WestVirginiaUniversity
THE RESEARCH REPOSITORY @ WVU

Graduate Theses, Dissertations, and Problem Reports

2020

The Host Gatekeeper: Using the Flagellar Pathway to Understand Symbiont Host Adaptation

Adam R. Pollio
arpollio@mix.wvu.edu

Follow this and additional works at: https://researchrepository.wvu.edu/etd

 Part of the Bioinformatics Commons, Computational Biology Commons, Entomology Commons, Environmental Microbiology and Microbial Ecology Commons, Evolution Commons, Genetics Commons, Genomics Commons, and the Systems Biology Commons

Recommended Citation

Pollio, Adam R., "The Host Gatekeeper: Using the Flagellar Pathway to Understand Symbiont Host Adaptation" (2020). *Graduate Theses, Dissertations, and Problem Reports*. 7712.
https://researchrepository.wvu.edu/etd/7712

This Thesis is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Thesis in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Thesis has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.

WestVirginiaUniversity
THE RESEARCH REPOSITORY @ WVU

Graduate Theses, Dissertations, and Problem Reports

2020

The Host Gatekeeper: Using the Flagellar Pathway to Understand Symbiont Host Adaptation

Adam R. Pollio

Follow this and additional works at: https://researchrepository.wvu.edu/etd

 Part of the Bioinformatics Commons, Computational Biology Commons, Entomology Commons, Environmental Microbiology and Microbial Ecology Commons, Evolution Commons, Genetics Commons, Genomics Commons, and the Systems Biology Commons

The Host Gatekeeper: Using the Flagellar Pathway to Understand Symbiont Host Adaptation

Adam Pollio

Thesis **submitted**

to the Eberly College
at West Virginia University

in partial fulfillment of the requirements for the degree of

Master's **in**
Biology

Rita Rio, Ph.D., Chair
Tim Driscoll, Ph.D.
Craig Barret, Ph.D.
Department of Biology

Morgantown, West Virginia
2020

Keywords: Symbiosis, Sodalis
Copyright 2018 Adam Pollio

ABSTRACT

The Host Gatekeeper: Using the Flagellar Pathway to Understand Symbiont Host Adaptation

Adam Pollio

The acquisition of microbial partners is a strategy used by a diverse group of arthropods to overcome ecological barriers that might normally make certain niches uninhabitable. The unique phylogenetic opportunities attainable from the natural experiment of the Sodalis-allied clade allow for better understanding of how molecular structures evolve through time. Here, we focus on the evolution of the flagellar synthesis pathway, due to its complexity and ability to diverge in response to ecological pressures. We used this molecular pathway and natural experiment to show that normal evolutionary outcomes associated with symbiosis (i.e., genome reduction) do not explain the predicted conservation of the flagella genes or lack thereof within ancestral nodes.

Table of Contents

Chapter 1: Introduction

A. Symbiosis and endosymbiotic theory

Symbiosis is the long-term interaction between two or more different species and can range from parasitic to mutualistic and from obligate to facultative (*1*, *2*). These types of relationships drive organismal evolution in a process known as symbiogenesis (Hereafter: endosymbiotic theory) (*3*–*5*). The endosymbiotic theory was first proposed to explain the evolution of organelles (i.e. the mitochodreia, chloroplast). The theory describes the process of integration of free-living bacteria by another cell to form a new organism that is better suited to compete in the environment (*6*, *7*). Endosymbiosis enables a variety of environmental niches to be occupied by insects. For example, limited diets consisting solely of blood, wood, phloem, or xylem are possible due to the complementary metabolic capabilities provided by endosymbionts (*8*–*11*). Through time, the insect may both support and protect their mutualists through the formation of specialized organs where endosymbionts take up residence (*12*). Likewise, symbolic of co-evolution, endosymbiont genomes may evolve to accommodate a purely exclusive host-associated lifestyle (*13*, *14*). These changes often lead to an extreme reduction in genome size of the endosymbiont (*15*) through the purging of unnecessary genes via relaxed selection (*16*–*18*). Over time these symbiont genomes may become too reduced due to a process known as Mullers ratchet (*19*). The acquisition of secondary symbionts can further impact the genome reduction process(*20*). Primary symbionts are replaced by their secondary partners after they stop giving an advantage to their host (*21*, *22*). The transient nature of acquiring, reducing, and purging symbionts is thought to be the intermediary steps in organelle acquisition (*23*). Our current understanding of plastid evolution assumes the integration and acquisition of symbiont genes by the host to be utilized by more recent partners (*24*).

Here we focus on the well-studied tsetse fly (Diptera: *Glossinidae*). The fly has developed a microbiota that enables metabolic complementation to compensate for its exclusive diet of vertebrate blood (*25*, *26*).

B. Tsetse fly

Tsetse flies are the obligate vectors of African trypanosomes that cause both Human African Trypanosomiasis (HAT, also known as sleeping sickness) (*27*), and Animal African Trypanosomiasis (AAT, also known as nagana). Both diseases have significant impact towards public health and the economic development within affected areas of Africa (*28*, *29*). In addition to potentially harboring trypanosomes, tsetse flies have a simple (i.e. nominally) microbiome, likely due to the exclusive diet of blood and their viviparous mode of reproduction which limits the microbial exposure of the immature stages (*30*). The viviparous mode of reproduction in tsetse means larvae develop through three larval instars '*in utero'* and then get deposited and quickly enter pupation. The simple microbiome consists of three bacteria within the genera (*31*, *32*)

Wigglesworthia (*33*), *Sodalis* (*34*, *35*), and *Wolbachia* (*36*), each with differing roles towards the fitness of the fly (*26*, *37*, *38*).

C. Tsetse microbiota

i. *Wigglesworthia*

The primary endosymbiont, "*Candidatus* Wigglesworthia glossinidia" (hereafter *Wigglesworthia*), plays a vital role in the B vitamin synthesis and provisioning to the tsetse fly host as blood is deficient in these nutrients (*25*, *39–42*). The reliance on *Wigglesworthia* by the tsetse fly is exhibited by the development of a specialized organ (i.e. bacteriome) to exclusively house this symbiont (*33*). The bacteriome found at the anterior gut is the primary location of the intracellular *Wigglesworthia* population (*34*, *43*). A second population of *Wigglesworthia* may also be found in an accessory milk gland within females (*44–47*). The milk gland population transmits *Wigglesworthia* (*30*, *38*, *44*, *46*) to the developing larva in utero. The removal of *Wigglesworthia* via antibiotic treatment can lead to a reduction in relative fitness defined by increased mortality and lower fecundity (*32*, *47–51*). Tsetse fly fitness in flies lacking their *Wigglesworthia* symbionts may be restored by introducing specific nutrients such as a cocktail of B vitamins (*50*).

Intracellular organisms face unique conditions that lead to reductions in genome capabilities. Endosymbionts, because of their small population size, lack of recombination and limited exposure to horizontal gene transfer, have a propensity for the accumulation of deleterious mutations in a process known as Muller's ratchet (*17*, *19*). Furthermore, positive selection will eliminate genes that are unnecessary for its association with the host due to the overall fitness advantage (*52*). Genes that remain within endosymbiotic genomes are likely essential for the association, particularly those that are necessary for the continued survival within the fly (*53*). The *Wigglesworthia* genome has undergone endosymbiont genome reduction for approximately 50-80 million years (*54*), leaving ~ 600 genes within its estimated 700 kb genome, extremely small as compared to *E. coli* at 4.8 MB (*43*, *55*). The *Wigglesworthia* genes that have persisted play roles in its continued survival within the host by complementing host biology (*25*, *41*, *42*, *55*) and performing functions necessary for survival in the microenvironment. Interestingly, both annotated *Wigglesworthia* genomes retain the genetic inventory for flagella assembly (*43*, *55*), suggesting that this pathway is essential for its survival within the host.

In this study we will be focusing on a flagellar pathway that is regulated and shaped, not necessarily for direct use within the fly, but by environmental pressures selecting for organisms with highly similar flagella gene retention. The genome of the ancient obligate endosymbiont of aphids, *Buchnera*, also contains genes associated (*56*) with the flagellum basal body structure. The basal body is hypothesized to be repurposed for protein secretion within the aphid host (*57*). The conservation of flagella genes for this

structure in all *Buchnera* independent of specific aphid lineage suggests it too has a benefit involved with the aphid serving as a host filter (*56*).

ii. *Sodalis*-allied bacteria

The *Sodalis*-allied group is mostly composed of insect symbionts (*43*, *58*–*63*). Each organism, often defined by their host (e.g. SOPE, *Sitophilus oryzae* primary endosymbiont (*64*)), plays a role specific to their microenvironment. Therefore, each independent *Sodalis*-allied organism is under unique selective pressures that impact their genomic capabilities (*61*). Unlike other symbionts whose origin is unknown, this class of symbionts have a fully annotated free-living relative, *Sodalis praecaptivus* (*60*). The availability of the genome sequence of this close relative enables changes likely adaptive to the host symbiosis to be identified and compared. Moreover, the unique relationships of these organisms are ideal to understand the effect of the host towards genome evolution.

D. Bacterial flagellum

The eubacteria flagellum is among the smallest known biological molecular machines, the formation of which is both highly conserved and tightly regulated (*65*–*68*). These little “propellers”, powered by a transmembrane potential force, are the primary means of locomotion in many bacteria (*69*–*71*). In the representative Gammaproteobacteria, *Salmonella enterica* serovar Typhimurium and *Escherichia coli*, the flagellar biosynthesis pathway consists of approximately 50 genes dispersed among at least 17 operons (*72*–*75*) (Table 1 and Fig. 1). Although flagellar assembly is well conserved, there is some diversity regarding the location and the number of flagella within bacteria. This diversity can arise from duplication events leading to alternative ways of motility (*76*). Flagella can be synthesized differently depending on specific changes within the genome selected by environmental settings (*77*–*81*). Differences can range from a single polar flagellum used for swimming (*67*) to a set of lateral flagella employed for swarming (*81*, *82*). Each individual phenotype has a distinct advantage, (e.g., swarming motility has been shown to be more effective within a viscous environment while swimming is advantageous within aqueous settings (*81*, *83*)).

The transcription of the flagellar genes occurs through a tight temporally regulated pathway (*75*, *76*, *84*). Due to the relative importance of timing, the flagellar genes are categorized into temporal classes; I, II, and III (*72*, *85*). Genes are activated differentially depending on the specific regulatory tools available within the genome; the *Sodalis*-allied organisms all share synteny of two flagellar clusters but with markedly different stages of conservation (Fig. 2 & 3). In *E. coli* and *S. enterica*, the transcriptional activators, *flhC* and *flhD* (Table 1) comprise the master controller for the entire flagella system (*86*–*89*). The class I genes form a hetero-tetrameric transcription factor responsible for the expression of class II genes (*90*–*92*). The FlhC dimer can interact with promoter regions independently but partnering with FlhD improves site specificity and affinity for the specific class II promoters (*93*). Class II genes comprise the major

assembly proteins within this cascade including genes for the basal body, the switch complex and the hook (*76*, *84*, *94*, *95*). The first step of morphogenesis involves the synthesis of the basal body. The basal body is comprised of the MS ring (composed of multiple copies of the FliF protein), a P ring (made up of multiple copies of FlgI), L ring (made up of multiple copies of FlgH) and the basal rod (comprised of FlgB, FlgC, FliE, FlgF, FlgG) (*96*, *97*) (Fig. 4). The MS ring serves as the anchor for the switch complex (also known as the C ring composed of FliG, FliM and FliN), essential for torque generation in the presence of the class III motor proteins (MotA and MotB) (*98*–*102*). Both the C ring and motor proteins interact with chemotaxis proteins that drive movement. Interestingly, in both the *Wigglesworthia* and intact *Sodalis glossinidius* flagella clusters, no chemotaxis proteins are present (Fig. 3) suggesting an unknown mechanism for motor movement is likely necessary and indicating evolutionary diversification from other flagella.

Following the formation of the inner membrane rings, the cell undergoes the development of the flagella specific TTSS. The TTSS is tied closely with the formation of the filament due to its regulation of the anti-sigma factor (FlgM) by way of its subsequent secretion into the medium (*67*, *92*, *103*). FliA (sigma factor 28), once uncoupled from FlgM, can transcribe the class III genes leading to the formation of the final components necessary to achieve motility. FlgM is absent within the *Wigglesworthia* genome, suggesting an alternative mechanism has arisen for FliA regulation. The hook protein (FlgE), also known as the universal joint, allows the transfer of torque from the motor complex to the large flagellin protein (FliC). The flagellin protein acts as the physical propeller within bacterial systems. Due to the retention of the abundant number of genes required for the synthesis of a flagellum within the *Wigglesworthia* genome as well as the specialization of these genes towards motility, it is inferred that a flagellum may be used for motion. Moreover, the retention of a similar pathway within *Sodalis* in both order and gene conservation further emphasizes the importance of these genes within the tsetse host.

E. Methodology

There are various selection filters, or environmental constraints, that play a role in determining the composition of microbial communities. These filters are thought to arise from the sum of biotic and abiotic features. Within bacteria, flagella, though generally ubiquitous, contain subtle differences that give rise to phenotypic change enabling novel ecological opportunities (*104*). *Sodalis praecaptivus*, the free-living relative of *Sodalis*-allied symbionts, has two complete flagella clusters. In contrast, the other *Sodalis*-allied species retain differing levels of conservation of these two clusters. The extent that the individual symbiotic setting regulates the conservation of these clusters is unknown, thus this presents a gap in knowledge that is necessary towards understanding forces which act on the genomes within host-associated microbiota. This is an important problem to solve because it will provide insights into how distinct species present within a microbiota may share features due to equivalent natural selection pressures present

within different host. We propose to use the natural experiment of the *Sodalis*-allied organisms due to their wide range of hosts and symbiotic relationships. Specifically, we focus on the bacterial flagellum due to its aforementioned ability to transition within unique environments. My long-term goal is to understand how symbiont strategies and host ecology impacts microbial evolution.

Here, I explore the conservation outcomes associated with the flagellar complex within a group of divergent symbionts. My central hypothesis is that within the tsetse fly, the two predominant bacterial symbionts ("*Candidatus* Wigglesworthia glossinidia" and *Sodalis glossinidius*; hereafter *Wigglesworthia* and *Sodalis*) conserve singular, unique flagellar pathways necessary for habitation within the fly. Further, that the *Sodalis*-allied symbionts show a similar conservation of the flagellar pathway to their co-symbiotic partner.

My overall aim is to identify trends related to host, and flagellar evolution. *Sodalis*-allied organisms share a great number of genomic features due to sharing a common environmental ancestor. Conspicuously, these organisms vary in genetic and metabolic capabilities likely arising from a combination of unique co-symbiotic mechanisms and host microenvironment. Flagella are important because they are directly in contact with the external environment, and act as a conduit between host and microorganism either through motility or secretory functions (*57*, *105*). Our working hypothesis is that organisms that have evolved from the same progenitor likely differ in genetic capabilities because of environmental conditions as opposed to solely genetic drift. Due to the inconsistent conservation of this pathway within known symbionts, and the presence of multiple versions of the flagella pathway within the progenitor species we hypothesize the flagella will give new insights into how environment drives the divergence of complex molecular pathways. To compare evolutionary divergence, we identify the presence and absence of flagella synthesis genes within *Sodalis*-allied symbionts. We also evaluate the ancestral state of presence/absence of these organisms flagella to give insights into the similar pressures between phylogenetically similar hosts or, alternatively a similar host diet. Based on our working hypothesis, we would expect that organisms with more related hosts, or similar dietary ecology, would see a similar conservation from the ancestor to the extant organism. Lastly, co-symbiotic partners within insect hosts will be evaluated to determine whether similar molecular alterations to flagella also occur. This proposed research is innovative, representing a unique line of inquiry in microbiota research. Using the host as a gatekeeper rather than a beneficiary is a novel perspective in evaluating the genome evolution of distinct bacteria present in a community. The interspecies comparison of *Sodalis*-allied organisms and their co-symbionts will emphasize how the host as a biological filter serves a selective mechanism for the structure of conserved pathways. These interspecies comparative analyses will reevaluate the conservation of a unique flagellar synthesis pathway. Upon completion of this aim we expect to show that host biology has a diversifying effect towards colonizing microorganisms. This will have an important positive impact on our understanding of microbial evolution, by demonstrating the impact a host has on the

evolution of individual members of a community that may give rise to highly similar retention patterns in molecular pathways.

F. Appendix.

Table 1

Master control	**Regulators**	**Chaperones**	**Motor control complex**
flhC	*fliK*	*fliJ*	*fliG*
flhD	*fliZ*	*flgA*	*fliM*
Flagellar export apparatus	*fliA*	*flgN*	*fliN*
flhA	*flgM*	*fliS*	*flgH*
flhB	**Hook and filament**	*fliT*	*motA*
flhE	*flgE*	**Chemotaxis**	*motB*
fliO	*flgD*	*cheA*	*flgB*
fliP	*flgK*	*cheB*	*flgC*
fliQ	*fliC*	*cheW*	*flgF*
fliR	*fliD*	*tsr/tar*	*flgG*
fliH		*tap*	*flgJ*
fliI		*cheY*	*flgI*
		cheZ	*fliF*
		cheR	*fliE*

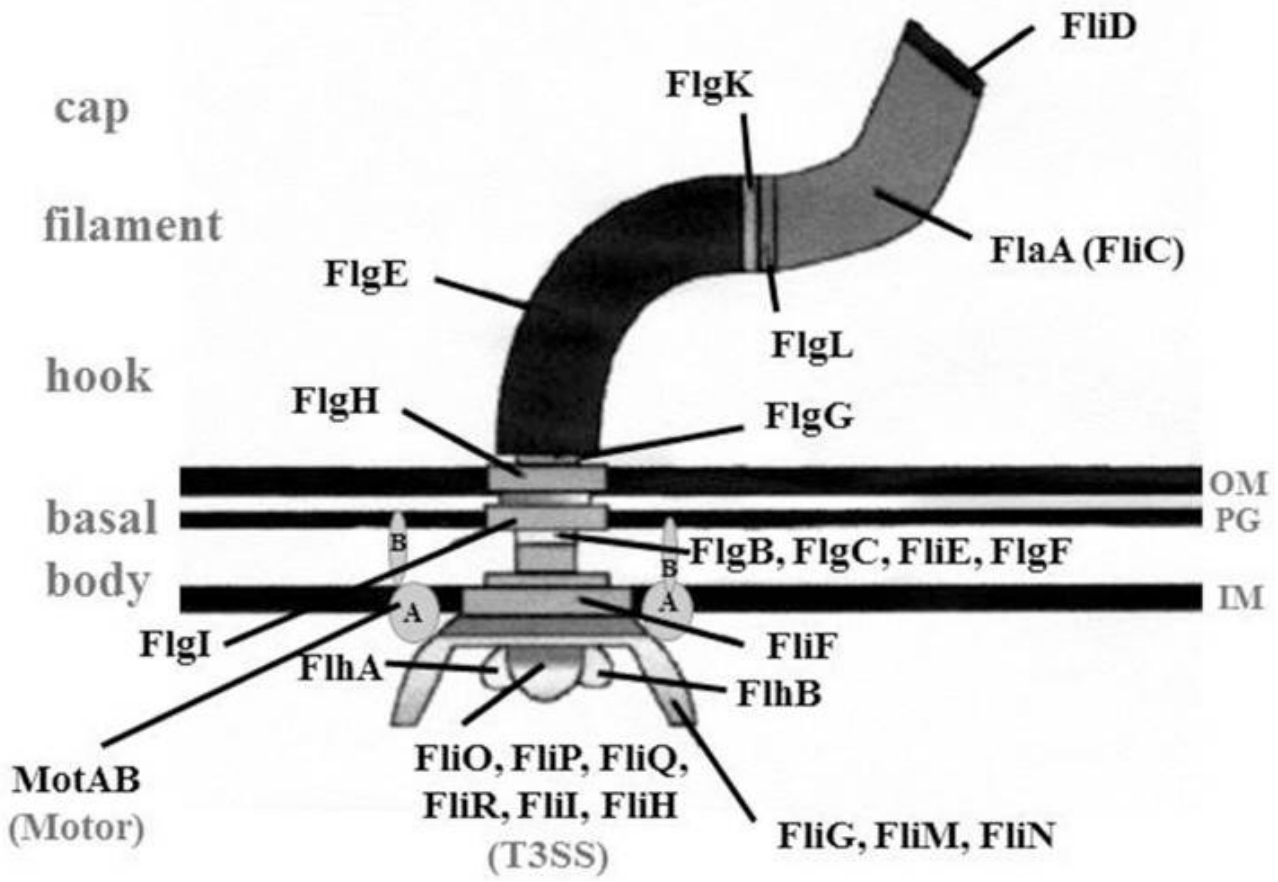

FliD
FlgK
cap
filament
FlaA (FliC)
FlgE
FlgL
hook
FlgH
FlgG
OM
basal
PG
body
B
B
FlgB, FlgC, FliE, FlgF
IM
A
A
FlgI
FliF
FlhA
FlhB
MotAB
(Motor)
FliO, FliP, FliQ,
FliR, FliI, FliH
(T3SS)
FliG, FliM, FliN

Figure 1.

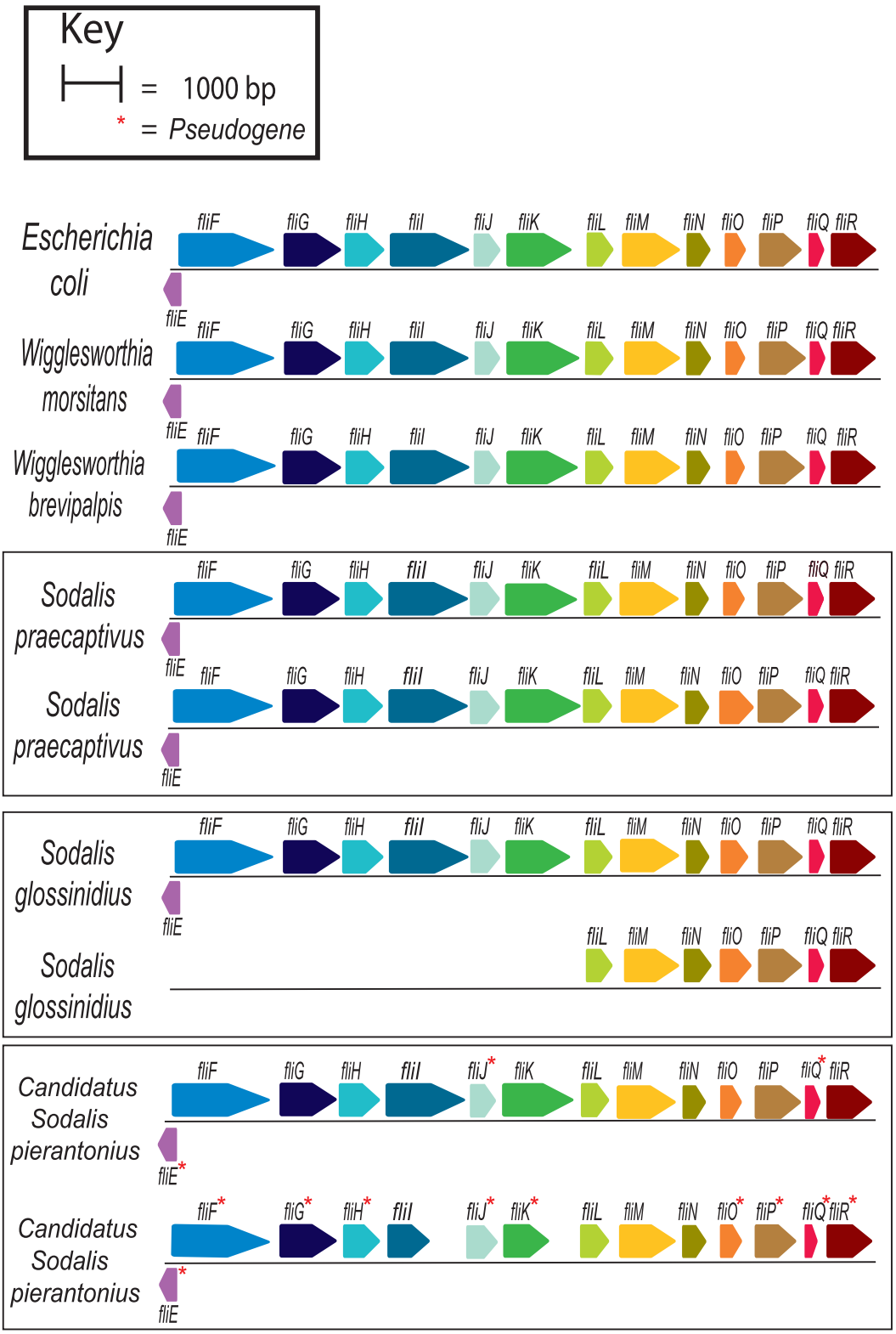

Key
= 1000 bp
* = Pseudogene
Escherichia coli
Wigglesworthia morsitans
Wigglesworthia brevipalpis
Sodalis praecaptivus
Sodalis praecaptivus
Sodalis glossinidius
Sodalis glossinidius
Candidatus Sodalis pierantonius
Candidatus Sodalis pierantonius
fliF fliG fliH fliI fliJ fliK fliL fliM fliN fliO fliP fliQ fliR
fliE

Figure 2.

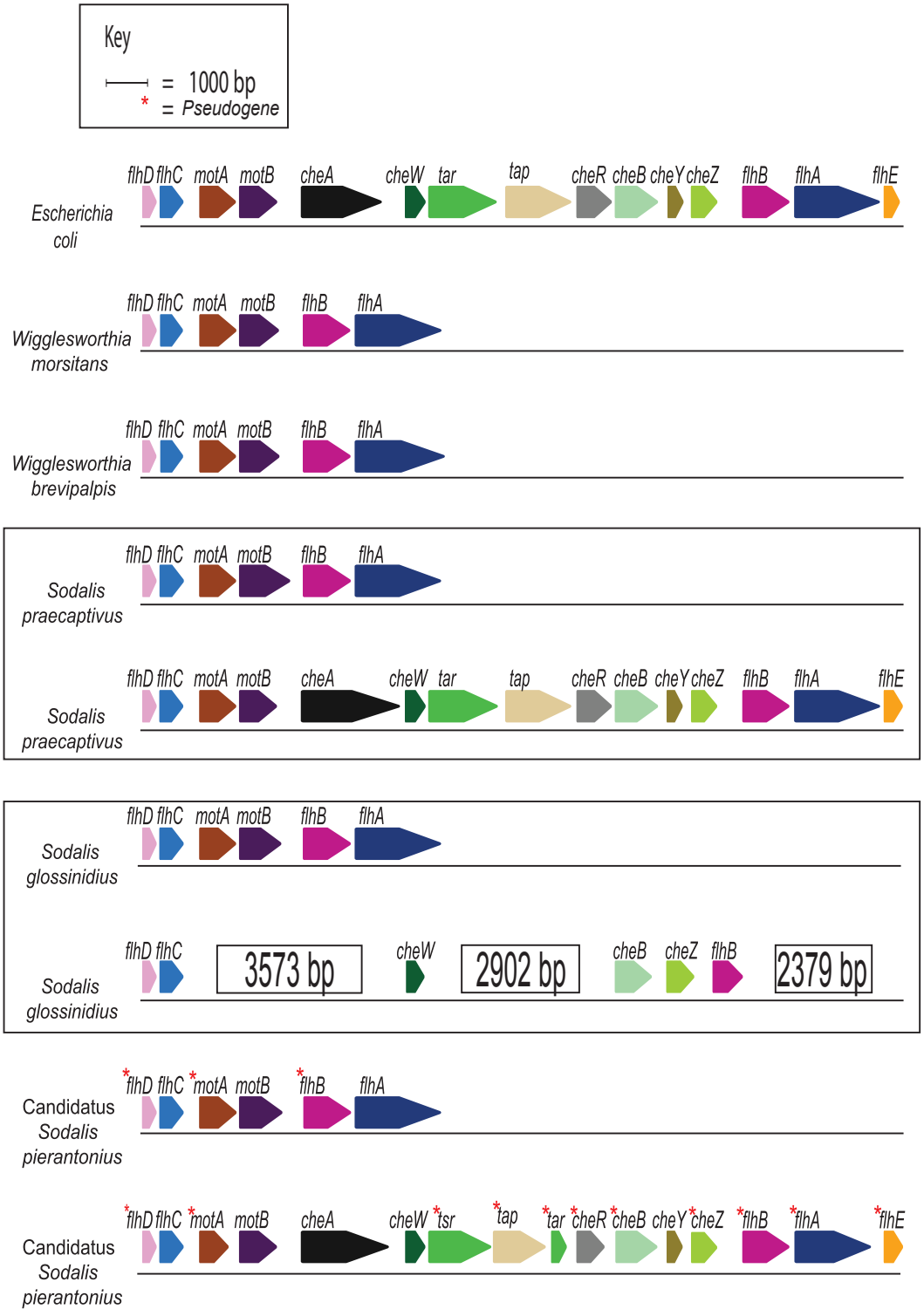

Key
= 1000 bp
* = Pseudogene
flhD flhC motA motB cheA cheW tar tap cheR cheB cheY cheZ flhB flhA flhE
Escherichia coli
flhD flhC motA motB flhB flhA
Wigglesworthia morsitans
flhD flhC motA motB flhB flhA
Wigglesworthia brevipalpis
flhD flhC motA motB flhB flhA
Sodalis praecaptivus
flhD flhC motA motB cheA cheW tar tap cheR cheB cheY cheZ flhB flhA flhE
Sodalis praecaptivus
flhD flhC motA motB flhB flhA
Sodalis glossinidius
flhD flhC
3573 bp
cheW
2902 bp
cheB cheZ flhB
2379 bp
Sodalis glossinidius
*flhD flhC *motA motB *flhB flhA
Candidatus Sodalis pierantonius
*flhD flhC *motA motB cheA cheW *tsr *tap *tar *cheR *cheB cheY *cheZ *flhB *flhA *flhE
Candidatus Sodalis pierantonius

Figure 3.

Table 1: List of genes, separated by functional description, comprising assembly of the canonical flagellar pathway within *Escherichia coli.* Chemotaxis genes are absent within the genomes of *Buchnera* and *Wigglesworthia spp..* All of the genes are present within the *Sodalis praecaptivus* genome.

Figure 1: The eubacteria flagella complex; *Buchnera* genomes have all the basal body genes present, while the *Wigglesworthia* genomes have the basal and additionally all genes listed above. Image was taken from Appelt and Heuner 2017.

Figure 2: A map of the fliEFGHIJKLMNOPQR operon within the genomes of representative bacteria. *Sodalis*-allied organisms have two separate clusters within each genome. Pseudogenes are designated with asterisks. Two *Wigglesworthia* species have the region extremely well conserved. The *Escherichia coli* flagella cluster is included for comparison.

Figure 3: A map of the master operon and the chemotaxis operon within the genomes of representative bacteria. Pseudogenes are designated with asterisks. Genes absent in the *Sodalis glossinidius* genome are represented with rectangles that also indicate the lengths of the intervening sequences between intact flagella genes. Both *Wigglesworthia* species are identical to the first cluster of *Sodalis glossinidius*. The *Escherichia coli* flagella cluster is also included for comparison.

