Sciencexpress Report

# Widespread Lateral Gene Transfer from Intracellular Bacteria to Multicellular Eukaryotes

Julie C. Dunning Hotopp,[1]*† Michael E. Clark,[2]* Deodoro C. S. G. Oliveira,[2] Jeremy M. Foster,[3] Peter Fischer,[4] Mónica C. Muñoz Torres,[5] Jonathan D. Giebel,[2] Nikhil Kumar,[1] Nadeeza Ishmael,[1] Shiliang Wang,[1] Jessica Ingram,[3] Rahul V. Nene,[1]‡ Jessica Shepard,[1]§ Jeffrey Tomkins,[5] Stephen Richards,[6] David J. Spiro,[1] Elodie Ghedin,[1,7] Barton E. Slatko,[3] Hervé Tettelin,[1]|| John H. Werren[2]||

[1]The Institute for Genomic Research, J. Craig Venter Institute, 9712 Medical Center Drive, Rockville, MD 20850, USA. [2]Department of Biology, University of Rochester, Rochester, NY 14627, USA. [3]Molecular Parasitology Division, New England Biolabs Inc., 240 County Road, Ipswich, MA 01938, USA. [4]Department of Internal Medicine, Infectious Diseases Division, Washington University School of Medicine, St. Louis, MO 63110, USA. [5]Clemson University Genomics Institute, 304 BRC, 51 New Cherry St, Clemson, SC 29634, USA. [6]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA. [7]Division of Infectious Diseases, University of Pittsburgh School of Medicine, Pittsburgh, PA 15261, USA.

*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: jdunning@jcvi.org

‡Present address: Brown University, Providence, RI 02912, USA.

§Present address: Pace University, New York, NY 10038, USA.

||These authors contributed equally to this work.

**Although common among bacteria, lateral gene transfer—the movement of genes between distantly related organisms—is thought to occur only rarely between bacteria and multicellular eukaryotes. However, the presence of endosymbionts, such as *Wolbachia pipientis*, within some eukaryotic germlines may facilitate bacterial gene transfers to eukaryotic host genomes. We therefore examined host genomes for evidence of gene transfer events from *Wolbachia* bacteria to their hosts. We found and confirmed transfers into the genomes of 4 insect and 4 nematode species that range from nearly the entire *Wolbachia* genome (>1 megabase) to short (<500 base pairs) insertions. Potential *Wolbachia* to host transfers were also detected computationally in three additional sequenced insect genomes. We also show that some of these inserted *Wolbachia* genes are transcribed within eukaryotic cells lacking endosymbionts. Therefore, heritable lateral gene transfer occurs into eukaryotic hosts from their prokaryote symbionts, potentially providing a mechanism for acquisition of new genes and functions.**

The transfer of DNA between diverse organisms, lateral gene transfer (LGT), facilitates the acquisition of novel gene functions. Among Eubacteria, LGT is involved in the evolution of antibiotic resistance, pathogenicity, and metabolic pathways (*1*). Rare LGT events have also been identified between higher eukaryotes with segregated germ cells (*2*), demonstrating that even these organisms can acquire novel DNA. Although most described LGT events occur within a single domain of life, LGT has been described both between Eubacteria and Archaea (*3*) and between prokaryotes and phagotrophic unicellular eukaryotes (*4*, *5*). However, few interdomain transfers involving higher multicellular eukaryotes have been found.

*Wolbachia pipientis* is a maternally inherited endosymbiont that infects a wide range of arthropods, including at least 20% of insect species, as well as filarial nematodes (*6*). It is present in developing gametes (*6*) and so provides circumstances conducive for heritable transfer of bacterial genes to the eukaryotic hosts. *Wolbachia*-host transfer has been described in the bean beetle *Callosobruchus chinensis* (*7*) and in the filarial nematode *Onchocerca* spp. (*8*).

