauB2, it is more than 12,000 bp larger than the original estimate of 43,016, includes multiple ankyrin repeat genes homologous to those in WOVitA1, and, like many other phage haplotypes (e.g., WORiC, WOVitA2, WOSuziC), integrates directly into Wolbachia’s magnesium chelatase (chll) gene.

We then analyzed each phage WO protein domain for homology and surrounding peptide architecture. Unlike the single domain architecture of phage WO’s structural genes, EAM genes are highly polymorphic and encompass fusions of both eukaryotic and bacterial protein domains. By extending the analysis to include homologous prophage regions from all sequenced Wolbachia chromosomes, ten types of protein domains with putative eukaryotic functions were revealed spanning four predicted functions: (i) toxins, (ii) host-microbe interactions, (iii) host cell suicide, and (iv) secretion of proteins through the cell membrane (Fig. 2). Notably, over half of these domain types (6/10; latrotoxin-CTD, PRANC, NACHT, SecA, gwv_1093-NTD, Octomom-NTD) share greater amino acid homology to eukaryotic invertebrates than to bacteria in GenBank. Among this subset with eukaryotic sequence homology, the protein domains are almost exclusively found in the EAM region (N=17) versus the Wolbachia chromosome (N=2). This pattern differs
from other EAM protein domains with bacterial homology, which are equally dispersed in phage WO (N=19) and the Wolbachia chromosome (N=18) (Fig. 2, Fisher’s Exact Test, p = 0.0072). This difference importantly indicates that the eukaryotic-like protein domains are highly enriched in the EAM, suggesting a near exclusive role in phage WO biology.

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

Latrotoxin C-terminal domain (CTD) is the most prevalent eukaryotic domain in phage WO. Originally described for its major role in the venom of widow spiders (*Latrodectus* species), latrotoxins cause the formation of membrane pores in their vertebrate or
invertebrate victims. Phylogenetic analysis indicates that the latrotoxin-CTD horizontally transferred between widow spiders and phage WO (Fig. 3). In addition, reciprocal search queries using homologous spider and phage CTDs return the same BLASTp hits shown in Fig. 3. These taxa occur in overlapping ecological niches (*Wolbachia* are known to infect spiders of the family *Theridiidae*) in which gene transfers are more likely to happen\textsuperscript{36,37}. We also confirmed the presence of *Wolbachia* in three independent *Latrodectus geometricus* samples by amplifying *Wolbachia* 16S rDNA and *wsp* membrane protein genes. The transfer event was apparently followed by a relatively more recent transfer from phage WO back to animals in the *Aedes aegypti* genome where the region is located between genes of mosquito origin [fibrinogen-related protein (AAEL004156) and GalE3 (AAEL004196)], or *A. aegypti* was the putative donor of the domain to phage WO, followed by a recursive transfer to black widow spiders.

Latrotoxin-CTD is universally located at the 3’-terminal ends of both conserved spider latrotoxin genes\textsuperscript{38} and enormous, polymorphic, and eukaryotic-like phage WO genes (up to 14,256 bp). Notably, phage WO CTD sequences have the highest amino acid similarity to black widow spider homologs that target invertebrates, which are the primary hosts of *Wolbachia*. There is also a high incidence of eukaryotic furin cleavage sites that immediately precede the latrotoxin-CTD. In spiders, cleavage at these sites by the eukaryotic furin protease in the trans-Golgi network or extracellular matrix is required for latrotoxin activation before the toxin exerts its effects upon the victim\textsuperscript{39-41}. We show that all phage WO EAMs contain at least one site for eukaryotic furin cleavage (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 cleavage 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.
Domains central to modifying animal proteins are also abundant in the phage EAM. The Pox protein repeats of ANkyrin C terminus (PRANC) domain in the WOVitA1 genome (gwv_1092) shares protein sequence homology with corresponding PRANC domains in multiple parasitic wasp hosts (Supplementary Information Table 3) and their eukaryotic viruses. Reciprocal BLASTp searches retrieve the same best hits and support previous findings that this protein domain horizontally transferred between eukaryotic viruses, animals, and Proteobacteria. The discovery here of the eukaryotic-like PRANC domain in phage WO parallels its presence in the Poxviridae virus family, in which it functions in evasion of eukaryotic immune responses via modification of host ubiquitination. PRANC is related to amino acid sequences in F-box proteins, which are eukaryotic proteins involved in protein degradation. The PRANC domain also occurs in vaccinia virus, ectromelia virus, cowpox virus and Orf virus and can regulate NF-κB signalling pathway to inhibit transcription of inflammatory cytokines.

