The Evolution of the Flagellar Assembly Pathway in Endosymbiotic Bacterial Genomes

Christina Toft and Mario A. Fares

Department of Genetics, Smurfit Institute of Genetics, University of Dublin, Trinity College, Dublin, Ireland

Genome shrinkage is a common feature of most intracellular pathogens and symbionts. Reduction of genome sizes is among the best-characterized evolutionary ways of intracellular organisms to save and avoid maintaining expensive redundant biological processes. Endosymbiotic bacteria of insects are examples of biological economy taken to completion because their genomes are dramatically reduced. These bacteria are nonmotile, and their biochemical processes are intimately related to those of their host. Because of this relationship, many of the processes in these bacteria have been either lost or have suffered massive remodeling to adapt to the intracellular symbiotic lifestyle. An example of such changes is the flagellum structure that is essential for bacterial motility and infectivity. Our analysis indicates that genes responsible for flagellar assembly have been partially or totally lost in most intracellular symbionts of gamma-Proteobacteria. Comparative genomic analyses show that flagellar genes have been differentially lost in endosymbiotic bacteria of insects. Only proteins involved in protein export within the flagella assembly pathway (type III secretion system and the basal body) have been kept in most of the endosymbionts, whereas those involved in building the filament and hook of flagella have only in few instances been kept, indicating a change in the functional purpose of this pathway. In some endosymbionts, genes controlling protein-export switch and hook length have undergone functional divergence as shown through an analysis of their evolutionary dynamics. Based on our results, we suggest that genes of flagellum have diverged functionally as to specialize in the export of proteins from the bacterium to the host.

Introduction

Bacterial genome sizes range between 9.2 Mb in the soilborne bacterium *Myxococcus xanthus* (Stepkowski and Legocki 2001) and 0.45 Mb in the smallest of the primary symbiotic bacteria of aphids, *Buchnera aphidicola* (Gil et al. 2002). This genome reduction is common for most obligate intracellular bacteria and parasites (Moran and Wernegreen 2000; Gil et al. 2002). The intimate relationship between the 2 organisms of the symbiotic system is believed to be responsible for the reduction in the bacterial genome size, saving thus energy through the removal of unnecessary redundant genes (Andersson et al. 1998). In addition, endosymbiotic bacteria go through severe population bottlenecks in the infection of new insect generations increasing the chance of passing mildly deleterious mutations into the next generations. These mutations are subsequently fixed by the lack of the recombination apparatus (Gil et al. 2003), and this increase in the mutational load inactivates protein-coding genes, which is followed by gene disintegration (Moran 1996; Andersson et al. 1998; Ochman and Moran 2001). The stable cellular environment provided by the host cell and the presence in some cases of secondary symbiotic bacteria providing biosynthetic components lacking in the primary symbiont (Perez-Brocal et al. 2006) renders most of the mechanisms associated with the free lifestyle redundant in endosymbiotic bacteria. An example is the assembly pathway for the flagellum that confers motility to free-living bacteria. The flagellum is characterized by a long rotating helical propeller called filament that is anchored to a basal body of proteins in the cell envelope through a flexible hook (Macnab 2003). The energetic cost of synthesizing the flagella apparatus is significant conferring a growth disadvantage of about 2% (e.g., a nonmotile population overtakes a motile bacterial population in 10 days [Macnab 1996]). This cost slows significantly the growth rate of bacteria (Kutsukake and Iino 1994). However, in free-living bacteria, these disadvantages are overcompensated by the increased capacity provided by the flagella to compete for resources and to avoid toxic chemicals through chemotaxis. In addition, many of the proteins of the flagella pathway are involved in protein export, especially in the export of virulence factors (Young et al. 1999). Flagella motility is an ancient system predating the divergence of archaebacteria and prokaryotes, and the export function may have hence evolved from proteins of the flagellum. However, in nonmotile bacteria, such as in obligate intracellular symbiotic bacteria of insects, the presence of flagella is unnecessary and energetically expensive unless proteins involved in flagella pathway are also involved in other essential functions for the bacterium or the host. Indeed, endosymbiotic bacteria such as *B. aphidicola* are nonmotile and have consequently lost most of the genes involved in the assembly of the flagellum (Maezawa et al. 2006). Many other endosymbionts having a similar endosymbiotic lifestyle and belonging to the gamma-Proteobacteria, such as *Blochmannia floridanus or Blochmannia pensylvanicus* (Gil et al. 2003) and *Baumannia* (Wu et al. 2006) have also lost most of the genes in this pathway. Other symbionts, such as *Wigglesworthia glossinidia*, which has been thought to have a motile phase when transmitted between host generations, retained most of the "flagellar" assembly pathway (Akman et al. 2002).

The 4 fully sequenced *B. aphidicola* genomes still have a large porting of the flagellar assembly genes retained in the genome (Shigenobu et al. 2000; Tamas et al. 2002; van Ham et al. 2003; Perez-Brocal et al. 2006). Many of the *fli* and *flg* gene homologs, involved in flagellar biosynthesis and protein export, show strikingly high amino acid divergence levels in the *Buchnera* lineage compared with its free-living relatives (Tamas et al. 2002). Their observation led them to the suggestion that these genes may have very

Key words: endosymbionts, flagellar pathway, *Buchnera aphidicola*, functional divergence, protein export.

E-mail: faresm@tcd.ie.