We have found *Wolbachia* inserts in the genomes of additional diverse invertebrate taxa, including fruit flies, wasps, and nematodes. A comparison of the published genome of the *Wolbachia* endosymbiont of *Drosophila melanogaster* (*9*) and assemblies of *Wolbachia* clone mates (*10*) from fruit fly whole genome shotgun sequencing data revealed a large *Wolbachia* insert in the genome of the widespread tropical fruit fly *Drosophila ananassae*. Numerous contigs were found that harbored junctions between *Drosophila* retrotransposons and *Wolbachia* genes.

The large number of these junctions and the deep sequencing coverage across the junctions indicated that these inserts were probably not due to chimeric libraries or assemblies. To validate these observations, five *Drosophila-Wolbachia* junctions were PCR amplified and three end-sequence verified. Fluorescence *in situ* hybridization (FISH) of banded polytene chromosomes with fluorescein-labeled probes of two *Wolbachia* genes (*11*) revealed the presence of *Wolbachia* genes on the 2L chromosome of *D. ananassae* (Fig. 1).

We found that nearly the entire *Wolbachia* genome was transferred to the fly nuclear genome as evidenced by the presence of PCR amplified products from 44/45 physically distant *Wolbachia* genes from cured strains of *D. ananassae* Hawaii verified by microscopy to be lacking the endosymbiont after treatment with antibiotics (fig. S1) (*11*). In contrast, only spurious, incorrectly sized, and weak amplification was detected from a cured control line lacking these inserts (Townsville). The 45 genes assayed (table S1) are spaced throughout the *Wolbachia* genome. Thus the high proportion of amplified genes suggests gene transfer of nearly the entire *Wolbachia* genome to the insect genome.

A 14 kb region containing four *Wolbachia* genes with two retrotransposon insertions was sequenced (*11*) from a single BAC, constituting an independent source of DNA as compared to the largely plasmid-derived whole genome sequence of *D. ananassae*. The two retroelements each contained 5 bp target site duplications (9/10 bp identical), long terminal repeats, and *gag-pol* genes (Fig. 2A) indicating that the *Wolbachia* insert is accumulating retroelements. Insertion of this region appears to be recent, as shown by the nearly identical target site duplications and >90% nucleotide identity between the corresponding endosymbiont genes and the sequenced homologs in the *D. ananassae* chromosome.

Crosses between *Wolbachia*-free Hawaii males (with the insert) and *Wolbachia*-free Mexico females (without the insert) revealed that the insert is paternally inherited by offspring of both sexes, confirming that *Wolbachia* genes are inserted into an autosome. Since *Wolbachia* infections are maternally inherited this also confirms that PCR amplification in the antibiotic treated line is not due to a low level infection. Furthermore, the Hawaii and Mexico crosses revealed Mendelian, autosomal inheritance of *Wolbachia* inserts (paternal N = 57, $k = 0.49$; maternal N = 40, $k = 0.58$). Six physically distant, inserted *Wolbachia* genes perfectly co-segregated in F2 maternal inheritance crosses (*11*), suggesting they also are closely linked.

PCR amplification and sequencing (*11*) of 45 *Wolbachia* loci in 14 *D. ananassae* lines from widely dispersed geographic locations revealed large *Wolbachia* inserts in lines from Hawaii, Malaysia, Indonesia, and India (table S2). Sequence comparisons of the amplicons from these four lines revealed that all ORFs remained intact with >99.9% identity between inserts. This is compared to an average 97.7% identity for the inserts compared to *w*Mel, the *Wolbachia* endosymbiont of *D. melanogaster*. These results indicate the widespread prevalence of *D. ananassae* strains with similar inserts of the *Wolbachia* genome, probably due to a single insertion from a common ancestor.

