Evolution of *Wolbachia* mutualism and reproductive parasitism: insight from two novel strains that co-infect cat fleas

Timothy P. Driscoll1,*, Victoria I. Verhoeve2,*, Cassia Brockway4, Darin L. Shrewsberry1, Mariah Plumer2, Spiridon E. Sevdalis2, John F. Beckmann3, Laura M. Krueger4, Kevin R. Macaluso5, Abdu F. Azad2 and Joseph J. Gillespie2

1 Biology, West Virginia University, Morgantown, WV, USA
2 Microbiology and Immunology, University of Maryland at Baltimore, Baltimore, MD, USA
3 Entomology and Plant Pathology, Auburn University, Auburn, AL, USA
4 Orange County Mosquito and Vector Control District, Garden Grove, CA, USA
5 Microbiology and Immunology, University of South Alabama, Mobile, AL, USA

* These authors contributed equally to this work.

**ABSTRACT**

Wolbachiae are obligate intracellular bacteria that infect arthropods and certain nematodes. Usually maternally inherited, they may provision nutrients to (mutualism) or alter sexual biology of (reproductive parasitism) their invertebrate hosts. We report the assembly of closed genomes for two novel wolbachiae, *w* *CfeT* and *w* *CfeJ*, found co-infecting cat fleas (*Ctenocephalides felis*) of the Elward Laboratory colony (Soquel, CA, USA). *w* *CfeT* is basal to nearly all described *Wolbachia* supergroups, while *w* *CfeJ* is related to supergroups C, D and F. Both genomes contain laterally transferred genes that inform on the evolution of *Wolbachia* host associations. *w* *CfeT* carries the Biotin synthesis Operon of Obligate intracellular Microbes (BOOM); our analyses reveal five independent acquisitions of BOOM across the *Wolbachia* tree, indicating parallel evolution towards mutualism. Alternately, *w* *CfeJ* harbors a toxin-antidote operon analogous to the *w* *Pip* cinAB operon recently characterized as an inducer of cytoplasmic incompatibility (CI) in flies. *w* *CfeJ* cinB and three adjacent genes are collectively similar to large modular toxins encoded in CI-like operons of certain *Wolbachia* strains and *Rickettsia* species, signifying that CI toxins streamline by fission of large modular toxins. Remarkably, the *C. felis* genome itself contains two CI-like antidote genes, divergent from *w* *CfeJ* cinA, revealing episodic reproductive parasitism in cat fleas and evidencing mobility of CI loci independent of WO-phage. Additional screening revealed predominant co-infection (*w* *CfeT*/*w* *CfeJ*) amongst *C. felis* colonies, though fleas in wild populations mostly harbor *w* *CfeT* alone. Collectively, genomes of *w* *CfeT*, *w* *CfeJ*, and their cat flea host supply instances of lateral gene transfers that could drive transitions between parasitism and mutualism.
INTRODUCTION

Wolbachiae (Alphaproteobacteria: Rickettsiales: Anaplasmataceae) comprise Gram-negative, obligate intracellular bacteria that infect over half the world’s described insect species as well as certain parasitic nematodes (Hilgenboecker et al., 2008). Unlike other notable rickettsial genera that contain human pathogens (e.g., Rickettsia, Orientia, Neorickettsia, Anaplasma, and Ehrlichia), wolbachiae do not infect vertebrates (Gillespie et al., 2012). A single species, Wolbachia pipientis, is formally recognized with numerous members designated as strains within 16 reported supergroups (Augustinos et al., 2011; Baldo & Werren, 2007; Bing et al., 2014; Glowska et al., 2015; Haegeman et al., 2009; Lo et al., 2002; Ros et al., 2009; Lefoulon et al., 2020). Genomic divergence indicates further species names are warranted (Chung et al., 2018), though increasing diversity and community consensus suggest caution regarding further Wolbachia classification at the species level (Lindsey et al., 2016; Ramírez-Puebla et al., 2015).

Like other obligate intracellular microbes, wolbachiae are metabolic parasites that complement a generally reduced metabolism with pilfering of host metabolites (Driscoll et al., 2017; Jiménez et al., 2019). Their ability to survive and flourish is also heavily influenced by the acquisition of key functions through lateral gene transfer (LGT). Several described Wolbachia strains demonstrate characteristics of limited mutualism with their invertebrate partner (Zug & Hammerstein, 2015; Hosokawa et al., 2010), through the synthesis and provisioning of riboflavin (Moriyama et al., 2015) and biotin (Ju et al., 2019; Nikoh et al., 2014). While riboflavin biosynthesis genes are highly conserved in wolbachiae (Newton & Rice, 2020), biotin biosynthesis genes are rare and likely originated via LGT with taxonomically divergent intracellular bacteria (Gillespie et al., 2012). Still other strains of Wolbachia exert varying degrees of reproductive parasitism (RP) on their insect host (Werren, Baldo & Clark, 2008), influencing host sexual reproduction via processes such as male-killing, feminization, parthenogenesis and cytoplasmic incompatibility (CI) (Hurst & Frost, 2015). Wolbachia genes underpinning CI and male killing have been characterized (Chen et al., 2019; Beckmann et al., 2019; LePage et al., 2017; Beckmann, Ronau & Hochstrasser, 2017; Beckmann & Fallon, 2013; Perlmutter et al., 2019) and occur predominantly in the eukaryotic association module (EAM) of Wolbachia prophage genomes (Bordenstein & Bordenstein, 2016). These genes highlight the role of LGT in providing wolbachiae with factors facilitating mutualism or RP, both of which are highly successful strategies for increasing infection frequency in invertebrate host populations.

Compared to reproductive parasites, Wolbachia mutualists appear to form more stable, long-term relationships with their hosts, as supported by Wolbachia-host codivergence in certain filarial nematodes (Bandi et al., 1998) and Nomada bees (Gerth & Bleidorn, 2017). In contrast to the stability of mutualists, relationships of reproductive parasites appear more ephemeral. RP can be a strong mechanism to increase infection frequency; for example, hosts of CI-inducing strains can rapidly replace uninfected members within populations (Turelli & Hoffmann, 1991; Schmidt et al., 2017). Despite this, CI is prone to neutralization through the evolution of host suppression (Hurst & Frost, 2015;
Cooper et al., 2017; Turelli, 1994) and purifying selection on the host does not preserve CI (Turelli, 1994).

Recent work proposes that CI genes, or CI factors (cifs), have genetic life cycles (Martinez et al., 2020). Initially, cifs might colonize new hosts by LGT, pushing said host to high frequency in populations through CI. Ultimately, CI collapses as cifs become pseudogenized. Thus, cifs are selfish elements that persist in a manner similar to transposons (Martinez et al., 2020; Cooper et al., 2019). The occurrence of duplicate cif copies (i.e., encoding the same biochemical mechanism), as well as cif pseudogenes, in many Wolbachia strains supports this model (Gillespie et al., 2018). Cif degradation would lead to a weak CI background, fomenting favorable conditions for Wolbachia strains with more active cif repertoires to invade and outpace resident strains carrying failing cifs. An alternative to autonomous cif jumping would be recombinant WO-phages or smaller elements actually carrying CI operons. If dosage is important to CI, all these events might shift the strength of CI leading to new bidirectional incompatibilities.

Such LGT partly explains why Wolbachia estimated phylogenies are often discordant with those of their hosts (Cooper et al., 2019; Raychoudhury et al., 2009; Turelli et al., 2018). Horizontal transmissions of Wolbachia strains, which can occur through direct host interactions, environmental contacts (shared habitat, food sources, etc.) or predator/parasitoid delivery, also cause tree discordance (Pietri, DeBruhl & Sullivan, 2016) and contribute to episodic invasions, replacements, and exchange of cifs. Horizontal transmission is evident in cases where hosts are infected with multiple divergent Wolbachia strains (Raychoudhury et al., 2009; Turelli et al., 2018; Pietri, DeBruhl & Sullivan, 2016). In host populations that vary from single- vs double-infections, co-infection might reflect a transitory phase (i.e., one strain replacing another) or stable coinfection. We envision that competition could arise between co-infections if the distinct reproductive parasites carry too similar RP-inducing genes and battle for control of parallel reproductive mechanisms. Alternatively, perhaps the stability of co-infections could be supported in part by different mechanistic approaches to RP (e.g., cin vs cid loci) or dosage effects from multiple strains collaborating to instantiate a new crossing type. Furthermore, stable coinfections might increase robustness by partitioning distinct strains to serve within distinct host micro-niches (i.e., nutrient provisioning vs RP); each might increase fitness distinctly or alternatively parasitize distinctly. Such a case is in part witnessed in doubly-infected cochineals (Dactylopius coccus), wherein a Supergroup B strain (wDacB) carries an arsenal of RP-inducing genes but a Supergroup A strain (wDacA) carries none (Raychoudhury et al., 2009). Clearly, documenting more cases of Wolbachia co-infections putatively serving under contrasting host relationships will be helpful to understand mechanisms of replacement and stability in host populations.