Mol. Biol. Evol. 25(9):2069–2076. 2008
doi:10.1093/molbev/msn153
Advance Access publication July 17, 2008

likely changed their function after the establishment of symbiosis. Later, Maezawa et al. (2006) reported the existence of hundreds of flagellar expressed hook and basal body structures but lacking the filament part of the flagellum, supporting previous suggestions of the possible specialization of these genes in protein export from the bacterium to the host (Shigenobu et al. 2000). A recent study has claimed the possible pathogenic and invasive role of kept flagellar genes in *Buchnera* (Moya et al. 2008), although most of the flagellar proteins likely involved in pathogenesis have been lost in these bacteria. The flagella pathway in endosymbiotic bacteria may represent therefore an example of reverse evolution dependent on the bacterium lifestyle whereby the ancient function of the flagella (cell motility) has been replaced by a new function (protein export) that mutational dynamics may be governed by the bacterium, but most likely by the host selection dynamics. It becomes hence crucial to uncover the role of flagellar genes in endosymbiotic bacteria to understand the biological way whereby bacterium and host communicate. However, the implication of flagellar proteins in the export system of proteins from the endosymbiont to the host remains to be investigated.

Here we test the hypothesis of reverse evolution of the flagellum biosynthesis pathway through comparative genomic analyses. We show that 1) there has been a progressive disintegration of genes correlated with the intracellular symbiosis process, 2) there is a differential loss of flagellar genes and functional gene divergence in the different primary symbiotic bacteria of aphids, and 3) the retained genes have been possibly selected for protein export to the host.

Materials and Methods

To test the hypothesis of functional divergence of flagellar genes and differential gene loss between the endosymbiotic bacteria of insects, we first conducted a comparative genomic analysis of the genomes available for the endosymbiotic bacteria of insects and then we studied the changes in the evolutionary dynamics of these genes.

Genomes, Genes, and Alignments

The full list of genes involved in flagellar assembly in *Escherichia coli* (*Ec*: NC_000913) was taken from table 1 in Macnab (1996). Orthologous genes were determined by reciprocal best hits performing BlastP searches of the amino acid sequence of these genes between the *Salmonella typhimurium* (*St*: NC_003197), *Buchnera aphidicola* endosymbionts of *Acyrthosiphon pisum* (*BAp*: NC_002528), *Schizaphis graminum* (*BSg*: NC_004061), *Baizongia pistaciae* (*BBp*: NC_004545), *Cinara cedri* (NC_008513), endosymbiotic bacteria of the carpenter ants *Candidatus Blochmannia pennsylvanicus* (*Bp*: NC_007292) and *Blochmannia floridanus* (*Bf*: NC_005061), and the endosymbiont of the tsetse fly *Wigglesworthia glossinidia* (*Wg*: NC_004344). Only reciprocal best top hits with scores of less or equal to 10^{-4} were accepted. We utilized the clusters of orthologous genes files from National Center for Biotechnology Information for the genomes to identify genes involved in flagellar assembly by looking at their gene names and products.

For each one of the genes, we subsequently built multiple protein sequence alignments using ClustalW (Thompson et al. 1994) using the default parameters. Then we obtained protein-coding sequence alignments concatenating nucleotide triplets according to their corresponding protein alignments. We also built multiple sequence alignments for the complete set of genes in common among the symbiotic and free-living bacterial genomes for downstream evolutionary analyses. All multiple sequence alignments were carefully inspected.

Analysis of Evolutionary Rates

We estimated the number of substitutions per nonsynonymous site (d_N) and number of mutations per synonymous site (d_S) using the modified Nei and Gojobori (1986) method implemented in the program PAML v4 (Yang 1997). Because of the bias in AT content in endosymbiotic bacteria of insects, we sought to obtain accurate estimates of these parameters by applying several maximum likelihood models implemented in PAML. The models applied were M0, M1, M2, M3, M7, and M8 (for detailed explanation of these models, see Yang and Nielsen 2000). We then obtained the mean values for d_N and d_S under the appropriate model. To determine the best model explaining our data and phylogenetic tree, we compared the models' log-likelihood values by the likelihood ratio test. Here we assumed that synonymous substitutions are neutrally fixed because they produce no changes in the amino acid composition of proteins. In theory, hence the number of synonymous substitutions per site is proportional to the time since the species diverged. Based on this, the ratio between d_N and d_S is a good measure of the force of selection acting on a particular protein. In order to identify shifts in the evolutionary rates due to the intracellular lifestyle of the endosymbiotic bacteria in each one of the genes, we compared the nonsynonymous-to-synonymous rates ratio (ω) for the pairwise comparison *BAp–BSg* with that for the comparison of *Ec–St*, by dividing both ratios ($R = \omega_{BAp-BSg}/\omega_{Ec-St}$). We implemented this comparison because both pairs of species have been estimated to present similar divergence dates. The hypothesis tested in these comparisons was whether R was maintained at 1 (no change in selective constraints), $R > 1$ (relaxed selective constraints in the endosymbiotic bacteria), and $R < 1$ (increased selective constraints in the endosymbiotic bacteria). It is noteworthy that saturation of synonymous sites due to nucleotide compositional bias in endosymbiotic genes make our analyses and conclusions conservatives because such saturation would lead to inflated ω values and would yield therefore significantly higher R values, which would be interpreted as evidence for no functional divergence in endosymbiotic proteins. To conduct this comparison in *BBp*, we estimated first the ω values for the pairwise sequence comparisons using the Nei and Gojobori method.