In addition, RT-PCR followed by sequencing (*11*) demonstrated that ~2% of *Wolbachia* genes (28 of 1206 genes assayed; table S3) are transcribed in cured adult males and females of *D. ananassae* Hawaii. The complete 5' sequence of one of the transcripts, WD_0336, was obtained with 5'-RACE on uninfected flies (*11*) suggesting that this transcript has a 5' mRNA cap, a form of eukaryotic post-transcriptional modification. Analysis of the transcript levels of inserted *Wolbachia* genes with qRT-PCR (*11*) revealed that they are $10^4$-fold to$10^7$-fold less abundant than the fly's highly transcribed Actin gene (*act5C*; table S3). There is no cutoff that defines a biologically relevant level of transcription, and assessment of transcription in whole insects can obscure important tissue specific transcription. Therefore, it is unclear whether these transcripts are biologically meaningful, and further work is needed to determine their significance.

Screening of public shotgun sequencing data sets has identified several additional cases of LGT in different invertebrate species. In *Wolbachia*-cured strains of the wasp *Nasonia*, six small *Wolbachia* inserts (<500 bp) were verified by PCR and sequencing (*11*) that have >96% nucleotide identity to native *Wolbachia* sequences, in some cases with short insertion site duplications. These include four in *Nasonia vitripennis;* one in *Nasonia giraulti;* and one in *Nasonia longicornis* (table S4; Fig. 2B). Amplification and sequencing of 14-18 geographically diverse strains of each species indicated that the inserts are species-specific. For example, three *Wolbachia* inserts in *N. vitripennis* are not found in the closely related species *N. giraulti* or *N. longicornis*, which diversified ~1 million years ago (*12*). These data suggest that the *Wolbachia* gene inserts are of relatively recent origin, similar to the inserts in *D. ananassae*.

Nematode genomes also contain inserted *Wolbachia* sequences. As *Wolbachia* infection is required for fertility and development of the worm *Brugia malayi*, the genomes of both organisms were sequenced simultaneously complicating assemblies and leading to the removal of *Wolbachia* reads during genome assembly [>98% identity over 90% of the read length on the basis of the independent BAC-based genome sequence of *w*Bm, the *Wolbachia* endosymbiont of *B. malayi* (*13*)]. Despite this, the genome of *B. malayi* contains 249 contigs with *Wolbachia* sequences (e-value < $10^{-40}$); nine of which were confirmed by long range PCR and end-sequencing (*11*). These include eight large scaffolds containing >1 kb *Wolbachia* fragments within 8 kb

of a *B. malayi* gene (table S5). Comparisons of *w*Bm homologs to these regions suggested that all of these *Wolbachia* genes within the *B. malayi* genome are degenerate. In addition, a single region <1 kb was examined that contains a degenerate fragment of the *Wolbachia* aspartate aminotransferase gene (Wbm0002). Its location was confirmed by PCR and sequencing in *B. malayi* as well as *B. timori* and *B. pahangi* (*11*).

Of the remaining 21 arthropod and nematode genomes in the trace repositories (*11*), we found six containing *Wolbachia* sequences. Potential host-*Wolbachia* LGT was detected in three: *Drosophila sechellia*, *Drosophila simulans*, and *Culex pipiens* (Table 1), as revealed by the presence of reads containing homology to both endosymbiont and host genomes (*11*).

The sequencing of *w*Bm also facilitated the discovery of a *Wolbachia* insertion in *Dirofilaria immitis* (dog heartworm). The *D. immitis Dg2* chromosomal region encoding the D34 immunodominant antigen (*14*, *15*) contains *Wolbachia* DNA within its introns and in the 5'-UTR (Fig. 2C). These *Wolbachia* genomic fragments have maintained synteny with the *w*Bm genome (*13*), suggesting they may have inserted as a single unit and regions were replaced by exons of *Dg2*. A second gene (*DgK*) has been identified in other *D. immitis* lines that has 91% nucleotide identity in the exon sequences but contains differing number, position, size, and sequence of introns (*16*) and has no homology to known *Wolbachia* sequences.