G. Bibliography

1. A. Frank, Uber die biologischen Verhältnisse des Thallus eineger Krustenflechnten Beitr. *Biol. Pflanz.* **2**, 123–200. (1887).
2. A. DeBary, Die Erscheinung der Symbiose. Vortrag auf der Versammlung der Naturforshung und Ärtze zu Cassel. *Verlag von Karl J. Trubn. Strassburg.*, 1–30 (1879).
3. A. F. W. Schimper, Ûber die Entwicklung der Schlorophyllkörner und Farbkörper. *Bot. Zeitung,*. **4**, 105–114. (1883).
4. S. Mereschkowsky, Über Natur und Ursprung der Chromatophoren im Pflanzenreiche. *Biol. Zentralbl.* **25**, 593– 604.
5. L. Sagan, On the origin of mitosing cells. *J. Theor. Biol.* **14**, 255–274 (1967).
6. M. W. Gray, The Endosymbiont Hypothesis Revisited. *Int. Rev. Cytol.* **141**, 233–357 (1992).
7. L. Margulis, Biodiversity: molecular biological domains, symbiosis and kingdom origins. *BioSystems*. **27**, 39–51 (1992).
8. T. Buchner, R. Pfeiffer, E. Stupperich, Reduplication behavior of chromosomes of group D (13-15) and identification of the extra-chromosome in trisomy D. **55**, 1–2 (1965).
9. A. E. Douglas, Nutritional interactions in insect-microbial symbioses: aphids and their symbiotic bacteria Buchnera. *Annu. Rev. Entomol.* **43**, 17–37 (1998).
10. N. A. Moran, D. B. Sloan, The Hologenome Concept: Helpful or Hollow? *PLoS Biol.* **13**, e1002311 (2015).
11. R. V. M. Rio, G. M. Attardo, B. L. Weiss, Grandeur Alliances: Symbiont Metabolic Integration and Obligate Arthropod Hematophagy. *Trends Parasitol.* **32**, 739–749 (2016).
12. W. F. Martin, S. Garg, V. Zimorski, Endosymbiotic theories for eukaryote origin. *Philos. Trans. R. Soc. B Biol. Sci.* **370** (2015), doi:10.1098/rstb.2014.0330.
13. N. A. Moran, J. P. McCutcheon, A. Nakabachi, Genomics and evolution of heritable bacterial symbionts. *Annu. Rev. Genet.* **42**, 165–190 (2008).
14. J. P. McCutcheon, B. R. McDonald, N. A. Moran, Origin of an alternative genetic code in the extremely small and GC-rich genome of a bacterial symbiont. *PLoS Genet.* **5** (2009), doi:10.1371/journal.pgen.1000565.
15. C. Kuo, N. a Moran, H. Ochman, The consequences of genetic drift for bacterial genome complexity The consequences of genetic drift for bacterial genome complexity. *Genome Res.*, 1450–1454 (2009).
16. H. Charles, D. Mouchiroud, J. Lobry, I. Goncalves, Y. Rahbe, Gene size reduction in the bacterial aphid endosymbiont, Buchnera. *Mol. Biol. Evol.* **16** (1999), pp.

1820–1822.

17. N. A. Moran, A. Mira, The process of genome shrinkage in the obligate symbiont Buchnera aphidicola. *Genome Biol.* **2**, RESEARCH0054 (2001).

18. J. P. McCutcheon, N. A. Moran, Extreme genome reduction in symbiotic bacteria. *Nat. Rev. Microbiol.* **10**, 13–26 (2011).

19. N. A. Moran, Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *Proc. Natl. Acad. Sci.* **93**, 2873–2878 (1996).

20. D. B. Sloan, N. A. Moran, Genome Reduction and Co-evolution between the Primary and Secondary Bacterial Symbionts of Psyllids Research article. **29**, 3781–3792 (2012).

21. F. Husnik, J. P. McCutcheon, Repeated replacement of an intrabacterial symbiont in the tripartite nested mealybug symbiosis. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E5416–E5424 (2016).

22. J. L. Morrow, A. A. G. Hall, M. Riegler, Symbionts in waiting: the dynamics of incipient endosymbiont complementation and replacement in minimal bacterial communities of psyllids. *Microbiome*. **5**, 58 (2017).

23. F. Husnik, P. J. Keeling, The fate of obligate endosymbionts: reduction, integration, or extinction. *Curr. Opin. Genet. Dev.* **58**–**59** (2019), pp. 1–8.

24. A. W. D. Larkum, P. J. Lockhart, C. J. Howe, Shopping for plastids. *Trends Plant Sci.* **12**, 189–195 (2007).

25. A. K. Snyder, J. W. Deberry, L. Runyen-Janecky, R. V. M. Rio, Nutrient provisioning facilitates homeostasis between tsetse fly (Diptera: Glossinidae) symbionts. *Proc. R. Soc. B Biol. Sci.* **277**, 2389–2397 (2010).

26. A. K. Snyder, R. V. M. Rio, Interwoven biology of the tsetse holobiont. *J. Bacteriol.* **195**, 4322–4330 (2013).

27. S. C. Welburn, E. M. Fevre, P. G. Coleman, M. Odiit, I. Maudlin, Sleeping sickness: a tale of two diseases. *Trends Parasitol.* **17**, 19–24 (2001).

28. P. P. Simarro, J. Jannin, P. Cattand, Eliminating human African trypanosomiasis: where do we stand and what comes next? *PLoS Med.* **5**, e55 (2008).

29. A. Stich, Human African trypanosomiasis. *BMJ*. **325**, 203–206 (2002).

30. W. Ma, D. L. Denlinger, Secretory discharge and microflora of milk gland in tsetse flies. *Nature*. **247**, 301–303 (1974).

31. E. Aksoy, E. L. Telleria, R. Echodu, Y. Wu, L. M. Okedi, B. L. Weiss, S. Aksoy, A. Caccone, Analysis of multiple tsetse fly populations in Uganda reveals limited diversity and species-specific gut microbiota. *Appl. Environ. Microbiol.* **80**, 4301–4312 (2014).

32. B. L. Weiss, J. Wang, S. Aksoy, Tsetse immune system maturation requires the

presence of obligate symbionts in larvae. *PLoS Biol.* **9**, e1000619 (2011).

33. S. Aksoy, Wigglesworthia gen. nov. and Wigglesworthia glossinidia sp. nov., Taxa Consisting of the Mycetocyte-Associated, Primary Endosymbionts of Tsetse Flies. *Int. J. Syst. Bacteriol.* **45**, 848–851 (1995).

34. I. Maudlin, S. C. Welburn, D. Mehlitz, The relationship between rickettsia-like-organisms and trypanosome infections in natural populations of tsetse in Liberia. *Trop. Med. Parasitol.* **41**, 265–267 (1990).

35. C. Dale, I. Maudlin, Sodalis gen. nov. and Sodalis glossinidius sp. nov., a microaerophilic secondary endosymbiont of the tsetse fly Glossina morsitans morsitans. *Int. J. Syst. Bacteriol.* **49 Pt 1**, 267–275 (1999).

36. S. L. O'Neill, R. H. Gooding, S. Aksoy, Phylogenetically distant symbiotic microorganisms reside in Glossina midgut and ovary tissues. *Med. Vet. Entomol.* **7**, 377–383 (1993).

37. J. Wang, B. Weiss, S. Aksoy, Tsetse fly microbiota: form and function . *Front. Cell. Infect. Microbiol.* . **3** (2013), p. 69.

38. S. Balmand, C. Lohs, S. Aksoy, A. Heddi, Tissue distribution and transmission routes for the tsetse fly endosymbionts. *J. Invertebr. Pathol.* **112 Suppl**, S116-22 (2013).

39. G. Nogge, Aposymbiotic tsetse flies, Glossina morsitans morsitans obtained by feeding on rabbits immunized specifically with symbionts. *J. Insect Physiol.* **24**, 299–304 (1978).

40. C. Cenik, E. S. Cenik, G. W. Byeon, F. Grubert, S. I. Candille, D. Spacek, B. Alsallakh, H. Tilgner, C. L. Araya, H. Tang, E. Ricci, M. P. Snyder, Integrative analysis of RNA, translation, and protein levels reveals distinct regulatory variation across humans. *Genome Res.* **25**, 1610–1621 (2015).

41. M. Medina Munoz, A. R. Pollio, H. L. White, R. V. M. Rio, Into the Wild: Parallel Transcriptomics of the Tsetse-Wigglesworthia Mutualism within Kenyan Populations. *Genome Biol. Evol.* **9**, 2276–2291 (2017).

42. V. Michalkova, J. B. Benoit, B. L. Weiss, G. M. Attardo, S. Aksoy, Vitamin B(6) Generated by Obligate Symbionts Is Critical for Maintaining Proline Homeostasis and Fecundity in Tsetse Flies. *Appl. Environ. Microbiol.* **80**, 5844–5853 (2014).

43. L. Akman, A. Yamashita, H. Watanabe, K. Oshima, T. Shiba, M. Hattori, S. Aksoy, Genome sequence of the endocellular obligate symbiont of tsetse flies, Wigglesworthia glossinidia. *Nat. Genet.* **32**, 402–407 (2002).

44. G. M. Attardo, C. Lohs, A. Heddi, U. H. Alam, S. Yildirim, S. Aksoy, Analysis of milk gland structure and function in Glossina morsitans: Milk protein production, symbiont populations and fecundity. *J. Insect Physiol.* **54**, 1236–1242 (2008).

45. G. M. Attardo, J. M. Ribeiro, Y. Wu, M. Berriman, S. Aksoy, Transcriptome analysis of reproductive tissue and intrauterine developmental stages of the

tsetse fly (Glossina morsitans morsitans). *BMC Genomics*. **11**, 160 (2010).

46. J. B. Benoit, G. M. Attardo, V. Michalkova, P. Takac, J. Bohova, S. Aksoy, Sphingomyelinase activity in mother's milk is essential for juvenile development: a case from lactating tsetse flies. *Biol. Reprod.* **87**, 1-10,17 (2012).

47. J. Wang, S. Aksoy, PGRP-LB is a maternally transmitted immune milk protein that influences symbiosis and parasitism in tsetse's offspring. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 10552–10557 (2012).

48. G. Nogge, Sterility in tsetse flies (Glossina morsitans Westwood) caused by loss of symbionts. *Experientia*. **32**, 995–996 (1976).

49. G. Nogge, A. Gerresheim, Experiments on the elimination of symbionts from the tsetse fly, Glossina morsitans morsitans (Diptera: Glossinidae), by antibiotics and lysozyme. *J. Invertebr. Pathol.* **40**, 166–179 (1982).

50. R. Pais, C. Lohs, Y. Wu, J. Wang, S. Aksoy, The Obligate Mutualist Wigglesworthia glossinidia Influences Reproduction, Digestion, and Immunity Processes of Its Host, the Tsetse Fly. *Appl. Environ. Microbiol.* **74**, 5965–5974 (2008).

51. U. Alam, J. Medlock, C. Brelsfoard, R. Pais, C. Lohs, S. Balmand, J. Carnogursky, A. Heddi, P. Takac, A. Galvani, S. Aksoy, Wolbachia symbiont infections induce strong cytoplasmic incompatibility in the tsetse fly Glossina morsitans. *PLoS Pathog.* **7**, e1002415 (2011).

52. F. J. Silva, A. Latorre, A. Moya, Genome size reduction through multiple events of gene disintegration in Buchnera APS. *Trends Genet.* **17**, 615–618 (2001).

53. L. Klasson, S. G. E. Andersson, Evolution of minimal-gene-sets in host-dependent bacteria. *Trends Microbiol.* **12**, 37–43 (2004).

54. X. Chen, S. Li, S. Aksoy, Concordant evolution of a symbiont with its host insect species: molecular phylogeny of genus Glossina and its bacteriome-associated endosymbiont, Wigglesworthia glossinidia. *J. Mol. Evol.* **48**, 49–58 (1999).

55. R. V. M. Rio, R. E. Symula, J. Wang, C. Lohs, Y.-N. Wu, A. K. Snyder, R. D. Bjornson, K. Oshima, B. S. Biehl, N. T. Perna, M. Hattori, S. Aksoy, Insight into the transmission biology and species-specific functional capabilities of tsetse (Diptera: glossinidae) obligate symbiont Wigglesworthia. *MBio*. **3**, 1–13 (2012).

56. C. Toft, M. A. Fares, The evolution of the flagellar assembly pathway in endosymbiotic bacterial genomes. *Mol. Biol. Evol.* **25**, 2069–2076 (2008).

57. K. Maezawa, S. Shigenobu, H. Taniguchi, T. Kubo, S. I. Aizawa, M. Morioka, Hundreds of flagellar basal bodies cover the cell surface of the endosymbiotic bacterium Buchnera aphidicola sp. strain APS. *J. Bacteriol.* **188**, 6539–6543 (2006).

58. N. Kaiwa, T. Hosokawa, Y. Kikuchi, N. Nikoh, X. Y. Meng, N. Kimura, M. Ito, T. Fukatsu, Primary gut symbiont and secondary, Sodalis-allied symbiont of the

Scutellerid stinkbug Cantao ocellatus. *Appl. Environ. Microbiol.* **76**, 3486–3494 (2010).

59. A. Chari, K. F. Oakeson, S. Enomoto, D. G. Jackson, M. A. Fisher, C. Dale, Phenotypic characterization of Sodalis praecaptivus sp. nov., a close non-insect-associated member of the Sodalis-allied lineage of insect endosymbionts. *Int. J. Syst. Evol. Microbiol.* **65**, 1400–1405 (2015).

60. A. L. Clayton, K. F. Oakeson, M. Gutin, A. Pontes, D. M. Dunn, A. C. von Niederhausern, R. B. Weiss, M. Fisher, C. Dale, A novel human-infection-derived bacterium provides insights into the evolutionary origins of mutualistic insect-bacterial symbioses. *PLoS Genet.* **8**, e1002990 (2012).

61. D. Santos-Garcia, F. J. Silva, S. Morin, K. Dettner, S. M. Kuechler, The all-rounder sodalis: A new bacteriome-associated endosymbiont of the lygaeoid bug henestaris halophilus (heteroptera: Henestarinae) and a critical examination of its evolution. *Genome Biol. Evol.* **9**, 2893–2910 (2017).

62. A. Zaidman-Remy, A. Vigneron, B. L. Weiss, A. Heddi, What can a weevil teach a fly, and reciprocally? Interaction of host immune systems with endosymbionts in Glossina and Sitophilus. *BMC Microbiol.* **18**, 150 (2018).

63. K. F. Oakeson, R. Gil, A. L. Clayton, D. M. Dunn, A. C. von Niederhausern, C. Hamil, A. Aoyagi, B. Duval, A. Baca, F. J. Silva, A. Vallier, D. G. Jackson, A. Latorre, R. B. Weiss, A. Heddi, A. Moya, C. Dale, Genome degeneration and adaptation in a nascent stage of symbiosis. *Genome Biol. Evol.* **6**, 76–93 (2014).

64. M. H. Pontes, M. Babst, R. Lochhead, K. Oakeson, K. Smith, C. Dale, Quorum sensing primes the oxidative stress response in the insect endosymbiont, Sodalis glossinidius. *PLoS One*. **3**, e3541 (2008).

65. K. Kutsukake, Y. Ohya, S. Yamaguchi, T. Iino, Operon structure of flagellar genes in Salmonella typhimurium. *Mol. Gen. Genet.* **214**, 11–15 (1988).

66. R. N. Doetsch, R. D. Sjoblad, Flagellar Structure and Function in Eubacteria. *Annu. Rev. Microbiol.* **34**, 69–108 (1980).

67. F. F. V. Chevance, K. T. Hughes, Coordinating assembly of a bacterial macromolecular machine. *Nat. Rev. Microbiol.* **6**, 455–465 (2008).

68. M. Erhardt, K. Namba, K. T. Hughes, Bacterial nanomachines: the flagellum and type III injectisome. *Cold Spring Harb. Perspect. Biol.* **2**, a000299 (2010).

69. Y. Imae, T. Atsumi, Na+-driven bacterial flagellar motors. *J. Bioenerg. Biomembr.* **21**, 705–716 (1989).

70. H. C. Berg, The rotary motor of bacterial flagella. *Annu. Rev. Biochem.* **72**, 19–54 (2003).

71. T. Minamino, K. Imada, The bacterial flagellar motor and its structural diversity. *Trends Microbiol.* **23**, 267–274 (2015).

72. G. J. Schoenhals, R. M. Macnab, Physiological and biochemical analyses of FlgH, a lipoprotein forming the outer membrane L ring of the flagellar basal body of Salmonella typhimurium. *J. Bacteriol.* **178**, 4200–4207 (1996).

73. R. M. Macnab, How Bacteria Assemble Flagella. *Annu. Rev. Microbiol.* **57**, 77–100 (2003).

74. R. M. Macnab, Type III flagellar protein export and flagellar assembly. *Biochim. Biophys. Acta*. **1694**, 207–217 (2004).

75. G. S. Chilcott, K. T. Hughes, Coupling of Flagellar Gene Expression to Flagellar Assembly in Salmonella enterica Serovar Typhimurium and Escherichia coli. *Microbiol. Mol. Biol. Rev.* **64**, 694–708 (2000).

76. R. Liu, H. Ochman, Origins of Flagellar Gene Operons and Secondary Flagellar Systems . *J. Bacteriol.* **189**, 7098–7104 (2007).

77. R. D. Allen, P. Baumann, Structure and arrangement of flagella in species of the genus Beneckea and Photobacterium fischeri. *J. Bacteriol.* **107**, 295–302 (1971).

78. E. C. McCoy, D. Doyle, H. Wiltberger, K. Burda, A. J. Winter, Flagellar ultrastructure and flagella-associated antigens of Campylobacter fetus. *J. Bacteriol.* **122**, 307–315 (1975).

79. K. W. Hranitzky, A. Mulholland, A. D. Larson, E. R. Eubanks, L. T. Hart, *Infect. Immun.*, in press (available at http://iai.asm.org/content/27/2/597.abstract).

80. S. Trachtenberg, D. J. DeRosier, Conformational switching in the flagellar filament of Salmonella typhimurium. *J. Mol. Biol.* **226**, 447–454 (1992).

81. L. L. McCarter, Dual flagellar systems enable motility under different circumstances. *J. Mol. Microbiol. Biotechnol.* **7**, 18–29 (2004).

82. D. B. Kearns, A field guide to bacterial swarming motility. *Nat. Rev. Microbiol.* **8**, 634–644 (2010).

83. P. G. Hall, N. R. Krieg, Swarming of Azospirillum brasilense on solid media. *Can. J. Microbiol.* **29**, 1592–1594 (1983).

84. S. Kalir, J. McClure, K. Pabbaraju, C. Southward, M. Ronen, S. Leibler, M. G. Surette, U. Alon, Ordering genes in a flagella pathway by analysis of expression kinetics from living bacteria. *Science*. **292**, 2080–2083 (2001).

85. L. D. B. Evans, C. Hughes, G. M. Fraser, Building a flagellum outside the bacterial cell. *Trends Microbiol.* **22**, 566–572 (2014).

86. A. Dufour, R. B. Furness, C. Hughes, Novel genes that upregulate the Proteus mirabilis flhDC master operon controlling flagellar biogenesis and swarming. *Mol. Microbiol.* **29**, 741–751 (1998).

87. O. Soutourina, A. Kolb, E. Krin, C. Laurent-Winter, S. Rimsky, A. Danchin, P. Bertin, Multiple control of flagellum biosynthesis in Escherichia coli: role of H-NS protein and the cyclic AMP-catabolite activator protein complex in transcription of

the flhDC master operon. *J. Bacteriol.* **181**, 7500–7508 (1999).

88. A. Francez-Charlot, B. Laugel, A. Van Gemert, N. Dubarry, F. Wiorowski, M.-P. Castanie-Cornet, C. Gutierrez, K. Cam, RcsCDB His-Asp phosphorelay system negatively regulates the flhDC operon in Escherichia coli. *Mol. Microbiol.* **49**, 823–832 (2003).

89. B. M. Prüß, X. Liu, W. Hendrickson, P. Matsumura, FlhD/FlhC-regulated promoters analyzed by gene array and lacZ gene fusions. *FEMS Microbiol. Lett.* **197**, 91–97 (2006).

90. K. Kutsukake, Y. Ohya, T. Iino, Transcriptional analysis of the flagellar regulon of Salmonella typhimurium. *J. Bacteriol.* **172**, 741–747 (1990).

91. R. B. Furness, G. M. Fraser, N. A. Hay, C. Hughes, Negative feedback from a Proteus class II flagellum export defect to the flhDC master operon controlling cell division and flagellum assembly. *J. Bacteriol.* **179**, 5585–5588 (1997).

92. G. M. Young, M. J. Smith, S. A. Minnich, V. L. Miller, *J. Bacteriol.*, in press (available at http://jb.asm.org/content/181/9/2823.abstract).

93. L. Claret, C. Hughes, Interaction of the atypical prokaryotic transcription activator FlhD2C2 with early promoters of the flagellar gene hierarchy. *J. Mol. Biol.* **321**, 185–199 (2002).

94. N. R. Francis, G. E. Sosinsky, D. Thomas, D. J. DeRosier, Isolation, characterization and structure of bacterial flagellar motors containing the switch complex. *J. Mol. Biol.* **235**, 1261–1270 (1994).

95. T. Kubori, S. Yamaguchi, S. Aizawa, Assembly of the switch complex onto the MS ring complex of Salmonella typhimurium does not require any other flagellar proteins. *J. Bacteriol.* **179**, 813–817 (1997).

96. X. Zhao, K. Zhang, T. Boquoi, B. Hu, M. A. Motaleb, K. A. Miller, M. E. James, Cryoelectron tomography reveals the sequential assembly of bacterial fl agella in Borrelia burgdorferi, 1–6 (2013).

97. A. M. Burrage, E. Vanderpool, D. B. Kearns, crossm Assembly Order of Flagellar Rod Subunits in Bacillus subtilis, 1–13 (2018).

98. A. G. Garza, L. W. Harris-Haller, R. A. Stoebner, M. D. Manson, Motility protein interactions in the bacterial flagellar motor. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 1970–1974 (1995).

99. S. A. Lloyd, F. G. Whitby, D. F. Blair, C. P. Hill, Structure of the C-terminal domain of FliG, a component of the rotor in the bacterial flagellar motor. *Nature*. **400**, 472–475 (1999).

100. D. Thomas, D. G. Morgan, D. J. DeRosier, *J. Bacteriol.*, in press (available at http://jb.asm.org/content/183/21/6404.abstract).

101. D. F. Blair, Flagellar movement driven by proton translocation. *FEBS Lett.* **545**,

86–95 (2003).

102. Y. V Morimoto, T. Minamino, Structure and function of the bi-directional bacterial flagellar motor. *Biomolecules*. **4**, 217–234 (2014).

103. J. E. Karlinsey, H. C. Tsui, M. E. Winkler, K. T. Hughes, Flk couples flgM translation to flagellar ring assembly in Salmonella typhimurium. *J. Bacteriol.* **180**, 5384–5397 (1998).

104. D. R. Mitchell, The evolution of eukaryotic cilia and flagella as motile and sensory organelles. *Adv. Exp. Med. Biol.* **607**, 130–140 (2007).

105. C. Dale, G. R. Plague, B. Wang, H. Ochman, N. A. Moran, Type III secretion systems and the evolution of mutualistic endosymbiosis. **99** (2002).

106. P. Baumann, Biology of Bacteriocyte-Associated Endosymbionts of Plant Sap-Sucking Insects. *Annu. Rev. Microbiol.* **59**, 155–189 (2005).

107. A. Nakabachi, R. Ueoka, K. Oshima, R. Teta, A. Mangoni, M. Gurgui, N. J. Oldham, G. Van Echten-Deckert, K. Okamura, K. Yamamoto, H. Inoue, M. Ohkuma, Y. Hongoh, S. Y. Miyagishima, M. Hattori, J. Piel, T. Fukatsu, Defensive bacteriome symbiont with a drastically reduced genome. *Curr. Biol.* **23**, 1478–1484 (2013).

108. M. E. Brentassi, E. Franco, P. Balatti, R. Medina, F. Bernabei, A. M. Marino, D. R. Lenicov, Bacteriomes of the corn leafhopper , Dalbulus maidis (DeLong Deltocephalinae) harbor Sulcia symbiont : molecular characterization , ultrastructure , and transovarial transmission. *Protoplasma* (2016), doi:10.1007/s00709-016-1033-4.

109. X. Bing, G. M. Attardo, A. Vigneron, E. Aksoy, F. Scolari, A. Malacrida, B. L. Weiss, S. Aksoy, Unravelling the relationship between the tsetse fly and its obligate symbiont Wigglesworthia : transcriptomic and metabolomic landscapes reveal highly integrated physiological networks. *Proc. R. Soc. B Biol. Sci.* **284**, 20170360 (2017).

110. B. M. Boyd, J. M. Allen, R. Koga, T. Fukatsu, A. D. Sweet, K. P. Johnson, D. L. Reed, Louse Proechinophthirus fluctus (Phthiraptera : Anoplura). *Appl. Environ. Microbiol.* **82**, 3185–3197 (2016).

111. S. S. Prado, R. P. P. Almeida, Phylogenetic placement of pentatomid stink bug gut symbionts. *Curr. Microbiol.* **58**, 64–69 (2009).

112. F. Husnik, N. Nikoh, R. Koga, L. Ross, R. P. Duncan, M. Fujie, M. Tanaka, N. Satoh, D. Bachtrog, A. C. C. Wilson, C. D. Von Dohlen, T. Fukatsu, J. P. McCutcheon, Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. *Cell*. **153**, 1567–1578 (2013).

113. R. V. M. Rio, Y.-N. Wu, G. Filardo, S. Aksoy, Dynamics of multiple symbiont density regulation during host development: tsetse fly and its microbial flora. *Proceedings. Biol. Sci.* **273**, 805–814 (2006).

114. G. M. Bennett, N. A. Moran, Small, smaller, smallest: The origins and evolution of ancient dual symbioses in a phloem-feeding insect. *Genome Biol. Evol.* **5**, 1675–1688 (2013).

115. J. Felsenstein, S. Yokoyama, The evolutionary advantage of recombination. II. Individual selection for recombination. *Genetics*. **83**, 845–859 (1976).

116. C. Rispe, N. A. Moran, Accumulation of deleterious mutations in endosymbionts: Muller's ratchet with two levels of selection. *Am. Nat.* **156**, 425–441 (2000).

117. M. Woolfit, L. Bromham, Increased rates of sequence evolution in endosymbiotic bacteria and fungi with small effective population sizes. *Mol. Biol. Evol.* **20**, 1545–1555 (2003).

118. J. J. Wernegreen, Genome evolution in bacterial endosymbionts of insects. *Nat. Rev. Genet.* **3**, 850–861 (2002).

119. U. Kutschera, K. J. Niklas, Endosymbiosis, cell evolution, and speciation. *Theory Biosci.* **124**, 1–24 (2005).

120. R. A. Chong, N. A. Moran, Evolutionary loss and replacement of Buchnera, the obligate endosymbiont of aphids. *ISME J.* (2018), doi:10.1038/s41396-017-0024-6.

121. C. D. von Dohlen, U. Spaulding, K. B. Patch, K. M. Weglarz, R. G. Foottit, N. P. Havill, G. R. Burke, Dynamic acquisition and loss of dual-obligate symbionts in the plant-sap-feeding adelgidae (Hemiptera: Sternorrhyncha: Aphidoidea). *Front. Microbiol.* **8**, 1–15 (2017).

122. F. Husnik, J. P. McCutcheon, Repeated replacement of an intrabacterial symbiont in the tripartite nested mealybug symbiosis. *Proc. Natl. Acad. Sci.* **113**, E5416–E5424 (2016).

123. M. L. Thao, M. A. Clark, L. Baumann, E. B. Brennan, N. A. Moran, P. Baumann, Secondary Endosymbionts of Psyllids Have Been Acquired Multiple Times. **41**, 300–304 (2000).

124. R. Koga, N. A. Moran, Swapping symbionts in spittlebugs: Evolutionary replacement of a reduced genome symbiont. *ISME J.* **8**, 1237–1246 (2014).

125. M. A. Clark, L. Baumann, M. L. Thao, N. A. Moran, P. Baumann, Degenerative minimalism in the genome of a psyllid endosymbiont. *J. Bacteriol.* **183**, 1853–1861 (2001).

126. H. Terashima, A. Kawamoto, Y. V Morimoto, K. Imada, T. Minamino, Structural differences in the bacterial flagellar motor among bacterial species. **14**, 191–198 (2017).

127. F. F. V Chevance, K. T. Hughes, Coordinating assembly of a bacterial macromolecular machine. *Nat. Rev. Microbiol.* **6**, 455–465 (2008).

128. I. A. Osterman, Y. Y. Dikhtyar, A. A. Bogdanov, O. A. Dontsova, P. V Sergiev,

Regulation of flagellar gene expression in Bacteria. *Biochem.* **80** (2015), pp. 1447–1456.

129. E. Nováková, F. Husník, E. Šochová, V. Hypša, Arsenophonus and sodalis symbionts in louse flies: An analogy to the wigglesworthia and sodalis system in tsetse flies. *Appl. Environ. Microbiol.* **81**, 6189–6199 (2015).

130. R. A. Chong, H. Park, N. A. Moran, Genome Evolution of the Obligate Endosymbiont Buchnera aphidicola. *Mol. Biol. Evol.* **36**, 1481–1489 (2019).

131. L. Li, C. J. J. Stoeckert, D. S. Roos, OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes -- Li et al. 13 (9): 2178 -- Genome Research. *Genome Res.* **13**, 2178–2189 (2003).

132. T. P. Driscoll, V. I. Verhoeve, M. L. Guillotte, S. S. Lehman, S. A. Rennoll, M. Beier-Sexton, M. S. Rahman, A. F. Azad, J. J. Gillespie, Wholly rickettsia! reconstructed metabolic profile of the quintessential bacterial parasite of eukaryotic cells. *MBio*. **8** (2017), doi:10.1128/mBio.00859-17.

133. A. E. Darling, B. Mau, N. T. Perna, Progressivemauve: Multiple genome alignment with gene gain, loss and rearrangement. *PLoS One*. **5** (2010), doi:10.1371/journal.pone.0011147.

134. R. K. Aziz, D. Bartels, A. Best, M. DeJongh, T. Disz, R. A. Edwards, K. Formsma, S. Gerdes, E. M. Glass, M. Kubal, F. Meyer, G. J. Olsen, R. Olson, A. L. Osterman, R. A. Overbeek, L. K. McNeil, D. Paarmann, T. Paczian, B. Parrello, G. D. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke, O. Zagnitko, The RAST Server: Rapid annotations using subsystems technology. *BMC Genomics*. **9**, 1–15 (2008).

135. R. C. Edgar, MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

136. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

137. S. Capella-Gutiérrez, J. M. Silla-Martínez, T. Gabaldón, trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. **25**, 1972–1973 (2009).

138. A. Stamatakis, RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. **22**, 2688–2690 (2006).

139. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. **30**, 1312–1313 (2014).

140. J. P. Hueslenbeck, B. Larget, R. E. Miller, F. Ronquist, Potential Applications and Pitfalls of Bayesian Inference of Phylogeny. **51**, 673–688 (2002).

141. F. Ronquist, M. Teslenko, P. Van Der Mark, D. L. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M. A. Suchard, J. P. Huelsenbeck, Mrbayes 3.2: Efficient bayesian

phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).

142. D. Glez-Peña, D. Gómez-Blanco, M. Reboiro-Jato, F. Fdez-Riverola, D. Posada, ALTER: Program-oriented conversion of DNA and protein alignments. *Nucleic Acids Res.* **38**, 14–18 (2010).

143. B. G. Hall, Building phylogenetic trees from molecular data with MEGA. *Mol. Biol. Evol.* **30**, 1229–1235 (2013).

144. S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).

145. A. Rambaut, A. J. Drummond, D. Xie, G. Baele, M. A. Suchard, Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* **67**, 901–904 (2018).

146. J. Huerta-Cepas, D. Szklarczyk, D. Heller, A. Hernández-Plaza, S. K. Forslund, H. Cook, D. R. Mende, I. Letunic, T. Rattei, L. J. Jensen, C. Von Mering, P. Bork, EggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **47**, D309–D314 (2019).

147. E. Paradis, J. Claude, K. Strimmer, APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*. **20**, 289–290 (2004).

148. L. J. Revell, phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012).

149. M. W. Pennell, J. M. Eastman, G. J. Slater, J. W. Brown, J. C. Uyeda, R. G. Fitzjohn, M. E. Alfaro, L. J. Harmon, Geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. **30**, 2216–2218 (2014).

150. D. T. Gillespie, Exact numerical simulation of the Ornstein-Uhlenbeck process and its integral. *Phys. Rev. E*. **54**, 2084–2091 (1996).

151. J. S. Farris, Phylogenetic Analysis Under Dollo's Law (1977).

152. A. G. Kluge, J. S. Farris, Quantitative Phyletics and the Evolution of Anurans (1967).

153. K. L. Gillen, K. T. Hughes, *J. Bacteriol.*, in press (available at http://jb.asm.org/content/173/7/2301.abstract).

154. F. Suaste-Olmos, C. Domenzain, J. C. Mireles-Rodríguez, S. Poggio, A. Osorio, G. Dreyfus, L. Camarena, The flagellar protein FliL is essential for swimming in Rhodobacter sphaeroides. *J. Bacteriol.* **192**, 6230–6239 (2010).