Table 1
Events of Gene Loss among the Endosymbiotic Bacteria of Aphids in the Flagellar Assembly Pathway

	BAp	*BSg*	*BBp*	*BCc*	Description
Master control					
FlhC	—	—	—	—	Master regulator
FlhD	—	—	—	—	Master regulator
Regulators					
FliK	BU079	BUsg072	bbp073	—	Hook-length control
FliZ	—	—	—	—	Regulator of FliA activity
FliA	—	—	—	—	σ 28 factor
FlgM	—	—	—	—	Anti-σ factor for FliA
Chaperones					
FliJ	BU077	BUsg071	bbp072	—	General chaperone
FlgA	BU336	BUsg324	—	—	Chaperone for FlgI
FlgN	BU335	BUsg323	—	—	Chaperone for FlgKL
FliS	—	—	—	—	Chaperone for FliC
FliT	—	—	—	—	Chaperone for FliD
Motor control complex					
FliG	BU074	BUsg068	bbp069	BCc_044	Motor/switch
FliM	BU080	BUsg073	bbp074	—	Motor/switch
FliN	BU081	BUsg074	bbp075	BCc_047	Motor/switch
FlgH	BU343	BUsg331	bbp314	BCc_212	Basel body L-ring
MotA	—	—	—	—	Motor protein
MotB	—	—	—	—	Motor rotation protein
FlgB	BU337	BUsg325	bbp310	—	Basal body rod
FlgC	BU338	BUsg326	bbp311	—	Basal body rod
FlgF	BU341	BUsg329	bbp312	BCc_211	Basal body rod
FlgG	BU342	BUsg330	bbp313	—	Basal body rod
FlgJ	BU345	BUsg333	bbp316	—	Temporary rod cap
FlgI	BU344	BUsg332	bbp315	BCc_213	Basel body P-ring
FliF	BU073	BUsg067	bbp068	BCc_043	Basal body MS-ring
FliE	BU072	BUsg066	bbp067	—	Basal body
Flagellar export apparatus					
FlhA	BU241	BUsg236	bbp223	BCc_151	Export pore protein
FlhB	BU240	BUsg235	bbp222	BCc_150	Export pore protein
FliO	—	—	—	—	Biosynthesis protein
FliP	BU082	BUsg075	bbp076	BCc_048	Biosynthesis protein
FliQ	BU083	BUsg076	bbp077	BCc_049	Biosynthesis protein
FliR	BU084	BUsg077	bbp078	BCc_050	Export pore protein
FliH	BU075	BUsg069	bbp070	BCc_045	Biosynthesis protein
FliI	BU076	BUsg070	bbp071	BCc_046	ATPase
Hook and filament					
FlgE	BU340	BUsg328	—	—	Hook protein
FlgD	BU339	BUsg327	—	—	Temporary hook cap
FlgK	BU346	BUsg334	—	—	Hook–filament junction
FlgL	—	—	—	—	Hook–filament junction
FliC	—	—	—	—	Filament
FliD	—	—	—	—	Filament cap

NOTE.—Presence of a gene is represented by its GenBank accession number for the corresponding species, whereas absence or loss is represented by (—).

Then we estimated the ω value for the branch leading to *BBp* as follows:

$$\omega_{BBp} = \frac{\left(\frac{1}{2}(\omega_{BBp-BAp} + \omega_{BBp-BSg})\right) + \left(\frac{1}{2}(\omega_{BBp-Ec} + \omega_{BBp-St})\right) - \left(\frac{1}{4}(\omega_{BAp-Ec} + \omega_{BAp-St} + \omega_{BSg-Ec} + \omega_{BSg-St})\right)}{2}$$

To test the significance of the *R* values for each one of the flagellar genes, we first estimated *R* values for the full set of genes present in free-living and endosymbiotic bacteria of aphids (see supplementary table S1, Supplementary Material online). Then we resampled 10,000 replicates from the distribution of *R* values and identified the median and threshold *R* values below which we consider *R* significant.

Results and Discussion

Differential Loss of Flagellar Genes in Endosymbiotic Bacteria

Comparative genomic analysis of the endosymbiotic bacteria of insects: *BAp; BSg; BBp; BCc; Bf; Bp; Wg;* and the free-living bacteria: Ec; St; indicates that the loss of flagellar genes is indeed associated to the intracellular life, with all the intracellular symbionts presenting lack of an important percentage of flagellar genes (fig. 1 and table 1). The different endosymbionts however showed different degrees of gene loss, going from complete lack of flagellar genes (in *Bf* and *Bp*) to a very partial gene content (in *BCc*) or to a greater flagellar genes representation (in *BAp* and *BSg*) (fig. 1 and table 1). In contrast, *Wg* have conserved most of the flagellum genes, suggesting that the flagellum is of