Whole eukaryote genome sequencing projects routinely exclude bacterial sequences on the assumption that these represent contamination. For example, the publicly available assembly of *D. ananassae* does not include any of the *Wolbachia* sequences described here. Therefore, the argument that the lack of bacterial genes in these assembled genomes indicates that bacterial LGT does not occur is circular and invalid. Recent bacterial LGT to eukaryotic genomes will continue to be difficult to detect if bacterial sequences are routinely excluded from assemblies without experimental verification. And these LGT events will remain understudied despite their potential to provide novel gene functions and impact arthropod and nematode genome evolution. Because *W. pipientis* is among the most abundant intracellular bacteria (*17*, *18*), and its hosts are among the most abundant animal phyla, the view that prokaryote to eukaryote transfers are uncommon and unimportant needs to be reevaluated.

**References and Notes**


1. Y. Boucher *et al.*, *Annu Rev Genet* **37**, 283 (2003).
2. S. B. Daniels, K. R. Peterson, L. D. Strausbaugh, M. G. Kidwell, A. Chovnick, *Genetics* **124**, 339 (1990).
3. K. E. Nelson *et al.*, *Nature* **399**, 323 (1999).
4. J. O. Andersson, *Cell Mol Life Sci* **62**, 1182 (2005).
5. W. F. Doolittle, *Trends Genet* **14**, 307 (1998).
6. R. Stouthamer, J. A. Breeuwer, G. D. Hurst, *Annu Rev Microbiol* **53**, 71 (1999).
7. N. Kondo, N. Nikoh, N. Ijichi, M. Shimada, T. Fukatsu, *Proc Natl Acad Sci U S A* **99**, 14280 (2002).
8. K. Fenn *et al.*, *PLoS Pathog* **2**, e94 (2006).
9. M. Wu *et al.*, *PLoS Biol* **2**, e69 (2004).
10. S. L. Salzberg *et al.*, *Genome Biol* **6**, R23 (2005).
11. Materials and methods are available as supporting material on *Science* Online.
12. B. C. Campbell, J. D. Steffen-Campbell, J. H. Werren, *Insect Mol Biol* **2**, 225 (1993).
13. J. Foster *et al.*, *PLoS Biol* **3**, e121 (2005).
14. S. Sun, K. Sugane, *J Helminthol* **68**, 259 (1994).
15. S. H. Sun, T. Matsuura, K. Sugane, *J Helminthol* **65**, 149 (1991).
16. K. Sugane, K. Nakayama, H. Kato, *J Helminthol* **73**, 265 (1999).
17. J. H. Werren, *Annu Rev Entomol* **42**, 587 (1997).
18. J. H. Werren, D. M. Windsor, *Proc Biol Sci* **267**, 1277 (2000).

19. The *D. ananassae* BAC 01L18 sequence is deposited in Genbank (EF426679), the *D. ananassae* sequence comparisons from the 4 lines are deposited in Genbank (EF611872-EF611985), the *D. ananassae* RACE sequence is deposited in dbEST (46867557) and Genbank (ES659088), the *Nasonia* sequences are deposited in Genbank (EF588824-EF588901), the *Brugia malayi* contigs are available in Genbank (DS237653, DS238272, DS238705, DS239028, DS239057, DS239291, DS239377, DS239315), the *Brugia malayi* scaffolds are available in Genbank (AAQA01000958, AAQA01000097, AAQA01001500, AAQA01000425, AAQA01001819, AAQA01001498, AAQA01000384, AAQA01001952, AAQA01000736, AAQA01000571, AAQA01000369), and the microarray primers sequences that were used for RT-PCR are deposited in ArrayExpress as (A-TIGR-28).

This work was supported by an NSF grant to JHW and HT, an NIAID grant to EG, and New England Biolabs Inc. support to BES.


**Supporting Online Material**

www.sciencemag.org/cgi/content/full/1142490/DC1
Materials and Methods
Fig. S1
Tables S1 to S5
References


13 March 2007; accepted 2 July 2007
Published online 30 August 2007; 10.1126/science.1142490
Include this information when citing this paper.