Adjacent to the PRANC-encoding gene in WOVitA1 is an ankyrin and tetratricopeptide repeat (TPR)-containing gwv_1093. Ankyrin repeats and TPRs mediate a broad range of protein-protein interactions (apoptosis, cell signaling, inflammatory response, etc.) within eukaryotic cells and are commonly associated with effector proteins of certain intracellular pathogens. While generally rare in viral genomes (Supplementary Information Fig. 2 and 3, respectively), they occur in all phage WO haplotypes from sequenced Wolbachia genomes (N=23). Phylogenetic analysis using reciprocal BLASTp hits (Fig. 4) shows that the N-terminus sequences of the TPR-containing gwv_1093 is 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
transduction molecules of the innate immune system. The polymorphic phage WO homolog encodes ankyrin repeats and a latrotoxin-CTD directly downstream from the conserved NTPase domain (Fig. 5a). NACHT domains have been identified in animals, fungi and bacteria and phylogenetic patterns indicate multiple instances of horizontal transfer. A NACHT-containing peptide was recently discovered in the Clostridium difficile-infecting phage phiCDHM1 genome although, in contrast to phage WO, the phiCDHM1 NACHT domain is bacterial in both amino acid homology and protein architecture. Similar to the phylogeny of the N-terminus of the TPR-containing gwv_1093, the NACHT domain sequence in phage WO is embedded within, and likely derived by horizontal transfer from, a deeper and more diverse set of ancestral variants in arthropods (Fig. 5b,c).
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,
these animal protein domains are central to anti-eukaryotic functions including the black widow latrotoxin, programmed cell death (NACHT), immune evasion (PRANC), and protein-protein interactions. They have never before been reported in bacteriophage genomes because phages have naturally been overlooked as recipients of eukaryotic DNA.

Bacteriophage WO frequently transfer between Wolbachia coinfections in the same animal host and to the host genome as part of large transfers of the Wolbachia chromosome. We previously reported that they were also capable of transferring adjacent flanking non-phage genes in the process of transfer between coinfections. For two of these flanking genes, sequence evidence indicated that Wolbachia genomes may be able to receive eukaryotic DNA. However, the nature of these lateral genetic transfers remained to be validated and elucidated as these regions were not previously known to be part of the packaged phage genome until now. Based on this work, we suspect that systematic surveys of phage genomes in intimate host-associated bacteria may uncover a broad range of eukaryotic protein domains involved in phage lifecycle adaptations and phage-eukaryote interactions.

The mechanisms by which eukaryotic protein domains integrate into phage WO are unknown, and could follow at least two models. First, animal genetic material could directly transfer to WO genomes during phage particle packaging in the cytoplasm of animal cells (Fig. 6a) or inside Wolbachia cells that are lysing and exposed to the eukaryotic cytoplasmic environment. Packaging of eukaryotic host RNAs, for instance,
occur in the virions of herpesvirus\textsuperscript{64,65} and cytomegalovirus\textsuperscript{66}. Second, genes may transfer from animal genomes to the \textit{Wolbachia} chromosome and then to prophage WO. However, for this scenario to be plausible, animal genetic material transferred presumably in random locations in the \textit{Wolbachia} genome would have to be preferentially lost in non-phage associated domains from the \textit{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).

\textbf{a. Direct DNA transfer}

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{direct_dna_transfer}
\caption{Direct DNA transfer between eukaryotes and \textit{Wolbachia} and prophage WO.}
\end{figure}

\textbf{b. Indirect DNA transfer}

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{indirect_dna_transfer}
\caption{Indirect DNA transfer between eukaryotes and \textit{Wolbachia} and prophage WO.}
\end{figure}

\textbf{Figure 6 | Models of DNA transfer between eukaryotes and bacteriophages}
Why are these protein domains present in the EAM of bacteriophage WO? Phages of obligate intracellular bacteria are contained within both bacterial and eukaryotic membranes and can possess an enigmatic "two-fold cell challenge". They may not only have to breach peptidoglycan and permeabilize bacterial membranes, but they may also have to exit (and enter) across the eukaryotic membrane(s) that directly encapsulates the bacteria. Functional studies of homologous domains (i.e., PRANC and NACHT) suggest that these proteins could have eukaryotic viral-like properties that are deployed in processes such as the lysis of eukaryotic cells and post-translational modification of host proteins. Phage WO can dwell in the eukaryotic cytoplasm and extracellular matrix that they encounter upon bacterial lysis, raising the possibility of direct interaction with the host's biology.