We recently sequenced the genome of the cat flea (Ctenocephalides felis) and concurrently closed the genomes of two novel wolbachiae, wCfeT and wCfeJ, that co-infect fleas of the Elward Laboratory (Soquel, CA, USA; hereafter EL felas) (Driscoll et al., 2020). Genome-based phylogeny estimation placed wCfeT on a basal divergent branch, while wCfeJ was found to subtend Wolbachia Supergroup C. Both genomes show evidence
of past WO phage infection, with LGT hotspots indicating recent acquisition of factors predicted to either underpin mutualism (wCfeT) or mediate CI (wCfeJ). Additionally, we describe two loci within the C. felis genome itself that support repeated acquisitions of CI antidote genes by the host. Infection surveys of cat flea populations and colonies from across the US indicate extensive spread of wCfeT among wild populations, and a strong bias for wCfeT/wCfeJ co-infection in colonies, highlighting the complex ecological relationship that drives this system. Finally, we apply an informatics approach including phylogenomics, synteny analysis, and functional alignment to construct a comprehensive picture revealing the recurrent evolution of nutritional symbiosis and RP across the Rickettsiales.

MATERIALS AND METHODS
Assembly and annotation of wCfeT and wCfeJ
Our recent report detailed (1) the assembly of Wolbachia reads generated from C. felis genome sequencing, (2) genome circularization and naming of two novel strains (wCfeT and wCfeJ), and (3) complete genome annotation (Driscoll et al., 2020). All relevant information pertaining to wCfeT and wCfeJ is available at NCBI under Bioproject PRJNA622233. Complete genome sequences are available in the NCBI RefSeq database at accession identifiers NZ_CP051156.1 (wCfeT) and NZ_CP051157.1 (wCfeJ).

Wolbachiae phylogenomics
Protein sequences (n = 84,836) for 66 sequenced Wolbachia genomes plus five additional Anaplasmataceae (Neorickettsia helminthoeca str. Oregon, Anaplasma centrale Israel, A. marginale Florida, Ehrlichia chaffeensis Arkansas, and E. ruminantium Gardel) were either downloaded directly from NCBI (n = 48), retrieved as genome sequences from the NCBI Assembly database (n = 13), contributed via personal communication (n = 8; Michael Gerth, Oxford Brookes University), or sequenced here (n = 2) (Table S1). For genomes lacking functional annotations (n = 15), gene models were predicted using the RAST v2.0 server (n = 12) or GeneMarkS-2 v1.10_1.07 (n = 3; (Lomsadze et al., 2018)). Ortholog groups (n = 3,038) were subsequently constructed using FastOrtho, an in-house version of OrthoMCL (Li, Stoeckert & Roos, 2003), using previously established criteria (i.e., an expect threshold of 0.01, percent identity threshold of 30%, and percent match length threshold of 50% for ortholog inclusion (Driscoll et al., 2017, 2020; Smith et al., 2015; Hagen et al., 2018)). A subset of single-copy families (n = 12) conserved across at least 60 of the 66 genomes were independently aligned with MUSCLE v3.8.31 (Edgar, 2004) using default parameters. Regions of poor alignment were masked with trimal v1.4.rev15 (Capella-Gutíérrez, Silla-Martínez & Gabaldón, 2009) using the “automated1” option. All modified alignments were concatenated into a single data set (2,803 positions) for phylogeny estimation using RAxML v8.2.4 (Stamatakis, 2014). The gamma model of rate heterogeneity and estimation of the proportion of invariant sites were used. Branch support was assessed with 1,000 pseudo-replications. Final ML optimization likelihood was −52350.085098.
**BOOM characterization**

**Phylogeny estimation**

Genomes \((n = 93)\) containing at least four of the six biotin synthesis genes were identified using Blastp searches against the NCBI nr database \((e\text{-value} < 0.01; \text{accessed 9 January 2020})\). Protein sequences for these loci were downloaded from NCBI, aligned with MUSCLE v3.8.31 \((Edgar, 2004)\) using default parameters, and regions of poor alignment masked with trimal v1.4.rev15 \((Capella-Gutiérrez, Silla-Martínez & Gabaldón, 2009)\) using the “automated1” option. All modified alignments were concatenated into a single data set \((1,457 \text{ positions})\) for phylogeny estimation using RAxML v8.2.4 \((Stamatakis, 2014)\), under the gamma model of rate heterogeneity and estimation of the proportion of invariant sites. Branch support was assessed with 1,000 pseudo-replications. Final ML optimization likelihood was \(-127140.955066\). See Table S2 for NCBI protein accession IDs for all sequences included in this analysis.

**Gene neighborhood analysis of wCfeT BOOM**

Genes in a 25-gene window around the wCfeT BOOM cluster were used to query (Blastp) the NCBI nr database \((\text{accessed 9 January 2020})\). The top 50 matches to each sequence \((e\text{-value} < 1, \text{BLOSUM45 matrix})\) were binned by taxonomic assignment as “Wolbachia”, “other Rickettsiales”, or “other (non-Rickettsiales) taxa.” Other (non-Rickettsiales) taxa were further binned by major taxonomic group to assess the boundaries and diversity of the wCfeT BOOM MGE.

**Synteny analysis of biotin genes in wolbachiae**

Rickettsiales genomes with at least 5 of the 6 biotin synthesis genes \((n = 28)\) were downloaded from NCBI and ortholog groups \((n = 2,527)\) constructed with FastOrtho, an in-house version of OrthoMCL \((Li, Stoeckert & Roos, 2003)\), using an expect threshold of 0.01, percent identity threshold of 30%, and percent match length threshold of 50% for ortholog inclusion. OGs containing biotin synthesis genes were identified using *Rickettsia buchneri* str. Wikel sequences as seeds. Genome locations for all genes were retrieved from NCBI in gene file format \((gff)\) and used to order the orthologs in each genome.

**Phylogeny estimation of the EamA protein of wCfeT**

Taxa were selected from Blastp searches against the NCBI Rickettsiales and nr \((\text{excluding Rickettsiales})\) databases using the wCfeT EamA protein \((WP_168464633.1)\) as a query. Proteins were aligned with MUSCLE v3.8.31 \((Edgar, 2004)\) using default parameters, with regions of poor alignment masked using Gblocks \((Talavera & Castresana, 2007)\). Phylogeny was estimated with the WAG substitution model using RAxML v8.2.4 \((Stamatakis, 2014)\) under the gamma model of rate heterogeneity and estimation of the proportion of invariant sites.

**CI gene characterization**

*Wolbachiae*

wCfeJ CinA and CinB were compared to other RP-inducing TA operons following our previous approach \((Gillespie et al., 2018)\). Another novel CndAB operon was identified on
a small scaffold (NCBI Nucleotide ID CABPRJ0100001055.1) in the aphid Cinara cedri, which has an associated Wolbachia symbiont, wCed (Gómez-Valero et al., 2004), and was also characterized in a similar manner. Briefly, EMBL’s Simple Modular Architecture Research Tool (SMART) (Letunic & Bork, 2017) and/or the Protein Homology/analogy Recognition Engine V 2.0 (Phyre2) (Kelley & Sternberg, 2009) were used to predict the following domains: OTU cysteine protease (Pfam OTU, PF02338) (Makarova, Aravind & Koonin, 2000), endoMS-like (Nakae et al., 2016) and NucS-like (Ren et al., 2009) PD-(D/E)XK nuclease (Pfam PDDEXK_1, PF12705) (Kinch et al., 2005), CE clan protease (Pfam Peptidase_C1, PF00112) (Pruneda et al., 2016), Latrotoxin-CTD (Zhang et al., 2012), and ankyrin repeats. Individual protein schemas were generated using Illustrator of Biological Sequences (Liu et al., 2015) with manual adjustment.

C. felis
A cidA-like gene was identified in the Ctenocephalides felis genome during initial genome decontamination (Driscoll et al., 2020). Briefly, a contig composed of a mosaic of flea- and Wolbachia-like sequence was identified by our comparative BLAST-based pipeline and flagged for manual inspection. Gene models on this contig were predicted using the RAST v2.0 server. The resulting proteins were used to create a custom Blast database that was queried using Blastp with CinAB sequences from Wolbachia endosymbiont of Culex quinquefasciatus Pel (CAQ54390, CAQ54391), and CidAB sequences Wolbachia pipiensis wAlbB (CCE77512, CCE77513). The contig was subsequently confirmed as belonging to the largest scaffold (NW_020538040) in the C. felis assembly. A second, cinA-like gene was identified by querying all CDS in the published C. felis assembly (NCBI accession ASM342690v1) with the same set of Wolbachia CinAB and CidAB sequences.