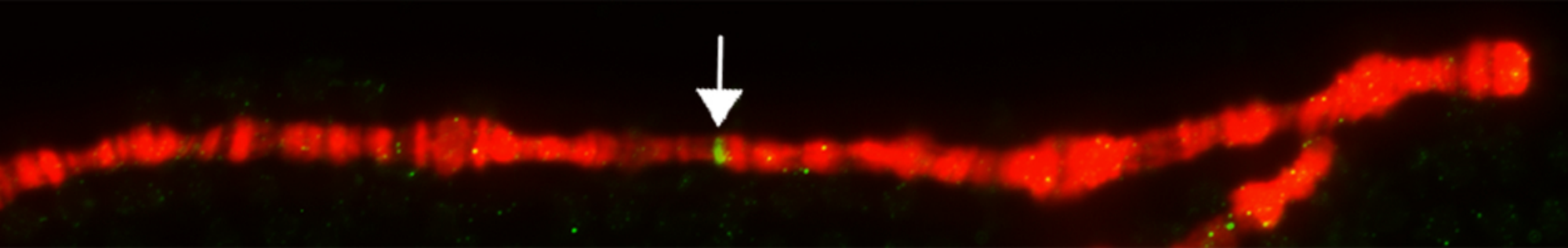
**Fig. 1.** Fluorescence microscopy evidence supporting *Wolbachia*/host LGT. DNA in the polytene chromosomes of *D. ananassae* are stained with propidium iodide (red) while a probe for the *Wolbachia* gene WD_0484 binds to a unique location (green, arrow) on chromosome 2L.

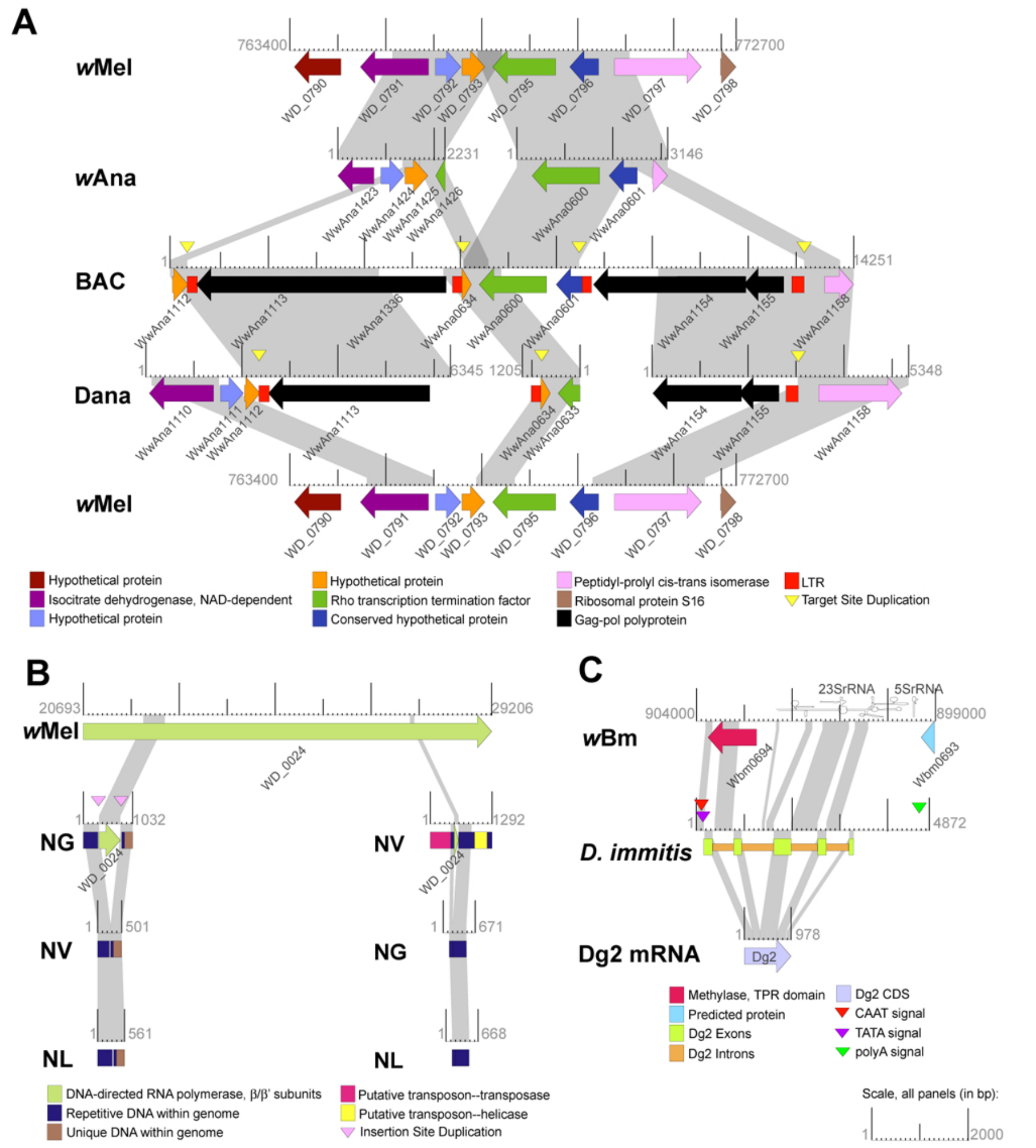
**Fig. 2.** Schematics of *Wolbachia* inserts in host chromosomes. A. Contigs containing *Wolbachia* sequences generated from the *D. ananassae* Hawaii shotgun sequencing project are segregated into sequences coming from the endosymbiont (*w*Ana) or from the *D. ananassae* chromosome (Dana) on the basis of the presence/absence of eukaryotic genes in the contigs. These are compared to those from the reference *D. melanogaster Wolbachia* genome (*w*Mel) and a *D. ananassae* BAC. B. Fragments of the *Wolbachia* gene WD_0024 gene have inserted into different positions in the *N. giraulti* (NG) and *N. vitripennis* (NV) genomes with unique insertions in each lineage, including *N. longicornis* (NL). C. A region in the *D. immitis* genome that is transcribed has introns similar to *Wolbachia* sequences. All matches in panels A and B have >90% nucleotide identity; those in panel C are >75% nucleotide identity.