155. N. Takekawa, M. Isumi, H. Terashima, S. Zhu, Y. Nishino, M. Sakuma, S. Kojima, M. Homma, K. Imada, Structure of vibrio flil, a new stomatin-like protein that assists the bacterial flagellar motor function. *MBio*. **10**, 1–13 (2019).

156. T. Minamino, Y. V. Morimoto, N. Hara, K. Namba, An energy transduction mechanism used in bacterial flagellar type III protein export. *Nat. Commun.* **2**, 475–479 (2011).

157. J. Kucera, E. M. Terentjev, FliI6-FliJ molecular motor assists with unfolding in the type III secretion export apparatus. *Sci. Rep.* **10**, 1–11 (2020).

158. F. A. Herlihey, P. J. Moynihan, A. J. Clarke, The essential protein for bacterial flagella formation FlgJ functions as a ??-N-acetylglucosaminidase. *J. Biol. Chem.* **289**, 31029–31042 (2014).

159. M. Erhardt, T. Hirano, Y. Su, K. Paul, D. H. Wee, S. Mizuno, S. I. Aizawa, K. T. Hughes, The role of the FliK molecular ruler in hook-length control in Salmonella enterica. *Mol. Microbiol.* **75**, 1272–1284 (2010).

160. P. Aldridge, J. Karlinsey, K. T. Hughes, The type III secretion chaperone FlgN regulates flagellar assembly via a negative feedback loop containing its chaperone substrates FlgK and FlgL. *Mol. Microbiol.* **49**, 1333–1345 (2003).

161. S. Maiti, P. Mitra, Bacterial flagellar switching: a molecular mechanism directed by the logic of an electric motor. *J. Mol. Model.* **24**, 280 (2018).

162. P. H. Degnan, Y. Yu, N. Sisneros, R. A. Wing, N. A. Moran, Hamiltonella defensa, genome evolution of protective bacterial endosymbiont from pathogenic ancestors. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 9063–9068 (2009).

163. S. Bhattacharya, A. K. Baidya, R. R. Pal, G. Mamou, Y. E. Gatt, H. Margalit, I. Rosenshine, S. Ben-Yehuda, A Ubiquitous Platform for Bacterial Nanotube Biogenesis. *Cell Rep.* **27**, 334-342.e10 (2019).

164. K. P. Williams, J. J. Gillespie, B. W. S. Sobral, E. K. Nordberg, E. E. Snyder, J. M. Shallom, A. W. Dickerman, Phylogeny of Gammaproteobacteria. *J. Bacteriol.* **192**, 2305–2314 (2010).

165. D. Darriba, D. Posada, A. M. Kozlov, A. Stamatakis, B. Morel, T. Flouri, ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. *Mol. Biol. Evol.* **37**, 291–294 (2020).

166. R. Lanfear, P. B. Frandsen, A. M. Wright, T. Senfeld, B. Calcott, PartitionFinder 2 : New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. **34**, 772–773 (2016).

167. R. A. Chong, H. Park, N. A. Moran, Genome Evolution of the Obligate Endosymbiont Buchnera aphidicola. *Mol. Biol. Evol.* **36**, 1481–1489 (2019).

168. R. Gil, S. Ramos, A. Lamelas, M. Postigo, J. M. Michelena, F. J. Silva, A Small Microbial Genome : The End of a Long Symbiotic Relationship. *Science (80-.).* **314**, 312–313 (2006).

169. A. Manzano-Marín, A. Latorre, Settling down: The genome of serratia symbiotica from the aphid cinara tujafilina zooms in on the process of accommodation to a cooperative intracellular life. *Genome Biol. Evol.* **6**, 1683–1698 (2014).

170. A. Lamelas, M. J. Gosalbes, A. Manzano-Marín, J. Peretó, A. Moya, A. Latorre, Serratia symbiotica from the aphid Cinara cedri: A missing link from facultative to obligate insect endosymbiont. *PLoS Genet.* **7** (2011), doi:10.1371/journal.pgen.1002357.

171. D. A. C. Bublitz, G. L. Chadwick, J. S. Magyar, K. M. Sandoz, D. M. Brooks, S. Mesnage, M. S. Ladinsky, A. I. Garber, P. J. Bjorkman, V. J. Orphan, J. P. McCutcheon, Peptidoglycan Production by an Insect-Bacterial Mosaic. *Cell*. **179**, 703-712.e7 (2019).

172. A. J. Drummond, S. Y. W. Ho, M. J. Phillips, A. Rambaut, Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**, 699–710 (2006).

Chapter 2: Comparative analysis of flagellar genes in *Sodalis*-allied organisms

Introduction

The acquisition of microbial partners, or endosymbiosis, is a strategy used by a diverse group of arthropods to overcome ecological barriers that might normally make certain niches uninhabitable (*106*). Here, symbionts often trade nutrients for protection that may be in short supply. For example, limited diets subsisting solely of blood, wood, phloem or xylem are possible due to the complementary metabolic capabilities provided by endosymbionts (*8*–*11*). In some cases, the host arthropod creates additional protection from environmental stress by synthesizing a unique set of cells called the bacteriocytes that collectively form an organ called a bacteriome (*38*, *54*, *61*, *107*, *108*). These specialized compartments may simplify transfer of nutrients between partners (*25*, *38*, *55*, *109*). Bacteria that have co-evolved to inhabit these organs are often acquired through vertical transmission (*8*, *9*, *110*–*112*) (i.e., the *Wigglesworthia glossinidius* symbiont of tsetse flies relocate to the milk glands allowing for inoculation of the larvae *in utero* (*55*, *113*)). Over time, due to smaller population sizes, relaxed selection on nonessential genes, and a loss of recombination mechanisms, some microbes are subject to strong genetic drift or the stochastic accumulation of deletional mutations over time (*114*).

Muller's ratchet, the process by which deleterious mutations accumulate in an irreversible way within populations lacking recombination (*115*), is a mechanism which drives the genome reduction often seen with bacterial symbionts (*23*, *116*). Through this evolutionary bottleneck, each subsequent generation will have unique mutations that multiply over time leading to pseudogenization, particularly as a result of relaxed selection on non-essential loci, and eventual gene loss (*17*, *117*). This process leads to specialized symbionts (*16*, *113*, *118*) at risk for genome deterioration. The eventual genome can become so reduced that that other symbionts can compete within the host system (*54*, *119*). In more extreme circumstances the initial symbiont is lost entirely and swapped with a more fit substitute(*21*, *120*, *121*). The nascent symbiont generally follow the same trajectory of the former symbiont.

Given the diversity of bacteria and the stochastic nature of endosymbiosis, it is difficult to compare evolutionary histories. The *Sodalis*-allied organisms are a closely related group of bacteria, many of which fulfill unique roles within diverse hosts (Table 1). Often, these organisms are acquired as secondary symbionts, sometimes supplementing a genome-reduced primary symbiont (*122*–*124*). Due to distinct symbiosis histories, each *Sodalis* species is at a different stage of the genome reduction continuum. Additionally, the diversity of host ecologies allows for the emergence of unique symbiont genomic capabilities. For example, the primary symbiont of the rice weevil (*Sitophilus oryzai)*, *Sodalis pierantonious* (SOPE) (*63*), has a larger genome (~ 4Mb) with many pathways relatively intact due to recent acquisition by the weevil host (*62*). Though the genome remains large there is evidence of ongoing genome reduction as demonstrated by the presence of an abundant number of pseudogenes (*63*). In contrast, *Moranella endobia* has a smaller genome (~ 600 Kb), and functions as a nested symbiont partner within the cells of the highly reduced Betaproteobacteria

Tremblaya princeps (*112*, *123*) within a mealybug host. This relationship is thought to be extremely ancestral. The *Moranella* genome has conserved only genes necessary for its unique role in amino acid synthesis (*112*, *123*, *125*). These two extremes demonstrate the diversity of trajectories found within this group of bacteria. The unique nature of this clade can further be demonstrated by *Sodalis praecaptivus* which is believed to represent a closely-related environmental antecedent to the *Sodalis*-allied symbionts found in many insects (*60*). By using *S. praecaptivus* as the closest free-living ancestor, this clade can be used as a natural experiment to understand the evolutionary process of symbiosis.

Here, we focus on the evolution of the flagellar synthesis pathway, due to its complexity and ability to diverge in response to ecological pressures (*68*, *71*, *126*). For example, organisms have evolved peritrichous flagella, or many flagella surrounding the cell, in response to viscous environments (*81*). The flagellar synthesis pathway is composed of many (~50) individual genes that are systematically transcribed to build an intricate mosaic structure (*76*). This propeller-like motility structure is composed of the basal body, the motor, the type III secretion system (T3SS), and the hook/filament complex (Figure 1). Each component is organized in co-transcribed groups, or operons, that are activated in an ordered molecular cascade (*127*). Thus, each operon plays a different role in the overall construction process (*128*). The ability of this system to adapt over evolutionary time, not just in free-living organisms but also within established endosymbionts (*55*–*57*), makes it a useful measure for examining evolutionary outcomes associated with the process of symbiosis.

In this study, we characterize the flagella diversification within the *Sodalis*-allied clade prior to and following symbiosis. Further, we use flagellar evolution as a proxy to associate molecular changes to host traits and diet. Lastly, we evaluate other co-occurring symbionts within these hosts to see if these molecular changes affect partners equally, providing further support for the significance of host biology towards selection. The unique phylogenetic opportunities attainable from the natural experiment of the *Sodalis*-allied clade allow for better understanding of how molecular structures evolve through time upon entering endosymbiosis with diverse insect hosts.

Materials and Methods

Organism selection. A representative set of organisms including *Buchnera* (representing ancestral gammaproteobacterial symbionts) and *Sodalis*-allied symbionts (NCBI accession numbers collected from Santos-Garcia et al. 2017) were included in analyses. To examine the effect of host phylogenetic relationships and dietary ecology towards flagellar gene evolution, other bacterial symbionts inhabiting the same hosts as the aforementioned bacteria were also included in our analyses. These cohabitating symbionts were identified by literature reviews. A total of 30 microorganisms (Table 1) were selected as either a *Sodalis*-allied organism or a co-symbiont (*22*, *112*, *122*, *129*) plus the additional *Buchnera* within the three clades: the large subfamily Aphidinae, the

subfamily Lachniinae, the tribe Fordini (*130*). Each of these organisms was recorded with a host species designation.

Gene presence verification. In order to verify the presence/absence of flagellar genes (Table 2) within each symbiont genome, we used three approaches : 1) ortholog group (OG) construction with OrthoMCL (*131*, *132*), 2) whole-genome alignment with ProgressiveMauve (*133*), and 3) manual sequence alignment of select OGs. Pseudogenes were identified by the appearance of large gaps (n $\geq$ 25 amino acids), loss of synteny, or altered protein length (i.e., arising from missense mutations disrupting start or stop sites). If two or more of these characteristics appeared the sequence was identified as a pseudogene (Figure 2).

Groups of orthologous genes were created using OrthoMCL (*131*, *132*) based on the top 100 hits of a self v. self BlastP with an e-value cutoff of 10^{-7} using a concatenated GenBank protein FASTA file of the genomes of interest. Any organism for which it was necessary was annotated using the RAST server with default settings (*134*). We identified flagellar ortholog groups (OGs) by searching our database file for *Sodalis praecaptivus* flagella proteins using the specific protein ID from genbank, incorporating separate OGs for duplicated genes when necessary. Gene status was then coded using a “0” (present), “1” (pseudogene) or “2” (missing), in order to inform character-based analyses (Figure 2).

Organisms with 1 or more flagellar genes (as identified by OGs) were aligned using the whole genome aligner ProgressiveMAUVE using the default parameters (Darling, Mau, and Perna 2010). Seed families were used with minimum LCB scores which calculates locally colinear, syntenic blocks (LCB). Double cut and join distance (DCJ) were used to estimate recombination and conservation of LCBs. Further, MUSCLE (*135*) was used as the “iterative refinement” tool to verify that aligned regions had secondary support.

OGs were manually aligned using MUSCLE (*135*) and MAFFT (*136*), both using the default parameters, to validate the robustness of any aligned feature. MAFFT and MUSCLE use distinct alignment methods and by using both tools, confidence in our predicted sequence similarities were increased. Genes were then manually compared to the other orthologs for abnormal traits (i.e., large gaps and/or disruption of start/stop codons). Genes that were identified as pseudogenes are described in Table 3.

We created an additional OrthoMCL database to further demonstrate the validity of the clustering method. Motile free-living organisms (*Salmonella enterica* enterica LT2, *Escherichia coli* K-12, *Alcaligenes aquatilis*, *Terribacillus goriensis)* were chosen to represent a wide phylogenetic spectrum. These organisms were included in combination with our 30 organisms of interest to build our test database.

Ancestral state reconstruction. A phylogenetic tree was constructed using a core gene set (n=35) and was used for the ancestral state reconstruction backbone (Table 4). Core genes were defined as single copy orthologs present in all the organisms within our data set. The resulting ortholog groups were aligned using MUSCLE (*135*), and

masked using trimAL (*137*). The automatic trimming heuristic optimized for Maximum Likelihood (ML) tree construction within trimAL was used. Trees were built using RAxML (*138*, *139*) and MrBayes (*140*, *141*) with both masked and unmasked data as input. Data sets were converted to Nexus files using the ALignment Transformation EnviRonment (ALTER) (*142*). Substitution models were identified based on Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) calculated using the model fit tool in MEGAX (*143*, *144*). Burnin for Bayesian analysis was determined based on the output from the individual runs visualized in Tracer (*145*). Core genes were binned into Cluster of Orthologous Groups of proteins (COGs) using EggNOG (*146*). Due to the weak branch support for the placement of *Sodalis*-clade organisms with Maximum Likelihood, a *Sodalis* sub-tree was constructed using the same process on the *Sodalis*-allied organisms.

In order to reconstruct the flagellar operon state at ancestral nodes we used the R packages “ape” (*147*), “phytools”(*148*), and “Geiger” (*149*). We estimated the state of each ancestral node based on several theoretical evolutionary models for both trait covariance and transitional probabilities (e.g. OU (*150*), Dollo (*151*) or Wagner (*152*). We verified the fit of each model by calculating Akaike information criterion using the command “fitdiscrete”. An ultrametric tree topology was estimated using the “Chronos” commands (*147*) within the “ape” package by making a penalized likelihood estimation. “Make.simmap”(*148*) was used to build 1000 stochastic character maps that were then summarized to create ancestral state probabilities. Finally, we superimposed those states onto the ultrametric tree to determine internal node state probabilities.

Pseudogene identification*.* Flagella genes were chosen based on function within the flagellar synthesis pathway. The sequential nature of this cascade required the inclusion of several major operons as well as the inclusion of key functional genes. The *fliE-fliR* and the *flgB-flgM* operons code for the synthesis of many of the proteins required for the formation of the basal body and type three secretion system (T3SS) both of which are essential for the final assembly of the extracellular components (Figure 1). It is also well known that *Buchnera aphidicola,* the primary symbiont of the aphid, has conserved a basal body structure hypothesized to be used for secretion (*57*). FliA, the sigma factor, was included due to its essential role in the synthesis of the flagellin proteins (*153*). We have also chosen to include the genes responsible for flagellin synthesis (FliC/D) to give insights into the conservation of motility genes (Figure 1). Further, we have included the flagellar motor proteins (MotA/B) to demonstrate the potential for motor function.

To identify a pseudogene, we built an OrthoMCL database based on our 30 organisms of interest. Additionally, we used the resulting OGs to characterize the state of each flagellar gene evaluated. Several genes were represented as multiple OGs, due to two copies in some of the genomes. We validated this divergence in OGs using manual alignments. We used a phylogenetic tree to further inform the evolutionary divergence of the two OGs (Figure S1a-f). Upon manual alignment of the FliP OG we noted two

distinguishable groups due to the differences in retention of an N-terminal block (Figure S1a) which was further supported by ML (Figure S1a). Therefore, we found that *fliP* was duplicated in SOPE, *S. melophagi,* and *S. praecaptivus* (Figure 4). Further, *fliM* and *flgM* were duplicated in SOPE, *S. praecaptivus*, and *S. glossinidius* while *flgK* was duplicated in only the former two (Figure 4). Lastly, *S. glossinidius,* and *S. praecaptivus* have duplicates of *flgN* (Figure 4).

We chose to run the Ancestral state reconstruction (ASR) using several of our constructed trees to further verify that subtle differences in topologies, and branch length did not affect the overall predicted state at each internal node. Before we ran the analysis, our input tree was converted into an ultrametric tree using a penalized likelihood estimation (Figure S5). ASRs were built using stochastic character mapping, utilizing a Bayesian approach to generate internal node probabilities with a given mutation model. We used a custom Dollo model that assumed unidirectional transitions (*151*). The pseudogenes were considered an intermediate step when one was present. We summarized this data by averaging probabilities at each leaf tip and node. Figures 8a and 8b shows the overall probability that a given internal node has a present, absent, or pseudogenized flagellar pathway.

Results

Presence and absence of flagellar genes across organisms of interest. The Firmicute *T. goriensis* had 36% of genes of interest fail to cluster (Figure 4). The Betaproteobacteria *A. aquatilis* only had one gene (*flgN*) fail to cluster (Figure 4) The Enterobacteriaceae *Sa. enterica* and *E. coli* had all the genes of interest cluster (Figure 4). These results validate our overall method for identifying OGs in the closely related clade that we tested. Due to the confirmation of our methodology with these controls we felt confident the genes that did not cluster within our OGs were justly classified as missing.

S. praecaptivus had an ortholog representative in every flagellar gene of interest (Figures 4 & 5). Likewise, *S. glossinidius* had these genes fully conserved but unlike *S. praecaptivus* lacked duplicates of *flgK, flgL,* and *fliP* (Figures 4 & 5). The remaining organisms ranked by conservation of flagella genes of interest were *S. melophagi* (85%), SOPE (80%), *S. fluctus* (59%). The symbiont of another dipteran, the sheep ked, *S. melophagi,* lacked *fliL,* which encodes a protein that enhances motility on more viscous surfaces (*154*, *155*), and did not have duplicates for *fliM, flgM, flgK, and flgL* (Figures 4 & 5) SOPE, the obligate symbiont of the grain weevil, specifically is missing the genes required for the synthesis of the flagellar filament (*fliC* and *fliD*), one of the hook/filament bridge proteins (*flgL*), and a critical motor component (*motA*) (Figures 4 & 5). SOPE is also missing several key genes for the formation of the T3SS, including the gene that encodes the L-ring (*flgH*) (*72*), a critical energy transducer (*fl*iJ) (*156*, *157*), and the N-acetylglucosaminidase important for bridging the cell membrane (*fl*gJ) (*158*) (Figure 4). *S. fluctus ,* seal louse symbiont, lacks many of the genes required for the regulation of flagellar synthesis lacking two chaperones (*flgN*, *flgA*), the hook-length controller (*fliK*) (*159*), and the alternate sigma-factor (*fliA*). We found *fliI*, an ATPase

essential for unfolding of key proteins to aid in secretion, to be pseudogenized (*156*), (Figure 4). Further, the genome is without key structural genes (*flgF*, *flgK*, *fliL*, and *fliM*) (Figure 4). *S. fluctus* was missing the *flgK* protein, a bridge protein for the hook/filament complex, as well as the rod protein *flgF* (Figure 4). The *flgK* and *flgN*, also missing, form an essential regulator for the transcription of the anti-sigma factor *f*lgM (*160*) (Figure 4). A key component of the motor switch complex that regulates the directionality of rotation *fliM*, is missing as well (*102*, *161*) (Figure 4). Lastly, a secondary component of the motor switch complex (*fliN*) was found to be pseudogenized. Despite the widespread losses, *S. fluctus* still retains the genes for both the hook and filament (*flgE* and *fliC*, respectively) as well as the genes for motor complex (*motA* and *motB*) (Figure 4).

We also explored organisms which coinhabited the hosts of the *Sodalis*-allied organisms to better understand the effects of insect phylogeny and dietary ecology on the evolution of the flagellar genes (Table 1). Further, we included *Buchnera* species, the obligate symbiont of the pea aphid, due to their conservation of the basal body-like structure (*56*, *57*). Bartonella *melophagi,* which shares the same host as *S. melophagi,* has all but lost the pathway, retaining only a pseudogenized version of *fliQ* (Figure 4). *Arsenophonous melophagi,* another symbiont present within the sheep ked, retains none of the genes required for flagellar assembly (Figure 4 & 5) *Hamiltonella defensa,* described within both white flies and aphids (*162*), had three intact genes *fliI*, *fliP,* and *fliQ,* each responsible for a separate component of the TTSS pathway (Figure 4). FliI, as described above, is an essential gene required secretion. FliP and FliQ are both essential genes for a dual-purpose nanotubule pore used for both flagellum structure and plasmid acquisition (*163*) (Figure 4).

Wigglesworthia glossinidia, the tsetse co-symbiont with *S. glossinidius,* has a well conserved flagellar pathway missing four regulatory genes *(fliL, fliO, fliJ, flgM)*. As stated above,FliL, regulates motility in differing viscous environments. FliO, a scaffolding protein, helps promote the formation of the nanotubule FliP (*163*). FliJ, as stated above, is an essential transferase. FlgM, the anti-sigma factor, prevents the activation of the sigma factor FliA, the regulator for the extracellular components. Though the regulators are missing, many of the genes that are directly impacted by FlgM are still present (e.g., the alternate sigma-factor *fliA*) as are many of the other components of the nanotubule mechanisms affected (e.g., FliQ).

Three *Buchnera aphidicola* species were included in this analysis to give representatives to the three large subclades (*130*). The Aphidinae representative (*B. aphidicola APS)* is the most divergent of the species sampled, but having the fewest genes lost over evolutionary time (*130*). In total *B. aphidicola* APS is missing 12 genes (*flgN*, *flgM*, *flgA*, *flgC*, *flgL*, *motA*, *motB*, *fliC, fliA*, *fliJ*, *fliL*, *fliO*). Much of the pathway required for transcription of the Class III genes are missing (*flgN*, *flgM*, *fliA*), as well as most of the genes associated with motility (*motA*, *motB*, *flgL*, *fliC*). An additional motility regulator, *fliL,* is also missing. *B. aphidicola* APS is also missing genes associated with the T3SS (*fliO*, *fliJ*). The basal body genes *flgA,* and *flgC* are also missing. The hook

length regulator, FliK, is pseudogenized, as is *flgJ. B. aphidicola Bp* has much the same structure with the presence of *flgN* and *flgC* and the lack of *flgD* and *flgE* which are both key parts of the flagellar rod complex. *B. aphidicola* BCc is missing all of the genes both other species of *B. aphidicola* are missing with the addition of *flgB, flgG, fliE, fliM. B. aphidicola BCc* also has *fliF* pseudogenized.

Gene orthology and phylogenetic analyses. Carsonella was removed from our analyses due to its lack of phylogenetic consistency which has been previously reported due to the phylogenetic attraction to the most A+T rich genome present within the phylogenetic analysis (*164*). he reduced genomes of *Carsonella* tend to be variable in their phylogenetic placement, often resulting in a monophyletic relationship to the most A+T rich genome within a selected tree. These inconsistencies have resulted in uncertainty in the phylogenetic placement of of *Carsonella* (*164*). A total of 35 universal single copy genes were identified with approximately 7000 sites for each organism. Using the annotation tool eggNOG, we were able to bin each shared gene into a COG category (Table S1, Figure S1). Genes were categorized as ‘translation, ribosomal structure and biogenesis’ (n=28), ‘transcription’ (n=2), ‘posttranslational modification, protein turnover, chaperones’ (n=2), ‘replication, recombination and repair’ (n=2), and ‘amino acid transport and metabolism’ (n=1). Based on modeltest we were able to identify the LG+G+F and Mtrev24 +G+F (*165*) models as the two with the best fit. We additionally verified the phylogeny based on the WAG and Blosum42 models due to their common use in studying bacteria. Importantly, we further validated that each individual core gene did not bias our overall topology by building individual gene trees. Any gene tree that showed unusual evolutionary patterns were removed from our core gene set and then rerun to verify that these did not skew phylogeny. As a supplementary step, to further address short branch lengths, we built a tree using PartitionFinder2 (*166*) to model the evolutionary change within each specific gene. Trees were generated with both MrBayes, and RAxML to further verify individual topologies. We used *Sulcia muelleri* as our root in trees with all organisms of interest. Our analysis showed that there was a monophyletic group of organisms with short branch lengths consistently identified within our phylogeny (Figure S2). Though the majority of this clade were *Sodalis*-allied organisms, there was a subclade of three non-*Sodalis* (*Hamiltonella defensa, Arsenophonous melophagi, Serratia symbiotica)* (Figure 5) (*61*). Further, we found two other consistent subclades within the monophyletic group. A subclade consisting of SOPE, *S. praecaptivus,* and *S. melophagi* and a subclade consisting of *S. fluctus,* the *Moranella endobia* polytomy, *Sodalis* endosymbiont of *Ctenarytaina eucalypti*, and the secondary endosymbiont of *Henestaris halophilus.* We found a branch length of zero between *S. praecaptivus* and *S. melophagi. Buchnera* species and *Wigglesworthia glossinidia* were the two closest ancestors to the *Sodalis*-clade. *Tremblaya princeps* species, *Zinderia insecticola, B. melophagi,* and *Sulcia muelleri* had a long branch length due to their distant relationship to the Enterobacericiae. The short branch lengths prohibited meaningful ancestral state reconstruction within the closer relationships of the *Sodalis*-clade (Figure 5). To address these short branches, we built a *Sodalis*-clade core gene set to strengthen the support for many of the internal nodes.

The *Sodalis*-clade shared 136 single copy genes with approximately 42,000 coding sites (Table S2). More than half (n=75) were binned into the “Translation, ribosomal structure and biogenesis” COG (Figure S3). A total of 12 genes were binned into COGs related to metabolism (Figure S3). To optimize branch support and branch length, we used the overall data and the separate individual COGs for independent phylogenetic analysis. Additionally, we evaluated all the COGs using the PartitionFinder2 algorithm to better model independent substitution rates of each OG. We found that this partitioned approach yielded the best overall topology in both RAxML and MrBayes (Figure 7 and S4). *Sodalis praecaptivus* was used as the outgroup due its proto-symbiotic evolutionary state. The resulting tree had an improved overall resolution, with the resulting topology having three monophyletic clades (Figure 7). The first most recently diverged clade was found to include three of the mealybug symbionts (*Mikella endobia, Gullanella endobia, Doolittlea endobia),* a symbiont of the scale insect (*Hoaglandella endobia),* and the secondary symbiont of the Leucaena psyllid (secondary endosymbiont of *Heteropsylla cubana)* (Figure 7). The second most recently diverged clade has two members, the secondary symbiont of the tsetse fly (*S. glossinidius),* and the secondary symbiont of the meadow froghopper (*Sodalis*-like symbiont of *Philaenus spumarius).* The least diverged clade has 4 members, a polytomy of the two mealybug symbiont species (*Moranella endobia)*, the primary symbiont of the Lygaeoid bug (*Sodalis* endosymbiont of *Henestaris halophilus*), and the secondary symbiont of the Blue gum psyllid (secondary endosymbiont of *Ctenarytaina eucalypti*). The branch length of zero between *S. praecaptivus,* and *S. melophagi* was resolved only within the RAxML tree (Figure S3).

Ancestral State Reconstruction. Within the aphid associated organisms (Table 1) we see the closest ancestor to the *B. aphidicola* species had a 50% chance of having all the genes that we explored. In contrast, *Se. symbiotica, A. symbiotica,* and *H. defensa* have an approximate 25% chance that the common ancestor has the flagellar proteins explored. Within the mealybug symbionts (Table 1) we see a uniform evolutionary outcome, the absence of all flagellar proteins, though the common ancestor appears to likely have different flagellar states. The recent ancestor of the *Mo. endobia* species is likely (58%) to have an intact flagellar pathway. *Mi. endobia,* and *G. endobia* likely evolved from a symbiont without the key flagellar pathway (25%). The common ancestor of *T. princeps* does not have any flagellar genes, but the more ancestral node does show that the genes are likely present here. Due to the long branch length it is unclear if this is an artifact of the organisms sampled. The sheep ked symbionts have a dichotomy of ancestral flagellar states. *S. melophagi,* and *B. melophagi* both have an ancestor likely to have a flagellar pathway. Like the *T. princeps, B. melophagi* ancestral state might be due to long branch length. *A. melophagi,* the other sheep ked symbiont, is likely to originate from an organism without the flagellar genes present. The meadow froghopper symbionts are both likely to have evolved from an ancestor who has all the flagellar genes. Both Psyllid species symbiont ancestors do not appear to share the same genotype. Our analyses demonstrate that secondary endosymbiont of *Heteropsylla cubana* likely evolved from an organism lacking any flagellar genes (28%

present), in contrast the secondary endosymbiont of *Ctenarytaina eucalypti* likely evolved from an organism that had the flagellar genes (60%). Both tsetse fly symbionts (*S. glossinidius,* and *W. glossinidia)* likely evolved from an organism who had all the flagellar genes examined.

Discussion

Flagellar conservation in tsetse symbionts. W. glossinidia is an ancient obligate symbiont with a reduced genome found within the tsetse fly. Through co-evolution with its host, the *W. glossinidia* genome has purged any genes that have not been advantageous for its perpetuation within the tsetse fly. Our results further emphasize this idea through the extreme levels of conservation throughout the flagellar pathway. The motility of the *W. glossinidia* has been suggested to be essential for the infection of the larva '*in utero'* within the tsetse fly (*55*). A comparison with other highly reduced symbionts, such as *C. ruddii,* and *T. princeps* suggest that the conservation of these genes are a result of selection. The presence of a majority of flagellar genes within all of the newly acquired *Sodalis* species (*61*) negates our ability to use *S. glossinidius'* conservation of the flagellar pathway as evidence of the essential role this pathway might play within all members of the tsetse fly micro-ecosystem. However, the lack of pseudogenes in *S. glossinidius* seems to support our central hypothesis that host pressures lead to conservation within the tsetse fly system.

Repurposing flagella in symbiotic lifestyles. The aphid symbionts (*Buchnera aphidicola, Serratia symbiotica,* and *Hamiltonella defensa),* each have a different level of conservation of genes associated with the flagellar pathway. The *B. aphidicola* species are known to have repurposed the basal body for secretion (*56*). Our analysis corroborates this, showing a common ancestor likely possessing a flagellar pathway. Our data calls into question the importance of this pathway among the extant *B. aphidicola* species. The three largest represented *B. aphidicola* lineages each have different levels of genomic reduction (*167*). We show that genomic reduction can be seen within pathways suggested to be essential within *B. aphdicola BCc* for the symbiotic lifestyle. *B. aphidicola BCc* has become extremely reduced over evolutionary time (*168*). The aphid (*Cinara cedri*), has acquired *S. symbiotica* to fill in functional gaps (*169*, *170*). We believe the more reduced version of this complex molecular machine is the result of the new co-symbiotic lifestyle. This demonstrates the continuous nature of symbiosis, showing ongoing evolutionary reduction through the acquisition of new symbiotic partners.

Repeated replacement of reduced symbionts. The mealybugs have evolved a tripartite symbiosis with *Tremblaya princeps* and an assortment of other *Sodalis* symbionts (*122*). Our analysis recapitulates the phylogenetic placement of these symbionts in the *Sodalis* clade, further emphasizing the independent establishment of each *T. princeps* partner. The ancestral state reconstruction supports the loss of the flagellar pathway in *Mi.*

endobia, and *G. endobia* prior to symbiosis establishment (Figure 8b). This is consistent with the hypothesis that *T. princeps* acquires already reduced partners from other host systems (*122*).

The symbiont of the Leucaena psyllid can illustrate the additional potential for an already reduced symbiotic partner. The 1.2 Mb genome (*61*) suggests a reduced genomic toolset, and the low percentage of coding regions indicate the potential for further reduction. These similar genomic features are scattered amongst these suspected host-jumping symbionts, many of which seem to have the like of an intermediate state.

This ongoing process of symbiont loss and acquisition might be better explained through the lens of the “Shopping bag” hypothesis of organelle evolution (*24*). This hypothesis gives insights to the chimeric nature of organelle genomes and provides an explanation for the host acquisition of symbiont genetic material. The symbiotic relationship within the mealybug has already begun to infer this trajectory with the acquisition of genes required for some symbiotic processes (*171*).