Screening colonies and wild populations of cat fleas for wCfeT and wCfeJ
Upon receipt from collaborators, cat fleas were stored at −20 °C in EtOH until processing for DNA extraction and qPCR. Fleas were sexed under a stereomicroscope and surface sterilized for 5 min with 1% bleach solution, followed by 5 min with 70% EtOH, and finally with three washes using molecular grade water. Individual fleas were flash-frozen in liquid N2 and ground to powder with sterile pestle in a 1.5 ml microcentrifuge tube. DNA was extracted from dissected tissues using the Zymo Quick-DNA Miniprep plus kit following the solid tissue protocol (Zymo Research; Irvine, CA, USA). The presence of wCfeT, wCfeJ, and C. felis DNA was determined via qPCR using the following primer sets: Cfe#18S|179, 5′-TGCTCACCGTTTGACTTGG-3′ and 5′-GTTTCTCAGGCTCCCTCTCC-3′ (Reif et al., 2008); wCfeT#ApaG|75: 5′-GCCGTCACTGGCAGGTAATA-3′ and 5′-GCTGTTCTCCAATAACGCCA-3′, wCfeJ#CinA|76, 5′-AGCAACACCAACATGCGATT-3′ and 5′-GAACCCCAGAGTTGGAGGG-3′. Each primer set amplicon was sequenced before use to ensure specificity. qPCR was performed using a QuantStudio 3 with the PowerUp SYBR green mastermix in a 20 µl reaction volume containing 2 µl of sample DNA with 400 nM of each primer. Cycling parameters were consistent with the “fast cycling mode” including a 2 min cycle at 50 °C, a 2 min cycle at 95 °C, 40 cycles with
3 s at 95 °C and 30 s at 60 °C followed by a melt curve analysis. Standard curves were generated to quantify the number of DNA copies present in each reaction. Results were expressed as the ratio of *Wolbachia* amplicons to *C. felis* 18S rDNA copies to normalize for varying efficiency of DNA extractions. Fleas were considered positive for wolbachiae if the flea 18S rDNA Ct value is 30 or lower and the *Wolbachia* Ct value is 37 or lower.

**Determining wCfeT and wCfeJ localization in *C. felis* tissues**

Elward Laboratories (EL, Soquel, CA, USA) cat fleas stored at −80 °C were used to assay tissue localization of wCfeT and wCfeJ. Ovaries and midguts of individual female fleas were dissected under a stereomicroscope. DNA extraction and qPCR were performed as described above.

**RESULTS AND DISCUSSION**

Fleas are understudied vectors of several human and animal diseases (*Rust, 2017*). While wolbachiae (mostly unnamed strains) have been detected in many flea species (*Gorham, Fang & Durden, 2003; Dittmar & Whiting, 2004; Lawrence et al., 2015*), it is pressing from a biocontrol perspective to know if these bacteria affect the ability of fleas to vector pathogens. The cat flea transmits pathogenic bacteria such as *Rickettsia typhi* (murine typhus), *R. felis* (murine typhus-like illness), *Bartonella* spp. (e.g., cat-scratch disease and various other bartonelloses), and to a lesser extent the plague agent *Yersinia pestis* (*Glickman et al., 2006; Nogueras et al., 2013; Angelakis et al., 2016; Leulmi et al., 2014*). Cat fleas are also intermediate hosts of the tapeworm *Dipylidium caninum* and the filarial nematode *Acanthocheilonema reconditum*, both of which can infect humans (*Fourie et al., 2012; Brianti et al., 2012; Traversa, 2013*). Other organisms observed in *C. felis* include undescribed Baculoviridae, amebae, trypanosomatids, cephaline gregarines (Apicomplexa), and microsporidia (*Beard, Butler & Hall, 1990*). However, nearly all epidemiological studies have focused on determining the frequency and distribution of only certain pathogens (e.g., *Bartonella* spp., *R. typhi*, *R. felis*), with few documenting wolbachiae co-occurrence with these pathogens (*Tay, 2013; Oteo et al., 2014*).

**Identification of novel *Wolbachia* parasites**

Recently, we assembled and annotated closed genomes from two divergent *Wolbachia* strains, named wCfeT and wCfeJ, that were captured during genome sequencing of the cat flea (EL Colony) (*Driscoll et al., 2020*). The 16S rDNA sequences of these two strains are identical to those previously identified in another cat flea colony (Louisiana State University) (*Pornwiroon et al., 2007; Sunyakumthorn et al., 2008; Gillespie et al., 2015*), which historically has been replenished with EL colony fleas. wCfeT and wCfeJ are broadly similar to other closed *Wolbachia* genomes in terms of length, GC content, number of protein coding genes, and coding density (*Table 1*), although it is interesting to note that wCfeT exhibits one of the highest pseudogene counts (second only to wCle) while wCfeJ boasts one of the lowest (second only to wOv). Coupled with other recently sequenced genomes for non-Supergroup A and B strains (e.g., wFol and wVulC), these cat flea-associated wolbachiae blur the lines demarcating genomic traits previously used to
distinguish basal and sister Wolbachia lineages from those of Supergroups A and B strains (Lefoulon et al., 2020), particularly regarding genome size and mobile element composition (Lefoulon et al., 2020). Prior reports have indicated the presence of multiple Wolbachia "types" infecting C. felis, with phylogenies inferred from partial gene sequences placing these unnamed Wolbachia strains in supergroups B or F, or in basal lineages (Gorham, Fang & Durden, 2003; Dittmar & Whiting, 2004; Casiraghi et al., 2005). Our robust genome-based phylogeny estimation reveals that both wCfeT and wCfeJ are divergent from Supergroup A and B wolbachiae (Fig. 1; Table S1). wCfeT is similar to undescribed C. felis-associated strains that branch basally to most other Wolbachia lineages (Gorham, Fang & Durden, 2003; González-Álvarez et al., 2017; Vasconcelos et al., 2018), while wCfeJ is similar to undescribed C. felis-associated strains closely related to Wolbachia supergroups C, D and F (Gorham, Fang & Durden, 2003; Casiraghi et al., 2005). The substantial divergence of

**Table 1** Summary statistics of selected Wolbachia genomes.

<table>
<thead>
<tr>
<th>Taxon</th>
<th>RefSeq ID</th>
<th>SG*</th>
<th>Length (nt)</th>
<th>GC (%)</th>
<th>Protein coding genes</th>
<th>Pseudogenes</th>
<th>Mean gene length (nt)</th>
<th>Coding density (coding/total nt)</th>
</tr>
</thead>
<tbody>
<tr>
<td>wFol</td>
<td>NZ_CP015510.2</td>
<td>E</td>
<td>1801626</td>
<td>34.4</td>
<td>1537</td>
<td>85</td>
<td>976</td>
<td>0.880</td>
</tr>
<tr>
<td>wCfeT</td>
<td>NZ_CP051156.1</td>
<td>?</td>
<td>1495538</td>
<td>35.2</td>
<td>1424</td>
<td>201</td>
<td>890</td>
<td>0.727</td>
</tr>
<tr>
<td>wOv</td>
<td>NZ_HG810405.1</td>
<td>C</td>
<td>960618</td>
<td>32.1</td>
<td>650</td>
<td>45</td>
<td>967</td>
<td>0.700</td>
</tr>
<tr>
<td>wOo</td>
<td>NC_018267.1</td>
<td>C</td>
<td>957990</td>
<td>32.1</td>
<td>639</td>
<td>73</td>
<td>968</td>
<td>0.719</td>
</tr>
<tr>
<td>wCfeJ</td>
<td>NZ_CP051157.1</td>
<td>?</td>
<td>1201647</td>
<td>35.6</td>
<td>1116</td>
<td>51</td>
<td>917</td>
<td>0.813</td>
</tr>
<tr>
<td>wCle</td>
<td>NZ_AP013028.1</td>
<td>F</td>
<td>1250060</td>
<td>36.3</td>
<td>1012</td>
<td>210</td>
<td>818</td>
<td>0.829</td>
</tr>
<tr>
<td>wBm</td>
<td>NZ_CP034333.1</td>
<td>D</td>
<td>1080064</td>
<td>34.2</td>
<td>803</td>
<td>178</td>
<td>868</td>
<td>0.789</td>
</tr>
<tr>
<td>wBm</td>
<td>NC_006833.1</td>
<td>D</td>
<td>1080084</td>
<td>34.2</td>
<td>839</td>
<td>167</td>
<td>854</td>
<td>0.795</td>
</tr>
<tr>
<td>wMau</td>
<td>NZ_CP034334.1</td>
<td>B</td>
<td>1273527</td>
<td>34.0</td>
<td>1045</td>
<td>136</td>
<td>944</td>
<td>0.876</td>
</tr>
<tr>
<td>wMau</td>
<td>NZ_CP034335.1</td>
<td>B</td>
<td>1273530</td>
<td>34.0</td>
<td>1047</td>
<td>131</td>
<td>946</td>
<td>0.876</td>
</tr>
<tr>
<td>wNo</td>
<td>NC_021084.1</td>
<td>B</td>
<td>1301823</td>
<td>34.0</td>
<td>1066</td>
<td>133</td>
<td>952</td>
<td>0.878</td>
</tr>
<tr>
<td>wAlbB FL2016</td>
<td>NZ_CP041923.1</td>
<td>B</td>
<td>1482279</td>
<td>34.5</td>
<td>1190</td>
<td>195</td>
<td>917</td>
<td>0.858</td>
</tr>
<tr>
<td>wAlbB HN2016</td>
<td>NZ_CP041924.1</td>
<td>B</td>
<td>1483853</td>
<td>34.4</td>
<td>1200</td>
<td>187</td>
<td>917</td>
<td>0.858</td>
</tr>
<tr>
<td>wAlbB</td>
<td>NZ_CP031221.1</td>
<td>B</td>
<td>1484007</td>
<td>34.4</td>
<td>1197</td>
<td>191</td>
<td>916</td>
<td>0.858</td>
</tr>
<tr>
<td>wMeg</td>
<td>NZ_CP021120.1</td>
<td>B</td>
<td>1376868</td>
<td>34.0</td>
<td>1136</td>
<td>125</td>
<td>948</td>
<td>0.869</td>
</tr>
<tr>
<td>wHa</td>
<td>NC_021089.1</td>
<td>A</td>
<td>1295804</td>
<td>35.1</td>
<td>1123</td>
<td>104</td>
<td>899</td>
<td>0.852</td>
</tr>
<tr>
<td>wCauA</td>
<td>NZ_CP041215.1</td>
<td>A</td>
<td>1449344</td>
<td>35.0</td>
<td>1260</td>
<td>134</td>
<td>861</td>
<td>0.860</td>
</tr>
<tr>
<td>wAna</td>
<td>NZ_CP042904.1</td>
<td>A</td>
<td>1401460</td>
<td>35.2</td>
<td>1218</td>
<td>107</td>
<td>893</td>
<td>0.857</td>
</tr>
<tr>
<td>wRi</td>
<td>NC_012416.1</td>
<td>A</td>
<td>1445873</td>
<td>35.2</td>
<td>1245</td>
<td>122</td>
<td>892</td>
<td>0.859</td>
</tr>
<tr>
<td>wAu</td>
<td>NZ_LK055284.1</td>
<td>A</td>
<td>1268461</td>
<td>35.2</td>
<td>1099</td>
<td>125</td>
<td>878</td>
<td>0.855</td>
</tr>
<tr>
<td>wMel</td>
<td>NZ_CC02978.6</td>
<td>A</td>
<td>1267782</td>
<td>35.2</td>
<td>1116</td>
<td>129</td>
<td>863</td>
<td>0.857</td>
</tr>
<tr>
<td>wMel I23</td>
<td>NZ_CP042444.1</td>
<td>A</td>
<td>1269137</td>
<td>35.2</td>
<td>1119</td>
<td>128</td>
<td>862</td>
<td>0.857</td>
</tr>
<tr>
<td>wMel N25</td>
<td>NZ_CP042446.1</td>
<td>A</td>
<td>1267781</td>
<td>35.2</td>
<td>1120</td>
<td>125</td>
<td>863</td>
<td>0.857</td>
</tr>
<tr>
<td>wMel ZH26</td>
<td>NZ_CP042445.1</td>
<td>A</td>
<td>1267436</td>
<td>35.2</td>
<td>1117</td>
<td>125</td>
<td>865</td>
<td>0.857</td>
</tr>
<tr>
<td>wIrr</td>
<td>NZ_CP037426.1</td>
<td>A</td>
<td>1352354</td>
<td>35.3</td>
<td>1232</td>
<td>115</td>
<td>827</td>
<td>0.855</td>
</tr>
</tbody>
</table>

