

RESEARCH

Open Access



Adaptation to a host-associated lifestyle is associated with convergent loss of flagella-related genes in Pseudomonadota

Roberto Siani^{1,2*}, Caroline Gutjahr³ and Michael Schloter^{1,2}

Abstract

Background Host-associated bacteria must balance the benefits of motility through flagella against the offsets of energetic costs and immune surveillance. Understanding the interplay of evolutionary forces shaping this complex trait can provide insights in the dynamics and extent of within-host adaptations of flagellar assembly. We compared prevalence, redundancy, and homology of 55 known flagellar assembly genes across genomes of free-living and host-associated bacteria from a collection covering the entire Pseudomonadota phylum.

Results Our results indicate that host-associated bacteria have a lower prevalence and reduced sequence conservation of flagellar genes than free-living taxa. However, genome size explains substantially more variation than lifestyle classification, suggesting that genome erosion and genetic drift are the primary factors associated with flagellar gene loss. While host-associated taxa display heterogeneous patterns of sequence divergence across flagellar components, these patterns are consistent with relaxed selective constraint rather than pervasive adaptive or diversifying selection.

Conclusions Our results provide valuable insights into the distribution of flagellar genes in the phylum Pseudomonadota and its relation to bacterial lifestyle.

Keywords Host-associated bacteria, Flagella, Within-host adaptation, Flagellar phylogenomics, Genome reduction

Introduction

Bacteria lose, re-purpose, and acquire novel genes, in remarkably short periods of time when adapting to novel environments [1–3]. Microenvironments inside and outside host organisms house rich communities of host-associated microorganisms, including archaea, bacteria, fungi and viruses [4]. In contrast to free-living bacteria, mutations through genetic drift may become quickly fixed in host-associated bacteria, due to population bottlenecks and reduced genetic diversity [5]. The accumulation of neutral and deleterious mutations [6] can lead to loss or diversification of genes and, over time, degradation of the genome [7]. Survival within a host also requires adaptation to host control mechanisms (for

*Correspondence:

Roberto Siani
roberto.siani@tum.de

¹Chair of Environmental Microbiology, TUM School of Life Sciences, Technical University Munich, Freising 85354, Germany

²Research Unit for Comparative Microbiome Analysis, Helmholtz Munich, Oberschleissheim 85764, Germany

³Max Planck Institute of Molecular Plant Physiology, Potsdam 14476, Germany



© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

instance, the immune system), which are hypothesized to be a driver of evolutionary changes within host-associated microbiota [8].

The bacterial flagellum can facilitate host association by enhancing movement toward hosts and colonization sites [9]. Beyond its primary role in motility, flagella are also involved in mechanosensing, adhesion, chemotaxis, and complex behaviors such as colonization and biofilm formation [10–12]. Multiple studies have explored the origins of the bacterial flagellum. Identification of a conserved set of flagellar genes, suggested the flagellum arose through duplication and diversification of a single sequence [13]. However, larger phylogenetic analysis of flagellar genes challenged this view, proposing diverse, independent origins for flagellar genes and highlighting the role of horizontal gene transfer in their spread across bacterial phyla [14].

Whether from simple or complex origins, the bacterial flagellum is an evolutionary ancient and carefully constructed nano-machine: the reference assembly pathway maintained by the KEGG ontology includes 55 ortholog genes [15] encoding proteins with structural and regulatory roles. The pathway regulating production of flagellar building blocks and their assembly was firstly mapped in *Escherichia coli* and *Salmonella typhimurium*, and is organized in hierarchical layers and regulatory loops that orchestrate the timely expression of its downstream genes [16]. Briefly, the master regulator, FlhDC (or species-specific alternatives such as FlrA and FleQ), initiate σ^{70} (RpoD) or σ^{54} (RpoN) dependent transcription of tier II operons. These operons encode components necessary for the assembly of the flagellar type III secretion system (fT3SS), the M/S ring, the switch complex, and the MS-rod junction. Once the fT3SS is assembled, the anti-sigma factor FlgM is exported, allowing FlhDC, together with σ^{70}/σ^{54} and σ^{28} (the flagellar-specific sigma factor FliA) to drive the transcription of mixed tier II and III operons. These operons encode structural components of the basal-body and hook. Finally, σ^{28} -dependent transcription of tier III operons completes the process, leading to the assembly of the motor/stator and the filament (flagellin proteins, primarily FliC).