**Table 1.** Summary of *Wolbachia* sequences and evidence for LGT in public databases.

| Organism | Total traces screened | *Wolbachia* traces | LGT | Junctions validated* | *Wolbachia* infection† |
|---|---|---|---|---|---|
| **Trace repository sequences‡** | | | | | |
| *Acyrthosiphon pisum* (aphid) | 4,285,120 | 0 | | | + |
| *Aedes aegypti* (mosquito) | 16238263 | 0 | | | - |
| *Anopheles gambiae* (mosquito) | 5,456,630 | 0 | | | - |
| *Apis mellifera* (honeybee) | 3,941,137 | 0 | | | - |
| *Brugia malayi* (filarial nematode) | 1,260,214 | 22524 | + | 10/12 | + |
| *Culex pipiens quinquefasciatus* (mosquito) | 7,380,430 | 21304 | + | 0/0 | + |
| *Daphnia pulex* (crustacean) | 2,724,768 | 0 | | | - |
| *Drosophila ananassae* (fruit fly) | 3,878,537 | 38605 | + | 6/7 | + |
| *Drosophila erecta* (fruit fly) | 2,916,936 | 0 | | | - |
| *Drosophila grimshawi* (fruit fly) | 2,874,111 | 0 | | | - |
| *Drosophila melanogaster* (fruit fly) | 1,001,855 | 0 | | | + |
| *Drosophila mojavensis* (fruit fly) | 3,130,180 | 107§ | - | | - |
| *Drosophila persimilis* (fruit fly) | 1,375,313 | 0 | | | - |
| *Drosophila pseudoobscura* (fruit fly) | 5,161,792 | 0 | | | - |
| *Drosophila sechellia* (fruit fly) | 1,203,722 | 1 | + | 0/0 | + |
| *Drosophila simulans* (fruit fly) | 2,321,958 | 7473 | + | 0/0 | + |
| *Drosophila virilis* (fruit fly) | 3,632,492 | 0 | | | - |
| *Drosophila willistoni* (fruit fly) | 2,332,565 | 2519 | - | | + |
| *Drosophila yakuba* (fruit fly) | 2,269,952 | 0 | | | + |
| *Ixodes scapularis* (tick) | 13,088,763 | 44 | - | | + |
| *Nasonia giraulti* (wasp) | 540,102 | 2 | + | 1/1 | + |
| *Nasonia longicornis* (wasp) | 447,736 | 1 | + | 1/1 | + |
| *Nasonia vitripennis* (wasp) | 3,360,694 | 30 | + | 4/4 | + |
| *Pediculus humanus* (head louse) | 1,480,551 | 0 | | | + |
| *Pristonchus pacificus* (nematode) | 2,292,543 | 0 | | | - |
| *Tribolium castaneum* (beetle) | 1,918,906 | 0 | | | - |
| **Genbank sequences** | | | | | |
| *Dirofilaria immitis* (filarial nematode) | NA | NA | + | | + |