Note:

* Supergroup designations for wCfeT and wCfeJ remain unclear.

Prior reports have indicated the presence of multiple Wolbachia “types” infecting C. felis, with phylogenies inferred from partial gene sequences placing these unnamed Wolbachia strains in supergroups B or F, or in basal lineages (Gorham, Fang & Durden, 2003; Dittmar & Whiting, 2004; Casiraghi et al., 2005). Our robust genome-based phylogeny estimation reveals that both wCfeT and wCfeJ are divergent from Supergroup A and B wolbachiae (Fig. 1; Table S1). wCfeT is similar to undescribed C. felis-associated strains that branch basally to most other Wolbachia lineages (Gorham, Fang & Durden, 2003; González-Álvarez et al., 2017; Vasconcelos et al., 2018), while wCfeJ is similar to undescribed C. felis-associated strains closely related to Wolbachia supergroups C, D and F (Gorham, Fang & Durden, 2003; Casiraghi et al., 2005). The substantial divergence of
Figure 1  *Wolbachia* genome-based phylogeny estimation. *Wolbachia* supergroups are within colored ellipses. *Ctenocephalides felis*-associated *Wolbachia* are within black boxes, with red (wCfeT) and blue (wCfeF) stars depicting the two novel *Wolbachia* parasites infecting *C. felis*. Gray inset: color scheme for nematode and arthropod hosts. Phylogeny was estimated for 66 *Wolbachia* and five outgroup genomes (see "Materials and Methods" for more details). Branch support was assessed with 1,000 pseudoreplications. Final ML optimization likelihood was −52350.085098.

PeerJ. DOI 10.7717/peerj.10646/fig-1
wCfeT and wCfeJ from a Wolbachia supergroup B strain associated with C. felis from Germany (wCte) indicates a diversity of wolbachiae that are capable of infecting cat fleas. Importantly, the nature of host-microbe relationships for any of these C. felis-associated wolbachiae is unknown.

**wCfeT is equipped with biotin synthesis genes**
The enzymatic steps for bacterial biosynthesis of biotin from malonyl-CoA are largely conserved (Lin & Cronan, 2011) and usually involve six bio enzymes and the fatty acid biosynthesis machinery (Lin, Hanson & Cronan, 2010) (Fig. 2A). We previously identified the first plasmid-borne bio gene operon, which occurs on plasmid pREIS2 of Rickettsia buchneri, an endosymbiont of the black-legged tick (Gillespie et al., 2012). We observed a rare gene order for the pREIS2 bio operon shared only with bio operons encoded on the chromosomes of obligate intracellular pathogens Neorickettsia spp. and Lawsonia intracellularis (Deltaproteobacteria). These unique bio operons formed a clade in estimated phylogenies, indicating LGT of the operon across divergent intracellular species. Subsequent studies, using our same dataset for phylogeny estimation, discovered this bio operon in certain Wolbachia (Nikoh et al., 2014; Gerth & Bleidorn, 2017; Balvín et al., 2018), Cardinium (Penz et al., 2012; Zeng et al., 2018) and Legionella (Ríhová et al., 2017) species. We refer hereafter to this intriguing LGT of bio genes as the Biotin synthesis Operon of Obligate intracellular Microbes (BOOM).

wCfeT carries a complete BOOM with greater similarity to Wolbachia BOOM than those of other bacteria (Fig. 2B). Phylogeny estimation of BOOM and other diverse bio gene sets indicates wCfeT BOOM is basal to other Wolbachia BOOM (Fig. 2C; Fig. S1). However, as the divergence pattern for other Wolbachia BOOM is discordant with genome-based phylogeny (Fig. 1), it is clear that multiple independent BOOM acquisitions have occurred in wolbachiae irrespective of lineage divergences. This is supported by a previous study positing recent BOOM acquisition in wNfla and wNleu (Gerth & Bleidorn, 2017) and a current report that detected partial BOOM loci in wApol, a Supergroup S Wolbachia strain infecting pseudoscorpions (Lefoulon et al., 2020).

The nature of BOOM flanking regions in the genomes of wCfeT, wVulC, wCle, wStr, and wLug, which collectively lack synteny and are riddled with transposases, recombinases and phage-related elements (Table S3), also attests to LGT as the origin for BOOM in these genomes. Furthermore, analysis of wCfeT BOOM flanking genes reveals hotspots for additional LGT (Fig. 2D), exemplified by an S-adenosylmethionine importer (EamA) gene (Fig. S1B) that is a signature among many diverse obligate intracellular bacteria (Gillespie et al., 2012; Klinges et al., 2019; Haferkamp et al., 2013; Driscoll et al., 2013).

As some BOOM-containing Wolbachia strains have been shown to provide biotin to their insect hosts (Ju et al., 2019; Nikoh et al., 2014), we posit that wCfeT has established an obligate mutualism with C. felis mediated by biotin-provisioning. Genes involved in CI (and putatively in MK) have been predominantly found in the EAM; the absence of such genes in wCfeT’s WO prophage (Fig. S2) provides no evidence that wCfeT can induce these two forms of RP, implying that its maintenance may indeed rely on nutritional supplementation to the cat flea. Wolbachia strains associated with bedbug (wCle),
**Figure 2** wCfeT carries the Biotin synthesis Operon of Obligate intracellular Microbes (BOOM). (A) Metabolism of biotin from malonyl-CoA (Lin & Cronan, 2011). (B) Comparisons of wCfeT BOOM to equivalent loci in Cardinium endosymbiont of Encarsia pergandiella (cEper1, CCM10336–CCM10341) and Wolbachia endosymbiont of Cimex lectularius (wCle, BAP00143–BAP00148). Red shading and numbers indicate % identity across pairwise protein alignments (Blastp). Gene colors correspond to enzymatic steps depicted in panel A, with all proteins drawn to scale (as a reference, wCfeT BioB is 316 aa). (C) Phylogeny of BOOM and other bio gene sets from diverse bacteria estimated from the concatenation of six
planthopper (wStr, wLug), and termites (wCtub) with evidence for BOOM (Hellemans et al., 2019) reside in bacteriocytes of their insect hosts, structures associated with nutritional symbiosis in many arthropods (Douglas, 2015). Interestingly, C. felis is not known to contain bacteriocytes, and the ability of wCfeT to synthesize and supplement biotin to cat fleas awaits experimentation.

wCfeJ lends insight on the evolution of reproductive parasitism

wCfeJ carries a putative toxin-antidote (TA) operon that is architecturally similar to the wPip_Pel CinA/B TA operon (Fig. 3A), which was previously characterized as a CI inducer in flies (Chen et al., 2019). Containing at least three distinct variants (LePage et al., 2017; Lindsey et al., 2018), all CinB toxins harbor dual nuclease (NUC) domains in place of the deubiquitinase (DUB) domain of CidB, another CI inducer in flies (LePage et al., 2017; Beckmann, Ronau & Hochstrasser, 2017; Beckmann & Fallon, 2013). Chimeric NUC-DUB toxins, termed CndB toxins (Beckmann et al., 2019), also occur as operons (cndAB) in certain Wolbachia and Rickettsia genomes (Beckmann, Ronau & Hochstrasser, 2017; Gillespie et al., 2018). Although their functions remain unknown, these larger CndB toxins belong to an extraordinarily diverse array of toxins with highly modular architectures that occur in divergent intracellular bacteria (Gillespie et al., 2018). Curiously, despite a size similar to the CinB toxins, wCfeJ’s CinB toxin has much higher sequence identity to larger CndB and NUC-containing toxins found mostly in Wolbachia Supergroups A and B (Fig. 3B). A similar pattern was found for wCfeJ’s CinA antidote, indicating the genes were likely acquired as an operon.