Despite the advantages of flagella for locomotion and biofilm formation, the assembly, operation and maintenance of flagella demand significant energy resources [17], compromising growth rates in nutrient-poor environments [18]. Besides, as flagella are associated with virulence in several pathogenic species [19–24], animal and plants have evolved specialized receptors to recognize the main constituent of the flagellum, the filament protein FliC, as a marker of potential pathogens [25–27]. Thus, in addition to the advantage of saving energy, counter-adaptations must balance the trade-off between retaining functional flagella while evading recognition

by the host immune system [28, 29]. In fact, in bacterial communities of the plant rhizosphere FliC has accumulated a number of polymorphisms, which play a potential role in immune evasion [30–32].

Apart from *fliC*, a broader pattern of flagellar gene mutations has been observed in various host-associated clades [33–35], to different effects. For instance, a study on convergent adaptations across 29 human pathogenic species identified the flagellar regulon as a prominent mutational hotspot [36], except for the filament protein, which was not mutated in all species. In fact, fine-tuning of flagellar function has been shown to predominantly arise from mutations at regulatory checkpoints, rather than at structural components [37]. On the other extreme, flagellar motility is entirely lost in aphid endosymbionts, for whom the process of degeneration likely started with loss of the regulatory genes *flhD* and *flhC* [38] and may have involved the transition of flagellar basal bodies into secretion systems [39, 40].

While adaptation to specific hosts is known to impact *fliC* and flagellar motility, how host-microbe interactions influence the full complement of flagellar genes is not understood. We conducted a phylogenomic analysis of Pseudomonadota and investigated whether the rates of gene loss differ between host-associated and free-living bacteria. We then assessed whether these patterns are consistent with genome erosion and heterogeneous selective constraints.

Method details

Pseudomonadota genome collection

1839 complete genomes of representatives Pseudomonadota (previously known as Proteobacteria) were retrieved from NCBI RefSeq [41] archives (date of accession: 14.10.2023) using NCBI datasets v14.1 [42]. Sequences from the Alpha-, Beta-, Gamma-proteobacteria were included as less than 10 genomes were available for the other Pseudomonadota classes. Sequences were decorated with one of two association states (host-associated, free-living), based on combined metadata from NCBI and BacDive [43] detailing the original or predominant conditions of isolation. Host-associated was operationally defined as sequences derived from samples obtained directly from a host organism [44, 45]. Contextual information deposited on NCBI was prioritized. When information on the isolation source was absent or conflicting, BacDive was consulted. Coding and translated amino-acid sequences were predicted de-novo with Prodigal v2.6.3 [46].

Phylogenomic tree reconstruction

A single-copy gene tree was constructed using GTóTree v1.8.4 [47] with the available hidden Markov model (HMM) set for Pseudomonadota, gene-hits exclusion

criteria of 0.33 and no filtering of genomes. Briefly, HMMER v3.2.2 [48] was used to identify 119 single-copy genes. Genes shorter than 0.67 x median length of the gene set or longer than 1.33 x median length of the gene set were considered spurious hits and removed. An individual alignment is produced for each gene set using muscle v5.1 [49]. The alignments were trimmed with trimal v1.4. rev15 [50] and concatenated. Phylogenetic estimation of the tree was conducted with FastTree2 v2.1.11 [51]. Protein alignments of the single-copy core genes were provided as input. Tree inference used BLOSUM45 amino-acid distances, balanced joining, and 1000 SH-like support replicates. The search strategy employed the default NNI and SPR optimizations (two SPR rounds with a radius of 10) and per-iteration ML-NNI optimization. FastTree's default "TopHits" heuristics were used (1.00x√N, refresh threshold 0.80). Model fitting applied the Jones–Taylor–Thornton (JTT) substitution model with CAT rate heterogeneity using 20 rate categories.