*Chlamydiomicroviridae* infect obligate intracellular bacteria, yet still do not directly contend with the eukaryotic membrane. Rather, they attach to dormant chlamydial cells (i.e., reticulate bodies) and enter via phagocytosis or endocytosis of the bacteria. The phages then alter development of their bacterial host, which leads to disintegration of the chlamydial inclusion and subsequent lysis of the eukaryotic host cell. The nature of phage WO’s lifestyle, on the other hand, may require a distinct interaction with multiple membranes and immune responses because lytic activity of phage WO has been associated with typical bacterial cell defects including degraded bacterial DNA, a detached inner membrane, and exit of the phage particles from inside *Wolbachia* and its host cell into the extracellular matrix of the reproductive tissues. Bacteriophages of free-living bacteria also regularly colonize eukaryotic environments, particularly those
associated with mucosal surfaces\textsuperscript{72}. They, however, do not infect or traverse the eukaryotic membrane and are still within the genomic boundaries of the bacterial virosphere.

Temperate dsDNA phages also occur in facultative symbionts of aphids\textsuperscript{73} and tsetse flies\textsuperscript{74}. While \textit{Wolbachia} has never successfully been cultured outside of host cells\textsuperscript{75}, these facultative symbionts can replicate both intra- and extracellularly (JW Brandt, personal communication, July 2015;\textsuperscript{76}) suggesting that their phages are not constrained by the same two-fold cell challenge. In addition, their phages encode a traditional lytic cassette (holin and lysozyme) that correlates with the need to deal only with bacterial membranes. In some cases, the phages harbor bacterial-derived toxins that target eukaryotic cells\textsuperscript{77}, and these function mutualistically in aphids by arresting parasitoid wasp larvae\textsuperscript{73}. Furthermore, unlike phage WO, these phages are readily lost in the absence of parasitoids during laboratory rearing, presumably due to the cost of their toxins\textsuperscript{78}.

In addition to providing new insights into the evolution of bacteriophages and showing phage WO to be far more complex than previously described, the findings here reveal that phage evolution in \textit{Wolbachia} leads to a novel example of phage-metazoan genomic chimerism. Acquisition and retooling of intact eukaryotic domains in phage WO appears to be analogous to the commandeering of host genes by eukaryotic viruses. Whether this newly discovered highway of lateral genetic transfer is common in the symbiotic virosphere remains to be determined.
Insect and Bacterial Strains

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

Phage Particle Purification

Phage particles were isolated according to Fujii et al with modifications. Approximately 4 g of adult insects were homogenized in 29.6 ml cold SM buffer (50mM Tris-HCl, pH 7.5, 0.1 M NaCl, 10mM MgSO₄·7H₂O, and 0.1% (w/v) gelatin). NaCl and RNase A were added to a final concentration of 1M and 1ug/ml, respectively. The homogenate was incubated on a shaker at 4°C for 1 h and then centrifuged at 13,000g for 10 min at 4°C. Polyethylene glycol (PEG) 6000 was added to a final concentration of 10% to precipitate phage particles, incubated at 4°C for 1 hr with gentle shaking and centrifuged at 13,000g for 10 min. The pellet was resuspended in 5 ml TM buffer (50 mM Tris-HCl, pH 7.5, 10 mM MgCl₂·6H₂O) and mixed with an equal volume chloroform. The suspension was centrifuged at 3,000g to remove PEG and the aqueous phase was filtered through a 0.22
um filter to remove bacterial cells. The suspension was centrifuged at 60,000g for 1 h at 4°C to collect phage particles. The pellet was suspended in 10 μl TM buffer.

**Phage DNA Extraction & Metagenomic Sequencing**

The phage suspension was treated with RQ1 RNase-Free DNase (Promega) for 30 min at 37°C, followed by heat inactivation for 10 min at 65°C, to remove host DNA contamination. Phage DNA was extracted from the suspension using the QIAamp MinElute Virus Spin Kit (Qiagen) and amplified using the REPLI-g Mini Kit (Qiagen).