The *Sodalis* allied symbionts are a clade of organisms that fulfill different roles in an assortment of different hosts (*61*). The presence of a *S. praecaptivus* proto-symbiont allowed us to explore evolutionary trajectories within a dynamic group of biological scenarios. We used this natural experiment to demonstrate evolutionary outcomes associated with symbiosis (i.e., genome reduction) do not explain the predicted conservation of the flagella genes or lack thereof within ancestral nodes. Further, we demonstrated that there are unexpected levels of conservation as compared to the known size of the genome in *W. glossinidia* (*61*), with a possible connection to host reproduction and infection through vertical transmission. We also demonstrated that certain organisms likely have undergone flagella loss prior to acquisition by their current host.

Table 1: *Sodalis*-clade organisms, co-symbionts and respective hosts. Organisms were chosen due to either belonging to the Sodalis-clade or residing within the same host. Included are the host scientific name and common name. The organisms are sorted by host taxonomy. Sodalis praecaptivus is believed to be free-living. (Clayton et al. 2012)

Organism	Host (scientific name)	Host (common name)	Genome citation
Sodalis-like endosymbiont of *Proechinophthirus fluctus* str. SPI-1	*Proechinophthirus fluctus*	Louse fly	Boyd et al 2016
Ca. *Mikella endobia*	*Paracoccus marginatus*	Papaya mealybug	Thao et al 2000
Tremblaya princeps PMAR	*Paracoccus marginatus*	Papaya mealybug	Husnik and McCutcheon 2016
Ca. *Moranella endobia* PCVAL	*Planococcus citri*	Citrus mealybug	Thao et al 2000
Tremblaya princeps PCVAL	*Planococcus citri*	Citrus mealybug	Thao et al 2000
Ca. *Moranella endobia* PCIT	*Planococcus citri*	Citrus mealybug	Thao et al 2000
Tremblaya princeps PCIT	*Planococcus citri*	Citrus mealybug	Thao et al 2000
Ca. *Hoaglandella endobia*	*Trionymus perrisii*	Scale insect	Husnik and McCutcheon 2016
Ca. *Doolittlea endobia*	*Maconellicoccus hirsutus*	Pink, Grape or Hibiscus mealybug	Husnik and McCutcheon 2016
Ca. *Gullanella endobia*	*Ferrisia virgata*	Striped mealybug	Husnik and McCutcheon 2016
Serratia symbiotica Cinara cedri	*Cinara cedri*	Conifer aphids	Lamelas et al 2011
Ca. *Hamiltonella defensa*	Several	White flies or aphids	Tatusova et al 2014
Buchnera Bp	*Baizongia pistaciae*	Pistacia horn gall aphid	van Ham et al. 2003
Buchnera BCc	*Cinara cedri*	Conifer aphids	Pérez-Brocal et al 2006
Buchnera APS	*Acyrthosiphon pisum*	Pea aphid	Shigenobu et al 2000
P-endosymbiont of *Henestaris halophilus*	*Henestaris halophilus*	Lygaeoid bug	Santos-Garcia 2017
Sodalis-like symbiont of *Philaneus spumarius* PSPU	*Philaneus spumarius*	Meadow Froghopper	Koga et al 2013
Sulcia muelleri PSPU	*Philaneus spumarius*	Meadow Froghopper	Koga et al 2013
Ca. *Zinderia insecticola* CARI	*Clastoptera arizonana*	Arizona Spittlebug	Mccutcheon et al 2010
Ca. *Carsonella ruddii* CE	*Ctenarytaina eucalypti*	Blue gum psyllid	Thao et al 2000
Ca. *Carsonella ruddii* HC	*Heteropsylla cubana*	Leucaena psyllid	Thao et al 2000
S-endosymbiont of *Ctenarytaina eucalypti*	*Ctenarytaina eucalypti*	Blue gum psyllid	Thao et al 2000
S-endosymbiont of *Heteropsylla cubana*	*Heteropsylla cubana*	Leucaena psyllid	Sloan and Moran 2012
Ca. *Sodalis pierantonius* str. SOPE	*Sitophilus oryzai*	Rice weevil	Oakeson et al 2014
Ca. *Sodalis melophagi*	*Melophagus ovinus*	Sheep Ked	Novakova et al 2015
Ca. *Arsenophenous melophagi*	*Melophagus ovinus*	Sheep Ked	Novakova et al 2015
Ca. *Bartonella melophagi*	*Melophagus ovinus*	Sheep Ked	Novakova et al 2015
Sodalis glossinidius str. "morsitans"	*Glossina morsitans*	Tsetse fly	Toh et al 2006
Wigglesworthia glossinidia	*Glossina morsitans*	Tsetse fly	Rio RV et al 2012
Sodalis praecaptivus	Free-living	Free-living	Clayton AL et al. 2012

Figure 1

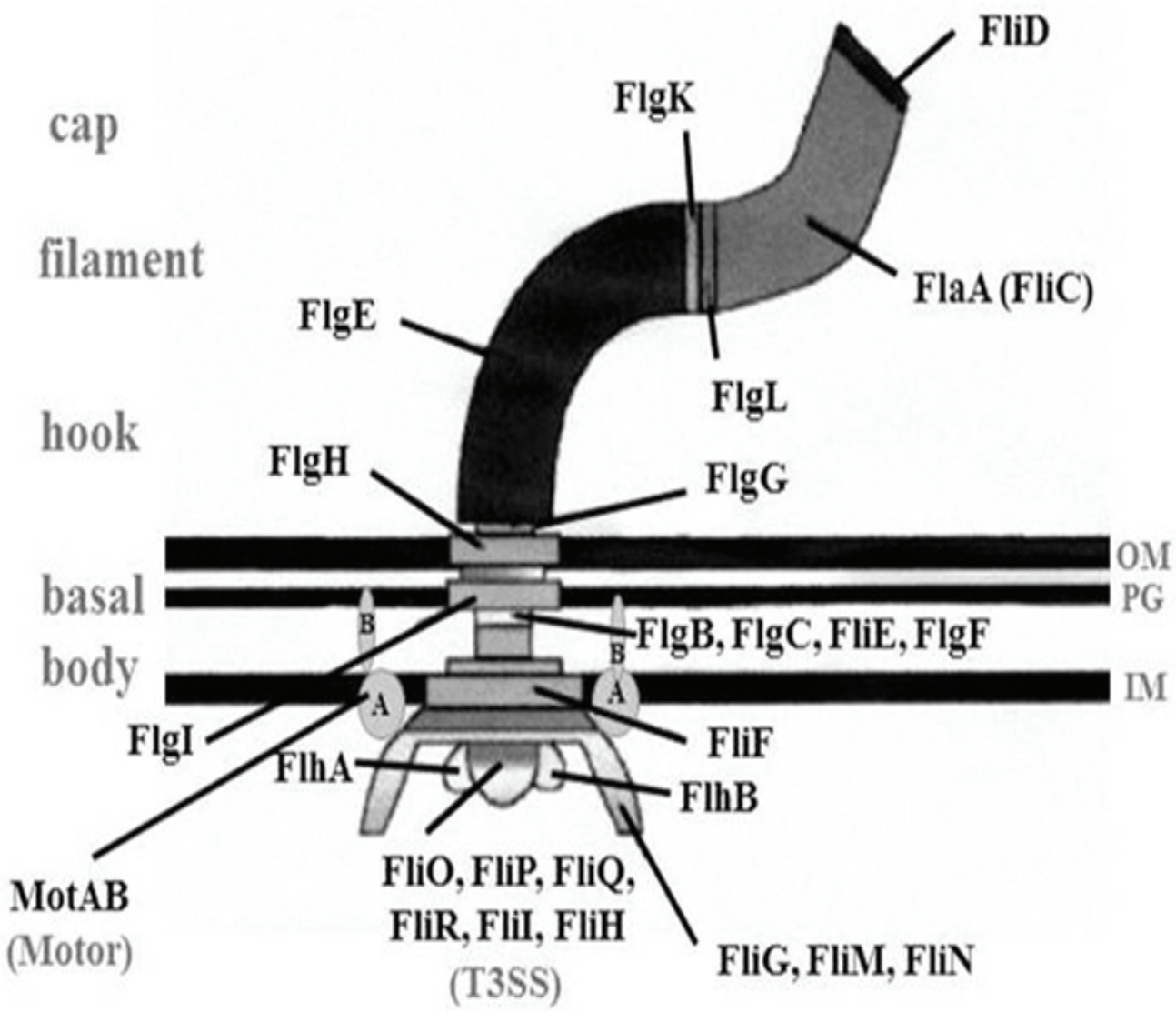

FliD
FlgK
cap
filament
FlaA (FliC)
FlgE
FlgL
hook
FlgH
FlgG
OM
basal
PG
body
B
B
FlgB, FlgC, FliE, FlgF
IM
A
A
FlgI
FliF
FlhA
FlhB
MotAB
(Motor)
FliO, FliP, FliQ,
FliR, FliI, FliH
(T3SS)
FliG, FliM, FliN

Table 2: List of genes, separated by functional description, comprising assembly of the canonical flagellar pathway within Escherichia coli (Toft & Fares, 2008).

Master control	**Regulators**	**Chaperones**	**Motor control complex**
flhC	*fliK*	*fliJ*	*fliG*
flhD	*fliZ*	*flgA*	*fliM*
Flagellar export apparatus	*fliA*	*flgN*	*fliN*
flhA	*flgM*	*fliS*	*flgH*
flhB	**Hook and filament**	*fliT*	*motA*
flhE	*flgE*	**Chemotaxis**	*motB*
fliO	*flgD*	*cheA*	*flgB*
fliP	*flgK*	*cheB*	*flgC*
fliQ	*fliC*	*cheW*	*flgF*
fliR	*fliD*	*tsr/tar*	*flgG*
fliH		*tap*	*flgJ*
fliI		*cheY*	*flgI*
		cheZ	*fliF*
		cheR	*fliE*

Figure 2

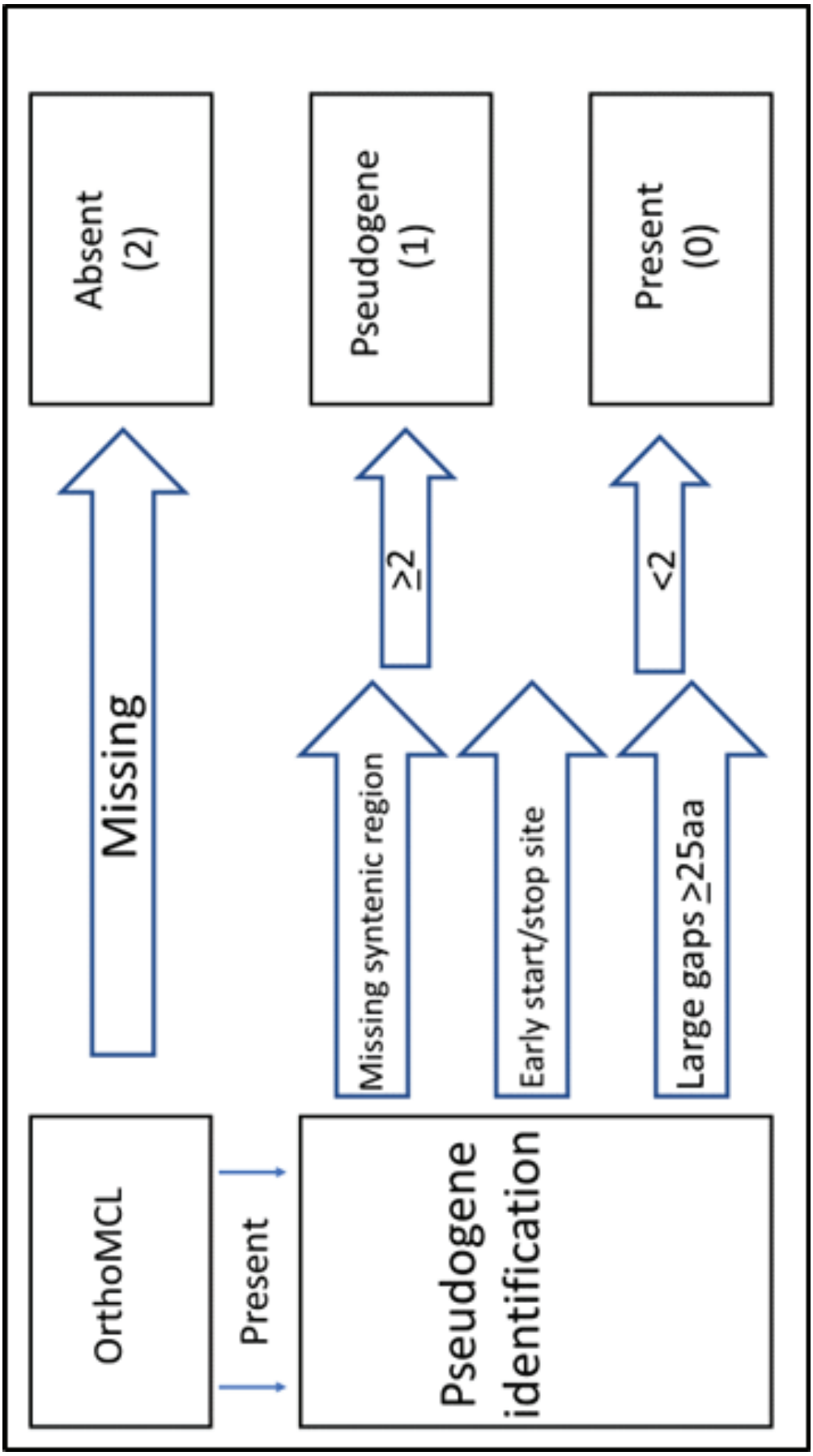

OrthoMCL
Present
Pseudogene identification
Missing
Missing syntenic region
Early start/stop site
Large gaps ≥25aa
≥2
<2
Absent (2)
Pseudogene (1)
Present (0)

Table 3: Pseudogenes identified in organisms of interest. A gene was labeled as a pseudogene based on the criteria described in Figure 2. Gene status provides the specifics as to why an individual gene was considered a pseudogene.

Organism	Protein	Pseudogene characterization
Sodalis like endosymbiont of *Proechinophthirus fluctus*	FliI	Truncation (~200 aa of ~300 aa length)
	FliN	Truncated (50 aa of 140 aa length)
Buchnera BCc	FliF	Truncated (489 aa of 560 aa average), Gap of 90 aa
Buchnera Bp	FlgJ	Truncated (270 aa of 360 aa average)
Buchnera APS	FliK	Truncated (235 aa of 385 aa average), Gap of 60 aa
	FlgJ	Truncated (270 aa of 360 aa average)
Ca. Bartonella melophagi	FliQ	Trucated (71 aa of 87 aa average), lack of LCB
	FlgC	two gaps of 30 aa, lack of LCB
	FlgD	two gaps >30 aa, lack of LCB
	FlgE	gap of 75 aa, lack of LCB
	FlgG	Truncated (176 of 260 average length), lack of LCB
Wigglesworthia morsitans	FlgJ	Truncated (270 aa of 360 average)

Figure 3

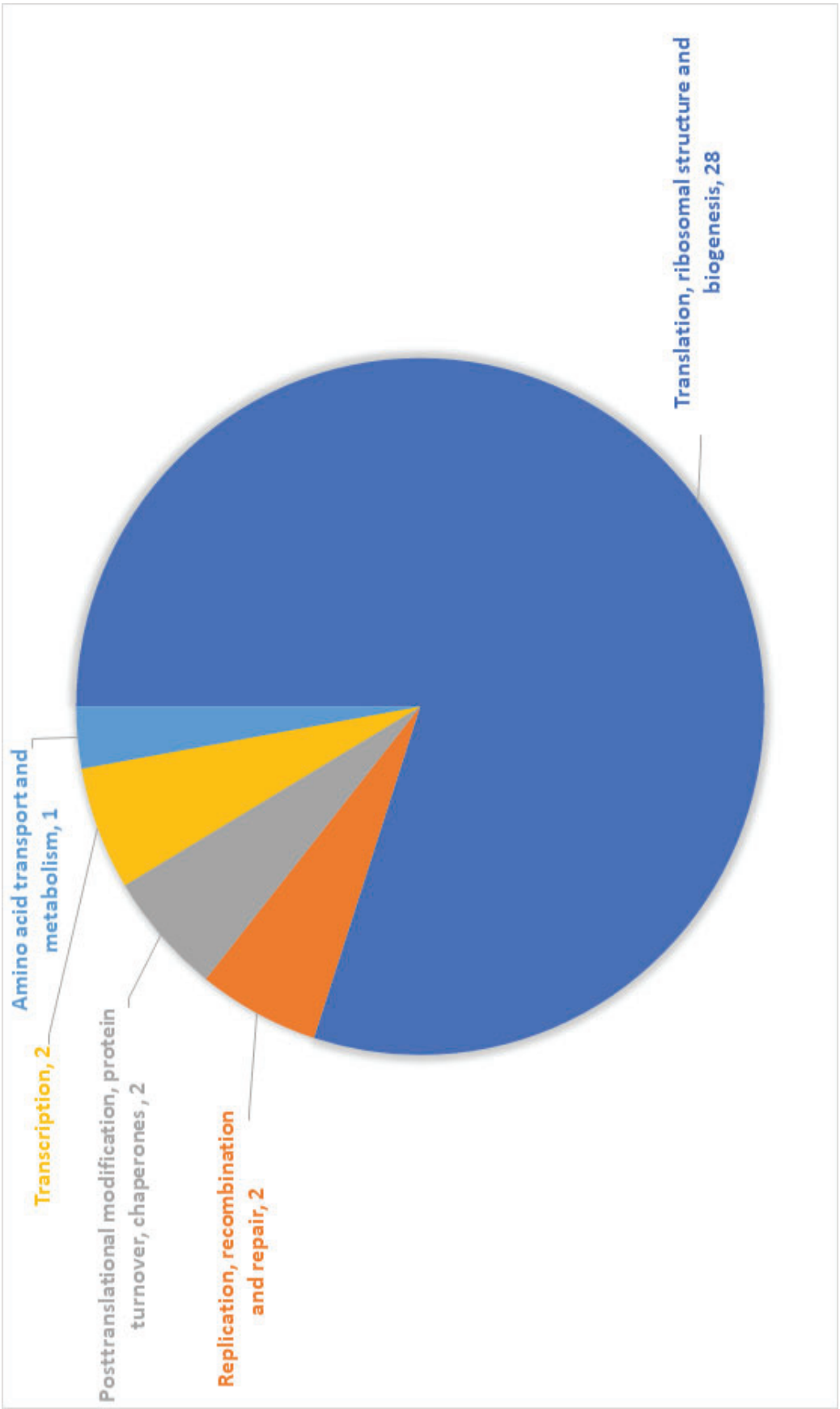

Translation, ribosomal structure and biogenesis, 28
Amino acid transport and metabolism, 1
Transcription, 2
Posttranslational modification, protein turnover, chaperones , 2
Replication, recombination and repair, 2

Figure 4

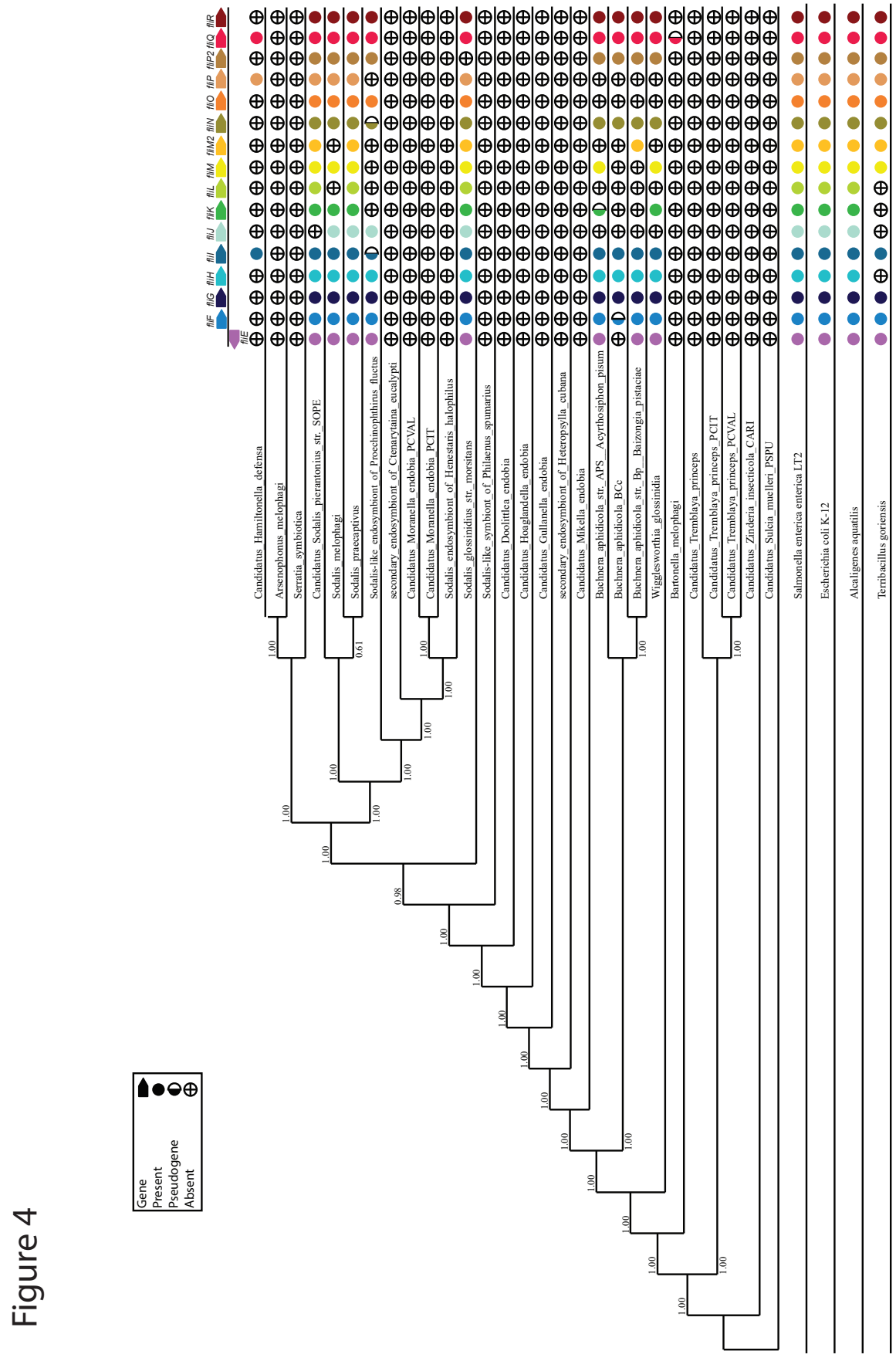

Gene
Present
Pseudogene
Absent
fliF
fliG
fliH
fliI
fliJ
fliK
fliL
fliM
fliM2
fliN
fliO
fliP
fliP2
fliQ
fliR
fliE
Candidatus_Hamiltonella_defensa
Arsenophonus_melophagi
Serratia_symbiotica
Candidatus_Sodalis_pierantonius_str._SOPE
Sodalis_melophagi
Sodalis_praecaptivus
Sodalis-like_endosymbiont_of_Proechinophthirus_fluctus
secondary_endosymbiont_of_Ctenarytaina_eucalypti
Candidatus_Moranella_endobia_PCVAL
Candidatus_Moranella_endobia_PCIT
Sodalis_endosymbiont_of_Henestaris_halophilus
Sodalis_glossinidius_str._morsitans
Sodalis-like_symbiont_of_Philaenus_spumarius
Candidatus_Doolittlea_endobia
Candidatus_Hoaglandella_endobia
Candidatus_Gullanella_endobia
secondary_endosymbiont_of_Heteropsylla_cubana
Candidatus_Mikella_endobia
Buchnera_aphidicola_str._APS_(Acyrthosiphon_pisum
Buchnera_aphidicola_BCc
Buchnera_aphidicola_str._Bp_(Baizongia_pistaciae
Wigglesworthia_glossinidia
Bartonella_melophagi
Candidatus_Tremblaya_princeps
Candidatus_Tremblaya_princeps_PCIT
Candidatus_Tremblaya_princeps_PCVAL
Candidatus_Zinderia_insecticola_CARI
Candidatus_Sulcia_muelleri_PSPU
Salmonella enterica enterica LT2
Escherichia coli K-12
Alcaligenes aquatilis
Terribacillus goriensis
1.00
0.98
0.61

Figure 5

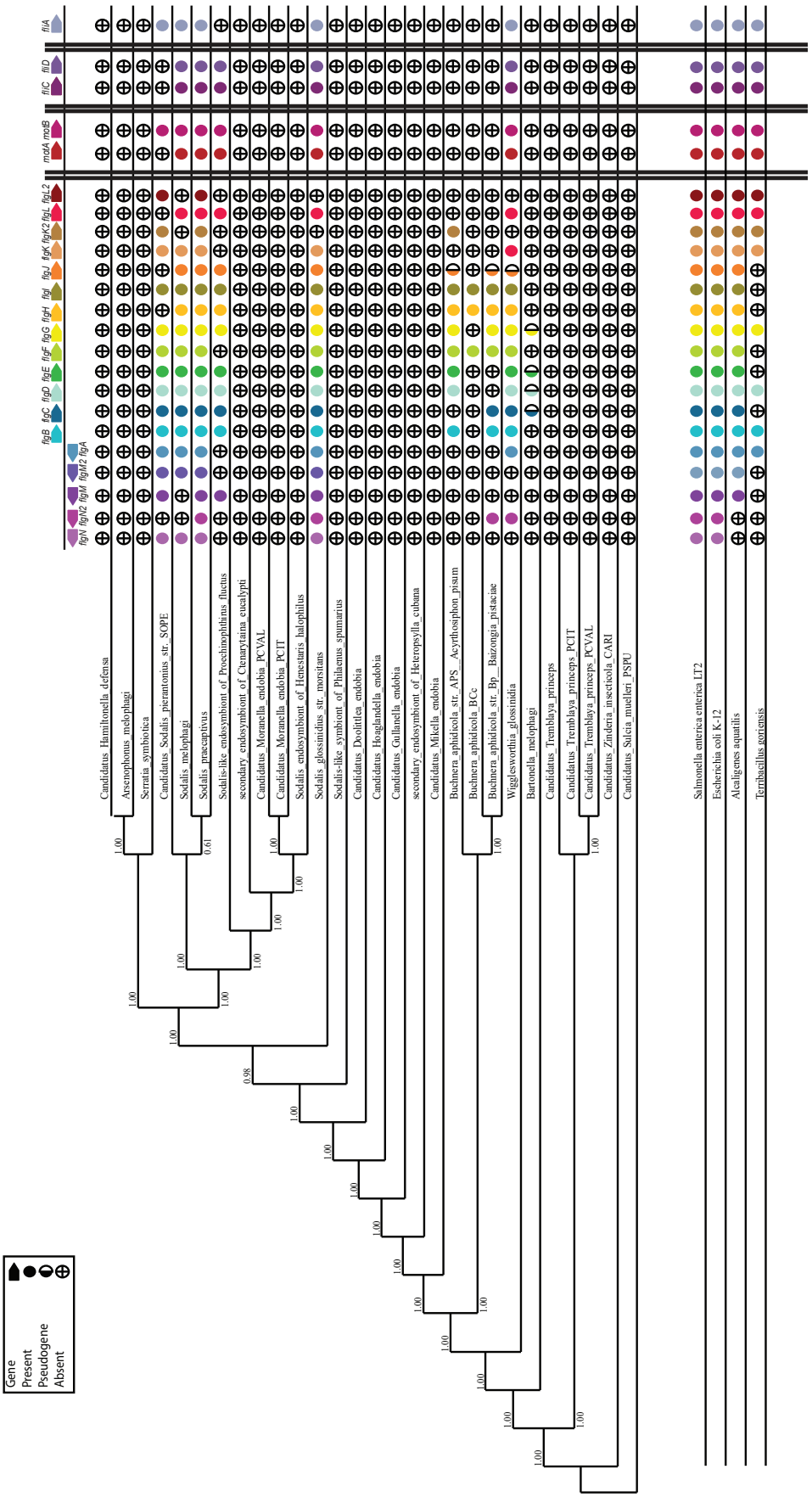

Gene
Present
Pseudogene
Absent
flgB
flgC
flgD
flgE
flgF
flgG
flgH
flgI
flgJ
flgK
flgK2
flgL
flgL2
flgN
flgN2
flgN4
flgM
flgM2
flgA
motA
motB
fliC
fliD
fliA
Candidatus_Hamiltonella_defensa
Arsenophonus_melophagi
Serratia_symbiotica
Candidatus_Sodalis_pierantonius_str._SOPE
Sodalis_melophagi
Sodalis_praecaptivus
Sodalis-like_endosymbiont_of_Proechinophthirus_fluctus
secondary_endosymbiont_of_Ctenarytaina_eucalypti
Candidatus_Moranella_endobia_PCVAL
Candidatus_Moranella_endobia_PCIT
Sodalis_endosymbiont_of_Henestaris_halophilus
Sodalis_glossinidius_str._morsitans
Sodalis-like_symbiont_of_Philaenus_spumarius
Candidatus_Doolittlea_endobia
Candidatus_Hoaglandella_endobia
Candidatus_Gullanella_endobia
secondary_endosymbiont_of_Heteropsylla_cubana
Candidatus_Mikella_endobia
Buchnera_aphidicola_str._APS_(Acyrthosiphon_pisum
Buchnera_aphidicola_BCc
Buchnera_aphidicola_str._Bp_(Baizongia_pistaciae
Wigglesworthia_glossinidia
Bartonella_melophagi
Candidatus_Tremblaya_princeps
Candidatus_Tremblaya_princeps_PCIT
Candidatus_Tremblaya_princeps_PCVAL
Candidatus_Zinderia_insecticola_CARI
Candidatus_Sulcia_muelleri_PSPU
Salmonella_enterica_enterica_LT2
Escherichia_coli_K-12
Alcaligenes_aquatilis
Terribacillus_goriensis
1.00
0.61
0.98