In contrast to wCfeT, wCfeJ harbors a highly degraded WO prophage genome with only a partial EAM, the CinA/B operon, and several other genes encoding hypothetical proteins and transposases (Fig. 3C). This is reminiscent of the CI- and male killer-inducing wRec strain, which lacks a functional WO phage yet possesses TA modules within remnant EAMs (Metcalf et al., 2014).

Synteny analysis with other Wolbachia phages revealed wCfeJ cinB and three downstream genes are colinear with a large (3707 aa) wDacB cndB toxin (Fig. 3D). We posit that gene fission of a large modular CndB toxin created the smaller CinB toxin. This interpretation is consistent with our prior hypothesis stating that more streamlined RP-inducing toxins (i.e., cinB and cidB) originate from larger modular toxins (i.e., cndB and others) that are widespread in the intracellular mobiome, particularly Wolbachia phage and Rickettsia plasmids and integrative and conjugative elements (ICEs) (Gillespie et al., 2018, 2015). We find the alternative hypothesis, that multiple gene fusions create...
Figure 3  wCfeJ lends insight on the evolution of reproductive parasitism. (A) Comparison of known and predicted RP-inducing toxin-antidote operons of diverse rickettsial species. Light green: CidA/B of wPip_Pel (CAQ54390, CAQ54391); gray: CndA/B of pLbAR_36/38 (KHO02168, KHO02170), Rickettsia gravesii (WP_017442886, WP_024547315), wDacB (OAM06111, OAM06112), wStr (OAB79364, OAB79365), wVulC (OJH31528, OJH31527), wCon (RDD34163, RDD34164); light brown, CinA/B of wCfeJ (blue star, WP_168456002 & WP_168456001), Orientia tsutsugamushi (CAM80637, CAM80639), wRi (ACN95432, ACN95431), wVita (ONI57866, ONI57867), and wPip_Pel (CAQ54402, CAQ54403). Type I-IV CI operons of Wolbachia parasites (LePage et al., 2017) are distinguished. Protein domains as follows: green, CE clan protease; brown, PD- and MutL HTH RadC. (B) Sequence similarity profiles for wCfeJ’s CinA antitode (top) and CinA toxin (bottom). Subjects are from a Blastp search against the NCBI “Wolbachia” database. Size, amino acids. Antidotes with cognate toxins (operon partners) also recovered in searches are noted with asterisks. Top seven hits are shown, with matches (where significant) to proteins of types II-IV TA operons shown below dashed lines. Toxin domains are from our prior report or predicted newly here using the EMBL Simple Modular Architecture Research Tool (SMART) (Letunic & Bork, 2017). Only NUC and DUB domains are listed, though larger toxins contain many other predicted domains. (C) Comparison of wCfeJ and wDacB degraded WO prophage genomes captures streamlining of a CI-like TA operon from a Cnd operon. WO prophage coordinates: wCfeJ, WP_168455994–WP_168456003; wDacB, OAM06111–OAM06118 (LSY01000098). Schema follows the description of the WOVitA prophage (Bordenstein & Bordenstein, 2016). (D) wCfeJ’s CndB-like toxin evolved from gene fission of a larger CndB-like toxin. NCBI protein accession numbers: wCfeJ (WP_168455999, WP_168456000, and WP_168456001); wDacB (OAM06111, OAM06112). Red shading and numbers indicate % identity across pairwise alignments. Proteins are drawn to scale; as a reference, wDacB CndB is 3707 aa.
larger modular toxins like CndB proteins, less parsimonious provided that RP toxins tend to degrade and pseudogenize rapidly in rickettsial genomes (Martinez et al., 2020; Gillespie et al., 2018).

\(wCfeJ\) is unique among described wolbachiae outside of Supergroups A and B by containing a CI-like TA operon. Some of the \(Wolbachia\) strains in closely related Supergroups C, F and D are considered mutualists with their nematode or arthropod hosts, supporting the absence of RP-inducing genes in the genomes of these strains. Still, other strains not included in our analysis, namely those belonging to Supergroups S (Lefoulon et al., 2020; Chafee et al., 2010), P (Glawska et al., 2015), and J (Lefoulon et al., 2016) may eventually be shown to carry RP-inducing genes, especially since some of these strains are not known as mutualists. \(wFol\), another basal \(Wolbachia\) lineage (Fig. 1), does carry a myriad of large modular candidate RP-inducing toxins but does not have \(cin\), \(cid\) or \(cnd\) TA operons (Faddeeva-Vakhrusheva et al., 2017; Kampfraath et al., 2019). This indicates LGT of candidate RP-inducing genes occurs across the breadth of \(Wolbachia\) diversity. Collectively, we speculate that \(wCfeJ\) acquired a WO prophage, possibly from a Supergroup A or B \(Wolbachia\) strain that also infects cat fleas (e.g., \(wCte\)), and that viral- or gene-level selection streamlined a \(cndAB\) locus into a \(cinAB\) operon (Gillespie et al., 2018). Given that the genomes of many \(Wolbachia\) reproductive parasites harbor diverse arrays of CinA/B-and CidA/B-like operons (Gillespie et al., 2018; Beckmann et al., 2019), we posit that \(wCfeJ\)’s CinA/B TA operon might function in CI or some other form of RP; however, further study is required to evaluate \(wCfeJ\) as a cat flea reproductive parasite. Furthermore, an intriguing conjecture is that functions of the larger modular toxins, exemplified by \(wDacB\) CndB, whatever they be, are likely the primal functions that gave rise to CI.

**Lateral transfer of \(Wolbachia\) CI-like antidotes to the cat flea genome**

Our previous work identified CI-like toxin and antidote genes on scaffolds in several arthropod genome assemblies, although antidote genes were found more frequently possibly due to utility in suppressing microbial invaders implementing RP (Gillespie et al., 2018). Many of these genomes contain multiple divergent antidote genes, though none were found to harbor large-scale \(Wolbachia\) gene transfer that is known to occur in certain invertebrates (Leclercq et al., 2016; Hotopp et al., 2007; Aikawa et al., 2009; Kondo et al., 2002; Dhaygude et al., 2019; McNulty et al., 2010; Werren et al., 2010). In light of this observation, we originally speculated that CI-like genes are mobilized from wolbachiae to host through WO phage insertion into host nuclear genomes; however, the absence of other WO phage genes in these genomes made this assertion tenuous. Instead, more recent work describing the first WO prophage carrying two sets of CI loci provides compelling evidence for transposon-mediated LGT of CI loci between divergent wolbachiae (Cooper et al., 2019). This phage-independent mechanism for RP gene jumping in wolbachiae may result in the inadvertent capture of these genes by arthropod host genomes.

When we analyzed the cat flea genome for CI-like genes, we discovered that the largest \(C. felis\) scaffold (185.5 Mb) was found to harbor two CI-like antidote genes (Fig. 4A).
C. felis CidA-like and CinA-like antidotes to CidA, CndA and CinA proteins from selected Rickettsiaceae species and O. felis strains. Schema at top depicts the C. felis scaffold (left), and a CinA-like exon encoded within a larger gene model with transcriptional support (right). Size, amino acids. (C) Alignment of CinA-like CDS (left), and a CinA-like exon encoded within a larger gene model with transcriptional support (right) independent of the cat C. felis genome. (B) Sequence similarity profiles for the C. felis CidA-like (left) and CinA-like (right) antidotes. Subjects above the black dashed line are the top hits from a Blastp search against the NCBI nr protein database; subjects below are weaker matches to CinA (left) and CidA (right) proteins; matches to wCfeJ CinA are below the orange line. Size, amino acids. (C) Alignment of C. felis CidA-like and CinA-like antidotes to CidA, CndA and CinA proteins from selected Rickettsiaceae species and Wolbachia strains. Schema at top depicts the wCfeJ CinA antidote and location of six conserved regions shown in the alignment below. Amino acid coloring as follows: black, hydrophobic; red, negatively charged; green, hydrophilic; purple, aromatic; blue, positively charged. Alignment generated using MUSCLE, default parameters (Edgar, 2004).