Identification of flagellar genes

GToTree optionally accepts a list of targets KEGG [52] ortholog (KO) families. Fifty-five KOs of orthologs belonging to the flagellar assembly pathway (flagellar) KO02040 were supplied as targets. GToTree sourced profile HMMs for the targets and used KOfamScan [53] to detect them in the genome collection, resulting in a count of number of gene copies per genome. KEGG profile HMMs used family-specific adaptive score threshold to control false positive rates.

For homology modelling, a methodology proposed by Wheeler and colleagues [54] was adapted. HMM log-odds bitscores captured goodness of fit of queries and targets and are provided as a metric of the quality of the match. Given that HMM models emphasize conserved regions in their scoring, differences in bitscores (delta-bitscore) between sequences were used as a measure of conservation. New HMM profiles were generated by aligning sequences detected by KOfamScan to their respective KOfam HMM profiles. The newly generated profiles were used to screen the collection using hmmer v3.4.

Phylogenetic analyses

All statistical analyses were conducted using R Statistical Software v4.3.2 [55] in RStudio v2023.12.0 + 369 [56], with the extensive use of tidyverse [57]. Firstly, the single-copy gene tree was coerced to ultrametric by extending all external edges using phytools v2.1-1 [58]. Multichotomies were resolved using the package ape v5.7.1 [59]. The tree was rooted at midpoint using phangorn v2.11-1 [60] and used for modelling phylogenetic correlation structures. Prevalence and number of copies per gene, as detected by KOfamScan, were modelled

using phylogenetic generalized (logistic and Poisson family) linear models from the package phylolm v2.6.2 [61]. For the bitscore modelling, a single best-hit per protein was retained for each genome. Bitscores were standardized and fit to a phylogenetic linear regression model from the package phylolm. In all three models, the logarithmic length of the genome was included as a covariate. For both the prevalence logistic model and the bitscore linear model, 999 independent bootstrap replicates were used to estimate 95% confidence intervals.

Results

Distribution of essential and accessory flagellar genes

To test whether the loss of flagella-associated genes differs between host-associated bacteria and free-living ones, we collected 1839 genomes including 599 Alpha-, 318 Beta-, and 922 Gamma-proteobacteria, spanning 597 genera of Pseudomonadota. Based on contextual metadata from NCBI and BacDive (see Methods for details), we classified 952 genomes (52%) as isolated from environmental sources (free-living bacteria) and 887 genomes (48%) as isolated from host-derived materials (host-associated bacteria), resulting in two comparably sized groups (Supplementary Data 1). Our collection-wide search of the 55 flagellar genes comprised in the reference KEGG flagellar assembly pathway retrieved matches for 53 of them, which were spread across the whole phylogeny. We did not find any homolog of 2 genes, coding for the flagellar biosynthesis fusion protein FliR/FlhB (K13820) and the flagellar motility protein FlgQ (K24346) which is described only in *Campylobacter jejuni* [62]. Among putatively motile bacteria, i.e. those with at least one copy of the flagellin-coding gene *fliC* ($n = 1234$ [63]), we identified three groups of genes that consistently, frequently, or rarely co-occur with *fliC* (Supplementary Data 2). Out of the genes retrieved in our collection, 14 occur in less than half of the *fliC*-positive bacteria (rare flagellar genes). Master regulators of the assembly pathway were found at surprisingly low prevalence (34% for *flhDC*, 36% for *flrA/flhE/flaK*). Other rare genes have regulatory functions (*flhE*, *flrC*, *fliZY*) or encode components of the H-ring and proteins associated with the outer membrane (*flgOPT*) and motor (*motCD*, *motXY*). Thirteen genes were found at moderate prevalence (accessory flagellar genes), between 50% and 80% of the *fliC*-positive bacterial genomes. These include the gene coding for the flagella-specific sigma-factor (*fliA*) and anti-sigma factor (*flgM*), for filament chaperones and cap (*fliDST*), for hook-filament junction, hook-length control and assembly chaperon (*fliK*, *flgL*, *flgN*), and for rod (*flgF*) proteins, the FT3SS ATPase, associated proteins (*fliHII*) and putative regulator/chaperone (*fliO/fliZ*). Finally, 25 protein-coding genes constituted a high prevalence cluster (core flagellar genes). This last group included genes

coding for non-flagellar specific sigma factors (*rpoN*, *rpoD*), most structural components and regulators of the σ^{54} (FliPQR, FliHAB), of the switch complex (FliLMN, FliFG), and basal body/hook (*flgA*, *flgBCDEGHIJK*, *fliE*), and the stator proteins (*motAB*).