Figure 6

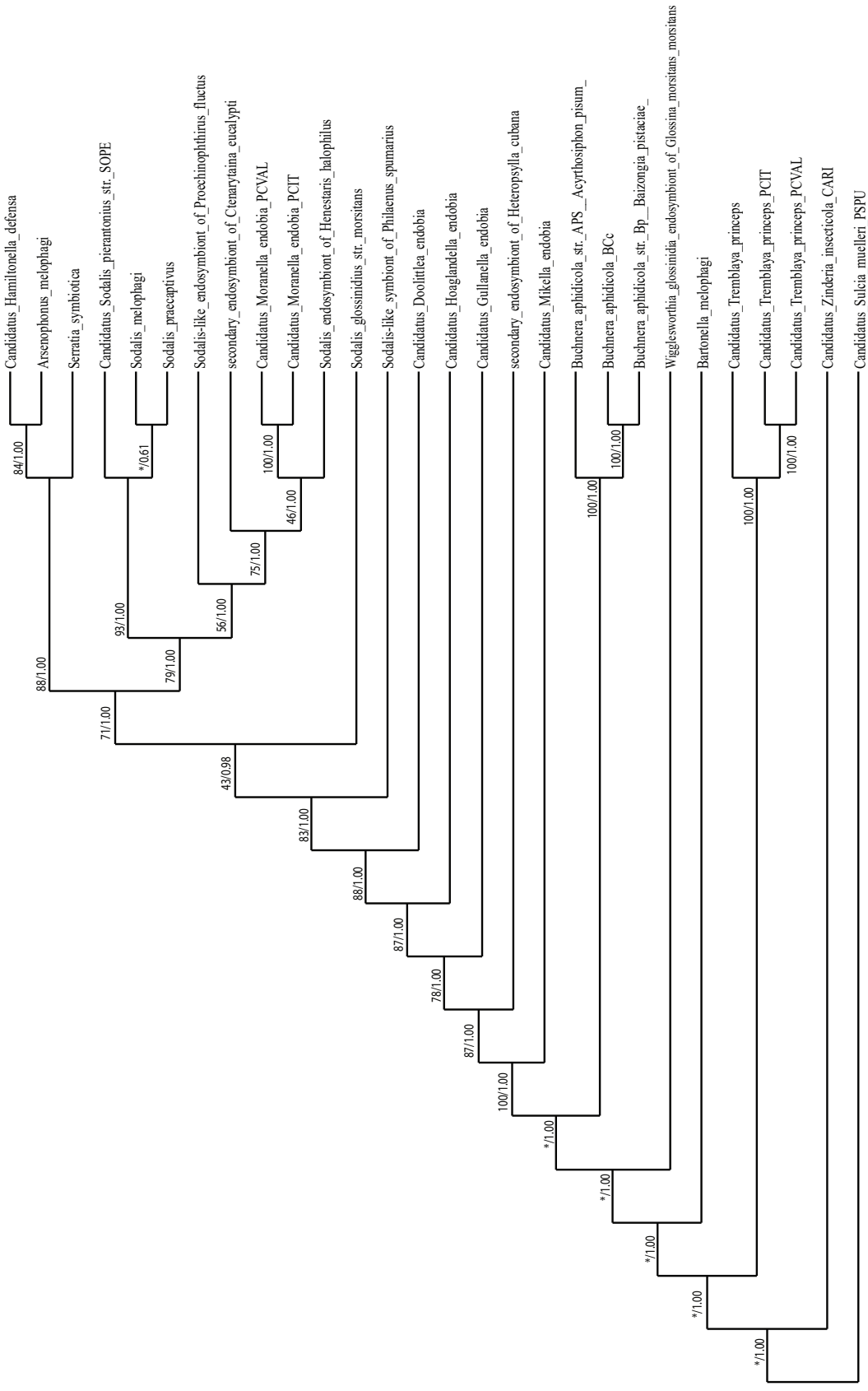

Candidatus_Hamiltonella_defensa
Arsenophonus_melophagi
Serratia_symbiotica
Candidatus_Sodalis_pierantonius_str._SOPE
Sodalis_melophagi
Sodalis_praecaptivus
Sodalis-like_endosymbiont_of_Proechinophthirus_fluctus
secondary_endosymbiont_of_Ctenarytaina_eucalypti
Candidatus_Moranella_endobia_PCVAL
Candidatus_Moranella_endobia_PCIT
Sodalis_endosymbiont_of_Henestaris_halophilus
Sodalis_glossinidius_str._morsitans
Sodalis-like_symbiont_of_Philaenus_spumarius
Candidatus_Doolittlea_endobia
Candidatus_Hoaglandella_endobia
Candidatus_Gullanella_endobia
secondary_endosymbiont_of_Heteropsylla_cubana
Candidatus_Mikella_endobia
Buchnera_aphidicola_str._APS_(Acyrthosiphon_pisum)
Buchnera_aphidicola_BCc
Buchnera_aphidicola_str._Bp_(Baizongia_pistaciae)
Wigglesworthia_glossinidia_endosymbiont_of_Glossina_morsitans_morsitans
Bartonella_melophagi
Candidatus_Tremblaya_princeps
Candidatus_Tremblaya_princeps_PCIT
Candidatus_Tremblaya_princeps_PCVAL
Candidatus_Zinderia_insecticola_CARI
Candidatus_Sulcia_muelleri_PSPU
84/1.00
*/0.61
100/1.00
46/1.00
93/1.00
56/1.00
75/1.00
88/1.00
79/1.00
71/1.00
43/0.98
83/1.00
88/1.00
87/1.00
78/1.00
87/1.00
100/1.00
*/1.00
100/1.00
100/1.00
*/1.00
*/1.00
*/1.00
100/1.00
100/1.00
*/1.00

Figure 7

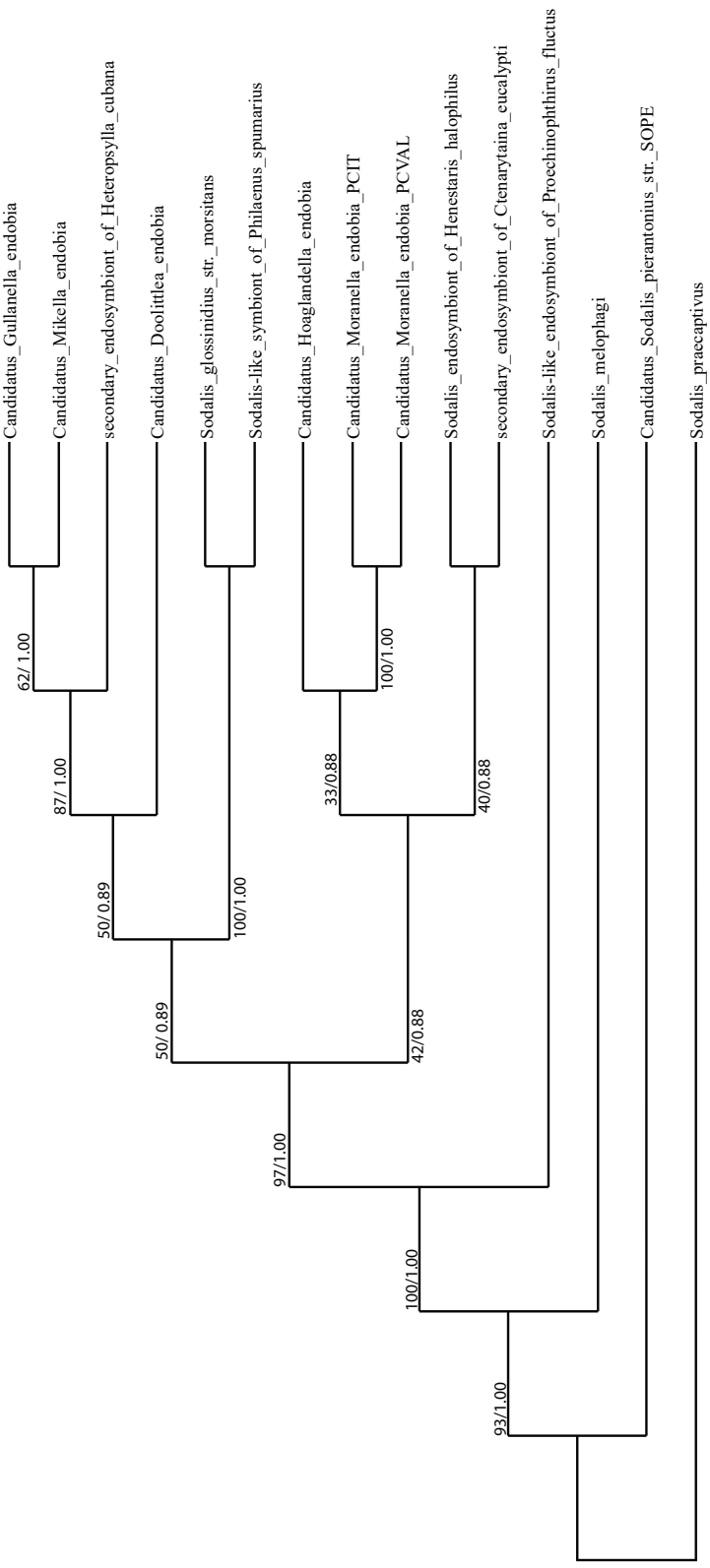

Candidatus_Gullanella_endobia
Candidatus_Mikella_endobia
secondary_endosymbiont_of_Heteropsylla_cubana
Candidatus_Doolittlea_endobia
Sodalis_glossinidius_str._morsitans
Sodalis-like_symbiont_of_Philaenus_spumarius
Candidatus_Hoaglandella_endobia
Candidatus_Moranella_endobia_PCIT
Candidatus_Moranella_endobia_PCVAL
Sodalis_endosymbiont_of_Henestaris_halophilus
secondary_endosymbiont_of_Ctenarytaina_eucalypti
Sodalis-like_endosymbiont_of_Proechinophthirus_fluctus
Sodalis_melophagi
Candidatus_Sodalis_pierantonius_str._SOPE
Sodalis_praecaptivus
62/ 1.00
87/ 1.00
50/ 0.89
100/1.00
50/ 0.89
42/0.88
33/0.88
100/1.00
40/0.88
97/1.00
100/1.00
93/1.00

Figure 8a

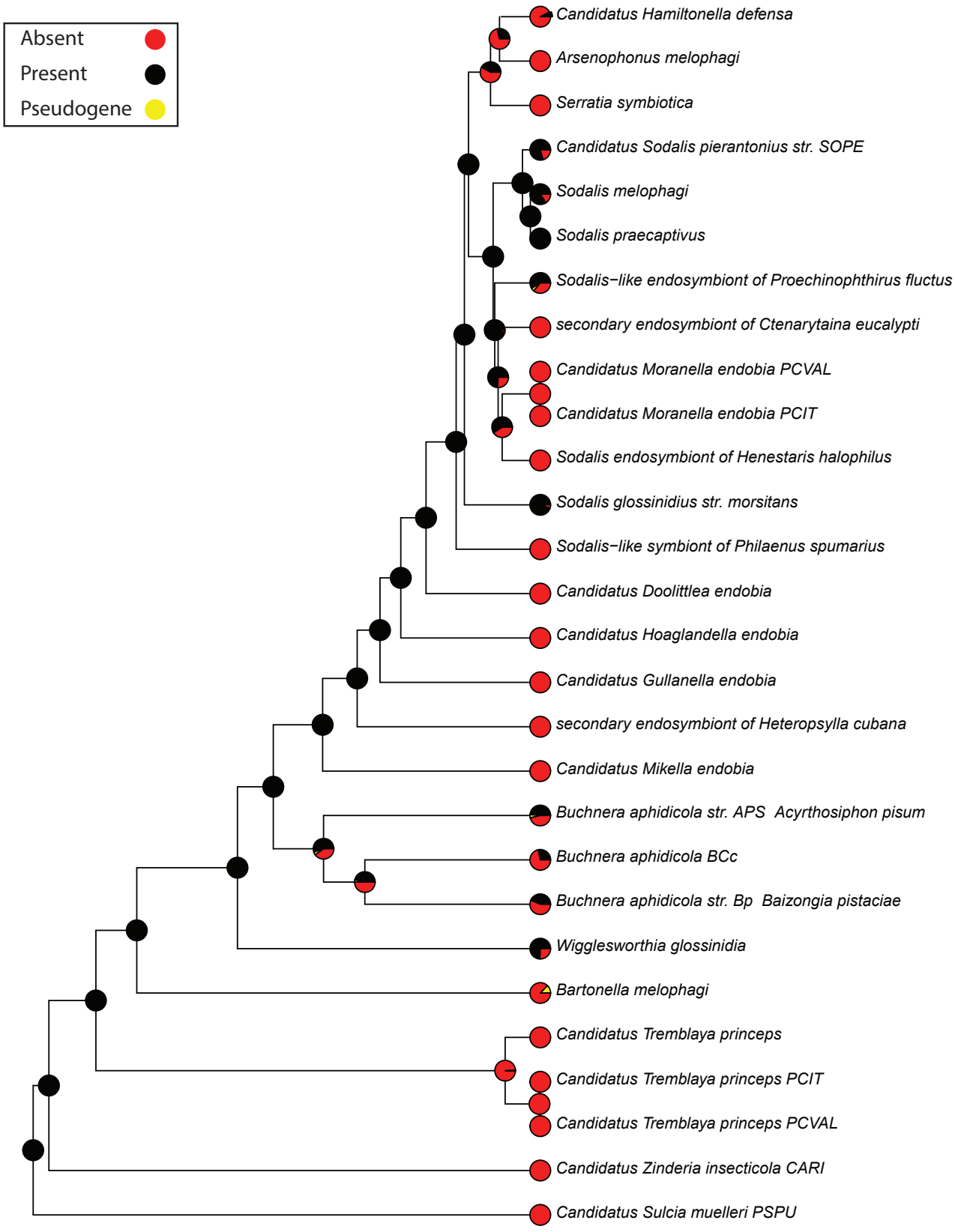

Absent
Present
Pseudogene
Candidatus Hamiltonella defensa
Arsenophonus melophagi
Serratia symbiotica
Candidatus Sodalis pierantonius str. SOPE
Sodalis melophagi
Sodalis praecaptivus
Sodalis−like endosymbiont of Proechinophthirus fluctus
secondary endosymbiont of Ctenarytaina eucalypti
Candidatus Moranella endobia PCVAL
Candidatus Moranella endobia PCIT
Sodalis endosymbiont of Henestaris halophilus
Sodalis glossinidius str. morsitans
Sodalis−like symbiont of Philaenus spumarius
Candidatus Doolittlea endobia
Candidatus Hoaglandella endobia
Candidatus Gullanella endobia
secondary endosymbiont of Heteropsylla cubana
Candidatus Mikella endobia
Buchnera aphidicola str. APS Acyrthosiphon pisum
Buchnera aphidicola BCc
Buchnera aphidicola str. Bp Baizongia pistaciae
Wigglesworthia glossinidia
Bartonella melophagi
Candidatus Tremblaya princeps
Candidatus Tremblaya princeps PCIT
Candidatus Tremblaya princeps PCVAL
Candidatus Zinderia insecticola CARI
Candidatus Sulcia muelleri PSPU

Figure 8b

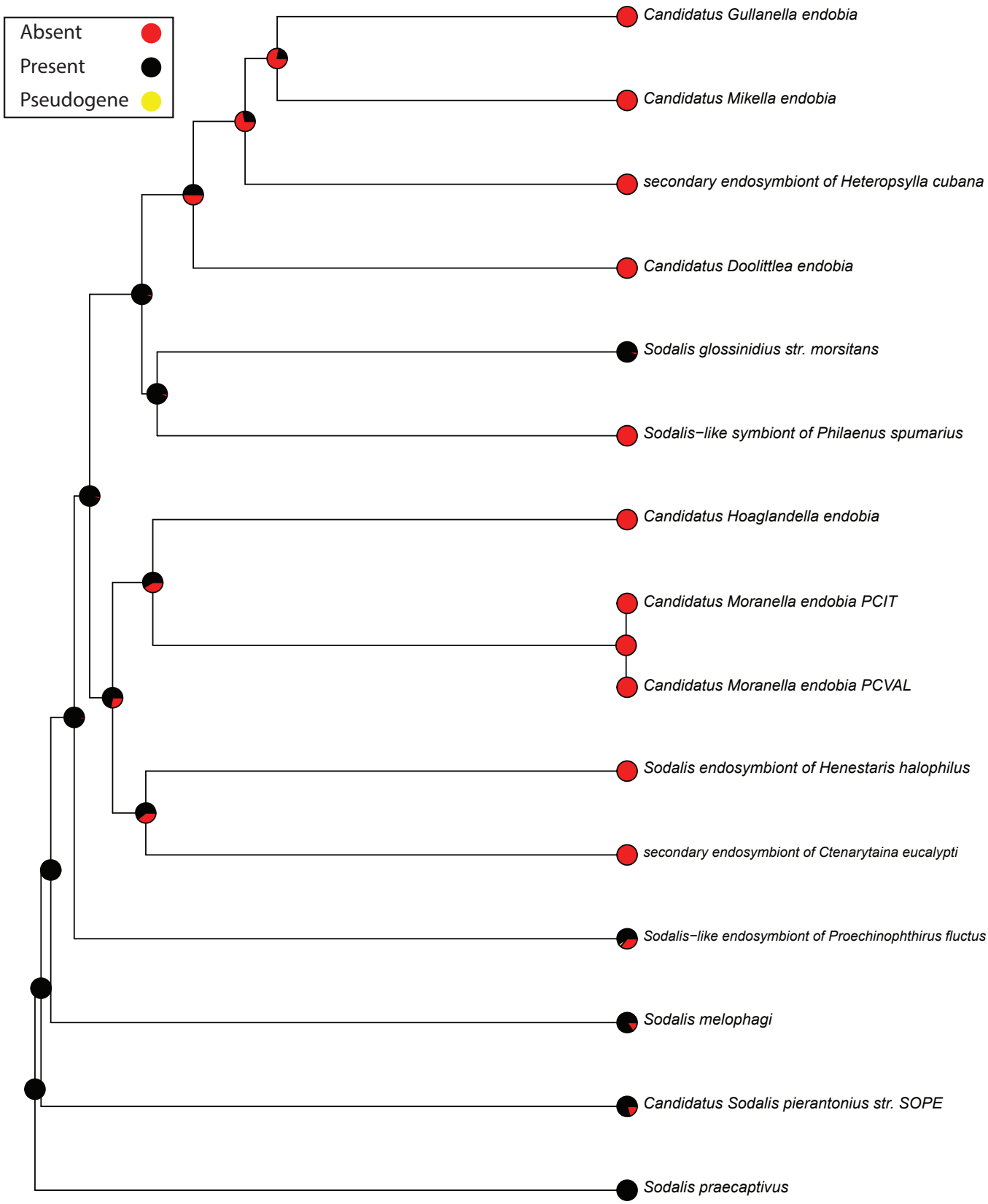

Absent
Present
Pseudogene
Candidatus Gullanella endobia
Candidatus Mikella endobia
secondary endosymbiont of Heteropsylla cubana
Candidatus Doolittlea endobia
Sodalis glossinidius str. morsitans
Sodalis−like symbiont of Philaenus spumarius
Candidatus Hoaglandella endobia
Candidatus Moranella endobia PCIT
Candidatus Moranella endobia PCVAL
Sodalis endosymbiont of Henestaris halophilus
secondary endosymbiont of Ctenarytaina eucalypti
Sodalis−like endosymbiont of Proechinophthirus fluctus
Sodalis melophagi
Candidatus Sodalis pierantonius str. SOPE
Sodalis praecaptivus

Figure Legends

Figure 1: The eubacteria flagella complex. Image was taken from Appelt and Heuner, 2017.

Figure 2: Visual representation of pipeline used to construct flagella trait table. Genes that fit ≥2 of the criteria were characterized as pseudogenes in all downstream analysis. Syntenic analysis was performed using ProgressiveMAUVE using *Sodalis praecaptivus* as the root sequence. Gene structure analyses were done manually using 2 different alignment formats (MAFFT and MUSCLE).

Figure 3: Distribution of core genes by Clusters of Orthologous Groups (COG) as determined through eggNOG. A total of 35 genes were found to be conserved across all the organisms in our data set (refer to Table 1). The majority of orthologs are grouped within Translation, ribosomal structure, and biogenesis.

Figure 4: Presence and absence of genes in the flagellar operon *fliEFGHIJKLMONPQR*. A Bayesian tree is used as a backbone with gene arrangement obtained from the *S. praecaptivus* genome. Duplicated OGs are presented in tandem within the image, but do not colocalize within the genome. Motile free-living representatives depicted at the bottom did not have duplicated genes as the distance from the *Sodalis*-clade resulted in a single OG. For this reason, we chose to show both OGs as present to make clear that these organisms do not have functional gaps. Motile free-living organisms are placed in order from most related to least related to the *Sodalis* symbionts (i.e. from top to bottom).

Figure 5: Presence and absence of genes in the flagellar operon *fl*g*NMABCDEFGHIJKL*. The presence and absence of *motAB, fliCD,* and *fliA* are also shown. A Bayesian tree is used as a backbone with gene arrangement obtained from the *S. praecaptivus* genome. Duplicated OGs are presented in tandem within the image, but do not colocalize within the genome. Motile free-living representatives depicted at the bottom did not have duplicated genes as the distance from the *Sodalis*-clade resulted in a single OG. For this reason, we chose to show both OGs as present to make clear that these organisms do not have functional gaps. Likewise, motile free-living organisms who were missing OGs that were duplicated in our general analysis are shown to be missing both OGs. Motile free-living organisms are placed in order from most related to least related to the *Sodalis* symbionts.

Figure 6: Phylogram for the organisms of interest not including *Carsonella.* Numbers denote RAxML bootstrap support/ MrBayes posterior probability. Tree topology is generated from the MrBayes analyses. "*" denotes a change in branch topology when compared to the RAxML generated tree. RAxML topology was built using the automatic heuristic model with 1000 cycles using the total core gene set (n=35). MrBayes analyses was set to 3000000 generations with Burnin decided based on tracer visualization (*145*) to build effective sample size over 200 (25% burnin) (*172*). A partitioned analysis gave similar overall topologies (*166*).

Figure 7: Phylogram for the *Sodalis*-allied organisms. Numbers denote RAxML bootstrap support/ MrBayes posterior probability values. The tree topology was generated using RAxML due to the lack of polytomy between *S. praecaptivus* and *S. melophagi.* Both trees were built with the concatenated *Sodalis* core gene set and using a partitioned model (*166*).

Figure 8: Ancestral state reconstruction for the retention of flagella genes of interest, (a) all organisms of interest, (b) *Sodalis* clade organism, derived from the presence/absence data. Probabilities at each node are the average the probability an individual node would be “present”, “absent”, or “pseudogenized” for each independent gene derived from the ancestral state reconstruction. Pie charts at the tip visualize the proabability that each organism has an overall “present”, “absent” or “pseudogenized” version form of the genes explored.

Bibliography

1. A. Frank, Uber die biologischen Verhältnisse des Thallus eineger Krustenflechnten Beitr. *Biol. Pflanz.* **2**, 123–200. (1887).
2. A. DeBary, Die Erscheinung der Symbiose. Vortrag auf der Versammlung der Naturforshung und Ärtze zu Cassel. *Verlag von Karl J. Trubn. Strassburg.*, 1–30 (1879).
3. A. F. W. Schimper, Ûber die Entwicklung der Schlorophyllkörner und Farbkörper. *Bot. Zeitung,*. **4**, 105–114. (1883).
4. S. Mereschkowsky, Über Natur und Ursprung der Chromatophoren im Pflanzenreiche. *Biol. Zentralbl.* **25**, 593– 604.
5. L. Sagan, On the origin of mitosing cells. *J. Theor. Biol.* **14**, 255–274 (1967).
6. M. W. Gray, The Endosymbiont Hypothesis Revisited. *Int. Rev. Cytol.* **141**, 233–357 (1992).
7. L. Margulis, Biodiversity: molecular biological domains, symbiosis and kingdom origins. *BioSystems*. **27**, 39–51 (1992).
8. T. Buchner, R. Pfeiffer, E. Stupperich, Reduplication behavior of chromosomes of group D (13-15) and identification of the extra-chromosome in trisomy D. **55**, 1–2 (1965).
9. A. E. Douglas, Nutritional interactions in insect-microbial symbioses: aphids and their symbiotic bacteria Buchnera. *Annu. Rev. Entomol.* **43**, 17–37 (1998).
10. N. A. Moran, D. B. Sloan, The Hologenome Concept: Helpful or Hollow? *PLoS Biol.* **13**, e1002311 (2015).
11. R. V. M. Rio, G. M. Attardo, B. L. Weiss, Grandeur Alliances: Symbiont Metabolic Integration and Obligate Arthropod Hematophagy. *Trends Parasitol.* **32**, 739–749 (2016).
12. W. F. Martin, S. Garg, V. Zimorski, Endosymbiotic theories for eukaryote origin. *Philos. Trans. R. Soc. B Biol. Sci.* **370** (2015), doi:10.1098/rstb.2014.0330.
13. N. A. Moran, J. P. McCutcheon, A. Nakabachi, Genomics and evolution of heritable bacterial symbionts. *Annu. Rev. Genet.* **42**, 165–190 (2008).
14. J. P. McCutcheon, B. R. McDonald, N. A. Moran, Origin of an alternative genetic code in the extremely small and GC-rich genome of a bacterial symbiont. *PLoS Genet.* **5** (2009), doi:10.1371/journal.pgen.1000565.
15. C. Kuo, N. a Moran, H. Ochman, The consequences of genetic drift for bacterial genome complexity The consequences of genetic drift for bacterial genome complexity. *Genome Res.*, 1450–1454 (2009).
16. H. Charles, D. Mouchiroud, J. Lobry, I. Goncalves, Y. Rahbe, Gene size reduction in the bacterial aphid endosymbiont, Buchnera. *Mol. Biol. Evol.* **16** (1999), pp.

1820–1822.

17. N. A. Moran, A. Mira, The process of genome shrinkage in the obligate symbiont Buchnera aphidicola. *Genome Biol.* **2**, RESEARCH0054 (2001).

18. J. P. McCutcheon, N. A. Moran, Extreme genome reduction in symbiotic bacteria. *Nat. Rev. Microbiol.* **10**, 13–26 (2011).

19. N. A. Moran, Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *Proc. Natl. Acad. Sci.* **93**, 2873–2878 (1996).

20. D. B. Sloan, N. A. Moran, Genome Reduction and Co-evolution between the Primary and Secondary Bacterial Symbionts of Psyllids Research article. **29**, 3781–3792 (2012).

21. F. Husnik, J. P. McCutcheon, Repeated replacement of an intrabacterial symbiont in the tripartite nested mealybug symbiosis. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E5416–E5424 (2016).

22. J. L. Morrow, A. A. G. Hall, M. Riegler, Symbionts in waiting: the dynamics of incipient endosymbiont complementation and replacement in minimal bacterial communities of psyllids. *Microbiome*. **5**, 58 (2017).

23. F. Husnik, P. J. Keeling, The fate of obligate endosymbionts: reduction, integration, or extinction. *Curr. Opin. Genet. Dev.* **58**–**59** (2019), pp. 1–8.

24. A. W. D. Larkum, P. J. Lockhart, C. J. Howe, Shopping for plastids. *Trends Plant Sci.* **12**, 189–195 (2007).

25. A. K. Snyder, J. W. Deberry, L. Runyen-Janecky, R. V. M. Rio, Nutrient provisioning facilitates homeostasis between tsetse fly (Diptera: Glossinidae) symbionts. *Proc. R. Soc. B Biol. Sci.* **277**, 2389–2397 (2010).

26. A. K. Snyder, R. V. M. Rio, Interwoven biology of the tsetse holobiont. *J. Bacteriol.* **195**, 4322–4330 (2013).

27. S. C. Welburn, E. M. Fevre, P. G. Coleman, M. Odiit, I. Maudlin, Sleeping sickness: a tale of two diseases. *Trends Parasitol.* **17**, 19–24 (2001).

28. P. P. Simarro, J. Jannin, P. Cattand, Eliminating human African trypanosomiasis: where do we stand and what comes next? *PLoS Med.* **5**, e55 (2008).

29. A. Stich, Human African trypanosomiasis. *BMJ*. **325**, 203–206 (2002).

30. W. Ma, D. L. Denlinger, Secretory discharge and microflora of milk gland in tsetse flies. *Nature*. **247**, 301–303 (1974).

31. E. Aksoy, E. L. Telleria, R. Echodu, Y. Wu, L. M. Okedi, B. L. Weiss, S. Aksoy, A. Caccone, Analysis of multiple tsetse fly populations in Uganda reveals limited diversity and species-specific gut microbiota. *Appl. Environ. Microbiol.* **80**, 4301–4312 (2014).

32. B. L. Weiss, J. Wang, S. Aksoy, Tsetse immune system maturation requires the

presence of obligate symbionts in larvae. *PLoS Biol.* **9**, e1000619 (2011).

33. S. Aksoy, Wigglesworthia gen. nov. and Wigglesworthia glossinidia sp. nov., Taxa Consisting of the Mycetocyte-Associated, Primary Endosymbionts of Tsetse Flies. *Int. J. Syst. Bacteriol.* **45**, 848–851 (1995).

34. I. Maudlin, S. C. Welburn, D. Mehlitz, The relationship between rickettsia-like-organisms and trypanosome infections in natural populations of tsetse in Liberia. *Trop. Med. Parasitol.* **41**, 265–267 (1990).

35. C. Dale, I. Maudlin, Sodalis gen. nov. and Sodalis glossinidius sp. nov., a microaerophilic secondary endosymbiont of the tsetse fly Glossina morsitans morsitans. *Int. J. Syst. Bacteriol.* **49 Pt 1**, 267–275 (1999).

36. S. L. O'Neill, R. H. Gooding, S. Aksoy, Phylogenetically distant symbiotic microorganisms reside in Glossina midgut and ovary tissues. *Med. Vet. Entomol.* **7**, 377–383 (1993).

37. J. Wang, B. Weiss, S. Aksoy, Tsetse fly microbiota: form and function . *Front. Cell. Infect. Microbiol.* . **3** (2013), p. 69.

38. S. Balmand, C. Lohs, S. Aksoy, A. Heddi, Tissue distribution and transmission routes for the tsetse fly endosymbionts. *J. Invertebr. Pathol.* **112 Suppl**, S116-22 (2013).

39. G. Nogge, Aposymbiotic tsetse flies, Glossina morsitans morsitans obtained by feeding on rabbits immunized specifically with symbionts. *J. Insect Physiol.* **24**, 299–304 (1978).

40. C. Cenik, E. S. Cenik, G. W. Byeon, F. Grubert, S. I. Candille, D. Spacek, B. Alsallakh, H. Tilgner, C. L. Araya, H. Tang, E. Ricci, M. P. Snyder, Integrative analysis of RNA, translation, and protein levels reveals distinct regulatory variation across humans. *Genome Res.* **25**, 1610–1621 (2015).

41. M. Medina Munoz, A. R. Pollio, H. L. White, R. V. M. Rio, Into the Wild: Parallel Transcriptomics of the Tsetse-Wigglesworthia Mutualism within Kenyan Populations. *Genome Biol. Evol.* **9**, 2276–2291 (2017).

42. V. Michalkova, J. B. Benoit, B. L. Weiss, G. M. Attardo, S. Aksoy, Vitamin B(6) Generated by Obligate Symbionts Is Critical for Maintaining Proline Homeostasis and Fecundity in Tsetse Flies. *Appl. Environ. Microbiol.* **80**, 5844–5853 (2014).

43. L. Akman, A. Yamashita, H. Watanabe, K. Oshima, T. Shiba, M. Hattori, S. Aksoy, Genome sequence of the endocellular obligate symbiont of tsetse flies, Wigglesworthia glossinidia. *Nat. Genet.* **32**, 402–407 (2002).

44. G. M. Attardo, C. Lohs, A. Heddi, U. H. Alam, S. Yildirim, S. Aksoy, Analysis of milk gland structure and function in Glossina morsitans: Milk protein production, symbiont populations and fecundity. *J. Insect Physiol.* **54**, 1236–1242 (2008).

45. G. M. Attardo, J. M. Ribeiro, Y. Wu, M. Berriman, S. Aksoy, Transcriptome analysis of reproductive tissue and intrauterine developmental stages of the

tsetse fly (Glossina morsitans morsitans). *BMC Genomics*. **11**, 160 (2010).

46. J. B. Benoit, G. M. Attardo, V. Michalkova, P. Takac, J. Bohova, S. Aksoy, Sphingomyelinase activity in mother's milk is essential for juvenile development: a case from lactating tsetse flies. *Biol. Reprod.* **87**, 1-10,17 (2012).

47. J. Wang, S. Aksoy, PGRP-LB is a maternally transmitted immune milk protein that influences symbiosis and parasitism in tsetse's offspring. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 10552–10557 (2012).

48. G. Nogge, Sterility in tsetse flies (Glossina morsitans Westwood) caused by loss of symbionts. *Experientia*. **32**, 995–996 (1976).

49. G. Nogge, A. Gerresheim, Experiments on the elimination of symbionts from the tsetse fly, Glossina morsitans morsitans (Diptera: Glossinidae), by antibiotics and lysozyme. *J. Invertebr. Pathol.* **40**, 166–179 (1982).

50. R. Pais, C. Lohs, Y. Wu, J. Wang, S. Aksoy, The Obligate Mutualist Wigglesworthia glossinidia Influences Reproduction, Digestion, and Immunity Processes of Its Host, the Tsetse Fly. *Appl. Environ. Microbiol.* **74**, 5965–5974 (2008).

51. U. Alam, J. Medlock, C. Brelsfoard, R. Pais, C. Lohs, S. Balmand, J. Carnogursky, A. Heddi, P. Takac, A. Galvani, S. Aksoy, Wolbachia symbiont infections induce strong cytoplasmic incompatibility in the tsetse fly Glossina morsitans. *PLoS Pathog.* **7**, e1002415 (2011).

52. F. J. Silva, A. Latorre, A. Moya, Genome size reduction through multiple events of gene disintegration in Buchnera APS. *Trends Genet.* **17**, 615–618 (2001).

53. L. Klasson, S. G. E. Andersson, Evolution of minimal-gene-sets in host-dependent bacteria. *Trends Microbiol.* **12**, 37–43 (2004).

54. X. Chen, S. Li, S. Aksoy, Concordant evolution of a symbiont with its host insect species: molecular phylogeny of genus Glossina and its bacteriome-associated endosymbiont, Wigglesworthia glossinidia. *J. Mol. Evol.* **48**, 49–58 (1999).

55. R. V. M. Rio, R. E. Symula, J. Wang, C. Lohs, Y.-N. Wu, A. K. Snyder, R. D. Bjornson, K. Oshima, B. S. Biehl, N. T. Perna, M. Hattori, S. Aksoy, Insight into the transmission biology and species-specific functional capabilities of tsetse (Diptera: glossinidae) obligate symbiont Wigglesworthia. *MBio*. **3**, 1–13 (2012).