Remarkably, neither antidote has any significant similarity to CinA of wCfeJ (Fig. 4B), indicating that these CI antidotes originated from other, as yet undescribed wolbachiae. The C. felis CinA-like CDS (XP_026474240.1) has greater similarity to CinA proteins than...
CidA proteins, while the CidA-like CDS (not annotated in the current C. felis assembly) has much greater similarity to CidA antidotes (Beckmann, Ronau & Hochstrasser, 2017; Beckmann & Fallon, 2013; Bonneau et al., 2018; Madhav et al., 2020; Morrow et al., 2020). The C. felis cinA-like gene is found as an exon within a larger flea gene model (LOC113377978), and is well supported (18× coverage, 7.2 TPM) by transcripts from the 1KITE project (Misof et al., 2014). Since these transcripts were generated in fleas from a different colony (Dr. Michael Dryden, Kansas State University), we conclude that this gene transfer occurred during an historical infection predating the establishment of these two colonies. Finally, the C. felis CinA- and CidA-like antidotes each possess the conserved motifs we previously defined for all CI antidotes (Gillespie et al., 2018), with the exception of a truncation of motifs 5 and 6 in the CinA-like protein (Fig. 4C). As other arthropod-encoded antidotes share this C-terminal truncation, it may signify an expendable domain in the host (i.e., a bacterial secretion signal, a motif involved in cognate toxin activation or delivery, etc.).

The incorporation of a Wolbachia antidote into an exon of a host gene provides one explanation for how such a laterally transferred gene would remain functional in the absence of the native regulatory elements encoded on WO prophage genomes. We previously hypothesized that CI antidote genes would be under strong selection in arthropod genomes if their expression could result in rescuing CI induction by reproductive parasites (Gillespie et al., 2018). In support of this supposition, it has been shown that: host genotypes modulate CI (Metcalf et al., 2014; Hornett et al., 2006, 2014; Bordenstein, Uy & Werren, 2003); some hosts show weak (Dmel and D. suzukii (Cooper et al., 2017; Conner et al., 2017; Hamm et al., 2014)) or strong (Culex pipiens and D. simulans (Poinso, Charlat & Mercot, 2003; Mercot & Charlat, 2004)) CI; and a Wolbachia strain’s CI can be suppressed when transfected across different hosts (Bordenstein, Uy & Werren, 2003). While none of the arthropod-encoded CI antidotes have been characterized, we propose they impart immunity to toxins secreted by intracellular parasites, curtailing chronic parasite drive into populations. At least four independent evolutionary mechanisms might lead to suppression of CI. Gene family expansion of, and/or overexpression of key factors capable of counteracting CI mechanisms is possible; similarly, it was postulated that evolution in Cid interacting proteins like karyopherin-α or P32 could lead to target site resistance (Beckmann et al., 2019). A third possibility contributing to weakening CI, though not suppression per se, is simply the mutational collapse of the operons themselves due to lack of purifying selection (Martinez et al., 2020; Meany et al., 2019). We supply evidence of a fourth possibility whereby CI might be counteracted by germline expression of acquired CI antidotes. Significantly, our finding raises awareness of a potential barrier to Wolbachia-mediated pathogen biocontrol measures.

**Characterization of wCfeT and wCfeJ infection of cat fleas**

Coupled with earlier reports indicating multiple early-branching wolbachiae lineages infecting C. felis (Gorham, Fang & Durden, 2003; Casiraghi et al., 2005), our characterization of two divergent Wolbachia strains infecting cat fleas in the EL colony
as well as fleas maintained for a decade at LSU (Pornwiroon et al., 2007; Sunyakumthorn et al., 2008; Gillespie et al., 2015) prompted us to test individual fleas for double infection. In addition to EL fleas, three other colony strains and three geographically diverse wild cat flea populations were also surveyed (Fig. 5A). Aside from the Modesto Wild strain, which is historically replenished with wild-caught cat fleas (Dr. B. Donahue, 2017, personal communication), strong wCfeT-wCfeJ co-infection was found in colony fleas. Conversely, single wCfeJ infection dominated wild cat fleas, with only a few fleas from a large sampling throughout Orange Co. CA harboring wCfeJ. While far greater sampling of both colony and wild populations is necessary, our results hint at a strong selection for wCfeT over wCfeJ in nature, which is concordant with our premise that wCfeT’s BOOM might drive mutualism. The ability of wCfeJ to highly infect colony fleas may arise as a consequence of bottlenecks within colonies, which are known to strengthen CI and maternal artificial transmission (Turelli, 1994). No sex ratio distortions were observed in colony or wild fleas, ruling out other forms of RP as driving factors. Nonetheless, the ability of wCfeT and wCfeJ to co-infect individual cat fleas provides a system to study contrasting forces (nutritional symbiosis and RP) that potentially drive multiple Wolbachia infections in a single invertebrate host.

To gain more insight into the nature of the wCfeT-wCfeJ co-infection, we analyzed the distribution of wCfeT and wCfeJ in EL flea tissues (Fig. 5B). Surmising that both Wolbachia strains are maternally transmitted, we dissected out the ovaries of unfed virgin female fleas and compared Wolbachia density in these tissues vs others (primarily midgut).
While wCfeT predominantly localizes in flea ovaries, the cif carrier, wCfeJ, mostly localizes to somatic tissues, although minimal co-localization was observed. The significance of this distribution is unclear. wCfeJ may relocate to ovaries upon a mating cue, or perhaps exert CI via some undefined somatic cell mechanism. Transmission of wCfeJ may also be predominantly horizontal via environmental or direct flea contact (per (Pietri, DeBruhl & Sullivan, 2016)). Sampling co-infected fleas at specific timepoints during development, as well as determining wCfeJ tissue localization in the absence of wCfeT infection, will be required to unravel the intertwined transmission dynamics of these Wolbachia strains in cat fleas.

**Evolution of Wolbachia mutualism and reproductive parasitism**

Order Rickettsiales is a highly diverse group of obligate intracellular bacteria (Gillespie et al., 2012; Driscoll et al., 2013; Darby et al., 2007). A rapid pace of newly identified lineages (Klinges et al., 2019; Castelli et al., 2019; Tashyrev et al., 2018; Kamm et al., 2019; Gruber-Vodicka et al., 2019; Martijn et al., 2015; Schulz et al., 2016; Yurchenko et al., 2018; Schrallhammer et al., 2013; Montagna et al., 2013; Bascaro et al., 2019; Muñoz-Gómez et al., 2019; Mediannikov et al., 2014; Szokoli et al., 2016; Prokopchuk et al., 2019) provides a rich source for evaluating mechanisms of evolution within constraints of the eukaryotic cell. Phage, plasmids, transposons, ICEs, and other insertion sequences are variably present across rickettsial genomes and mobilize numerous genes to offset reductive genome evolution and provide lifestyle altering traits (Gillespie et al., 2012). Our analyses of Wolbachia BOOM and RP genes prompted us to evaluate the recurrent evolution of nutritional-mediated mutualism and reproductive parasitism across diverse rickettsial lineages (Fig. 6).

**Nutritional-mediated mutualism**

Phylogeny estimation supports three independent acquisitions of biotin biosynthesis genes by rickettsial species (Fig. 6A; Fig. S1A). The largest clade (group 1) contains all BOOM-containing species, though only Neorickettsia spp., certain wolbachiae, R. buchneri, and the Peranema trichophorum endosymbiont carry the conserved BOOM gene arrangement (highlighted yellow in Fig. 6A). Other species in this group, as well as Groups 2 and 3, show different gene arrangements indicating recombination and/or possible replacement of specific bio genes through additional LGT (Fig. S3). BOOM and other bio gene clusters have repeatedly lost bioH, which encodes the methyl esterase generating pimeloyl-CoA, the last step in the synthesis of the pimelate moiety of biotin. Bacteria have evolved numerous strategies for catalyzing this reaction (Lin & Cronan, 2011; Shapiro, Chakravartty & Cronan, 2012; Bi et al., 2016; Feng et al., 2014; Zeng et al., 2020; Chow et al., 2018; Shi et al., 2016; Estrada et al., 2017; Sullivan et al., 2001). Interestingly, a novel methyl esterase of Ehrlichia chaffeensis (BioU) was recently shown to complement an E. coli ΔbioH mutant (Hang et al., 2019); however, we determined that BioU orthologs are present in many rickettsial species, including some that harbor a BioH already and others with no bio synthesis genes at all (Figs. S4A–S4C). This indicates that while BioU may generate pimeloyl-CoA in certain rickettsial species—namely,
Figure 6 Recurrent evolution of nutritional-mediated mutualism and reproductive parasitism across diverse rickettsial lineages. (A) Diverse strategies for biotin biosynthesis, acquisition and utilization in Rickettsiales. Top, biotin synthesis genes (BOOM arrangement), the BioY biotin transporter, the biotin ligase BirA, and biotin utilization enzymes (biotin-dependent propionyl-CoA carboxylase complex and FabD) involved in malonyl-CoA synthesis. The distribution of these genes across major rickettsial lineages is summarized, with phylogeny of Rickettsiales drawn as a consensus from recent studies (Driscoll et al., 2017; Castelli et al., 2019; Muñoz-Gómez et al., 2019; Szokoli et al., 2016). The acquisition of biotin synthesis operons is depicted by yellow, blue and orange highlighting over branches (see estimated phylogeny in Fig. 1A). Question marks indicate missing BioH enzymes in otherwise complete biotin synthesis pathways. These genomes were scanned for other functionally equivalent methyl esterases and found to contain none (Fig. S4D). Inset highlights the gatekeeper role of BioH in shunting X Pimeloyl-CoA methyl ester (metabolite 2) away from fatty acid biosynthesis for the production of biotin. All other metabolites in the biotin synthesis pathway are described in Fig. 2A. X, complete loss of biotin utilization in Driscoll et al. (2020), PeerJ, DOI 10.7717/peerj.10646.
those with complete bio synthesis modules except for BioH—it is likely to have a broad range of functions in bacteria (Fig. S4D). Furthermore, some species lack the genes for both BioH and BioU, indicating that alternative pathways for generating pimeloyl-CoA await discovery.