Following amplification, paired-end DNA libraries were prepared according to manufacturer’s (Illumina) instructions and samples were sequenced with an Illumina HiSeq 2000 (2×100-nt read length).

**Bioinformatics & Statistics**

Metagenomic sequences (reads) were trimmed, paired and assembled into contigs using the CLC Assembler (CLC bio) with bubble size = 50, insertion and deletion cost = 3, mismatch cost = 2, length fraction = 0.6, minimum contig size = 130, similarity = 0.5, minimum distance = 90 and maximum distance = 200. Contigs were compared to the GenBank non-redundant database using NCBI’s BLASTN (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and those with similarity to phage WO and/or *Wolbachia* (E-value <10^-10) were manually annotated using Geneious (Biomatters Ltd.). Individual reads were mapped to reference sequences using Geneious. Open reading frame (ORF) homology searches were performed to determine putative function using NCBI’s BLASTP (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and Wellcome Trust Sanger
Institute’s pfam database (http://pfam.sanger.ac.uk). Coiled coil domains were predicted with EMBL’s Simple Modular Architecture Research Tool (SMART, http://smart.embl-heidelberg.de). Furin cleavage sites were identified using PiTou (http://www.nuolan.net/reference.html). The number of genes with and without furin cleavage sites was analyzed with respect to phage-region using Fisher’s Exact Test (GraphPad Software). Phylogenetic trees were built using the Bayes plugin in Geneious and model selection for each Bayes analysis was estimated using ProtTest.

Confirmation of Phage WO Terminal Genes

Genomic DNA was extracted from wVitA-infected N. vitripennis (strain 12.1) and wCauB-infected E. kuehniella individuals using the Gentra Puregene Tissue Kit (Qiagen). Primers were designed for both WOVitA1 and WOCauB3 att sites, respectively:

VitA1_attF (5’- CGA AGA ACC AGC ACA GGG TGG -3’), VitA1_attR (5’- GCT GGA AGA GGG CAT CTG CAT C -3’), CauB3_attF (5’- TCG TGA CTG CCC TAT TGC TGC T – 3’) and CauB3_attR (5’- ATG CGG CCA AAG CTG GGT GT -3’).

Amplification was performed in a Veriti thermal cycler (Applied Biosystems) using GoTaq green master mix (Promega) under the following conditions: 94C for 2 min; 35 cycles of 94C for 30 s, 53C for 30 s, 72C for 1 min; and a final elongation cycle of 72C for 10 min. PCR products were sequenced via Sanger sequencing (Genewiz, Inc).
REFERENCES


**FIGURE LEGENDS**

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

**Figure 2 | Eukaryotic-like EAM genes are enriched in prophage WO regions in the Wolbachia chromosome.** EAM genes with (a) eukaryotic homology are most likely to be associated with prophage WO while those with (b) bacterial homology are both phage-associated and found scattered throughout the *Wolbachia* chromosome. (*) The two chromosomal latrotoxin-CTD domains (wNo_10650 and wHa_05390) are located within phage-associated genes and transposases, indicating a potential genomic rearrangement. (†) SecA represents one “domain type” but is listed separately because phage WO contains two different homologs (i.e., wHa_3920 and wHa_3930). Putative functional categories are: anti-eukaryotic toxins (orange); host-microbe interactions (green); host cell suicide (blue); secretion of virulence factors (pink); and unknown (black). Octomom refers to WD0513 of the wMel genome.
Figure 3 | Latrotoxin-CTD phylogeny and protein architecture reveal lateral genetic transfers between black widow spiders and bacteriophage WO.

(a) Phylogeny of phage WO latrotoxin-CTD protein domains and their eukaryotic homologs was constructed by Bayesian analysis of 74 amino acids using the JTT model of evolution. Consensus support values are shown at the nodes. Comparative protein architecture shows that spider venom (b) vertebrate-specific alpha-latrotoxins and (c) invertebrate-specific alpha- and delta-latrotoxins are highly conserved, whereas (d) phage WO are not. Bolded nomenclature in (d) denotes the specific phage WO haplotype (listed as WO). Genome locus tags are listed in parentheses. Predicted furin cleavage sites, listed in Supplementary Information Table 2, are illustrated with gray triangles. (*) A second *L. hesperus* sequence represents a recently-described downstream paralog with unknown toxin activity. (†) *wNo_10650* is located within phage-associated genes and transposases, indicating a potential genomic rearrangement of a phage region. (‡) Architecture is not shown for sequences on incomplete contigs (WOBol1-b, WOAlbB, WOCit, WOPipMol, WOVitB) because complete peptide information and specific phage association are unknown.