Host association correlates with loss of flagellar genes

For both free-living and host-associated bacteria, larger genomes retained more detectable flagellar genes, with a significant association for 40 out of 53 genes. Host-association correlated with significantly lower prevalence of 12 of the 53 tested flagellar genes (*flgN*, *flgHIL*, *flhA*, *flhC*, *fliDS*, *fliHK*, *fliR*, *fliQ*), after accounting for genome size (Fig. 1, Supplementary Data 3). During preliminary analyses we explored the contribution of obligate endosymbionts to the loss of flagellar genes by explicitly modelling obligate endosymbionts as a separate category. However obligate endosymbiosis was strongly correlated with reduced genome size, which already accounted for a substantial proportion of the observed variation. For this reason, we retained genome size as a continuous covariate rather than introducing an additional categorical class.

We found that only *fliY*, was significantly more conserved across host-associated than in free-living bacteria. After accounting for phylogeny and genome size, host-association had a marginally positive effect on the conservation of the σ^{54} factor coding gene *rpoN*. In addition, we modelled dependence of number of gene copies on host-association using a phylogenetic Poisson regression (Supplementary Fig. 1, Supplementary Data 4). We observed that *fliC*, *fliL*, *motA* and *motB*, were often present in more than one copy (respectively, 53%, 49%, 60% and 53% of the times). For these four orthologs, our model predicted a significant and negative effect of host-association on number of copies.

Sequence divergence of flagellar genes in host-associated bacteria

We investigated the effect of host-association on the degree of conservation of the detected flagellar gene sequences by phylogenetic linear regression. Host-association is associated with significantly lower conservation of 22 out of 53 flagellar genes (Fig. 2, Supplementary Data 5). However, the effect of host-association, on average, had a ten times smaller estimated impact than genome size on the sequence conservation. This suggests that, also in this case, genetic drift was the primary correlate of sequence polymorphism differences. Host-association was linked to greater sequence divergence of sigma-factor encoding genes (*rpoD*, *rpoN* and *fliA*), and of 5 genes of the σ^{54} : *fliH*, *fliI*, *fliJ*, *fliQ* and *flhB*. Other affected genes are the alternative master regulator *flrA/fleQ/flaK*, the regulator *flrC*, the switch complex genes *fliF* and *fliL*,

the hook-filament junction gene *flgL*, the basal-body gene *flgP*, the filament cap gene *fliD*, the stator gene *motB* and the adaptor protein gene *fliE*, and the accessory motor genes *motXY*.

Discussion

For host-associated bacteria, loss of flagellar genes may help to evade immune detection while reallocating resources from motility to growth [11, 18]. To explore whether convergent adaptive dynamics influence the evolution of the flagellar assembly pathway in host-associated bacteria, we examined the relationship between lifestyle, genome size, and the conservation of flagellar genes across a large phylogenetically diverse dataset.

Among bacteria harbouring *fliC*, encoding the essential filament protein, we identified three broad clusters of flagellar orthologs: rare, accessory, and core genes. Key regulatory genes (*flhC*, *flhD*, *flrA/fleQ/flaK*) exhibited low prevalence. Transcriptional regulators are known to be more evolutionary flexible than their targets [64], a condition that preserves the overall network structure even when the regulatory components themselves may vary [65]. For instance, previous studies have shown that *Pseudomonas fluorescens* can reintegrate flagellar functionality by recruiting a nitrogen uptake regulator following inactivation of *fleQ* [66]. A set of 23 orthologs, including components of the σ^{54} , the switch complex, and parts of the basal body, hook, and stator, consistently co-occurred with *fliC*. The composition of this set of flagellar core genes aligns largely with prior studies [13].