56. C. Toft, M. A. Fares, The evolution of the flagellar assembly pathway in endosymbiotic bacterial genomes. *Mol. Biol. Evol.* **25**, 2069–2076 (2008).

57. K. Maezawa, S. Shigenobu, H. Taniguchi, T. Kubo, S. I. Aizawa, M. Morioka, Hundreds of flagellar basal bodies cover the cell surface of the endosymbiotic bacterium Buchnera aphidicola sp. strain APS. *J. Bacteriol.* **188**, 6539–6543 (2006).

58. N. Kaiwa, T. Hosokawa, Y. Kikuchi, N. Nikoh, X. Y. Meng, N. Kimura, M. Ito, T. Fukatsu, Primary gut symbiont and secondary, Sodalis-allied symbiont of the

Scutellerid stinkbug Cantao ocellatus. *Appl. Environ. Microbiol.* **76**, 3486–3494 (2010).

59. A. Chari, K. F. Oakeson, S. Enomoto, D. G. Jackson, M. A. Fisher, C. Dale, Phenotypic characterization of Sodalis praecaptivus sp. nov., a close non-insect-associated member of the Sodalis-allied lineage of insect endosymbionts. *Int. J. Syst. Evol. Microbiol.* **65**, 1400–1405 (2015).

60. A. L. Clayton, K. F. Oakeson, M. Gutin, A. Pontes, D. M. Dunn, A. C. von Niederhausern, R. B. Weiss, M. Fisher, C. Dale, A novel human-infection-derived bacterium provides insights into the evolutionary origins of mutualistic insect-bacterial symbioses. *PLoS Genet.* **8**, e1002990 (2012).

61. D. Santos-Garcia, F. J. Silva, S. Morin, K. Dettner, S. M. Kuechler, The all-rounder sodalis: A new bacteriome-associated endosymbiont of the lygaeoid bug henestaris halophilus (heteroptera: Henestarinae) and a critical examination of its evolution. *Genome Biol. Evol.* **9**, 2893–2910 (2017).

62. A. Zaidman-Remy, A. Vigneron, B. L. Weiss, A. Heddi, What can a weevil teach a fly, and reciprocally? Interaction of host immune systems with endosymbionts in Glossina and Sitophilus. *BMC Microbiol.* **18**, 150 (2018).

63. K. F. Oakeson, R. Gil, A. L. Clayton, D. M. Dunn, A. C. von Niederhausern, C. Hamil, A. Aoyagi, B. Duval, A. Baca, F. J. Silva, A. Vallier, D. G. Jackson, A. Latorre, R. B. Weiss, A. Heddi, A. Moya, C. Dale, Genome degeneration and adaptation in a nascent stage of symbiosis. *Genome Biol. Evol.* **6**, 76–93 (2014).

64. M. H. Pontes, M. Babst, R. Lochhead, K. Oakeson, K. Smith, C. Dale, Quorum sensing primes the oxidative stress response in the insect endosymbiont, Sodalis glossinidius. *PLoS One*. **3**, e3541 (2008).

65. K. Kutsukake, Y. Ohya, S. Yamaguchi, T. Iino, Operon structure of flagellar genes in Salmonella typhimurium. *Mol. Gen. Genet.* **214**, 11–15 (1988).

66. R. N. Doetsch, R. D. Sjoblad, Flagellar Structure and Function in Eubacteria. *Annu. Rev. Microbiol.* **34**, 69–108 (1980).

67. F. F. V. Chevance, K. T. Hughes, Coordinating assembly of a bacterial macromolecular machine. *Nat. Rev. Microbiol.* **6**, 455–465 (2008).

68. M. Erhardt, K. Namba, K. T. Hughes, Bacterial nanomachines: the flagellum and type III injectisome. *Cold Spring Harb. Perspect. Biol.* **2**, a000299 (2010).

69. Y. Imae, T. Atsumi, Na+-driven bacterial flagellar motors. *J. Bioenerg. Biomembr.* **21**, 705–716 (1989).

70. H. C. Berg, The rotary motor of bacterial flagella. *Annu. Rev. Biochem.* **72**, 19–54 (2003).

71. T. Minamino, K. Imada, The bacterial flagellar motor and its structural diversity. *Trends Microbiol.* **23**, 267–274 (2015).

72. G. J. Schoenhals, R. M. Macnab, Physiological and biochemical analyses of FlgH, a lipoprotein forming the outer membrane L ring of the flagellar basal body of Salmonella typhimurium. *J. Bacteriol.* **178**, 4200–4207 (1996).

73. R. M. Macnab, How Bacteria Assemble Flagella. *Annu. Rev. Microbiol.* **57**, 77–100 (2003).

74. R. M. Macnab, Type III flagellar protein export and flagellar assembly. *Biochim. Biophys. Acta*. **1694**, 207–217 (2004).

75. G. S. Chilcott, K. T. Hughes, Coupling of Flagellar Gene Expression to Flagellar Assembly in Salmonella enterica Serovar Typhimurium and Escherichia coli. *Microbiol. Mol. Biol. Rev.* **64**, 694–708 (2000).

76. R. Liu, H. Ochman, Origins of Flagellar Gene Operons and Secondary Flagellar Systems . *J. Bacteriol.* **189**, 7098–7104 (2007).

77. R. D. Allen, P. Baumann, Structure and arrangement of flagella in species of the genus Beneckea and Photobacterium fischeri. *J. Bacteriol.* **107**, 295–302 (1971).

78. E. C. McCoy, D. Doyle, H. Wiltberger, K. Burda, A. J. Winter, Flagellar ultrastructure and flagella-associated antigens of Campylobacter fetus. *J. Bacteriol.* **122**, 307–315 (1975).

79. K. W. Hranitzky, A. Mulholland, A. D. Larson, E. R. Eubanks, L. T. Hart, *Infect. Immun.*, in press (available at http://iai.asm.org/content/27/2/597.abstract).

80. S. Trachtenberg, D. J. DeRosier, Conformational switching in the flagellar filament of Salmonella typhimurium. *J. Mol. Biol.* **226**, 447–454 (1992).

81. L. L. McCarter, Dual flagellar systems enable motility under different circumstances. *J. Mol. Microbiol. Biotechnol.* **7**, 18–29 (2004).

82. D. B. Kearns, A field guide to bacterial swarming motility. *Nat. Rev. Microbiol.* **8**, 634–644 (2010).

83. P. G. Hall, N. R. Krieg, Swarming of Azospirillum brasilense on solid media. *Can. J. Microbiol.* **29**, 1592–1594 (1983).

84. S. Kalir, J. McClure, K. Pabbaraju, C. Southward, M. Ronen, S. Leibler, M. G. Surette, U. Alon, Ordering genes in a flagella pathway by analysis of expression kinetics from living bacteria. *Science*. **292**, 2080–2083 (2001).

85. L. D. B. Evans, C. Hughes, G. M. Fraser, Building a flagellum outside the bacterial cell. *Trends Microbiol.* **22**, 566–572 (2014).

86. A. Dufour, R. B. Furness, C. Hughes, Novel genes that upregulate the Proteus mirabilis flhDC master operon controlling flagellar biogenesis and swarming. *Mol. Microbiol.* **29**, 741–751 (1998).

87. O. Soutourina, A. Kolb, E. Krin, C. Laurent-Winter, S. Rimsky, A. Danchin, P. Bertin, Multiple control of flagellum biosynthesis in Escherichia coli: role of H-NS protein and the cyclic AMP-catabolite activator protein complex in transcription of

the flhDC master operon. *J. Bacteriol.* **181**, 7500–7508 (1999).

88. A. Francez-Charlot, B. Laugel, A. Van Gemert, N. Dubarry, F. Wiorowski, M.-P. Castanie-Cornet, C. Gutierrez, K. Cam, RcsCDB His-Asp phosphorelay system negatively regulates the flhDC operon in Escherichia coli. *Mol. Microbiol.* **49**, 823–832 (2003).

89. B. M. Prüß, X. Liu, W. Hendrickson, P. Matsumura, FlhD/FlhC-regulated promoters analyzed by gene array and lacZ gene fusions. *FEMS Microbiol. Lett.* **197**, 91–97 (2006).

90. K. Kutsukake, Y. Ohya, T. Iino, Transcriptional analysis of the flagellar regulon of Salmonella typhimurium. *J. Bacteriol.* **172**, 741–747 (1990).

91. R. B. Furness, G. M. Fraser, N. A. Hay, C. Hughes, Negative feedback from a Proteus class II flagellum export defect to the flhDC master operon controlling cell division and flagellum assembly. *J. Bacteriol.* **179**, 5585–5588 (1997).

92. G. M. Young, M. J. Smith, S. A. Minnich, V. L. Miller, *J. Bacteriol.*, in press (available at http://jb.asm.org/content/181/9/2823.abstract).

93. L. Claret, C. Hughes, Interaction of the atypical prokaryotic transcription activator FlhD2C2 with early promoters of the flagellar gene hierarchy. *J. Mol. Biol.* **321**, 185–199 (2002).

94. N. R. Francis, G. E. Sosinsky, D. Thomas, D. J. DeRosier, Isolation, characterization and structure of bacterial flagellar motors containing the switch complex. *J. Mol. Biol.* **235**, 1261–1270 (1994).

95. T. Kubori, S. Yamaguchi, S. Aizawa, Assembly of the switch complex onto the MS ring complex of Salmonella typhimurium does not require any other flagellar proteins. *J. Bacteriol.* **179**, 813–817 (1997).

96. X. Zhao, K. Zhang, T. Boquoi, B. Hu, M. A. Motaleb, K. A. Miller, M. E. James, Cryoelectron tomography reveals the sequential assembly of bacterial fl agella in Borrelia burgdorferi, 1–6 (2013).

97. A. M. Burrage, E. Vanderpool, D. B. Kearns, crossm Assembly Order of Flagellar Rod Subunits in Bacillus subtilis, 1–13 (2018).

98. A. G. Garza, L. W. Harris-Haller, R. A. Stoebner, M. D. Manson, Motility protein interactions in the bacterial flagellar motor. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 1970–1974 (1995).

99. S. A. Lloyd, F. G. Whitby, D. F. Blair, C. P. Hill, Structure of the C-terminal domain of FliG, a component of the rotor in the bacterial flagellar motor. *Nature*. **400**, 472–475 (1999).

100. D. Thomas, D. G. Morgan, D. J. DeRosier, *J. Bacteriol.*, in press (available at http://jb.asm.org/content/183/21/6404.abstract).

101. D. F. Blair, Flagellar movement driven by proton translocation. *FEBS Lett.* **545**,

86–95 (2003).

102. Y. V Morimoto, T. Minamino, Structure and function of the bi-directional bacterial flagellar motor. *Biomolecules*. **4**, 217–234 (2014).

103. J. E. Karlinsey, H. C. Tsui, M. E. Winkler, K. T. Hughes, Flk couples flgM translation to flagellar ring assembly in Salmonella typhimurium. *J. Bacteriol.* **180**, 5384–5397 (1998).

104. D. R. Mitchell, The evolution of eukaryotic cilia and flagella as motile and sensory organelles. *Adv. Exp. Med. Biol.* **607**, 130–140 (2007).

105. C. Dale, G. R. Plague, B. Wang, H. Ochman, N. A. Moran, Type III secretion systems and the evolution of mutualistic endosymbiosis. **99** (2002).

106. P. Baumann, Biology of Bacteriocyte-Associated Endosymbionts of Plant Sap-Sucking Insects. *Annu. Rev. Microbiol.* **59**, 155–189 (2005).

107. A. Nakabachi, R. Ueoka, K. Oshima, R. Teta, A. Mangoni, M. Gurgui, N. J. Oldham, G. Van Echten-Deckert, K. Okamura, K. Yamamoto, H. Inoue, M. Ohkuma, Y. Hongoh, S. Y. Miyagishima, M. Hattori, J. Piel, T. Fukatsu, Defensive bacteriome symbiont with a drastically reduced genome. *Curr. Biol.* **23**, 1478–1484 (2013).

108. M. E. Brentassi, E. Franco, P. Balatti, R. Medina, F. Bernabei, A. M. Marino, D. R. Lenicov, Bacteriomes of the corn leafhopper , Dalbulus maidis (DeLong Deltocephalinae) harbor Sulcia symbiont : molecular characterization , ultrastructure , and transovarial transmission. *Protoplasma* (2016), doi:10.1007/s00709-016-1033-4.

109. X. Bing, G. M. Attardo, A. Vigneron, E. Aksoy, F. Scolari, A. Malacrida, B. L. Weiss, S. Aksoy, Unravelling the relationship between the tsetse fly and its obligate symbiont Wigglesworthia : transcriptomic and metabolomic landscapes reveal highly integrated physiological networks. *Proc. R. Soc. B Biol. Sci.* **284**, 20170360 (2017).

110. B. M. Boyd, J. M. Allen, R. Koga, T. Fukatsu, A. D. Sweet, K. P. Johnson, D. L. Reed, Louse Proechinophthirus fluctus (Phthiraptera : Anoplura). *Appl. Environ. Microbiol.* **82**, 3185–3197 (2016).

111. S. S. Prado, R. P. P. Almeida, Phylogenetic placement of pentatomid stink bug gut symbionts. *Curr. Microbiol.* **58**, 64–69 (2009).

112. F. Husnik, N. Nikoh, R. Koga, L. Ross, R. P. Duncan, M. Fujie, M. Tanaka, N. Satoh, D. Bachtrog, A. C. C. Wilson, C. D. Von Dohlen, T. Fukatsu, J. P. McCutcheon, Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. *Cell*. **153**, 1567–1578 (2013).

113. R. V. M. Rio, Y.-N. Wu, G. Filardo, S. Aksoy, Dynamics of multiple symbiont density regulation during host development: tsetse fly and its microbial flora. *Proceedings. Biol. Sci.* **273**, 805–814 (2006).

114. G. M. Bennett, N. A. Moran, Small, smaller, smallest: The origins and evolution of ancient dual symbioses in a phloem-feeding insect. *Genome Biol. Evol.* **5**, 1675–1688 (2013).

115. J. Felsenstein, S. Yokoyama, The evolutionary advantage of recombination. II. Individual selection for recombination. *Genetics*. **83**, 845–859 (1976).

116. C. Rispe, N. A. Moran, Accumulation of deleterious mutations in endosymbionts: Muller's ratchet with two levels of selection. *Am. Nat.* **156**, 425–441 (2000).

117. M. Woolfit, L. Bromham, Increased rates of sequence evolution in endosymbiotic bacteria and fungi with small effective population sizes. *Mol. Biol. Evol.* **20**, 1545–1555 (2003).

118. J. J. Wernegreen, Genome evolution in bacterial endosymbionts of insects. *Nat. Rev. Genet.* **3**, 850–861 (2002).

119. U. Kutschera, K. J. Niklas, Endosymbiosis, cell evolution, and speciation. *Theory Biosci.* **124**, 1–24 (2005).

120. R. A. Chong, N. A. Moran, Evolutionary loss and replacement of Buchnera, the obligate endosymbiont of aphids. *ISME J.* (2018), doi:10.1038/s41396-017-0024-6.

121. C. D. von Dohlen, U. Spaulding, K. B. Patch, K. M. Weglarz, R. G. Foottit, N. P. Havill, G. R. Burke, Dynamic acquisition and loss of dual-obligate symbionts in the plant-sap-feeding adelgidae (Hemiptera: Sternorrhyncha: Aphidoidea). *Front. Microbiol.* **8**, 1–15 (2017).

122. F. Husnik, J. P. McCutcheon, Repeated replacement of an intrabacterial symbiont in the tripartite nested mealybug symbiosis. *Proc. Natl. Acad. Sci.* **113**, E5416–E5424 (2016).

123. M. L. Thao, M. A. Clark, L. Baumann, E. B. Brennan, N. A. Moran, P. Baumann, Secondary Endosymbionts of Psyllids Have Been Acquired Multiple Times. **41**, 300–304 (2000).

124. R. Koga, N. A. Moran, Swapping symbionts in spittlebugs: Evolutionary replacement of a reduced genome symbiont. *ISME J.* **8**, 1237–1246 (2014).

125. M. A. Clark, L. Baumann, M. L. Thao, N. A. Moran, P. Baumann, Degenerative minimalism in the genome of a psyllid endosymbiont. *J. Bacteriol.* **183**, 1853–1861 (2001).

126. H. Terashima, A. Kawamoto, Y. V Morimoto, K. Imada, T. Minamino, Structural differences in the bacterial flagellar motor among bacterial species. **14**, 191–198 (2017).

127. F. F. V Chevance, K. T. Hughes, Coordinating assembly of a bacterial macromolecular machine. *Nat. Rev. Microbiol.* **6**, 455–465 (2008).

128. I. A. Osterman, Y. Y. Dikhtyar, A. A. Bogdanov, O. A. Dontsova, P. V Sergiev,

Regulation of flagellar gene expression in Bacteria. *Biochem.* **80** (2015), pp. 1447–1456.

129. E. Nováková, F. Husník, E. Šochová, V. Hypša, Arsenophonus and sodalis symbionts in louse flies: An analogy to the wigglesworthia and sodalis system in tsetse flies. *Appl. Environ. Microbiol.* **81**, 6189–6199 (2015).

130. R. A. Chong, H. Park, N. A. Moran, Genome Evolution of the Obligate Endosymbiont Buchnera aphidicola. *Mol. Biol. Evol.* **36**, 1481–1489 (2019).

131. L. Li, C. J. J. Stoeckert, D. S. Roos, OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes -- Li et al. 13 (9): 2178 -- Genome Research. *Genome Res.* **13**, 2178–2189 (2003).

132. T. P. Driscoll, V. I. Verhoeve, M. L. Guillotte, S. S. Lehman, S. A. Rennoll, M. Beier-Sexton, M. S. Rahman, A. F. Azad, J. J. Gillespie, Wholly rickettsia! reconstructed metabolic profile of the quintessential bacterial parasite of eukaryotic cells. *MBio*. **8** (2017), doi:10.1128/mBio.00859-17.

133. A. E. Darling, B. Mau, N. T. Perna, Progressivemauve: Multiple genome alignment with gene gain, loss and rearrangement. *PLoS One*. **5** (2010), doi:10.1371/journal.pone.0011147.

134. R. K. Aziz, D. Bartels, A. Best, M. DeJongh, T. Disz, R. A. Edwards, K. Formsma, S. Gerdes, E. M. Glass, M. Kubal, F. Meyer, G. J. Olsen, R. Olson, A. L. Osterman, R. A. Overbeek, L. K. McNeil, D. Paarmann, T. Paczian, B. Parrello, G. D. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke, O. Zagnitko, The RAST Server: Rapid annotations using subsystems technology. *BMC Genomics*. **9**, 1–15 (2008).

135. R. C. Edgar, MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

136. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

137. S. Capella-Gutiérrez, J. M. Silla-Martínez, T. Gabaldón, trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. **25**, 1972–1973 (2009).

138. A. Stamatakis, RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. **22**, 2688–2690 (2006).

139. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. **30**, 1312–1313 (2014).

140. J. P. Hueslenbeck, B. Larget, R. E. Miller, F. Ronquist, Potential Applications and Pitfalls of Bayesian Inference of Phylogeny. **51**, 673–688 (2002).

141. F. Ronquist, M. Teslenko, P. Van Der Mark, D. L. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M. A. Suchard, J. P. Huelsenbeck, Mrbayes 3.2: Efficient bayesian

phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).

142. D. Glez-Peña, D. Gómez-Blanco, M. Reboiro-Jato, F. Fdez-Riverola, D. Posada, ALTER: Program-oriented conversion of DNA and protein alignments. *Nucleic Acids Res.* **38**, 14–18 (2010).

143. B. G. Hall, Building phylogenetic trees from molecular data with MEGA. *Mol. Biol. Evol.* **30**, 1229–1235 (2013).

144. S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).

145. A. Rambaut, A. J. Drummond, D. Xie, G. Baele, M. A. Suchard, Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* **67**, 901–904 (2018).

146. J. Huerta-Cepas, D. Szklarczyk, D. Heller, A. Hernández-Plaza, S. K. Forslund, H. Cook, D. R. Mende, I. Letunic, T. Rattei, L. J. Jensen, C. Von Mering, P. Bork, EggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **47**, D309–D314 (2019).

147. E. Paradis, J. Claude, K. Strimmer, APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*. **20**, 289–290 (2004).

148. L. J. Revell, phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012).

149. M. W. Pennell, J. M. Eastman, G. J. Slater, J. W. Brown, J. C. Uyeda, R. G. Fitzjohn, M. E. Alfaro, L. J. Harmon, Geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. **30**, 2216–2218 (2014).

150. D. T. Gillespie, Exact numerical simulation of the Ornstein-Uhlenbeck process and its integral. *Phys. Rev. E*. **54**, 2084–2091 (1996).

151. J. S. Farris, Phylogenetic Analysis Under Dollo's Law (1977).

152. A. G. Kluge, J. S. Farris, Quantitative Phyletics and the Evolution of Anurans (1967).

153. K. L. Gillen, K. T. Hughes, *J. Bacteriol.*, in press (available at http://jb.asm.org/content/173/7/2301.abstract).

154. F. Suaste-Olmos, C. Domenzain, J. C. Mireles-Rodríguez, S. Poggio, A. Osorio, G. Dreyfus, L. Camarena, The flagellar protein FliL is essential for swimming in Rhodobacter sphaeroides. *J. Bacteriol.* **192**, 6230–6239 (2010).

155. N. Takekawa, M. Isumi, H. Terashima, S. Zhu, Y. Nishino, M. Sakuma, S. Kojima, M. Homma, K. Imada, Structure of vibrio flil, a new stomatin-like protein that assists the bacterial flagellar motor function. *MBio*. **10**, 1–13 (2019).

156. T. Minamino, Y. V. Morimoto, N. Hara, K. Namba, An energy transduction mechanism used in bacterial flagellar type III protein export. *Nat. Commun.* **2**, 475–479 (2011).

157. J. Kucera, E. M. Terentjev, FliI6-FliJ molecular motor assists with unfolding in the type III secretion export apparatus. *Sci. Rep.* **10**, 1–11 (2020).

158. F. A. Herlihey, P. J. Moynihan, A. J. Clarke, The essential protein for bacterial flagella formation FlgJ functions as a ??-N-acetylglucosaminidase. *J. Biol. Chem.* **289**, 31029–31042 (2014).

159. M. Erhardt, T. Hirano, Y. Su, K. Paul, D. H. Wee, S. Mizuno, S. I. Aizawa, K. T. Hughes, The role of the FliK molecular ruler in hook-length control in Salmonella enterica. *Mol. Microbiol.* **75**, 1272–1284 (2010).

160. P. Aldridge, J. Karlinsey, K. T. Hughes, The type III secretion chaperone FlgN regulates flagellar assembly via a negative feedback loop containing its chaperone substrates FlgK and FlgL. *Mol. Microbiol.* **49**, 1333–1345 (2003).

161. S. Maiti, P. Mitra, Bacterial flagellar switching: a molecular mechanism directed by the logic of an electric motor. *J. Mol. Model.* **24**, 280 (2018).

162. P. H. Degnan, Y. Yu, N. Sisneros, R. A. Wing, N. A. Moran, Hamiltonella defensa, genome evolution of protective bacterial endosymbiont from pathogenic ancestors. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 9063–9068 (2009).

163. S. Bhattacharya, A. K. Baidya, R. R. Pal, G. Mamou, Y. E. Gatt, H. Margalit, I. Rosenshine, S. Ben-Yehuda, A Ubiquitous Platform for Bacterial Nanotube Biogenesis. *Cell Rep.* **27**, 334-342.e10 (2019).

164. K. P. Williams, J. J. Gillespie, B. W. S. Sobral, E. K. Nordberg, E. E. Snyder, J. M. Shallom, A. W. Dickerman, Phylogeny of Gammaproteobacteria. *J. Bacteriol.* **192**, 2305–2314 (2010).

165. D. Darriba, D. Posada, A. M. Kozlov, A. Stamatakis, B. Morel, T. Flouri, ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. *Mol. Biol. Evol.* **37**, 291–294 (2020).

166. R. Lanfear, P. B. Frandsen, A. M. Wright, T. Senfeld, B. Calcott, PartitionFinder 2 : New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. **34**, 772–773 (2016).

167. R. A. Chong, H. Park, N. A. Moran, Genome Evolution of the Obligate Endosymbiont Buchnera aphidicola. *Mol. Biol. Evol.* **36**, 1481–1489 (2019).

168. R. Gil, S. Ramos, A. Lamelas, M. Postigo, J. M. Michelena, F. J. Silva, A Small Microbial Genome : The End of a Long Symbiotic Relationship. *Science (80-.).* **314**, 312–313 (2006).

169. A. Manzano-Marín, A. Latorre, Settling down: The genome of serratia symbiotica from the aphid cinara tujafilina zooms in on the process of accommodation to a cooperative intracellular life. *Genome Biol. Evol.* **6**, 1683–1698 (2014).

170. A. Lamelas, M. J. Gosalbes, A. Manzano-Marín, J. Peretó, A. Moya, A. Latorre, Serratia symbiotica from the aphid Cinara cedri: A missing link from facultative to obligate insect endosymbiont. *PLoS Genet.* **7** (2011), doi:10.1371/journal.pgen.1002357.

171. D. A. C. Bublitz, G. L. Chadwick, J. S. Magyar, K. M. Sandoz, D. M. Brooks, S. Mesnage, M. S. Ladinsky, A. I. Garber, P. J. Bjorkman, V. J. Orphan, J. P. McCutcheon, Peptidoglycan Production by an Insect-Bacterial Mosaic. *Cell*. **179**, 703-712.e7 (2019).

172. A. J. Drummond, S. Y. W. Ho, M. J. Phillips, A. Rambaut, Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**, 699–710 (2006).

Supplementary Material

Figure S1: Amino Alignment colors indicate polarity (top) and Maximum Likelihood tree (a) FliP, (b) FlgK, (c) FliM, (d) FlgL, (e) FlgM, (f) FlgN OGs. The boxes enclose one of the designated groupings. The combination of alignment and tree made us consider these as two different OGs.

Figure S2: A phylogenetic tree for all organisms included in this study for the core gene (n=35). (a) MrBayes and (b) RAxML tree topology with branch lengths are shown. Both trees demonstrate short branch lengths within the *Sodalis*-clade. We attempted to increase the branch lengths and resolve any polytomies by doing additional analyses on the *Sodalis*-clade organisms. Both trees were built using the auto heuristic model. Branch lengths are shown as (a) number of changes per site, (b) substitutions per site.

Figure S3: The distribution of core genes (n=136) retained by all *Sodalis*-allied organisms. These genes are organized by Clusters of Orthologous Groups (COG) as determined through eggNOG. A total of 35 genes were found to be conserved across all the organisms in our data set (refer to Table 1). The majority of orthologs are grouped within Translation, ribosomal structure, and biogenesis.

Figure S4: *Sodalis*-clade partitioned core gene tree. (a) MrBayes and (b) RAxML tree topology with branch lengths. Both trees show identical topologies with similar branch lengths. The RAxML analysis was able to resolve the polytomy within *S. melophagi* and *S. praecaptivus.* Both trees were constructed using the partitioned model obtained through partitionfinder2 (*166*)? using the *Sodalis* core gene set. Both trees were built using the auto heuristic model. Branch lengths are shown as (a) number of changes per site, (b) substitutions per site.

Figure S5: Ultrametric tree (a) all organisms of interest, (b) *Sodalis*-allied organisms. Estimates for branch lengths were created using “Chronos” penalized likelihood estimation.

Table S2: *Sodalis*-clade core gene set. Single copy genes identified within the *Sodalis* subset of organisms. All genes are organized by COG category as identified by EGGNOG. A total of 136 single copy genes were identified as retained within the genomes of all Sodalis-allied organisms