Analysis of the genomic distribution of biotin synthesis genes, the biotin transporter BioY, the biotin ligase BirA, and the enzymes involved in malonyl-CoA synthesis (biotin-dependent propionyl-CoA carboxylase complex and FabD) revealed several strategies for biotin utilization in Rickettsiales (Fig. 6A). Species that harbor a complete set of bio genes (as discussed above) also contain ligase and utilization enzymes, although (not surprisingly) most do not carry the BioY importer. Genera Wolbachia and Rickettsia largely rely on parasitic scavenging of host biotin to fuel fatty acid biosynthesis: they are dominated by species that lack bio genes but carry the BioY importer, BirA ligase, and utilization enzymes. Two species of Transitional Group Rickettsiae (R. felis and R. hoogstraalii) may be able to utilize imported dethiobiotin as well, as they retain the biotin synthase BioB; alternatively, this may represent the frayed end of a decaying metabolic pathway similar to the glycerol phospholipid and folate biosynthesis pathways in Rickettsiae (Driscoll et al., 2017; Audia, 2012). While characterized as two-component transportsers (with BioM/N), BioY subunits alone can efficiently transport biotin into cells (Hebbeln et al., 2007) and also participate with host transporters to relocate cytosolic biotin to vacuoles occupied by pathogens (Fisher et al., 2012). However, certain rickettsial species lacking synthesis genes also lack BioY importers yet retain the ligase and utilization enzymes, indicating the presence of alternative transport mechanisms. Remarkably, Orientia spp. lack all genes involved in biotin synthesis, import, ligation and utilization, exhibiting the ultimate form of parasitism for host-dependent fatty acid synthesis: these Rickettsiae must pilfer malonyl-CoA from the host (Nakayama et al., 2008), using a transport system yet to be described.

The acquisition of BOOM by certain wolbachiae and R. buchneri provides a theoretical framework for how LGT facilitates a transition from metabolite thievery to nutritional-mediated symbiosis. However, it is important to consider that these endoparasites may be utilizing the BOOM primarily for selfish purposes, as almost all bacteria require biotin. For example, Neorickettsia spp. and the deltaproteobacterium Lawsonia intracellularis are pathogens that also harbor the BOOM, suggesting a more complex ecological role surrounding biotin acquisition. If Neorickettsia spp. do indeed actively provision biotin to their trematode and insect vectors, as has been shown for wCle (Nikoh et al., 2014),
such host-specific supplementation would necessitate as-yet unidentified genes encoding export factors as well as stringent regulation of the BOOM.

Canonical regulation of biotin synthesis is accomplished via a repressor, fused to the N-terminal domain of the biotin ligase gene, which down-regulates transcription of the bio gene cluster in response to high cellular biotin (Barker & Campbell, 1981; Satiaputra et al., 2016). Interestingly, all rickettsial BirA biotin ligases lack this repressor domain, possibly indicating alternative mechanisms for regulating biotin synthesis and host supplementation by these bacteria. “Candidatus Aquarickettsia rohweri” and the Trichoplax sp. H2 endosymbiont (Midichloriaceae) lack BirA ligases altogether (Fig. 6A), and no other biotin ligases were found in these genomes. Regarding the Trichoplax sp. H2 endosymbiont, our prior analysis of the closely related Rickettsiales endosymbiont of Trichoplax adhaerens (RETA) discovered two biotin-protein ligase N-terminal domain-containing proteins (BPL-N) encoded in the T. adhaerens genome (Driscoll et al., 2013). These host-encoded genes contain introns and eukaryotic regulatory elements yet are strongly supported as rickettsial in origin. Homologs of these T. adhaerens proteins are also present in the Trichoplax sp. H2 genome (Kamm et al., 2018) (NCBI acc. nos. RDD42697 and RDD42436), supporting their cross-domain LGT. This raises the possibility for a metabolic interdependence involving biotin synthesis and utilization in placozoans and their rickettsial symbionts, similar to the mosaic metabolic networks underpinning nutritional symbioses in other systems (Bublitz et al., 2019; McCutcheon, 2010; Wilson & Duncan, 2015). For example, the nested tripartite symbiosis between Planococcus citri mealybugs, Tremblaya princeps, and Moranella endobia shows evidence for LGT of several bio genes to the mealybug genome (Husnik et al., 2013); these genes share highest similarity to Wolbachia BOOM counterparts. This is reminiscent of the pea aphid-Buchnera symbiosis, wherein several rickettsial LGTs have been implicated in facilitating metabolic interdependence between host and symbiont (Nikoh et al., 2010). For rickettsial endosymbionts of placozoans, host acquisition of BPL-N genes (among others (Driscoll et al., 2013)) may have established an inextricable obligatory relationship (McCutcheon, 2016; McCutcheon, Boyd & Dale, 2019) and may hint at the evolutionary fate of other BOOM-containing rickettsiae. However, while our analyses of the BOOM and other bio genes in rickettsial genomes are suggestive, in the absence of experimental validation we urge caution against implicating these genes as factors underpinning nutritional-mediated mutualism.

Reproductive parasitism

Among other bacterial reproductive parasites of arthropods (e.g., species of Cardinium, Arsenophonus and Spiroplasma) wolbachiae are perhaps the best-studied for their roles in RP induction (Hurst & Frost, 2015). Yet within the Rickettsiales, wolbachiae do not walk alone in inducing RP in their hosts, as three other lineages contain species known to induce RP phenotypes (Orientia tsutsugamushi, basal Rickettsia species and Transitional Group Rickettsiae). Inspection of genomic data indicates SFG Rickettsiae may also harbor reproductive parasites (Fig. 6B). These analyses reveal four discrete molecular determinants for RP induction that have evolved throughout rickettsial evolution.
(cinAB, cidAB, cndAB, wmk), though others likely exist based on previous description of numerous gene models sharing commonalities across toxins and antidote architectures (Gillespie et al., 2018). We posit at least four independent gains of RP in the Rickettsiaceae have occurred, coupled with widespread RP seen in wobachiae (Fig. 6B). This speculative interpretation is more parsimonious than treating RP as an ancestral rickettsial trait, given that RP-inducing genes strongly associate with mobile genetic elements (e.g., WO prophage in Wolbachia strains, RAGE ICEs and plasmids in Rickettsia species). Additionally, were RP an ancestral rickettsial trait, one would expect to see cifs in Rickettsia genomes more frequently than what is observed based on analysis of available genomes (Gillespie et al., 2018).

In contrast to rickettsiae, RP-inducing genes are found in greater numbers and diversity in wobachiae, undoubtedly due to their strong association with EAMs of WO prophage genomes and mobile elements (Bordenstein & Bordenstein, 2016; Lindsey et al., 2018). In our previous work, we uncovered a link between the characterized CI-inducing CinA/B and CidA/B operons of wobachiae and a putative TA operon encoded on the pLbAR plasmid of Rickettsia felis str. LSU-Lb, an obligate endosymbiont of the booklouse Liposcelis bostrychophila (Gillespie et al., 2015). We posited that this TA operon induces parthenogenesis in the booklouse: pLbAR is unknown in other strains of R. felis which do not induce parthenogenesis, in booklice or other arthropods (Gillespie et al., 2015), and booklice lacking R. felis str. LSU-Lb reproduce sexually (Yang et al., 2015; Mockford & Krushelnycky, 2008). Aside from being a CndB toxin (harboring both the NUC and DUB domains of wobachiae CinB and CidB toxins, respectively) the large pLbAR toxin (3,241 aa) was found to carry several other domains present in other Wolbachia proteins, most of which are also EAM-associated. These domains, together with the large central domain (LCD) of the pLbAR toxin, were utilized to identify an extraordinarily diverse array of candidate RP-inducing toxins in a wide-range of intracellular bacteria, some of which are known reproductive parasites (Gillespie et al., 2018). Linked within the intracellular mobilome, we consider this vast family of proteins a source for the recurrent evolution of RP in invertebrate-borne bacteria.