Based on our comparative analysis, patterns of flagellar gene loss and sequence divergence in host-associated bacteria are consistent with genetic drift, resulting in differential retention and divergence relative to free-living taxa. This is evidenced by the near-complete absence of flagellar genes in several exclusively host-associated orders from our genomic collection, such as Pasteurellales, Oceanospirillales, and Rickettsiales. Even in orders containing both host-associated and free-living strains, we found that flagellar gene prevalence is consistently lower in host-associated bacteria. Notable examples include Enterobacterales, Legionellales, Rhodospirillales, and Pseudomonadales.

We also identified exceptions where free-living strains lack flagellar genes altogether (Moraxellales, Kangiellales, Acidiferrobacterales) or exhibit reduced prevalence compared to their host-associated relatives (Xanthomonadales, Aeromonadales). As our sample only included representative genomes, we cannot exclude that strain-level variability within these orders might produce some exceptions due to, for instance, horizontal gene transfer. Members of the order Moraxellales contain known opportunistic pathogens, and, at least in the case of the genus *Psychrobacter*, their evolution might have involved

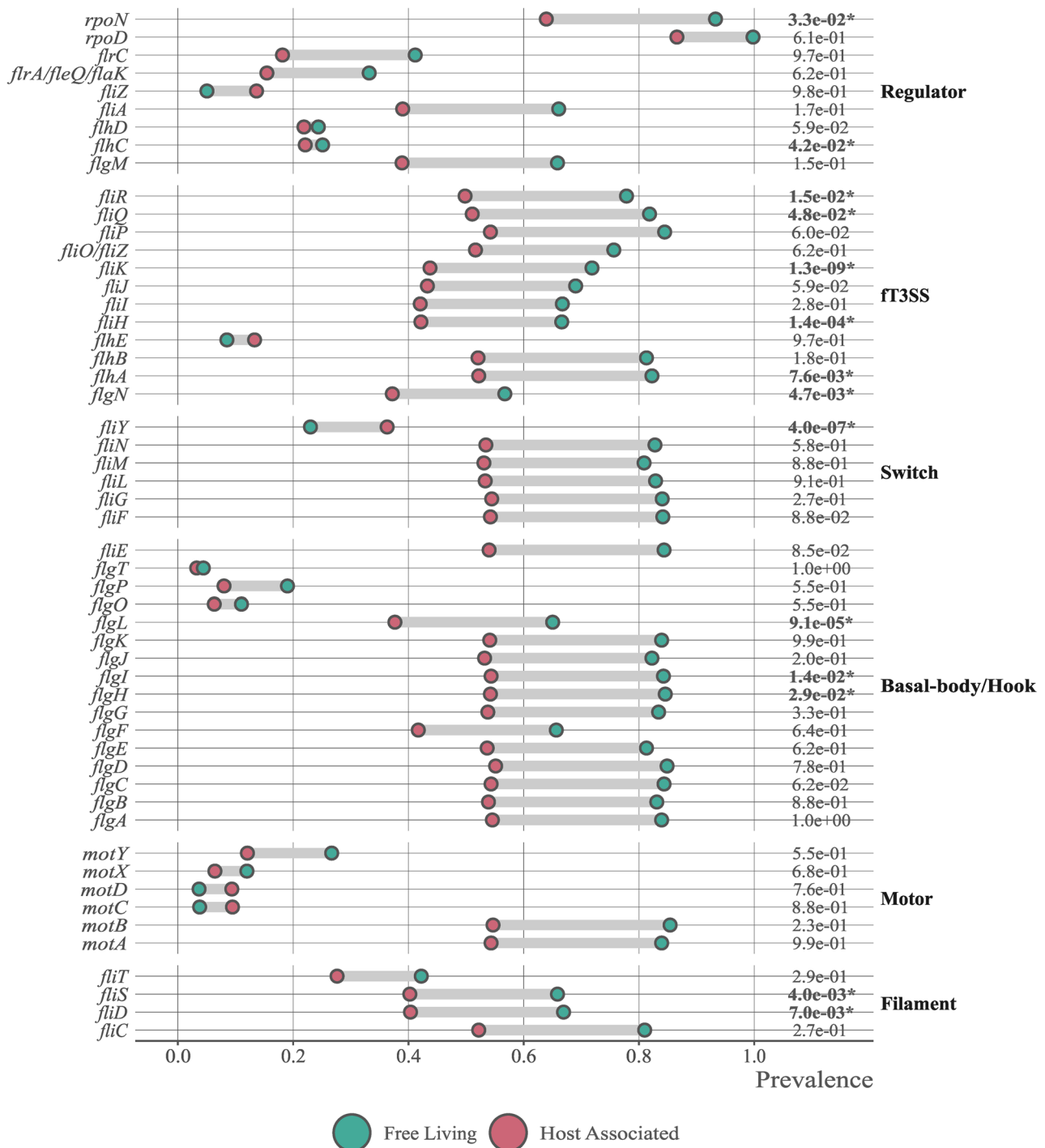


Fig. 1 Detected prevalence of flagellar genes in host-associated and free-living bacteria. Observed prevalence of each flagellar gene in genomes of free-living and host-associated bacteria (including endosymbionts). The genes were detected by homology to the reference KEGG orthologs families. In almost all cases, flagellar genes are more prevalent in free-living bacteria. False discovery rates from the phylogenetic logistic regression on the right side of the graph show the significance of the effects of host-association on the prevalence of each gene. FDRs lower than 0.05 are marked with an asterisk

phases of host-association [67]. Several species of marine bacteria have experienced extensive losses of genetic material, although the drivers of reduction are likely different than those suggested for host-associated bacteria [68]. The Kangiellales are represented in our collection by

4 genomes of marine bacteria, with a genome size ranging from 2.4 to 2.8 Mb (for reference, *Escherichia coli* K12 has a genome of 4.6 Mb).

Our models predict that both host-associated and free-living bacteria with smaller genomes are less likely to be