Gene Name, Annotation	Mean Length (aa)	Mean Length (aa)	COG category
polA, DNA polymerase I	968	792.83	Replication, recombination and repair
dnaN, DNA polymerase III subunit beta	376	367.56	Replication, recombination and repair
holA, DNA polymerase III subunit delta	361	342.5	Replication, recombination and repair
dnaq, DNA polymerase III subunit epsilon	316	245.41	Replication, recombination and repair
dnax, DNA polymerase III subunit gamma/tau	820	558.11	Replication, recombination and repair
gyrb, DNA topoisomerase (ATP-hydrolyzing) subunit B	817	797.76	Replication, recombination and repair
ycfh, metal-dependent hydrolase	279	264.3	Replication, recombination and repair
yqgf, Holliday Translation, ribosomal structure and biogenesisunction resolvase RuvX	159	137.18	Replication, recombination and repair
liga, NAD-dependent DNA ligase LigA	718	654.17	Replication, recombination and repair
dnab, replicative DNA helicase DnaB	500	458.37	Replication, recombination and repair
iscs, IscS subfamily cysteine desulfurase	404	389.62	Amino acid transport and metabolism
arok, shikimate kinase	195	173.86	Amino acid transport and metabolism
tils, tRNA lysidine(34) synthetase TilS	577	476.76	Cell cycle control, cell division, chromosome partitioning
mnmg, tRNA uridine-5-carboxymethylaminomethyl(34) synthesis enzyme MnmG	653	612	Cell cycle control, cell division, chromosome partitioning
rsmh, 16S rRNA (cytosine(1402)-N(4))-methyltransferase	336	314.77	Cell wall/membrane/envelope biogenesis
lepa, elongation factor 4	610	586.92	Cell wall/membrane/envelope biogenesis

era, GTPase Era	324	292.91	Cell wall/membrane/envelope biogenesis
ribh, 6,7-dimethyl-8-ribityllumazine synthase	162	156.22	Coenzyme transport and metabolism
lipa, lipoyl synthase	329	316.4	Coenzyme transport and metabolism
lipb, lipoyl(octanoyl) transferase LipB	246	219.44	Coenzyme transport and metabolism
lpda, dihydrolipoyl dehydrogenase	481	475.12	Energy production and conversion
flda, flavodoxin FldA	184	173.45	Energy production and conversion
acee, pyruvate dehydrogenase (acetyl-transferring), homodimeric type	895	876.05	Energy production and conversion
acef, pyruvate dehydrogenase complex dihydrolipoyllysine-residue acetyltransferase	638	491.63	Energy production and conversion
sufa, Fe-S cluster assembly scaffold SufA	135	125.59	Function unknown
mnme, tRNA uridine-5-carboxymethylaminomethyl(34) synthesis GTPase MnmE	465	452.37	Function unknown
engb, YihA family ribosome biogenesis GTP-binding protein	236	211.27	Function unknown
tusb, sulfurtransferase complex subunit TusB	101	93.33	Inorganic ion transport and metabolisms
tusd, sulfurtransferase complex subunit TusD	137	127.57	Inorganic ion transport and metabolisms
tuse, sulfurtransferase TusE	114	107.75	Inorganic ion transport and metabolisms
sece, preprotein translocase subunit SecE	154	127.36	Intracellular trafficking, secretion, and vesicular transport
seca, preprotein translocase subunit SecA	1030	890.04	Intracellular trafficking, secretion, and vesicular transport

smpb, SsrA-binding protein SmpB	162	153.57	Posttranslational modification, protein turnover, chaperones
yeaz, tRNA (adenosine(37)-N6)-threonylcarbamoyltransferase complex dimerization subunit type 1 TsaB	255	227.13	Posttranslational modification, protein turnover, chaperones
dksa, RNA polymerase-binding transcription factor DksA	158	150.71	Signal transduction mechanisms
yrba, BolA family transcriptional regulator	107	84.04	Transcription
rpoa, DNA-directed RNA polymerase subunit alpha	337	313.3	Transcription
rpob, DNA-directed RNA polymerase subunit beta	1382	1,327.20	Transcription
rpoc, DNA-directed RNA polymerase subunit beta'	1436	1,363.77	Transcription
rho, transcription termination factor Rho	421	416.5	Transcription
nusg, transcription termination/antitermination protein NusG	184	179.92	Transcription
rnc, ribonuclease III	236	226.54	Transcription
rpod, RNA polymerase sigma factor RpoD	667	562.57	Transcription
nusb, transcription antitermination factor NusB	158	138.65	Transcription
grea, transcription elongation factor GreA	160	158.12	Transcription
nusa, transcription termination/antitermination protein NusA	537	497.78	Transcription
rsma, 16S rRNA (adenine(1518)-N(6)/adenine(1519)-N(6))-dimethyltransferase RsmA	310	270.52	Translation, ribosomal structure and biogenesis
rlme, 23S rRNA (uridine(2552)-2'-O)-methyltransferase RlmE	238	207.15	Translation, ribosomal structure and biogenesis

rluc, 23S rRNA pseudouridine(955/2504/2580) synthase RluC	329	316.83	Translation, ribosomal structure and biogenesis
rpsm, 30S ribosomal protein S13	135	115.8	Translation, ribosomal structure and biogenesis
rpsb, 30S ribosomal protein S2	258	235.2	Translation, ribosomal structure and biogenesis
rpsf, 30S ribosomal protein S6	143	127.44	Translation, ribosomal structure and biogenesis
rpsh, 30S ribosomal protein S8	135	130.5	Translation, ribosomal structure and biogenesis
rbfa, 30S ribosome-binding factor RbfA	138	131.91	Translation, ribosomal structure and biogenesis
rpla, 50S ribosomal protein L1	234	232.48	Translation, ribosomal structure and biogenesis
rplj, 50S ribosomal protein L10	173	165.92	Translation, ribosomal structure and biogenesis
rplk, 50S ribosomal protein L11	149	141.77	Translation, ribosomal structure and biogenesis
rplo, 50S ribosomal protein L15	155	137.9	Translation, ribosomal structure and biogenesis
rplq, 50S ribosomal protein L17	142	124.29	Translation, ribosomal structure and biogenesis
rplt, 50S ribosomal protein L20	133	117.61	Translation, ribosomal structure and biogenesis
rplx, 50S ribosomal protein L24	109	103.28	Translation, ribosomal structure and biogenesis
rpma, 50S ribosomal protein L27	101	85.93	Translation, ribosomal structure and biogenesis
rplc, 50S ribosomal protein L3	246	204.97	Translation, ribosomal structure and biogenesis
rpld, 50S ribosomal protein L4	208	199.87	Translation, ribosomal structure and biogenesis
rplf, 50S ribosomal protein L6	190	176.97	Translation, ribosomal structure and biogenesis
rpll, 50S ribosomal protein L7/L12	125	121.57	Translation, ribosomal structure and biogenesis
rpli, 50S ribosomal protein L9	190	151.71	Translation, ribosomal structure and biogenesis
pth, aminoacyl-tRNA hydrolase	207	194.04	Translation, ribosomal structure and biogenesis

asns, asparagine--tRNA ligase	468	466	Translation, ribosomal structure and biogenesis
asps, aspartate--tRNA ligase	618	573.48	Translation, ribosomal structure and biogenesis
tsf, elongation factor Ts	307	278.72	Translation, ribosomal structure and biogenesis
gltx, glutamate--tRNA ligase	500	460.85	Translation, ribosomal structure and biogenesis
glns, glutamine--tRNA ligase	571	537.7	Translation, ribosomal structure and biogenesis
glyq, glycine--tRNA ligase subunit alpha	337	302.27	Translation, ribosomal structure and biogenesis
glys, glycine--tRNA ligase subunit beta	740	679.21	Translation, ribosomal structure and biogenesis
hiss, histidine--tRNA ligase	491	420.62	Translation, ribosomal structure and biogenesis
lyss, lysine--tRNA ligase	526	499.04	Translation, ribosomal structure and biogenesis
fmt, methionyl-tRNA formyltransferase	323	311.83	Translation, ribosomal structure and biogenesis
cca, multifunctional CCA addition/repair protein	442	415.91	Translation, ribosomal structure and biogenesis
rpsa, 30S ribosomal protein S1	576	539.6	Translation, ribosomal structure and biogenesis
rpsj, 30S ribosomal protein S10	107	102.27	Translation, ribosomal structure and biogenesis
rpsn, 30S ribosomal protein S14	101	99.27	Translation, ribosomal structure and biogenesis
rpsq, 30S ribosomal protein S17	88	82.33	Translation, ribosomal structure and biogenesis
rpsr, 30S ribosomal protein S18	425	90.93	Translation, ribosomal structure and biogenesis
rpss, 30S ribosomal protein S19	96	91.17	Translation, ribosomal structure and biogenesis
rpst, 30S ribosomal protein S20	109	88.09	Translation, ribosomal structure and biogenesis
rpsc, 30S ribosomal protein S3	263	231.97	Translation, ribosomal structure and biogenesis
rpsd, 30S ribosomal protein S4	208	204.8	Translation, ribosomal structure and biogenesis
rpse, 30S ribosomal protein S5	186	164.43	Translation, ribosomal structure and biogenesis

rpsg, 30S ribosomal protein S7	158	156.07	Translation, ribosomal structure and biogenesis
rpsi, 30S ribosomal protein S9	165	130.3	Translation, ribosomal structure and biogenesis
rplm, 50S ribosomal protein L13	157	141.46	Translation, ribosomal structure and biogenesis
rpln, 50S ribosomal protein L14	124	122.63	Translation, ribosomal structure and biogenesis
rplp, 50S ribosomal protein L16	139	136.07	Translation, ribosomal structure and biogenesis
rplr, 50S ribosomal protein L18	122	115.32	Translation, ribosomal structure and biogenesis
rplv, 50S ribosomal protein L22	130	113.52	Translation, ribosomal structure and biogenesis
rplw, 50S ribosomal protein L23	112	101	Translation, ribosomal structure and biogenesis
rply, 50S ribosomal protein L25	204	103.21	Translation, ribosomal structure and biogenesis
rpmb, 50S ribosomal protein L28	97	78.52	Translation, ribosomal structure and biogenesis
rpmc, 50S ribosomal protein L29	77	64.65	Translation, ribosomal structure and biogenesis
rpmd, 50S ribosomal protein L30	68	59.83	Translation, ribosomal structure and biogenesis
rpmf, 50S ribosomal protein L32	116	59.63	Translation, ribosomal structure and biogenesis
rpmg, 50S ribosomal protein L33	60	55.07	Translation, ribosomal structure and biogenesis
rpmi, 50S ribosomal protein L35	67	64.31	Translation, ribosomal structure and biogenesis
rnpa, ribonuclease P protein component	187	115.05	Translation, ribosomal structure and biogenesis
infa, translation initiation factor IF-1	107	74.23	Translation, ribosomal structure and biogenesis
prfa, peptide chain release factor 1	362	341.93	Translation, ribosomal structure and biogenesis
prmc, peptide chain release factor N(5)-glutamine methyltransferase	295	281.96	Translation, ribosomal structure and biogenesis
def, peptide deformylase	178	165	Translation, ribosomal structure and biogenesis

phes, phenylalanine--tRNA ligase subunit alpha	364	321.42	Translation, ribosomal structure and biogenesis
phet, phenylalanine--tRNA ligase subunit beta	808	789.52	Translation, ribosomal structure and biogenesis
pros, proline--tRNA ligase	583	543.2	Translation, ribosomal structure and biogenesis
ychf, redox-regulated ATPase YchF	368	362.52	Translation, ribosomal structure and biogenesis
sers, serine--tRNA ligase	451	422.48	Translation, ribosomal structure and biogenesis
thrs, threonine--tRNA ligase	659	635.08	Translation, ribosomal structure and biogenesis
infb, translation initiation factor IF-2	902	795.17	Translation, ribosomal structure and biogenesis
miaa, tRNA (adenosine(37)-N6)-dimethylallyltransferase MiaA	328	311.65	Translation, ribosomal structure and biogenesis
mnma, tRNA 2-thiouridine(34) synthase MnmA	409	365.48	Translation, ribosomal structure and biogenesis
trps, tryptophan--tRNA ligase	356	333.56	Translation, ribosomal structure and biogenesis
map, type I methionyl aminopeptidase	281	263.5	Translation, ribosomal structure and biogenesis
tyrs, tyrosine--tRNA ligase	429	411.3	Translation, ribosomal structure and biogenesis

secy, preprotein translocase subunit SecY	449	439.6	Intracellular trafficking, secretion, and vesicular transport
lepb, signal peptidase I	403	311.46	Intracellular trafficking, secretion, and vesicular transport
fabg, 3-oxoacyl-ACP reductase FabG	247	242.96	Lipid transport and metabolism
acpp, acyl carrier protein	100	79.58	Lipid transport and metabolism
der, ribosome biogenesis GTPase Der	498	470.63	Nucleotide transport and metabolism
ybey, rRNA maturation RNase YbeY	160	150.76	Nucleotide transport and metabolism
clpx, ATP-dependent protease ATP-binding subunit ClpX	440	412.72	Posttranslational modification, protein turnover, chaperones
ftsh, ATP-dependent zinc metalloprotease FtsH	677	635.91	Posttranslational modification, protein turnover, chaperones
lon, endopeptidase La	816	786.71	Posttranslational modification, protein turnover, chaperones
nfua, Fe-S biogenesis protein NfuA	221	191.56	Posttranslational modification, protein turnover, chaperones
dnak, molecular chaperone DnaK	666	634.67	Posttranslational modification, protein turnover, chaperones
dnaj, molecular chaperone DnaTranslation, ribosomal structure and biogenesis	388	375.07	Posttranslational modification, protein turnover, chaperones
grxd, monothiol glutaredoxin, Grx4 family	117	112.43	Posttranslational modification, protein turnover, chaperones
clpp, ATP-dependent Clp endopeptidase, proteolytic subunit ClpP	264	207.38	Posttranslational modification, protein turnover, chaperones
trxa, thioredoxin TrxA	114	108.52	Posttranslational modification, protein turnover, chaperones

Table S1: Core gene set used for tree construction. A total of 35 single copy genes were identified in our organisms of interest. All genes were binned into Clusters of orthologous groups of proteins (COGs) using eggNOG.

Gene name, annotation	Max. length (aa)	Mean length (aa)	COG
aroK, shikimate kinase	263	231	Translation, ribosomal structure and biogenesis
dnaJ, molecular chaperone	135	116	Translation, ribosomal structure and biogenesis
dnaK, molecular chaperone	158	156	Translation, ribosomal structure and biogenesis
dnaB, replicative DNA helicase	124	123	Translation, ribosomal structure and biogenesis
dnaX, DNA polymerase III subunit gamma/tau	96	91	Translation, ribosomal structure and biogenesis
rpoA, DNA-directed RNA polymerase subunit alpha	246	205	Translation, ribosomal structure and biogenesis
rpoB, DNA-directed RNA polymerase subunit beta	208	205	Translation, ribosomal structure and biogenesis
infA, translation initiation factor IF-1	101	99	Translation, ribosomal structure and biogenesis
infB, translation initiation factor IF-2	208	200	Translation, ribosomal structure and biogenesis
rplC, 50S ribosomal protein L3	186	164	Translation, ribosomal structure and biogenesis
rplD, 50S ribosomal protein L4	165	130	Translation, ribosomal structure and biogenesis
rplF, 50S ribosomal protein L6	337	313	Transcription
rplK, 50S ribosomal protein L11	88	82	Translation, ribosomal structure and biogenesis
rplL, 50S ribosomal protein L7/L12	155	138	Translation, ribosomal structure and biogenesis
rplM, 50S ribosomal protein L13	190	177	Translation, ribosomal structure and biogenesis
rplN, 50S ribosomal protein L14	139	136	Translation, ribosomal structure and biogenesis
rplO, 50S ribosomal protein L15	125	122	Translation, ribosomal structure and biogenesis
rplP, 50S ribosomal protein L16	107	74	Translation, ribosomal structure and biogenesis

rplQ, 50S ribosomal protein L17	101	86	Translation, ribosomal structure and biogenesis
rplR, 50S ribosomal protein L18	258	235	Translation, ribosomal structure and biogenesis
rplT, 50S ribosomal protein L20	576	540	Translation, ribosomal structure and biogenesis
rpmA, 50S ribosomal protein L27	149	142	Translation, ribosomal structure and biogenesis
rpmG, 50S ribosomal protein L33	1382	1327	Transcription
rpsA, 30S ribosomal protein S1	666	635	Posttranslational modification, protein turnover, chaperones
rpsB, 30S ribosomal protein S2	902	795	Translation, ribosomal structure and biogenesis
rpsC, 30S ribosomal protein S3	500	458	Replication, recombination and repair
rpsD, 30S ribosomal protein S4	135	131	Translation, ribosomal structure and biogenesis
rpsE, 30S ribosomal protein S5	142	124	Translation, ribosomal structure and biogenesis
rpsG, 30S ribosomal protein S7	388	375	Posttranslational modification, protein turnover, chaperones
rpsH, 30S ribosomal protein S8	122	115	Translation, ribosomal structure and biogenesis
rpsI, 30S ribosomal protein S9	195	174	Amino acid transport and metabolism
rpsM, 30S ribosomal protein S13	60	55	Translation, ribosomal structure and biogenesis
rpsN, 30S ribosomal protein S14	133	118	Translation, ribosomal structure and biogenesis
rpsQ, 30S ribosomal protein S17	820	558	Replication, recombination and repair
rpsS, 30S ribosomal protein S19	157	141	Translation, ribosomal structure and biogenesis

FliP

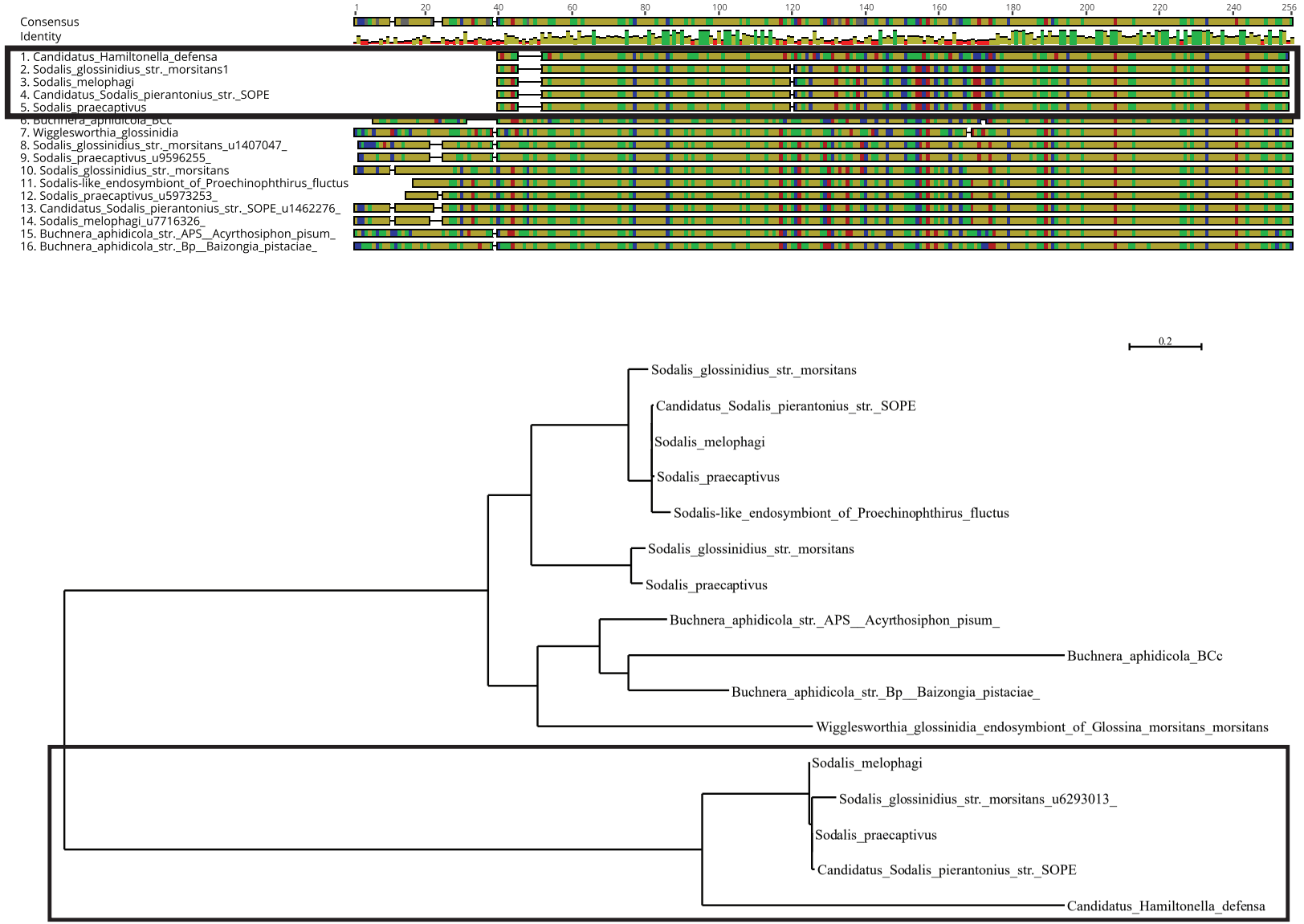
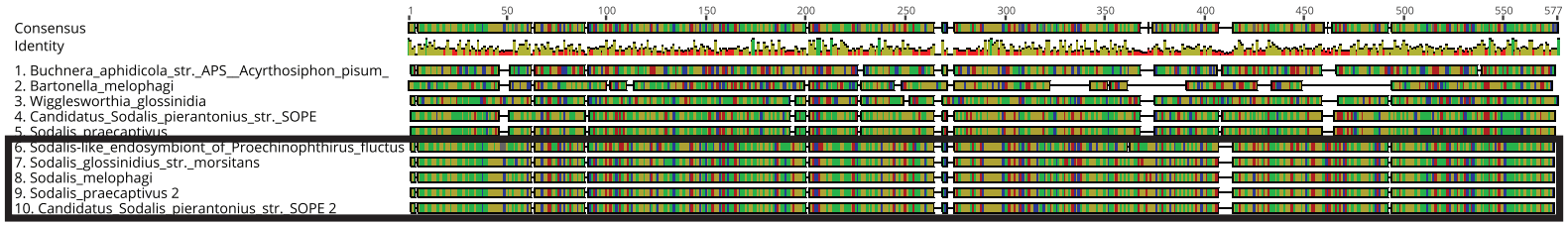


Figure S1a

FlgK



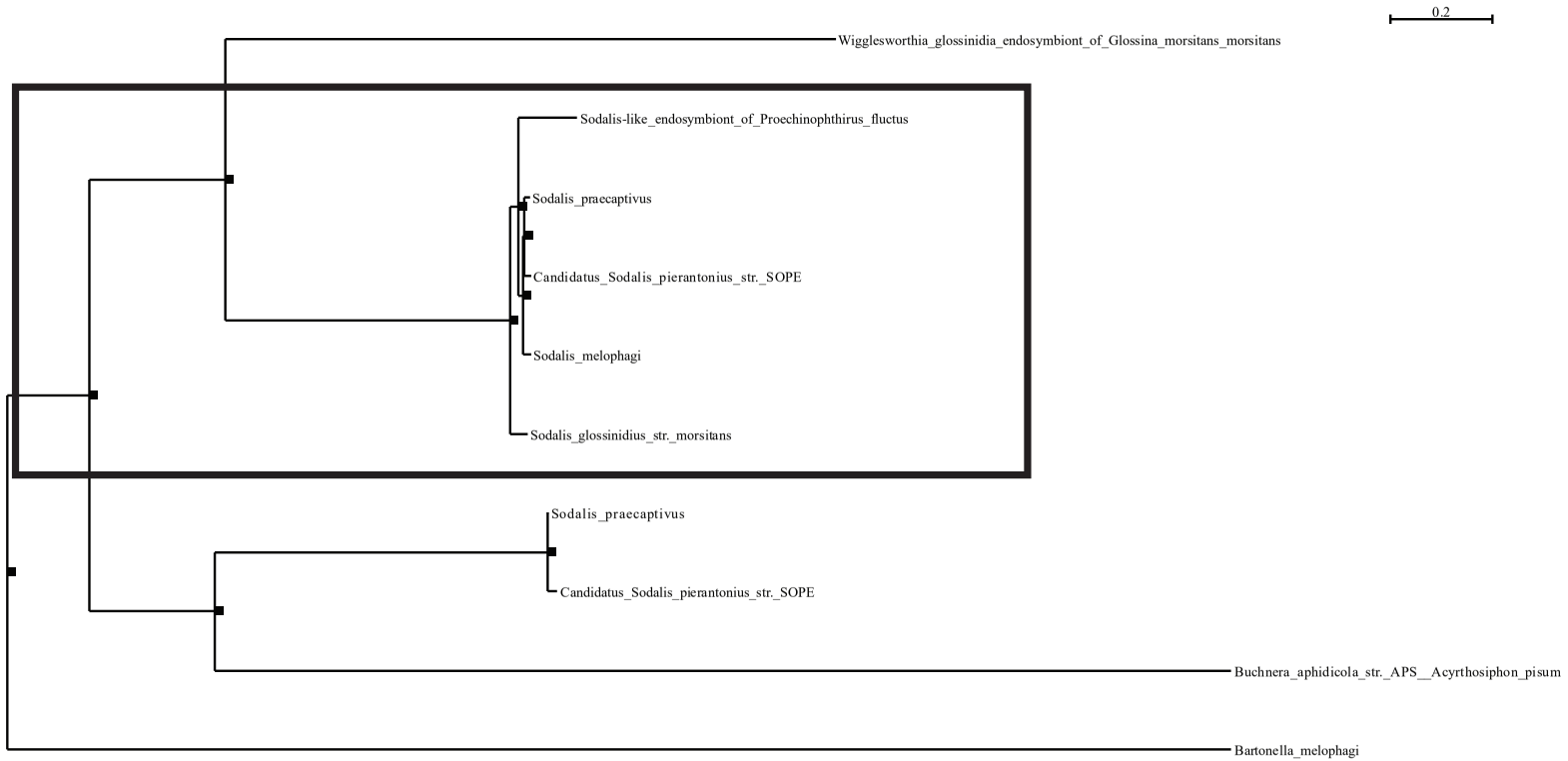
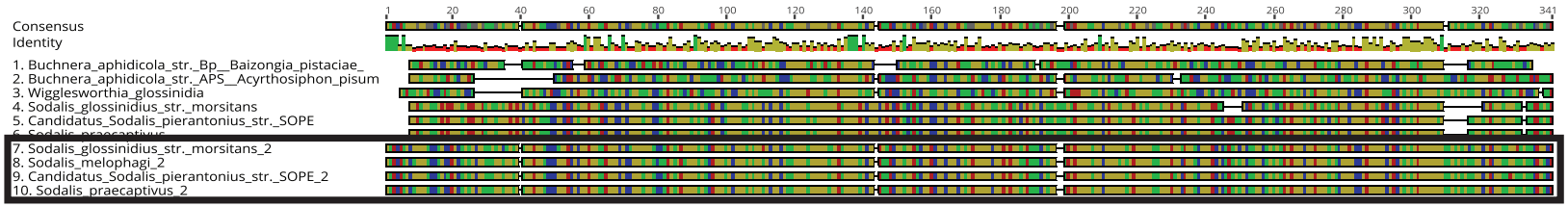


Figure S1b

FliM



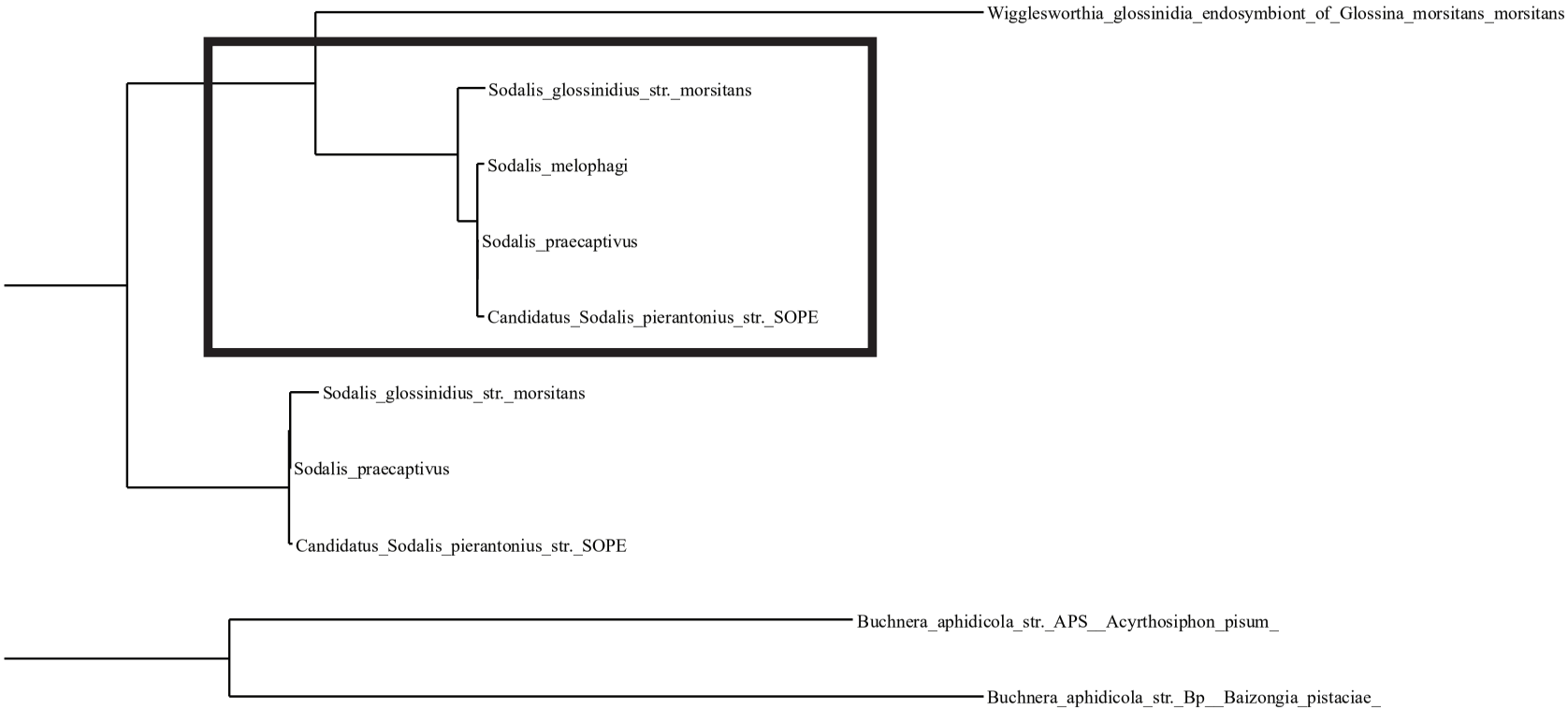
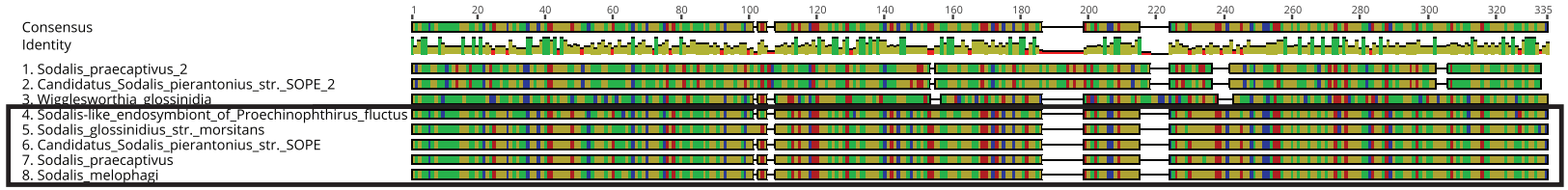


Figure S1c

FlgL



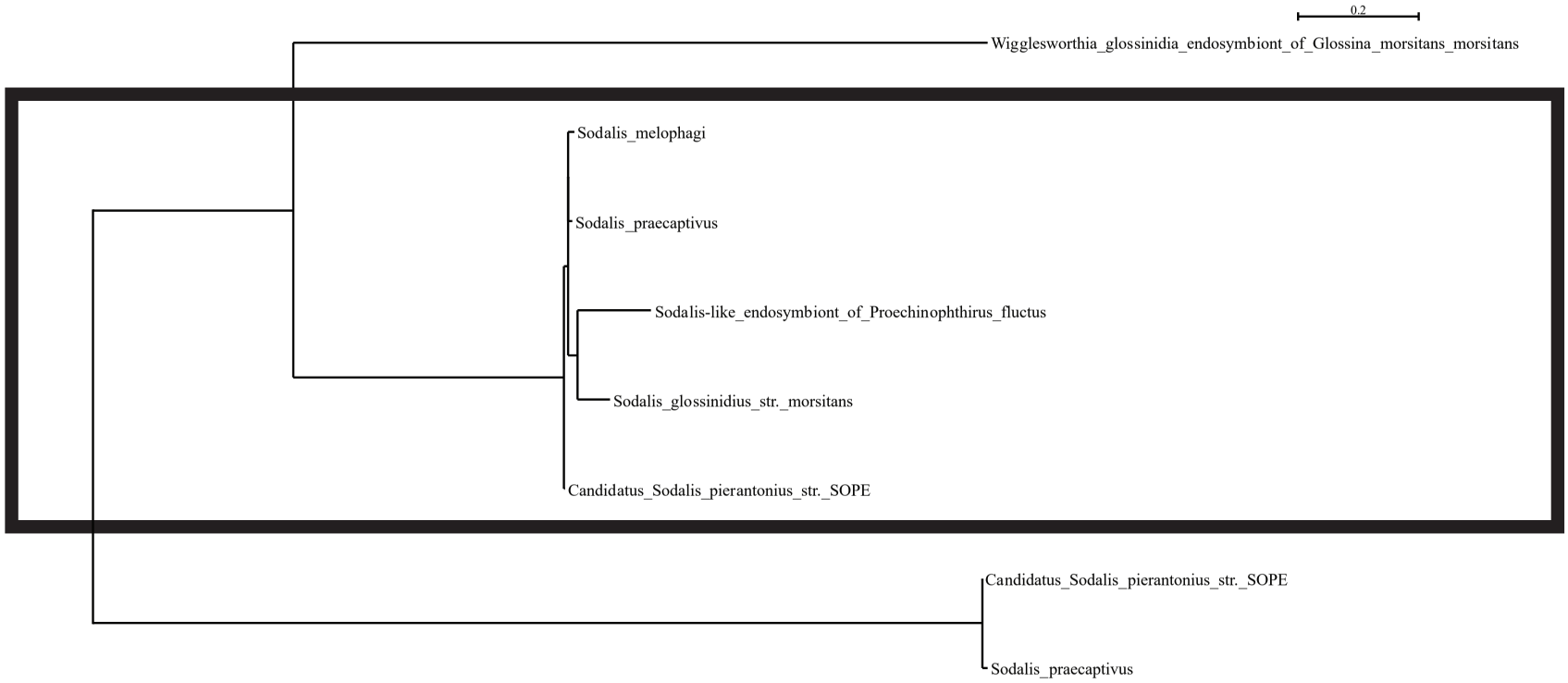
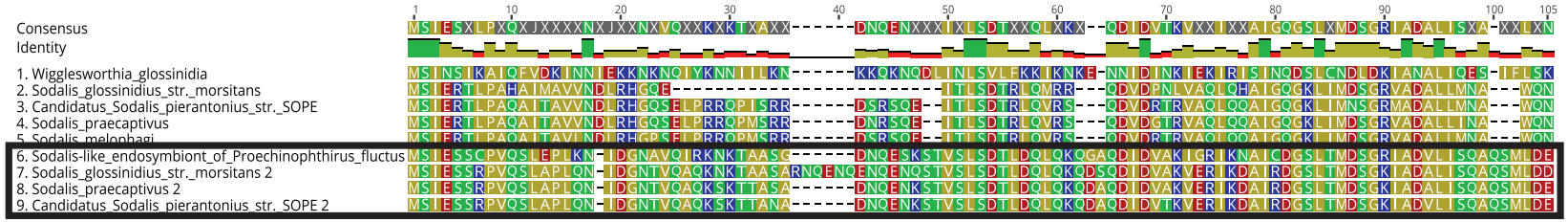