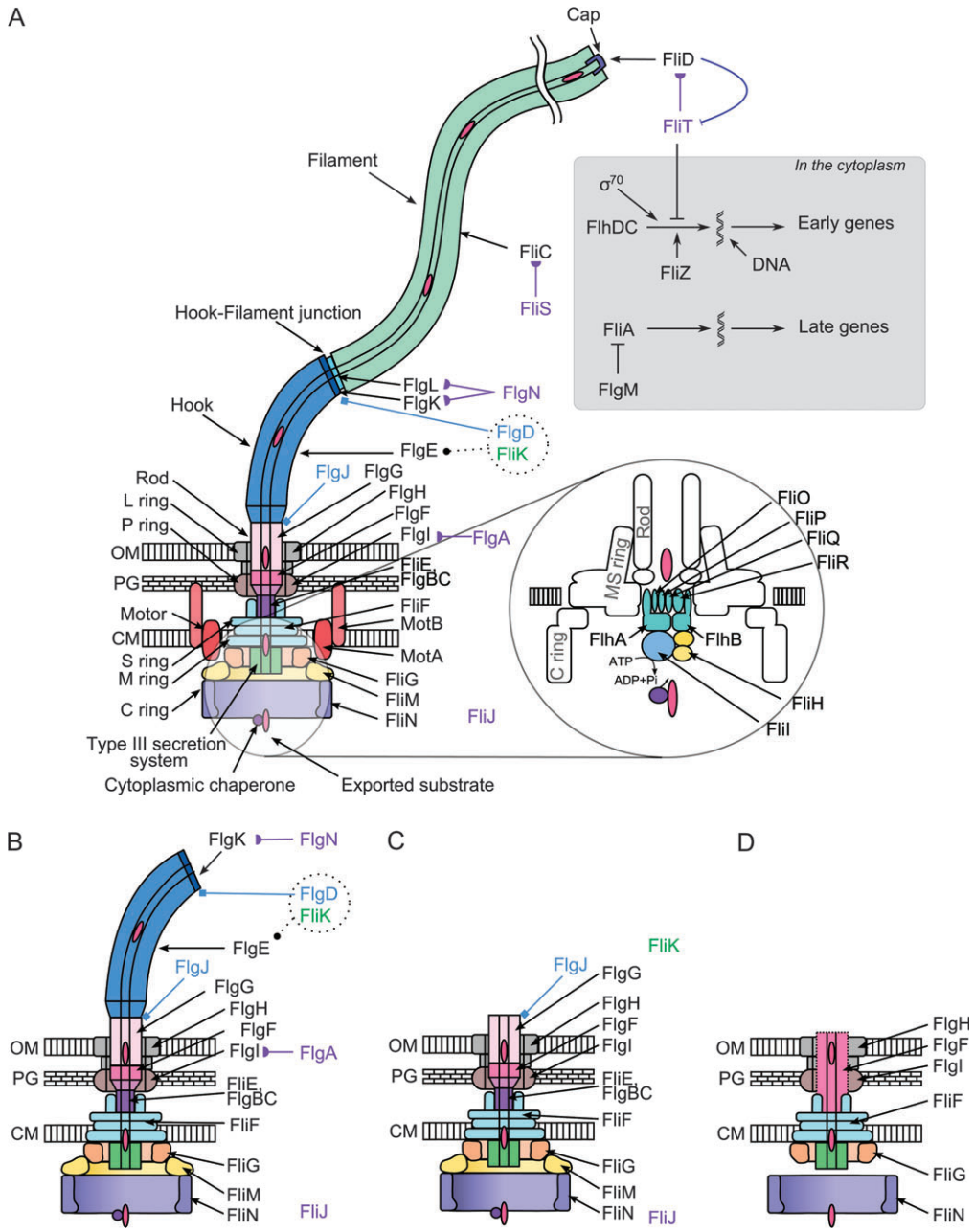

FIG. 1.—Schematic diagram of the bacterial flagellar assembly pathway, excluding the bacteria chemotaxic pathway. (*A*) The flagellar assembly pathway as observed in *Escherichia coli* (*Ec*) and *Salmonella typhimurium* (*St*). The 4 fully sequenced genomes of the endosymbiont *Buchnera aphidicola* only contain part of this pathway/structure. All 4 endosymbionts have lost the regulatory genes of the pathway, and they have all retained most of the type III export apparatus proteins. (*B*) *Buchnera Acyrthosiphon pisum* (*BAp*) and *Buchnera Schizaphis graminum* (*BSg*) have retained the basal body and hook. (*C*) *Buchnera Bayzongia pistaciae* (*BBp*) has farther reduced the pathway to only the basal body. (*D*) *Buchnera Cinara cedri* (*BCc*), the smallest of the 4 *B. aphidicola* genomes, has reduced the gene number codifying for the basal body. Outer membrane (OM); peptidoglycal layer (PG); and cytoplasmic membrane (CM) are indicated. The purple proteins (FliJST, FlgADN) are chaperones, and they are linked through connectors to their specific client proteins. Proteins names in blue (FlgDJ) are those forming the temporary caps. This figure is redrawn with permission from the original authors (Minamino and Namba 2004).

importance for the lifestyle of this bacterium and could facilitate the transmission to intrauterine progeny (Aksoy and Rio 2005).

Genes involved in the biosynthesis of flagella are organized into 3 classes of operons (classes 1, 2, and 3) with the expression of the next class (e.g., class 2) requiring the expression of the previous transcriptional class (e.g., class 1) (Kutsukake et al.1990). The first class, also named master operon (flhDC), includes 2 genes, and they are essential for positive transcriptional activation of class 2 operons that contains genes whose products are required for the morphogenesis of the hook and basal body (Jones and Macnab 1990). Finally, class 3 operons include late-expression genes such as the motor torque generator subunits MotA

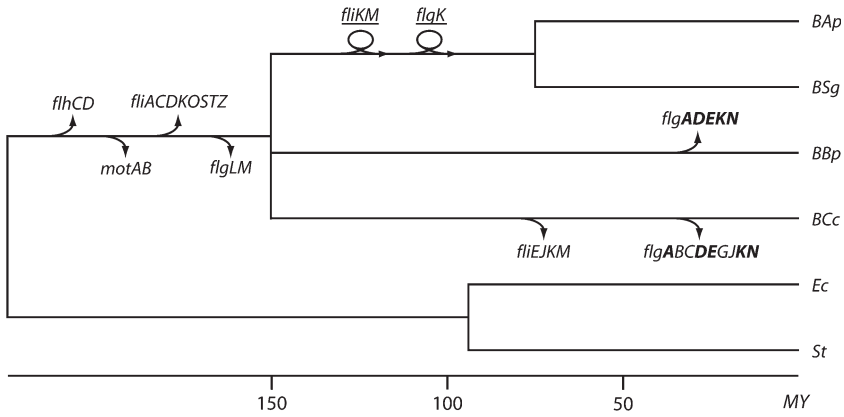

FIG. 2.—Schematic representation of events of gene loss or functional divergence for the flagellar assembly pathway in aphids' endosymbionts. Arrows leading off branches indicate gene loss, whereas an arrow looping back onto the branch indicate that the genes (are underlined) have possibly undergone functional diverges. The genes in bold are those that are lost in both the branch leading to *Buchnera Byzongia pistaciae* (*BBp*) and *Buchnera Cinara cedri* (*BCc*). Caution however must be taken because of the phylogenetic position ambiguity of *BCc*. The dates of the splits are only approximate dates: *Buchnera Acyrthosiphon pisum* (*BAp*) and *Buchnera Schizaphis graminum* (*BSg*) are estimated to have split 50–70 MYA, the most common symbiotic ancestor is thought to date back to 150–250 MYA. The free-living bacteria *Escherichia coli* (*Ec*) *and Salmonella typhimurium* (*St*) are thought to have diverged approximately100 MYA.