Figure 4 | A conserved TPR and anyrin-repeat protein horizontally transferred from eukaryotes to bacteriophage WO. (a) An 800-aa BLASTp query of WOVitA1’s gwv_1093 N-terminus reveals homologs throughout moquitoes, ants, beetles, a mealybug and one obligate intracellular gammaproteobacteria. Bayesian phylogenetic trees were constructed based on (b) a 137-aa alignment of all homologs with E-value less than $e^{-40}$ using the LG +G model of evolution. (c) To resolve taxa closest to phage WO, trees were
reconstructed based on a 627-aa alignment of all homologs with an E-value of 0 using the JTT +I+G model of evolution. Isoforms were removed from each alignment. Both trees are unrooted. Consensus support values are shown at the nodes. Chromosomal neighborhood analyses of available animal genome sequences indicate that animal homologs to the phage WO protein are on contigs with other animal genes.

**Figure 5 | The programmed cell death domain, NACHT, horizontally transferred from eukaryotes to bacteriophage WO.** (a) A 450-aa BLASTp query of phage WO’s NACHT region reveals homologs throughout arthropods and crustaceans. (b) Bayesian phylogenetic trees were constructed based on a 271-aa alignment of all homologs with E-value less than $e^{-15}$ and coverage greater than 70% using the cpREV+G model of evolution. To resolve taxa closest to phage WO, all *Daphnia* sequences were removed from the alignment and clusters of highly divergent residues (i.e., 5 or more sequential residues with less than 15% pairwise identity) were trimmed. Trees were reconstructed based on this 262-aa alignment using the LG +G model of evolution. Consensus support values are shown at the nodes. Both trees are unrooted. Chromosomal neighborhood analyses of available animal genome sequences indicate that animal homologs to the phage WO protein are on contigs with other animal genes.

**Figure 6 | Models of DNA transfer between eukaryotes and bacteriophages.** Genetic transfers between eukaryotes and bacteriophages can occur either (a) directly via incorporation into the phage genome followed by subsequent inclusion in the chromosomal prophage region or (b) indirectly via the transfer of eukaryotic DNA to the
bacterial chromosome or vice versa. Since phage EAM genes carrying protein domains central to eukaryotic functions primarily occur in phage/prophage genomes (see Fig. 2), transferred DNA from eukaryotes to non-phage regions in *Wolbachia* is likely eliminated from the bacterial genome. Prophage genomes adjacent to these EAM genes then incorporate the DNA into their packaged genomes and pass it on to new copies of the phage.
**SUPPLEMENTARY INFORMATION**

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

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

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

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

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

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

**ACKNOWLEDGEMENTS**

We thank Tetsuhiko Sasaki and Takema Fukatsu for providing the *wCauB*-infected moths; Rini Pauly for computational assistance; Merri Lynn Casem for providing widow spiders; and Kerry Oliver for information on *Hamiltonella*. This research was funded by NIH Awards R01 GM085163 and R21 HD086833 and NSF Awards DEB 1046149 and IOS 1456778 to SRB. The funders had no role in study design, data collection and interpretation or the decision to submit the work for publication.

**AUTHOR CONTRIBUTION**

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

**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\textsuperscript{a} and Seth R. Bordenstein\textsuperscript{* \textit{a,b}}

Affiliations:

\textsuperscript{a}Department of Biological Sciences, Vanderbilt University, Nashville, TN 37232, USA.

\textsuperscript{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., wAu, wMel, and wRi).
Bacteriophages

Ankyrin-containing protein records in NCBI Length (# aa residues)

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

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.
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.
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; wMel - AE017196; wPip - AM999887; wRI - CP001391; wVitA - PRJDB1504; wHa - CP003884; wNo - CP003883.