Figure S1d

FlgM



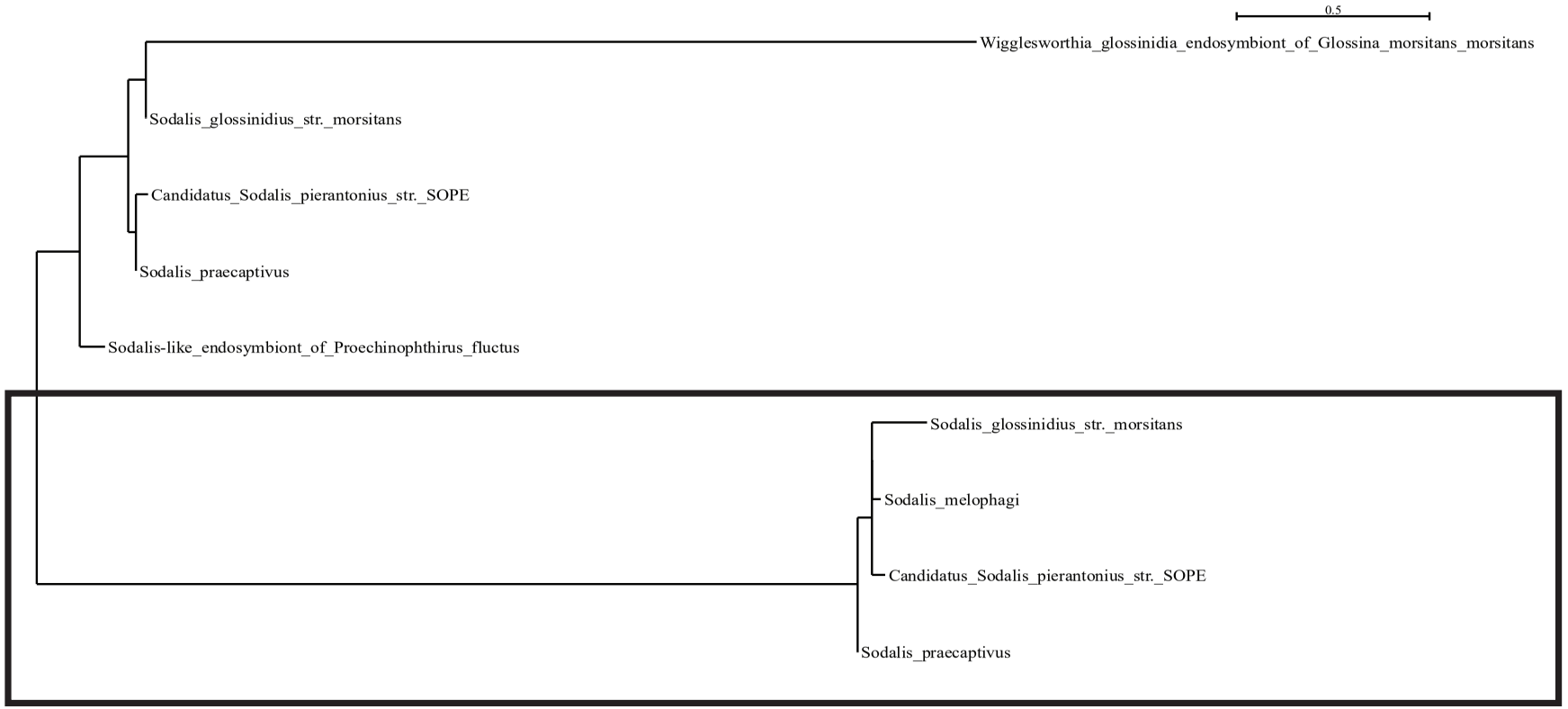
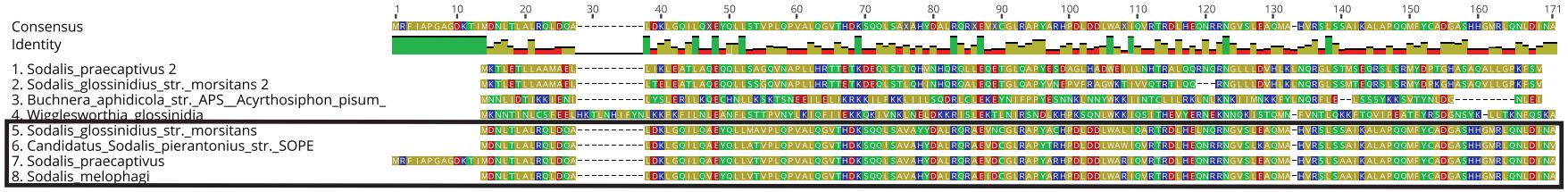


Figure S1e

FlgN



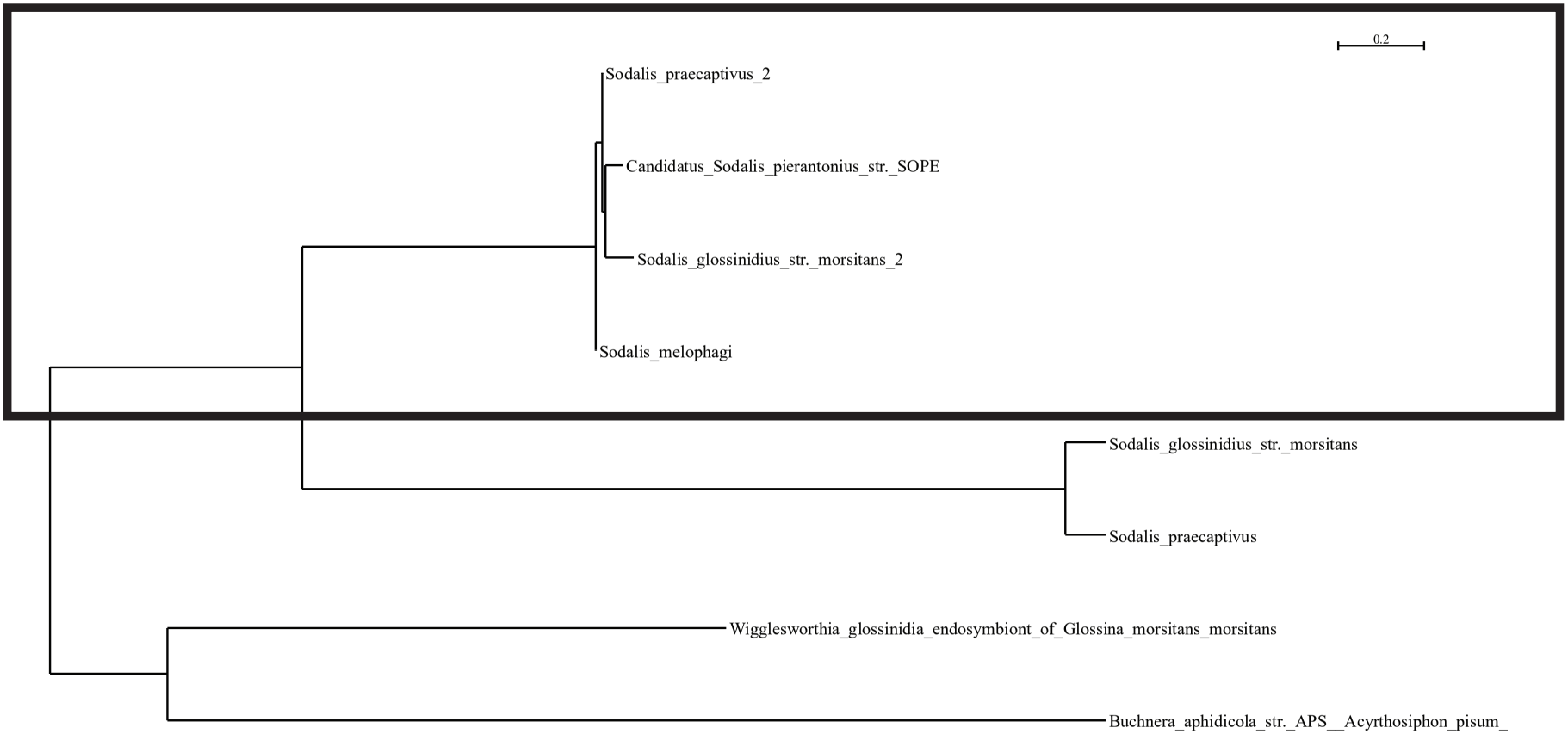
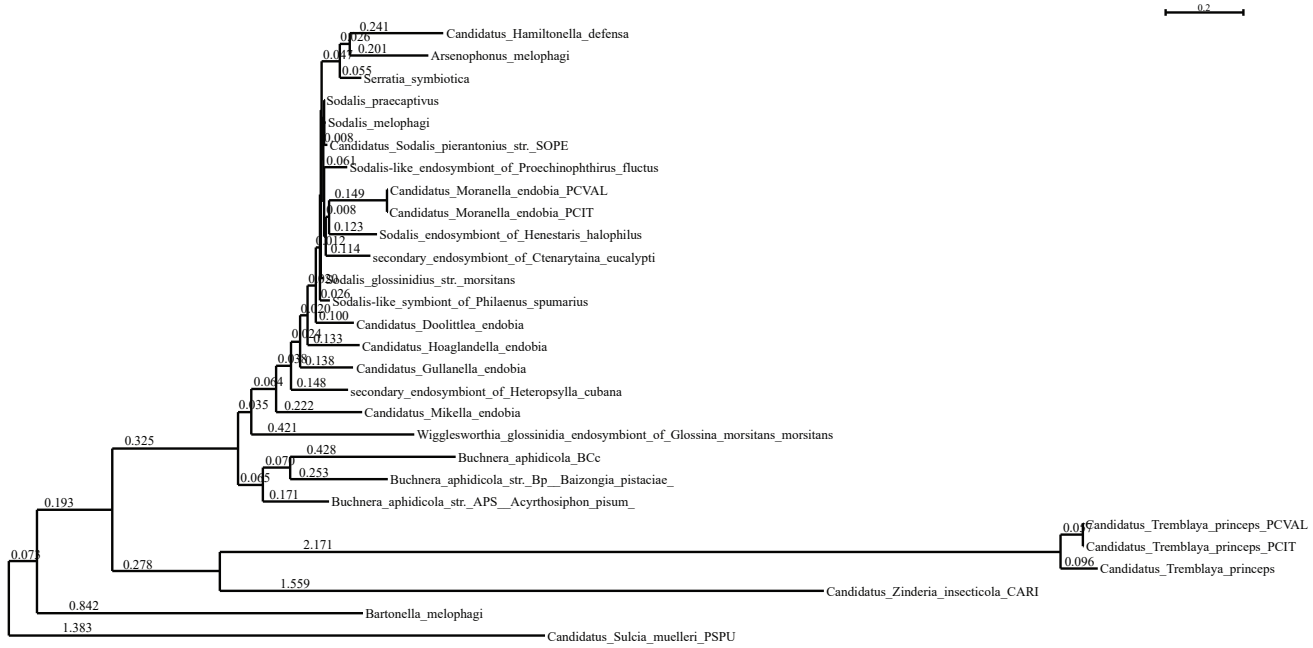
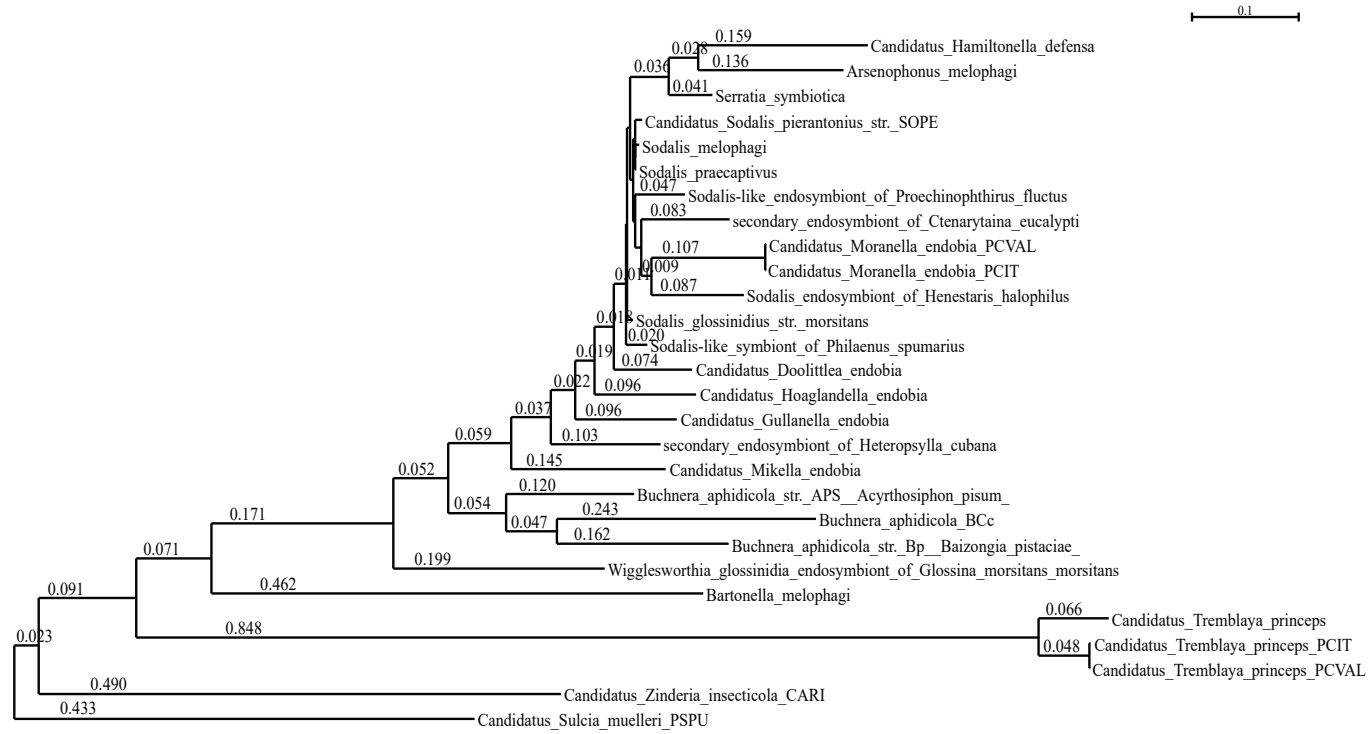


Figure S1f

Figure S2

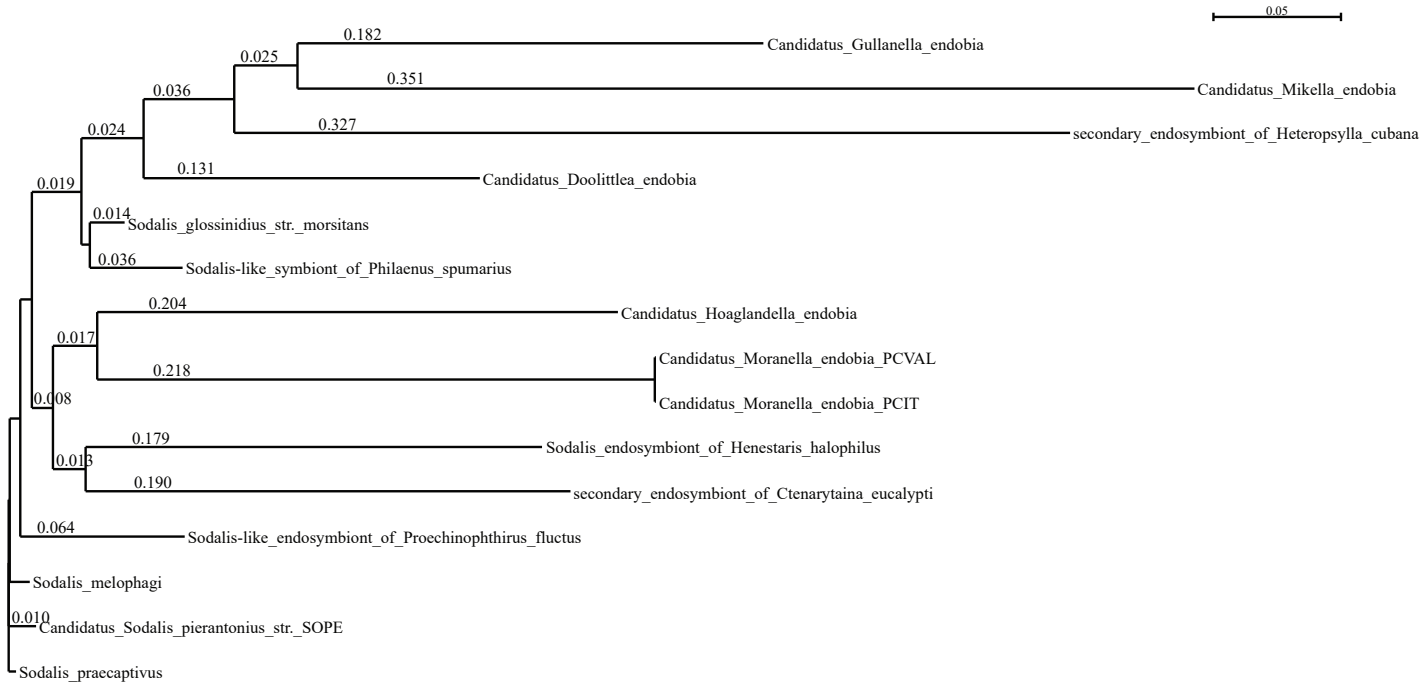


(a)

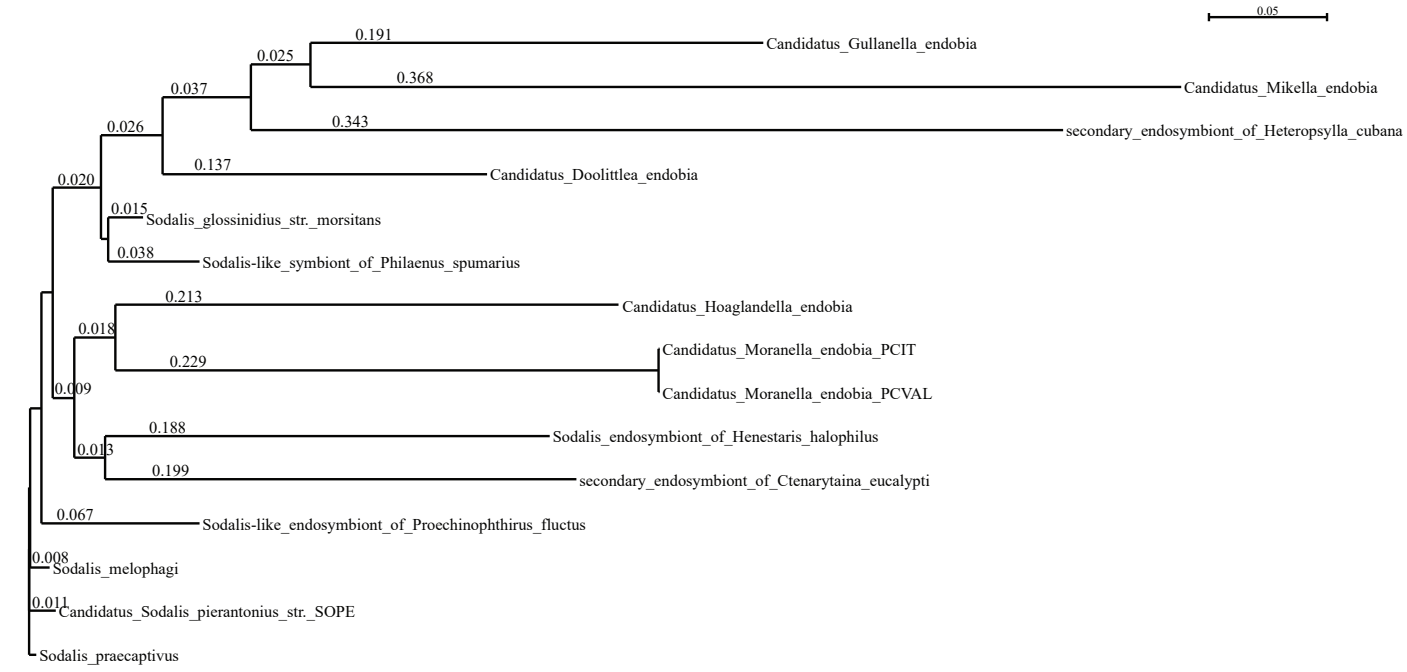


(b) RAxML tree

Figure S3



(a)



(b) RAxML tree

Figure S4

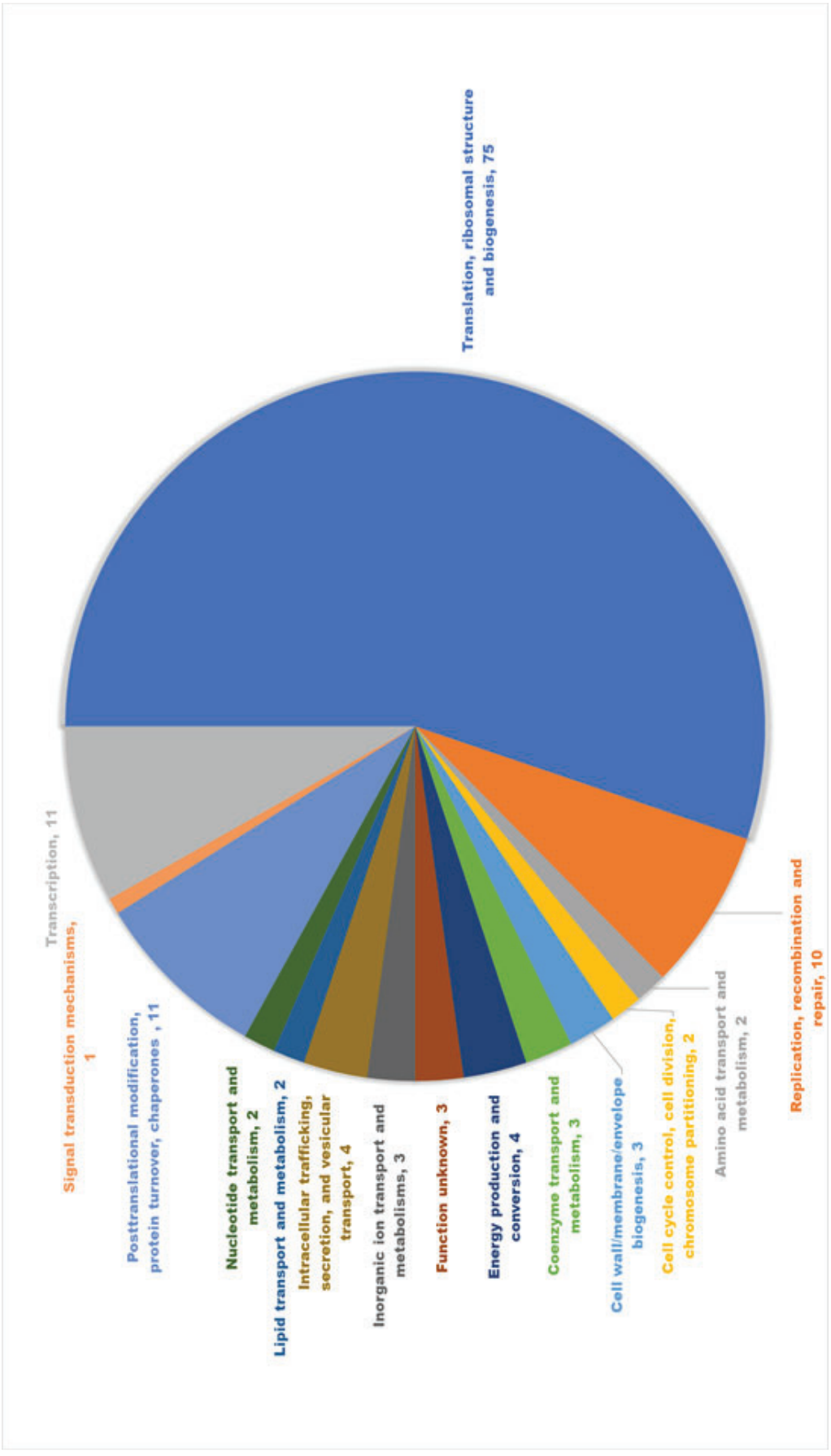

Translation, ribosomal structure and biogenesis, 75
Replication, recombination and repair, 10
Amino acid transport and metabolism, 2
Cell cycle control, cell division, chromosome partitioning, 2
Cell wall/membrane/envelope biogenesis, 3
Coenzyme transport and metabolism, 3
Energy production and conversion, 4
Function unknown, 3
Inorganic ion transport and metabolisms, 3
Intracellular trafficking, secretion, and vesicular transport, 4
Lipid transport and metabolism, 2
Nucleotide transport and metabolism, 2
Posttranslational modification, protein turnover, chaperones , 11
Signal transduction mechanisms, 1
Transcription, 11

Figure S5a

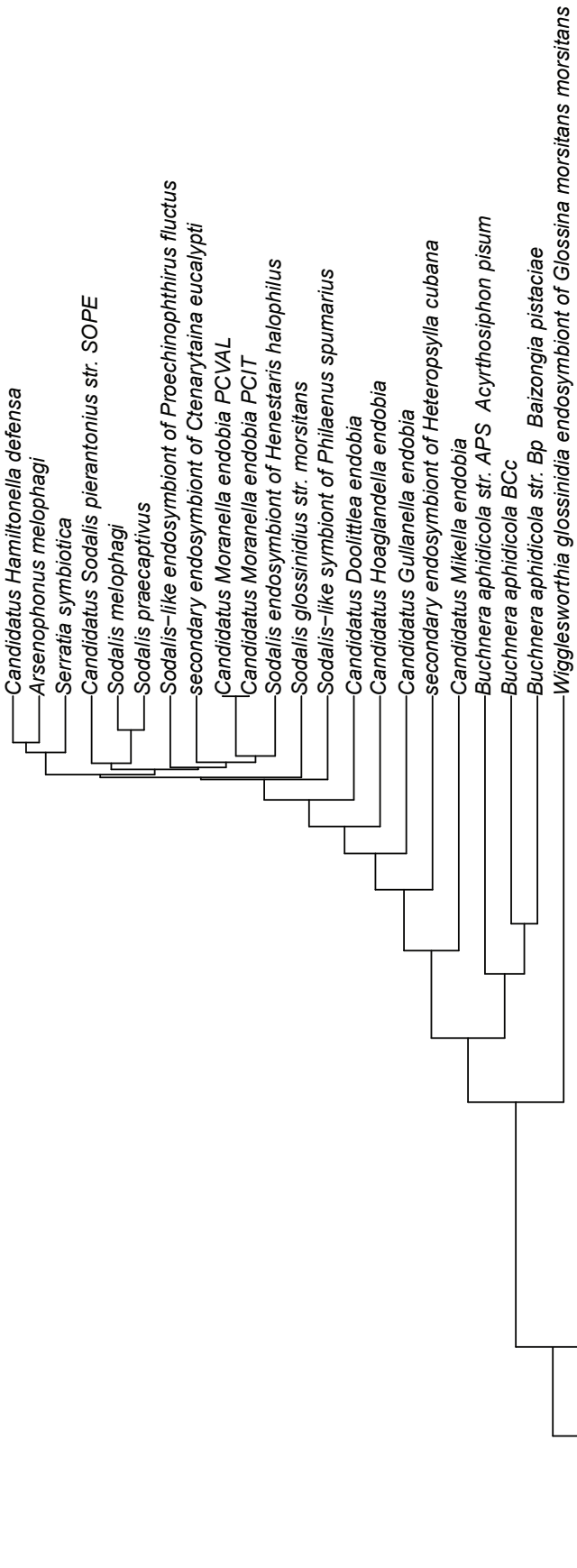
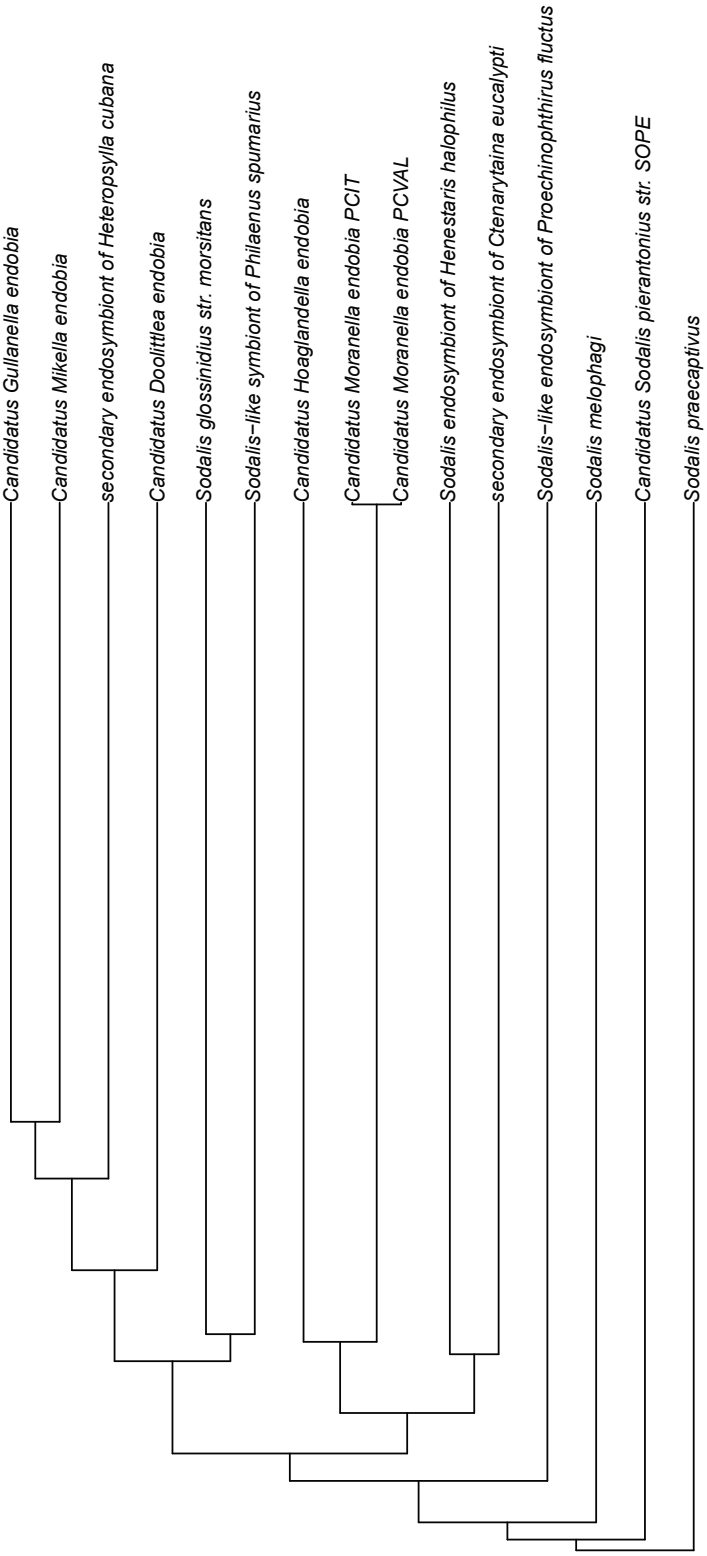

Candidatus Hamiltonella defensa
Arsenophonus melophagi
Serratia symbiotica
Candidatus Sodalis pierantonius str. SOPE
Sodalis melophagi
Sodalis praecaptivus
Sodalis−like endosymbiont of Proechinophthirus fluctus
secondary endosymbiont of Ctenarytaina eucalypti
Candidatus Moranella endobia PCVAL
Candidatus Moranella endobia PCIT
Sodalis endosymbiont of Henestaris halophilus
Sodalis glossinidius str. morsitans
Sodalis−like symbiont of Philaenus spumarius
Candidatus Doolittlea endobia
Candidatus Hoaglandella endobia
Candidatus Gullanella endobia
secondary endosymbiont of Heteropsylla cubana
Candidatus Mikella endobia
Buchnera aphidicola str. APS Acyrthosiphon pisum
Buchnera aphidicola BCc
Buchnera aphidicola str. Bp Baizongia pistaciae
Wigglesworthia glossinidia endosymbiont of Glossina morsitans morsitans
Bartonella melophagi
Candidatus Tremblaya princeps
Candidatus Tremblaya princeps PCIT
Candidatus Tremblaya princeps PCVAL
Candidatus Zinderia insecticola CARI
Candidatus Sulcia muelleri PSPU

Figure S5b


Candidatus Gullanella endobia
Candidatus Mikella endobia
secondary endosymbiont of Heteropsylla cubana
Candidatus Doolittlea endobia
Sodalis glossinidius str. morsitans
Sodalis−like symbiont of Philaenus spumarius
Candidatus Hoaglandella endobia
Candidatus Moranella endobia PCIT
Candidatus Moranella endobia PCVAL
Sodalis endosymbiont of Henestaris halophilus
secondary endosymbiont of Ctenarytaina eucalypti
Sodalis−like endosymbiont of Proechinophthirus fluctus
Sodalis melophagi
Candidatus Sodalis pierantonius str. SOPE
Sodalis praecaptivus