and MotB, chemotaxis proteins, and the flagellin proteins FliC and FljB (Macnab 1996). The 2 genes of the master operon have been lost in all *Buchnera* lineages (fig. 1 and table 1). It has been shown that when the genes *flhC* and *flhD* are mutated in *Ec* and *St*, cells become nonmotile and nonflagellated (Wang et al. 2006). Most of the genes belonging to class 2 operons have been retained in *Buchnera*, as well as some of the structural flagellar proteins (table 1). Regarding the structural proteins of the hook, we observed differential conservation between the different *Buchnera* primary symbionts (fig. 1 and table 1). For example, FlgE, FlgD, and FlgK have been retained only in *BAp* and *BSg* but not in *BBp* or *BCc*. Also, the protein determining the length of the hook (FliK) has been retained in the largest *Buchnera* primary endosymbionts: *BAp*, *BSg*, and *BBp*. Most of the genes therefore belonging to class 3 operons have been lost in the 4 *Buchnera* endosymbiotic lineages including (MotA, MotB) together with all the genes encoding the proteins of the hook, hook–filament junction, and filament. The lineage formed by *BAp* and *BSg* however represents an intermediate stage with some of the genes encoding for the hook and hook–filament junction, belonging to class 3 operons, having been retained (fig. 1 and table 1). Genes from the class 3 operons have a positive transcriptional control over the master operon FlhDC through the protein FliZ. FliZ however is not present in any of the *Buchnera* endosymbionts, coinciding with that FlhDC has also been lost in these bacteria. Further, the sigma factor 28 (FliA) that has a negative transcriptional control over the master operon and its anti-sigma factor FlgM has also been lost in all *Buchnera* endosymbiont lineages. So in general, they have lost all the genes involved in the regulation of the flagellar assembly pathway.

In sharp contrast to the case of genes involved in biosynthesis of the flagellum, the protein export system of the flagellar proteins has been almost completely retained in the *Buchnera* endosymbionts analyzed. Because most of the genes involved in hook–filament junction and filament biosynthesis have been lost in *Buchnera* endosymbionts, the export system may be more specialized in exporting proteins to the host. However, this mechanism does not seem to be a general feature in endosymbiotic bacteria of insects because neither *Bf* nor *Bp* retained any of these export proteins. In addition, analysis of the distribution of gene loss events adopting a maximum parsimony criterion in the phylogenetic tree of Buchnera (fig. 2) puts forward the conclusion that most of these gene losses have occurred in the most common symbiotic ancestor as well as in the lineages leading to *BBp* and *BCc*. We could not find events specific to *BAp* or *BSg*, which is in agreement with the genome stasis previously demonstrated for these bacteria (Tamas et al. 2002). The question remaining to be answered is why *BAp* and *BSg* present a differential gene loss in comparison with *BBp* or *BCc*?

Differential Selective Pressures among Flagellar Genes

Endosymbiotic bacteria of insects have small population sizes, do not undergo recombination, and are maternally transmitted in a strictly clonal manner through tight population bottlenecks (Funk et al. 2000, 2001). The consequence of this transmission dynamic is the fixation of deleterious mutations due to genetic drift and the irreversible decline in fitness (Muller 1964). This decline in fitness may be compensated by the overexpression of the chaperonin GroEL that buffers the effects that mutations have on the protein folding (Fares et al. 2002). Small populations of asexual organisms, such as the endoymbiont of *B. aphidicola*, show increased rates of sequence evolution when the amount of mildly deleterious mutations is substantial (Ohtaka and Ishikawa 1993). As shown by Moran (1996), the increased rates of evolution should only affect amino acid sites under selection because neutral sites are independent of the population structure. We therefore expect selective constraints to be relaxed over functional sites and consequently the ratio of nonsynonymous-to-synonymous mutations rate may have increased. Although this may be