*Junctions were validated by PCR amplification and sequencing (*11*) with the number of successful reactions compared to the number attempted. †These species are described in the literature as being infected with *Wolbachia*. ‡All whole genome shotgun sequencing reads were downloaded for 26 arthropod and nematode genomes (*11*). Organisms identified as lacking *Wolbachia* sequences either had no match or matches only to the prokaryotic rRNA. Since the *Nasonia* genomes are from antibiotic-cured insects, they were identified as having a putative LGT event merely on identification of *Wolbachia* sequences in a read. All other organisms were considered to have putative LGT events if the trace repository contained ≥1 read with (i) >80% nucleotide identity over 10% of the read to a characterized eukaryotic gene, (ii) >80% identity over 10% of the read to a *Wolbachia* gene, and (iii) manual review of the BLAST results for 1 to 20 reads to ensure significance (*11*). §This isolate was previously shown to have *Wolbachia* reads in its trace repositories that are contaminating reads from the *D. ananassae* genome sequencing project (*10*).




A
wMel
763400
772700
WD_0790
WD_0791
WD_0792
WD_0793
WD_0795
WD_0796
WD_0797
WD_0798
wAna
1
2231
1
3146
WwAna1423
WwAna1424
WwAna1425
WwAna1426
WwAna0600
WwAna0601
BAC
1
14251
WwAna1112
WwAna1113
WwAna1336
WwAna0634
WwAna0600
WwAna0601
WwAna1154
WwAna1155
WwAna1158
Dana
1
6345
1205
1
1
5348
WwAna1110
WwAna1111
WwAna1112
WwAna1113
WwAna0634
WwAna0633
WwAna1154
WwAna1155
WwAna1158
wMel
763400
772700
WD_0790
WD_0791
WD_0792
WD_0793
WD_0795
WD_0796
WD_0797
WD_0798
Hypothetical protein
Isocitrate dehydrogenase, NAD-dependent
Hypothetical protein
Hypothetical protein
Rho transcription termination factor
Conserved hypothetical protein
Peptidyl-prolyl cis-trans isomerase
Ribosomal protein S16
Gag-pol polyprotein
LTR
Target Site Duplication
B
20693
29206
wMel
WD_0024
1
1032
NG
WD_0024
1
501
NV
1
561
NL
1
1292
NV
WD_0024
1
671
NG
1
668
NL
DNA-directed RNA polymerase, β/β' subunits
Repetitive DNA within genome
Unique DNA within genome
Putative transposon--transposase
Putative transposon--helicase
Insertion Site Duplication
C
23SrRNA
5SrRNA
904000
899000
wBm
Wbm0694
Wbm0693
1
4872
D. immitis
1
978
Dg2 mRNA
Dg2
Methylase, TPR domain
Predicted protein
Dg2 Exons
Dg2 Introns
Dg2 CDS
CAAT signal
TATA signal
polyA signal
Scale, all panels (in bp):
1
2000