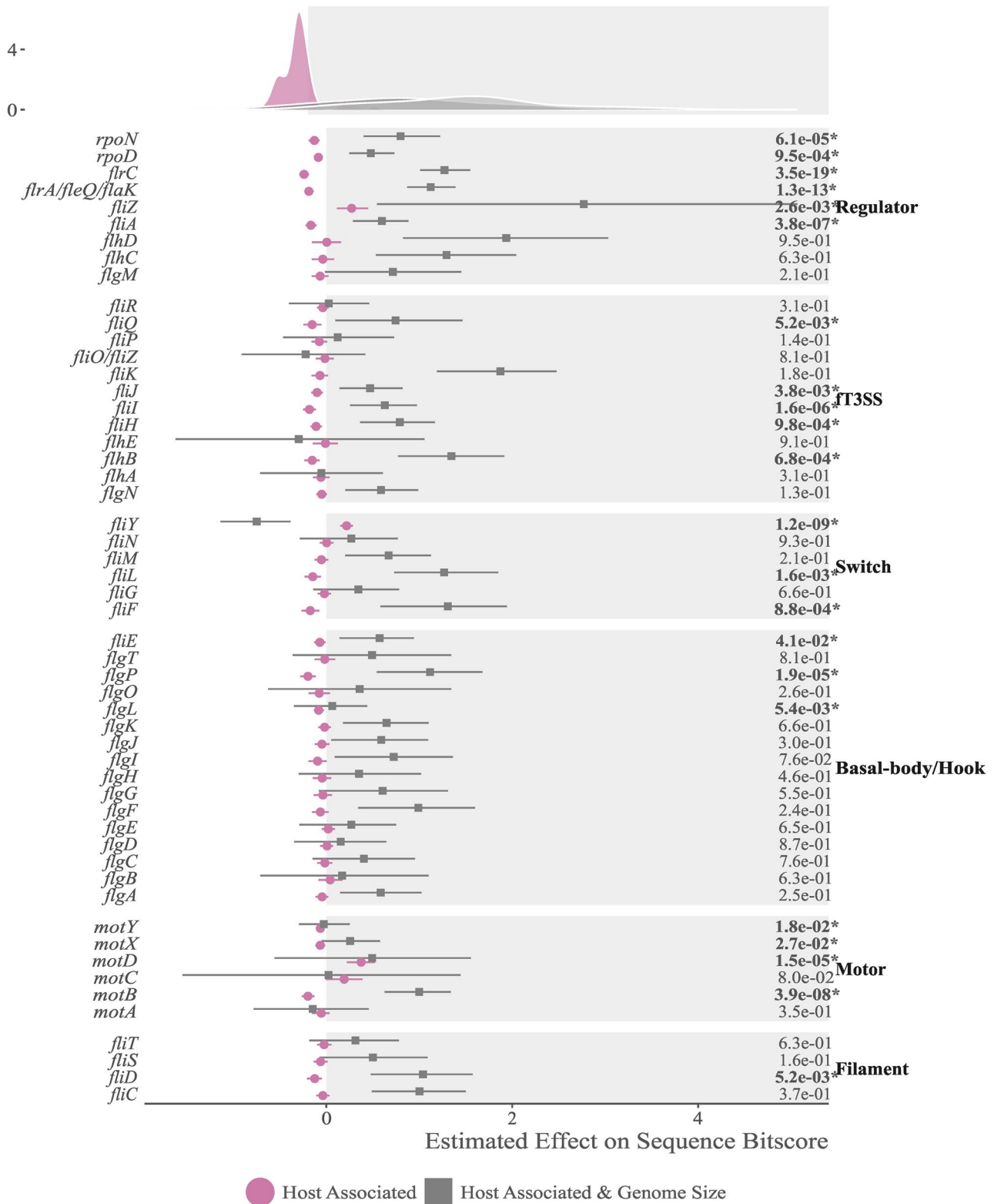


Fig. 2 Estimates of host-association effects on flagellar gene conservation. Estimated effects and standard error of host-association, genome size and their interaction on the standardized bitscores of each gene, calculated using a phylogenetic linear regression. Lower bitscores indicate lower sequence similarity and conservation. False discovery rates on the right side of graph show the significance of the effects of host-association, with FDRs lower than 0.05 marked with an asterisk. Density curves on top of the graph show the distribution of the effects for each term, highlighting the lower but consistently negative effects of host-association on sequence conservation

flagellated, underscoring the link between genome size and bacterial lifestyle. This does not only apply to marine bacteria, but also to [68] Acidiferrobacterales, which are autotrophic bacteria, able to oxidise sulphur and partly iron [69, 70]. Their ecological niche imposes a careful energetic budgeting and a predominantly non-motile lifestyle, as biofilms are fundamental for the leaching process [71].

Finally, the orders Xanthomonadales and Aeromonadales contain known plant and animal pathogenic and opportunistic species, for whom the evolution of flagellar genes is intertwined with the characteristic of their lifecycles. For instance, we detected no flagellar genes in the genome of *Xylella fastidiosa* Temecula1, which is an insect-borne plant pathogen. On the contrary, flagellar genes were commonly present in *Xanthomonas* spp., which is typically found in the phyllosphere, thus not transmitted by a symbiotic insect [72]. Flagellated host-associated bacteria that rely on motility for both survival outside the host and for host infection can adapt to evade the immune system [28, 73]. While some pathogenic species from the genus *Yersinia* have lost their flagellar genes, others regulate the expression of flagellar assembly in a temperature-dependent manner [74], enabling them to evade the immune system upon entering a mammalian host.

Our models link most differences in flagellar gene content to decreases in genome size, highlighting the role of genetic drift in the degeneration of the flagellar pathway. Importantly, while often considered neutral, genetic drift may include adaptive biases, as suggested by the positional and functional characteristics of lost genes [6]. In addition, host-association explains residual variation in both the prevalence and sequence conservation of flagellar genes after accounting for genome size and phylogeny, indicating lifestyle-associated differences in selective constraint. However, the present analyses do not allow us to distinguish relaxed purifying selection from positive selection. Given the strong association between reduced genome size and decreased gene conservation in our analyses, relaxed purifying selection and genetic drift provide a more parsimonious explanation.