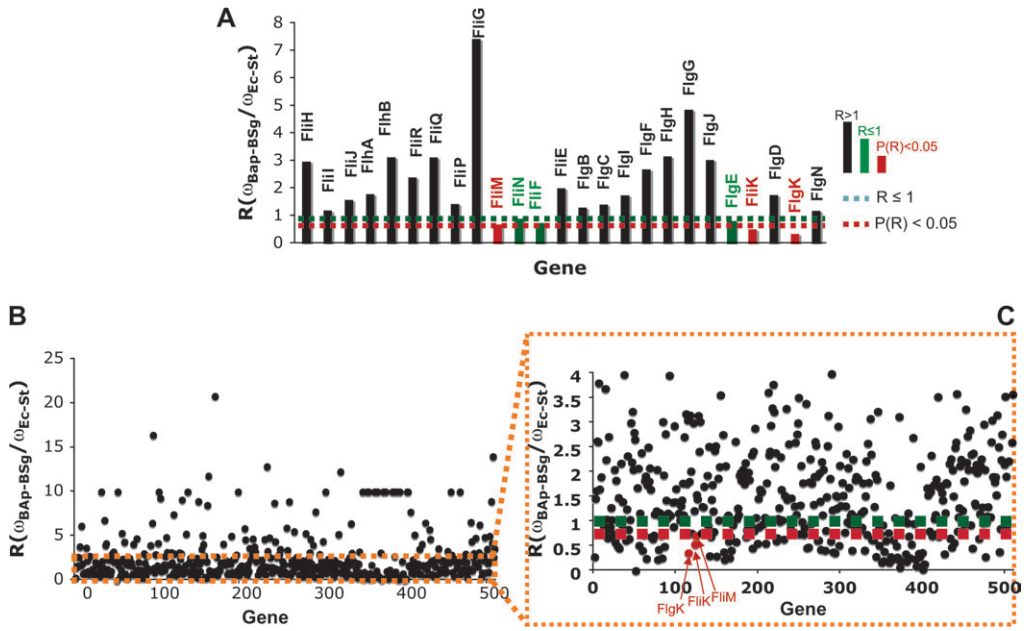

FIG. 3.—Comparative genomic analysis of selective constraints between endosymbiotic bacteria and their free-living relatives. We divided the nonsynonymous-to-synonymous rates ratio estimated for the comparison of *Buchnera Acyrthosiphon pisum* (*BAp*) and *Buchnera Schizaphis graminum* (*BSg*) by that estimated for the comparison of *Escherichia coli* (*Ec*) and *Salmonella typhimurium* (*St*) ($R = \omega_{BAp-BSg}/\omega_{Ec-St}$) for the genes of the flagellar assembly pathway (*A*). Then we estimated *R* for the complete set of genes in the genome of endosymbionts (*B*) and tested the significance of the *R* values of the flagellar genes against a distribution of 10,000 pseudorandomly sampled *R* from the 520 genes examined. We then identified significant *R* values at the 5% confidence (*C*).

true for the vast majority of genes in the endosymbiont, several circumstances may challenge this outcome. For example, in highly essential genes that present no selective flexibility to mutations, the evolutionary rate may be maintained in endosymbionts in comparison with free-living bacteria due to the unavoidable deleterious effects that mutations have on these genes. In addition, genes that have functionally diverged in endosymbionts to accommodate their function to a new lifestyle (intracellular life) may have undergone selective shifts presenting evolutionary rates that are equal or lower than those in free-living bacterial genes. We tested the functional divergence of kept flagellum genes in *Buchnera* toward other functions different from their original ones, for example, divergence for protein export. For that purpose, we compared the strength of selection in endosymbiotic genes in comparison with their free-living cousins by dividing ω values estimated for the comparison *BAp* and *BSg* by that estimated for the comparison of *Ec* and *St* ($R = \omega_{BAp-BSg}/\omega_{Ec-St}$).

Due to that *Buchnera* cells are nonmotile and flagella have lost components associated to the hook and the entire set of filament proteins, we expect a change in the function of the proteins in that pathway toward the export of proteins from the bacteria to the host. Most of the genes examined in this pathway showed greater increase (lower selective pressures) in the ω values in endosymbionts compared with their free-living relatives ($\omega_{BAp-BSg} >> \omega_{Ec-St}$) (fig. 3*A*). However, in few instances, the rate of evolution was slower in endosymbiont than in the free-living relatives. Such was the case of genes encoding for the C-ring (FliMN), hook–filament junction and hook proteins (FlgK and FlgE), basal body MS-ring protein (FliF), and FliK protein responsible for the hook-length control (fig. 3*A* and table 2). Proteins from the C-ring and FliK are intimately coordinated during the export of hook proteins in free-living flagellated bacteria. Aside from its role as hook-length ruler, FliK has also been shown to be involved in the initiation of the switch in export substrate specificity (Hirano et al. 1994; Koroyasu et al. 1998). However, the detailed role of FliK and its coordinated function with proteins from the C-ring is under continual debate. The involvement of both types of proteins in protein export is supported by several data. For instance, the C-terminal 87 residues of FliN have sequence homology to Spa33, a protein implicated in the protein transmembrane export in *Shigella flexneri* (Tang et al. 1995). Furthermore, FliM and FliN form a stable $FliM_1FliN_4$ solution complex (Brown et al. 2005), and FliM is known to have a chemotactic activity important for the orientation of the bacterial movement in the medium (Bren and Eisenbach 1998). A change in the selective constraints in FliM may have conducted its functional divergence toward sensing the concentration of exported proteins from *Buchnera* cells and maintaining thus a balance between exported and produced proteins by the cell. FliK regulates FlgK and FlgE and the functional divergence of these proteins may have conferred them separate but related functions. Finally, FliF performs a structural link between the S-ring and M-rings through which proteins are exported. To test the significance of low *R* values for these genes, we conducted a genomic comparison of *BAp* and *BSg* versus *Ec* and *St* and calculated the *R* values for each one of the genes present in all 4 genomes (supplementary table S1, Supplementary Material online) and plotted these values along the genome (fig. 3*B*). Plotting R_{FliK}, R_{FliM}, R_{FliN}, R_{FlgE}, R_{FliF}, and R_{FlgK}

Table 2
Analysis of Functional Divergence in Flagellar Genes in the Endosmbiont of *Buchnera aphidicola*