<table>
<thead>
<tr>
<th>GENOME</th>
<th>CORE PHAGE REGION</th>
<th>EAM REGION</th>
<th>(WD0611-WD0621)-like</th>
</tr>
</thead>
<tbody>
<tr>
<td>WOVitA1</td>
<td>gwv_1104-gwv_1156</td>
<td>gwv_1089-gwv_1103</td>
<td>na</td>
</tr>
<tr>
<td>WOVitA4</td>
<td>gwv_426-gwv_459</td>
<td>gwv_458-gwv_472; gwv_484-gwv_496</td>
<td>gwv_473-gwv_483</td>
</tr>
<tr>
<td>WOCauB2</td>
<td>gp1-gp45</td>
<td>gp46-(partial sequence)</td>
<td>na</td>
</tr>
<tr>
<td>WOCauB3</td>
<td>gp1-gp44</td>
<td>gp45-GF2gp25</td>
<td>na</td>
</tr>
<tr>
<td>WOSol</td>
<td>So0001-So0022</td>
<td>wSo0003(T)-wSo0014; So0023-So0025; So0026-wSo0028(T)</td>
<td>wSo0015-wSo0026</td>
</tr>
<tr>
<td>WOMelA</td>
<td>WD0259-WD0288</td>
<td>WD0289-WD0296(T); WD0253(T)-WD0258</td>
<td>na</td>
</tr>
<tr>
<td>WOMelB</td>
<td>WD0634-WD0647(T); WD0563(T)-WD0604</td>
<td>WD0605-WD0610; WD0622-WD0633</td>
<td>WD0611-WD0621</td>
</tr>
<tr>
<td>wMel WO-Island</td>
<td>na</td>
<td>WD0507-WD0514</td>
<td>na</td>
</tr>
<tr>
<td>WOPip1</td>
<td>WP0236(T)-WP0272</td>
<td>WP0273-WP0293</td>
<td>na</td>
</tr>
<tr>
<td>WOPip2</td>
<td>WP0297-WP0322</td>
<td>WP0294-WP0296</td>
<td>na</td>
</tr>
<tr>
<td>WOPip3</td>
<td>WP0332-WP0342</td>
<td>WP0343(T)-WP0354</td>
<td>na</td>
</tr>
<tr>
<td>WOPip4</td>
<td>WP0411-WP0455</td>
<td>WP0405(T)-WP0410; WP0456-WP0465</td>
<td>na</td>
</tr>
<tr>
<td>WOPip5</td>
<td>WP1294(T)-WP1340</td>
<td>WP1289(T)-WP1323; WP1341-WP1352(T)</td>
<td>na</td>
</tr>
<tr>
<td>WORA</td>
<td>wRi_012450-wRi_012680(T)</td>
<td>na</td>
<td>na</td>
</tr>
<tr>
<td>WORB</td>
<td>wRi_005400-wRi_005660</td>
<td>wRi_005310(T)-wRi_005390; wRi_005670-wRi_005720; wRi_005860-wRi_005930(T)</td>
<td>wRi_005730-wRi_005830</td>
</tr>
<tr>
<td>WORB(2)</td>
<td>wRi_010060-wRi_010320</td>
<td>wRi_009980(T)-wRi_010050; wRi_010330-wRi_010380; wRi_010500-wRi_010580(T)</td>
<td>wRi_010390-wRi_010490</td>
</tr>
<tr>
<td>WORC</td>
<td>wRi_006880-wRi_007230(T); wRi_007550-wRi_007860(T)</td>
<td>wRi_006620-wRi_006870</td>
<td>na</td>
</tr>
<tr>
<td>WOHa1</td>
<td>wHa_02360-wHa_02620</td>
<td>wHa_02010(T)-wHa_02050; wHa_02170-wHa_02350; wHa_02630-wHa_02760(T)</td>
<td>wHa_02060-wHa_02160</td>
</tr>
<tr>
<td>WOHa2</td>
<td>wHa_03390-wHa_03840</td>
<td>wHa_03850-wHa_03930</td>
<td>na</td>
</tr>
<tr>
<td>WONO1</td>
<td>wNo_01400-wNo_01060</td>
<td>wNo_01000(T)-wNo_01050(T)</td>
<td>na</td>
</tr>
<tr>
<td>WONO2</td>
<td>wNo_07170(T)-wNo_07380(T)</td>
<td>wNo_01090(T)-wNo_01050(T)</td>
<td>na</td>
</tr>
<tr>
<td>WONO3</td>
<td>wNo_09000-wNo_09120</td>
<td>wNo_09130-wNo_09150(T)</td>
<td>na</td>
</tr>
<tr>
<td>WONO4</td>
<td>wNo_10070(T)-wNo_10280</td>
<td>wNo_10290-wNo_10350</td>
<td>na</td>
</tr>
<tr>
<td>wNo WO-Islands</td>
<td>na</td>
<td>wNo_01980-wNo_1990; wNo_2110-wNo_02200; wNo_05070-wNo_05150(T)</td>
<td>wNo_2000-wNo_2100</td>
</tr>
<tr>
<td>GENOME</td>
<td>EAM FURIN CLEAVAGE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>-------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WOCauB2</td>
<td>(partial sequence)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WOCauB3</td>
<td>gp45, GF2gp18-GF2gp22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WOSol</td>
<td>wSo0011, So0023, So0025</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WOMEIA</td>
<td>WD0257-WD0258</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WOMEIB</td>