Aside from CI induction by CinA/B and CidA/B operons (Chen et al., 2019; Beckmann et al., 2019; LePage et al., 2017; Beckmann, Ronau & Hochstrasser, 2017), virtually nothing is known about the functions of related candidate RP-inducing genes. The function of the CinA/B-like operon of O. tsutsugamushi is unclear. O. tsutsugamushi is the pathogen causing scrub typhus and is capable of inducing parthenogenesis in Leptotrombidium mites (Roberts, Rapmund & Cadigan, 1977; Takahashi et al., 1997). Given this fact, exploration of its corresponding CinA/B module is warranted. Though the connection to RP in mites is unclear given that CinA/B only occurs in the Boryong strain; a myriad of other proteins encoding the pLbAR LCD are amplified in all Orientia genomes. Genome sequences are needed to determine toxin profiles for Rickettsia species either in basal lineages (Perotti et al., 2006; Gualtieri et al., 2017; Giorgini et al., 2010) or Transitional Group rickettsiae (Perotti et al., 2006; Hagimori et al., 2006; Nugnes et al., 2015) that induce parthenogenesis in eulophid wasps or trogiid booklice. Genome sequences are also lacking for Rickettsia species from basal lineages known to induce male-killing in
buprestid and ladybird beetles (Werren et al., 1994; Von der Schulenburg et al., 2001; Lawson et al., 2001; Majerus & Majerus, 2010). We previously mined the genome of the male killer Rickettsia parasite of the ladybird beetle Adalia bipunctata (Murray et al., 2016) and found a toxin highly similar in architecture to the Spiroplasma male killer toxin (Harumoto & Lemaitre, 2018). Both toxins carry an ovarian tumor (OTU) cysteine protease domain (Makarova, Aravind & Koonin, 2000), which is a deubiquitylating enzyme. Further searching with the OTU domain resulted in a hodgepodge of chimeric toxin architectures from bacteria known to induce other RP phenotypes, such as parthenogenesis (wFol (Faddeeva-Vakhrusheva et al., 2017)) and feminization (wVulC (Badawi et al., 2018)). Thus, it is clear that additional high-quality (closed) genomes are sorely needed to correlate RP-inducing protein domains and RP phenotypes.

New correlations will lead to downstream functional analyses of target’s abilities to induce RP.

A large number of candidate proteins from the pLbAR-derived assemblage occur in species that are not known as reproductive parasites. These include CndB toxins, NUC- and DUB-domain containing proteins that are much larger than characterized CinB and CidB toxins, diverse OTU-domain containing toxins, and especially proteins harboring the LCD that may or may not exhibit toxin-like profiles (Gillespie et al., 2018). Some of these candidate toxins appear to have cognate antidotes, and some genomes carry antidotes only, highlighting the complex diversity of this biological function. Recently, an EAM-encoded transcriptional factor in wMel termed WO-mediated killing (wmk) was shown to induce male-specific lethality during early embryogenesis when expressed transgenically in fruit flies (Perlmuter et al., 2019). This male-killer gene was not detected as part of our pLbAR-derived pool, indicating that RP-inducing gene arsenals are probably much larger than have been described thus far, particularly considering that cin, cid and cnd loci are not present in the genomes of Cardinium and Arsenophonus reproductive parasites (Gillespie et al., 2018; Beckmann et al., 2019; Gebiola et al., 2017) for which the mechanisms for RP induction remain unknown (Siozios et al., 2019; Duron et al., 2008).

Candidate RP-inducing toxins with similar features to the pLbAR toxin (large, modular, and associated with mobile genetic elements) are found in several Wolbachia genomes (e.g., wFol, wStr, wDacB, wVulC, wCon, wCed, and wCauA). They are often in conjunction with other cin and/or cid loci, yet not always associated with WO prophage. Consistent with prior reports identifying LGT between Rickettsia and Wolbachia genomes (Gillespie et al., 2012, 2015; Ishmael et al., 2009; Baldridge et al., 2016), similar toxins in Rickettsia genomes (R. felis str. LSU-Lb, R. gravesii, male killer of Adalia bipunctata) indicate a dynamic platform for disseminating RP-inducing genes across diverse rickettsial species that includes Rickettsia plasmids and WO prophage. Growing numbers of identified Rickettsia plasmids (Gillespie et al., 2015), proliferative RAGE ICEs in Rickettsia and Orientia genomes (Gillespie et al., 2012; Hagen et al., 2018; Nakayama et al., 2008; Batty et al., 2018; Cho et al., 2007; Nakayama et al., 2010), and the recent discovery of the first Wolbachia plasmid (Reveillaud et al., 2019), attest to a highly dynamic mobilome not previously realized for obligate intracellular bacteria. The proposed fission of a
CndB-like toxin into a wCfeJ CinB toxin, described in the current study (Fig. 3D), suggests larger toxins undergo streamlining into discrete inducers of CI or other RP phenotypes. As with the evolution of biotin utilization, Rickettsiales provides a framework to study the nature of RP in a diverse array of bacteria that enjoy different relationships with their invertebrate and vertebrate hosts.

CONCLUSION

The remarkable opposing features of wCfeT and wCfeJ offer testable hypotheses for contrasting host relationships (nutritional symbiosis vs RP, respectively). They also provide a framework to study the dynamics of Wolbachia co-infections, both in colonies and wild populations. The wCfeT and wCfeJ genomes yield lessons on the role of LGT in shaping host associations, highlighting the ephemeral nature of nutritional-based mutualism and RP over a general rickettsial platform of obligate intracellular parasitism. Host genome sequences provide clues to on-going mutualism and historical RP, and our hypothesis for host CI-gene capture leading to suppression of RP illuminates possible limitations for deployment of Wolbachia reproductive parasites to combat arthropod transmission of human pathogens. In summary, our robust characterization of two divergent strains of C. felis-associated wolbachiae, as well as the C. felis genome itself (Driscoll et al., 2020), provides essential information for determining the impact of these bacteria on cat flea biology and pathogen vectoral dynamics.

ACKNOWLEDGEMENTS

We acknowledge Dr. Michael Gerth (University of Liverpool) for sharing contigs from sequencing projects for Wolbachia endosymbiont of Zootermopsis nevadensis (wZoo), Wolbachia endosymbiont of Mengenilla moldrzyki (wMen), and Wolbachia endosymbiont of Ctenocephalides felis (wCte). We are grateful to Dr. Michael Dryden (Kansas State University), Dr. Bill Donahue (Sierra Research Laboratories, Modesto, CA), Dr. Lance Durden (Georgia Southern University), and Eric Dotseth (Division of Infectious Disease Epidemiology, West Virginia Department of Health and Human Resources) for contributing cat fleas.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was supported with funds from the National Institutes of Health National Institute of Allergy and Infectious Diseases grants (R21AI26108 and R21AI146773 to Joseph J. Gillespie, R01AI017828 and R01AI126853 to AFA, and R01AI122672 to Kevin R. Macaluso). Timothy P. Driscoll and Victoria I. Verhoeve were supported by start-up funding provided to Timothy P. Driscoll by West Virginia University. John F. Beckmann was supported by start-up funds from Auburn University, USDA Hatch Grant (1015922) and an Alabama Agricultural Experiment Station SEED grant. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.
Grant Disclosures
The following grant information was disclosed by the authors:
National Institutes of Health National Institute of Allergy and Infectious Diseases:
R21AI26108, R21AI146773, R01AI017828, R01AI126853 and R01AI122672.
West Virginia University.
Auburn University, USDA Hatch: 1015922.
Alabama Agricultural Experiment Station SEED.

Competing Interests
Timothy P. Driscoll and Joseph J. Gillespie are Academic Editors for PeerJ.

Author Contributions
- Timothy P. Driscoll conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Victoria I. Verhoeve conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Cassia Brockway performed the experiments, prepared figures and/or tables, and approved the final draft.
- Darin L. Shrewsberry performed the experiments, prepared figures and/or tables, and approved the final draft.
- Mariah Plumer performed the experiments, prepared figures and/or tables, and approved the final draft.
- Spiridon E. Sevdalis performed the experiments, prepared figures and/or tables, and approved the final draft.
- John F. Beckmann conceived and designed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Laura M. Krueger performed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Kevin R. Macaluso conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Abdu F. Azad conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Joseph J. Gillespie conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.

DNA Deposition
The following information was supplied regarding the deposition of DNA sequences:
All relevant information pertaining to wCfeT and wCfeJ is available at NCBI: PRJNA622233. Complete genome sequences are available in the NCBI RefSeq database: NZ_CP051156.1 (wCfeT) and NZ_CP051157.1 (wCfeJ).
Data Availability
The following information was supplied regarding data availability:

Data generated for this project, including raw PCR data for Fig. 5, are available at GitHub: https://www.github.com/wvuvectors/wCfe_genomes.

Supplemental Information
Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.10646#supplemental-information.

REFERENCES


Barker DF, Campbell AM. 1981. Genetic and biochemical characterization of the birA gene and its product: evidence for a direct role of biotin holoenzyme synthetase in repression of the biotin

DOI 10.1371/journal.pntd.0006566.


DOI 10.1038/nmicrobiol.2017.7.


DOI 10.1017/S0031182011002198.

DOI 10.1016/j.cell.2019.08.054.


of arthropod and nematode symbionts in the F supergroup, and evidence for further diversity in the Wolbachia tree. Microbiology 151(12):4015–4022 DOI 10.1099/mic.0.28313-0.


Leulmi H, Socolovsky C, Laudisio A, Houemenou G, Davoust B, Bitam I, Raoult D, Parola P. 2014. Detection of *Rickettsia felis*, *Rickettsia typhi*, *Bartonella* Species and *Yersinia pestis* in Fleas...