Operon	Gene	ω_{Ec-St}[a]	$\omega_{BAp-BSg}$[b]	R[c]	Structure[d]
Class 2	*fliH*	0.0764	0.2286	2.9924	Export
	fliI	0.0383	0.0465	1.2154	Export
	fliJ	0.0256	0.0417	1.6289	General chaperone
	flhA	0.0247	0.0445	1.8063	Export
	flhB	0.0434	0.1367	3.1482	Export
	fliR	0.0577	0.1395	2.4173	Export
	fliQ	0.0131	0.0372	3.1409	Export
	fliP	0.0287	0.0417	1.4521	Export
	fliG	0.0252	0.1874	7.4486	Rotor/switch protein
	fliF	0.0807	0.0609	0.7549	MS-ring
	fliM	0.2080	0.1468	0.7058	C-ring
	fliN	0.0832	0.0756	0.9093	C-ring
	fliE	0.0885	0.1786	2.0197	Proximal rod
	flgA	0.1673	0.1705	1.0191	Chaperone for P-ring protein
	flgB	0.0672	0.0884	1.3153	Rod
	flgC	0.0366	0.0525	1.4321	Rod
	flgI	0.0355	0.0626	1.7625	P-ring
	flgF	0.0421	0.1141	2.7109	Rod
	flgH	0.0237	0.0753	3.1784	L-ring
	flgG	0.0028	0.0139	4.8694	Distal rod
	flgJ	0.0386	0.1177	3.0529	Temporal rod cap
	flgE	0.0547	0.0618	0.8661	Hook
	fliK	0.2225	0.1181	0.5310	Control hook length
	flgD	0.0672	0.1192	1.7728	Temporal hook cap
Class 3	*flgK*	0.0988	0.0355	0.3597	First hook–filament junction
	flgN	0.0470	0.0566	1.2043	Chaperone for hook–filament junction proteins

[a] Nonsynonymous-to-synonymous rates ratio estimated by the modified method of Nei and Gojobori for the comparison between the sequence of *Escherichia coli* (*Ec*) and *Salmonella typhimurium* (*St*).

[b] Nonsynonymous-to-synonymous rates ratio estimated by the modified method of Nei and Gojobori for the comparison between the sequence of the endosymbiont of *Buchnera aphidicola Acyrthosiphon pisum* (*BAp*) and *Schizaphis graminum* (*BSg*).

[c] The ratio between the ratios of nonsynonymous-to-synonymous rates of free-living bacteria to that of endosymbiotic bacteria.

[d] The structural role of the protein codified by that gene in the flagella.

in the distribution of R values shows that some of these values are significantly smaller than expected (fig. 3*C*).

To determine whether these selective constraints are general among endosymbionts of aphids, we measured ω for the branch leading to *BBp* and compared this value with that obtained for free-living bacteria *Ec* and *St*. The analysis showed that all those flagellar genes that presented low R values in the comparison of *BAp–BSg* to *Ec–St* had values of $R > 1$ on the *BBp* lineage. Interestingly, some of the genes presenting very high R values in *BBp* (FliM, FlgG) have been lost in the most reduced *Buchnera* genome *BCc*. In addition, FlgD and FlgE that interact with FliK have been lost in *BBp* where FliK present values of $R > 1$, suggesting that no functional divergence has occurred in FliK in this *Buchnera* lineage. Furthermore, BlastP searches of FliK in *BBp* against the other bacteria only found homologs in other *Buchnera* but not in the free-living relatives, suggesting that FliK diverged functionally after the speciation event giving the lineages of *BAp* and *BSg*. Placing the events in the phylogenetic tree of Buchnera (fig. 2) shows that the 3 putative functional divergence events have occurred in the lineage leading to the ancestor of *BSg* and *BAp* and affecting FliK, FliM, and FlgK.

In conclusion, this work suggests that flagellar genes in endosymbiotic bacteria of insects belonging to the gamma-proteobacterium group seem to have undergone species-specific functional divergence events to adapt to the new environment and to become specialized in exporting proteins from the bacterium to the host. Our results however only support this hypothesis and do not definitively demonstrate such role. This work provides further support to the possible tight metabolic and biochemical communication between the endosymbiotic bacterium and its insect host. Further experimental work that targets specifically genes shown here to be under functional divergence (*fliK*, *fliM*, *fliN*, and *flgK*) may shed light on the veracity of these hypotheses.

Supplementary Material

Supplementary table S1 is available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This work was supported by a grant from the Irish Research Council for Science, Engineering, and Technology: funded by the National Development Plan to C.T. and a grant from Science Foundation Ireland to M.A.F. Finally, we would like to thank the reviewers of this manuscript for their important contribution in improving its quality.

Literature Cited

Akman L, Yamashita A, Watanabe H, Oshima K, Shiba T, Hattori M, Aksoy S. 2002. Genome sequence of the

endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*. Nat Genet. 32:402–407.
Aksoy S, Rio RV. 2005. Interactions among multiple genomes: tsetse, its symbionts and trypanosomes. Insect Biochem Mol Biol. 35:691–698.
Andersson SG, Zomorodipour A, Andersson JO, Sicheritz-Ponten T, Alsmark UC, Podowski RM, Naslund AK, Eriksson AS, Winkler HH, Kurland CG. 1998. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. Nature. 396:133–140.
Bren A, Eisenbach M. 1998. The N terminus of the flagellar switch protein, FliM, is the binding domain for the chemotactic response regulator, CheY. J Mol Biol. 278:507–514.
Brown PN, Mathews MA, Joss LA, Hill CP, Blair DF. 2005. Crystal structure of the flagellar rotor protein FliN from *Thermotoga maritima*. J Bacteriol. 187:2890–2902.
Fares MA, Ruiz-Gonzalez MX, Moya A, Elena SF, Barrio E. 2002. Endosymbiotic bacteria: groEL buffers against deleterious mutations. Nature. 417:398.
Funk DJ, Helbling L, Wernegreen JJ, Moran NA. 2000. Intraspecific phylogenetic congruence among multiple symbiont genomes. Proc Biol Sci. 267:2517–2521.
Funk DJ, Wernegreen JJ, Moran NA. 2001. Intraspecific variation in symbiont genomes: bottlenecks and the aphid-buchnera association. Genetics. 157:477–489.
Gil R, Sabater-Munoz B, Latorre A, Silva FJ, Moya A. 2002. Extreme genome reduction in *Buchnera* spp.: toward the minimal genome needed for symbiotic life. Proc Natl Acad Sci USA. 99:4454–4458.
Gil R, Silva FJ, Zientz E, et al. (13 co-authors). 2003. The genome sequence of *Blochmannia floridanus*: comparative analysis of reduced genomes. Proc Natl Acad Sci USA. 100:9388–9393.
Hirano T, Yamaguchi S, Oosawa K, Aizawa S. 1994. Roles of FliK and FlhB in determination of flagellar hook length in *Salmonella typhimurium*. J Bacteriol. 176:5439–5449.
Jones CJ, Macnab RM. 1990. Flagellar assembly in *Salmonella typhimurium*: analysis with temperature-sensitive mutants. J Bacteriol. 172:1327–1339.
Koroyasu S, Yamazato M, Hirano T, Aizawa SI. 1998. Kinetic analysis of the growth rate of the flagellar hook in *Salmonella typhimurium* by the population balance method. Biophys J. 74:436–443.
Kutsukake K, Iino T. 1994. Role of the FliA-FlgM regulatory system on the transcriptional control of the flagellar regulon and flagellar formation in *Salmonella typhimurium*. J Bacteriol. 176:3598–3605.
Kutsukake K, Ohya Y, Iino T. 1990. Transcriptional analysis of the flagellar regulon of *Salmonella typhimurium*. J Bacteriol. 172:741–747.
Macnab RM. 1996. Flagella and motility. In: Neidhardt FC, Curtis R III, Ingraham JL, Lin ECC, Low KB, Magasanik B, Reznikoff WS, Riley M, Schaechter M, Umbarger HE, editors. *Escherichia coli* and *Salmonella: cellular and molecular biology*. Washington (DC): ASM Press. p. 123–145.
Macnab RM. 2003. How bacteria assemble flagella. Annu Rev Microbiol. 57:77–100.
Maezawa K, Shigenobu S, Taniguchi H, Kubo T, Aizawa S, Morioka M. 2006. Hundreds of flagellar basal bodies cover the cell surface of the endosymbiotic bacterium *Buchnera aphidicola* sp. strain APS. J Bacteriol. 188:6539–6543.
Minamino T, Namba K. 2004. Self-assembly and type III protein export of the bacterial flagellum. J Mol Microbiol Biotechnol. 7:5–17.
Moran NA. 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. Proc Natl Acad Sci USA. 93:2873–2878.
Moran NA, Wernegreen JJ. 2000. Lifestyle evolution in symbiotic bacteria: insights from genomics. Trends Ecol Evol. 15:321–326.
Moya A, Pereto J, Gil R, Latorre A. 2008. Learning how to live together: genomic insights into prokaryote-animal symbioses. Nat Rev Genet. 9:218–229.
Muller HJ. 1964. The relation of recombination to mutational advance. Mutat Res. 106:2–9.
Nei M, Gojobori T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol. 3:418–426.
Ochman H, Moran NA. 2001. Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. Science. 292:1096–1099.
Ohtaka C, Ishikawa H. 1993. Accumulation of adenine and thymine in a groE-homologous operon of an intracellular symbiont. J Mol Evol. 36:121–126.
Perez-Brocal V, Gil R, Ramos S, Lamelas A, Postigo M, Michelena JM, Silva FJ, Moya A, Latorre A. 2006. A small microbial genome: the end of a long symbiotic relationship? Science. 314:312–313.
Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H. 2000. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. Nature. 407:81–86.
Stepkowski T, Legocki AB. 2001. Reduction of bacterial genome size and expansion resulting from obligate intracellular lifestyle and adaptation to soil habitat. Acta Biochim Pol. 48:367–381.
Tamas I, Klasson L, Canback B, Naslund AK, Eriksson AS, Wernegreen JJ, Sandstrom JP, Moran NA, Andersson SG. 2002. 50 million years of genomic stasis in endosymbiotic bacteria. Science. 296:2376–2379.
Tang H, Billings S, Wang X, Sharp L, Blair DF. 1995. Regulated underexpression and overexpression of the FliN protein of *Escherichia coli* and evidence for an interaction between FliN and FliM in the flagellar motor. J Bacteriol. 177:3496–3503.
Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22: 4673–4680.
van Ham RC, Kamerbeek J, Palacios C, et al. (16 co-authors). 2003. Reductive genome evolution in *Buchnera aphidicola*. Proc Natl Acad Sci USA. 100:581–586.
Wang S, Fleming RT, Westbrook EM, Matsumura P, McKay DB. 2006. Structure of the *Escherichia coli* FlhDC complex, a prokaryotic heteromeric regulator of transcription. J Mol Biol. 355:798–808.
Wu D, Daugherty SC, Van Aken SE, et al. (12 co-authors). 2006. Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. PLoS Biol. 4:e188.
Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci. 13:555–556.
Yang Z, Nielsen R. 2000. Estimating synonymous and non-synonymous substitution rates under realistic evolutionary models. Mol Biol Evol. 17:32–43.
Young GM, Schmiel DH, Miller VL. 1999. A new pathway for the secretion of virulence factors by bacteria: the flagellar export apparatus functions as a protein-secretion system. Proc Natl Acad Sci USA. 96:6456–6461.

Michele Vendruscolo, Associate Editor

Accepted July 7, 2008