<td>WD0610, WD0630-WD0632</td>
<td></td>
<td></td>
</tr>
<tr>
<td>wMel WO-Island</td>
<td>WD0652-WD0654</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WOPip1</td>
<td>WP0280, WP0283, WP0290, WP0292-WP0293</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WOPip2</td>
<td>WP0294, WP0319-WP0320</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WOPip3</td>
<td>WP0338</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WOPip4</td>
<td>WP0407, WP0410, WP0428, WP0433, WP0457, WP0460, WP0462-WP0463</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WOPip5</td>
<td>WP1291, WP1341, WP1344, WP1346, WP1349, WP1351</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WORIA</td>
<td>(haplotype does not have an EAM)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WORIB</td>
<td>wRi_005330, wRi_005360, wRi_005670, wRi_005720</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WORIB(2)</td>
<td>wRi_009990, wRi_010020, wRi_010330, wRi_010380, wRi_010570</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WORIC</td>
<td>wRi_006630-wRi_006660, wRi_006740</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WOVIA1</td>
<td>gwv_1095</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WOVIA2</td>
<td>gwv_464, gwv_484, gwv_489, gwv_491-gwv_495</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WOVIA4</td>
<td>gwv_134, gwv_141-gwv_142, gwv_144-gwv_145</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WOHa1</td>
<td>wHa_02170, wHa_02200, wHa_02290, wHa_02350, wHa_02690, wHa_02730</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WOHa2</td>
<td>wHa_03920</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WOno1</td>
<td>wNo_01030, wNo_01060</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WOno2</td>
<td>(haplotype does not have an EAM)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WOno3</td>
<td>wNo_09000, wNo_09080, wNo_09140</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WOno4</td>
<td>wNo_10290, wNo_10320-wNo_10340</td>
<td></td>
<td></td>
</tr>
<tr>
<td>wNo WO-Islands</td>
<td>wNo_01990, wNo_02030, wNo_02070, wNo_02120, wNo_02130, wNo_05080-wNo_05090, wNo_05110-wNo_05130</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**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; wMel – AE017196; wPip – AM999887; wRi – CP001391; wViA – PRJDB1504; wHa – CP003884; wNo – CP003883.
<table>
<thead>
<tr>
<th>EUKARYOTIC HOMOLOG</th>
<th>ACCESSION</th>
<th>E-VALUE</th>
<th>QUERY COVERAGE</th>
<th>IDENTITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microplitis demolitor</td>
<td>XP_014298115.1</td>
<td>8.00E-43</td>
<td>75%</td>
<td>49%</td>
</tr>
<tr>
<td>Nasonia vitripennis</td>
<td>XP_003426146.1</td>
<td>5.00E-23</td>
<td>76%</td>
<td>37%</td>
</tr>
<tr>
<td>Glypta fumiferanae</td>
<td>AKD28025.1</td>
<td>4.00E-21</td>
<td>71%</td>
<td>39%</td>
</tr>
<tr>
<td>Trichogramma pretiosum</td>
<td>XP_014232168.1</td>
<td>2.00E-16</td>
<td>73%</td>
<td>33%</td>
</tr>
<tr>
<td>Ceratosolen solmsi marchali</td>
<td>XP_011505281.1</td>
<td>8.00E-15</td>
<td>69%</td>
<td>31%</td>
</tr>
<tr>
<td>Copidosoma floridanum</td>
<td>XP_014206311.1</td>
<td>3.00E-12</td>
<td>58%</td>
<td>32%</td>
</tr>
<tr>
<td>Diaphorina citri</td>
<td>XP_008470724.1</td>
<td>9.00E-10</td>
<td>49%</td>
<td>31%</td>
</tr>
</tbody>
</table>

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.