Our findings are consistent with previous studies that reported higher *fliC* loss rates in host-associated bacteria [8], and extend these observations to the full flagellar assembly pathway. Together, these results support the generality of flagellar gene erosion in host-associated bacteria while underscoring the dominant influence of genome size and demographic processes [8]. Genes essential for filament elongation (*fliS*, *fliA*) were less prevalent in host-associated bacteria than in free-living bacteria. Furthermore, we found a link between host association and lower copy numbers of commonly supernumerary flagellar genes (*motAB*, *fliL*, *fliC*), and

increased diversification of a substantial fraction of flagellar orthologs. We found that host-associated bacteria disproportionately lose *fliC*, a mutation that would eliminate transcription of the entire flagellar regulon and might therefore predate the loss of *fliC*. However, the minimal difference in *fliDC* prevalence between lifestyles suggests that loss of the master regulator is likely part of, rather than a primary driver of, the stepwise loss of flagellar genes. They also exhibit diversification of *fliA*, which governs the expression of late-stage flagellar assembly genes and allows the fine-tuning of the motility phenotype, as suggested by a previous study on *Escherichia coli* [37]. In contrast, we did not observe direct effects of host-association on *flgM*, another regulatory gene proposed by the same study [37] to play a significant role in motility remodelling. Targeted in vivo validation would be necessary to assess whether the observed genomic patterns have functional consequences for motility or host interactions in host-associated bacteria. For example, loss of *fliC* has been reported to affect plant colonization in *Acidovorax* [75]. Interestingly, *fliY* is one of the few genes to be more prevalent in host-associated than in free-living bacteria, which hints at an unknown role in the context of associations. FliY is a phosphatase that is part of the switch complex, which plays a role in the transduction of chemotactic signals [76], and its inactivation results in impaired motility and virulence [22].

This study has limitations. First, lifestyle classifications derived from isolation metadata are inherently uncertain and necessarily coarse, and annotations reflect only the conditions at the time of sampling rather than the full ecological breadth of each lineage. Furthermore, the observed patterns might be disproportionately affected by the heterogeneity of association forms (for example, pathogenic species might have higher retention rates than obligate endosymbionts). Second, gene detection based on KEGG HMMs provides conservative presence/absence estimates and has limited sensitivity for poorly conserved flagellar components such as FliJ, FlgE, FliK, and FliD/ST. In addition, our pipeline does not incorporate synteny, gene-neighbourhood information, or remote homology searches, and therefore cannot distinguish complete flagellar systems from partial or degenerate remnants. Relatedly, polar and lateral flagellar systems are not resolved, and deep horizontal acquisition events cannot be identified with confidence. Finally, the evolutionary patterns reported here cannot fully separate genetic drift from relaxed selection, nor do they address potential neofunctionalization or adaptive diversification of retained genes. Future work combining targeted mutagenesis, experimental evolution, and structural or synteny-based annotation would help to refine and validate these inferences.

Conclusions

In host-associated bacteria, reduced reliance on motility in nutrient-rich and relatively stable host environments, combined with demographic processes that facilitate the fixation of mutations under genetic drift and relaxed selection, likely contributes to the extensive loss of flagellar genes. Our results describe broad evolutionary patterns consistent with genome erosion and heterogeneous constraint in host-associated lineages. The patterns described in this study underscore the evolutionary flexibility of bacterial systems and provide a basis for future work examining the molecular and ecological forces that shape microbial adaptation within host environments.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-026-12835-3>.

Supplementary Figure 1.

Supplementary Data 1.

Supplementary Data 2.

Supplementary Data 3.

Supplementary Data 4.

Supplementary Data 5.

Acknowledgements

We thank Karin Pritsch for her contribution to revising the manuscript.

Authors' contributions

R.S. designed the research study, conducted the analysis and wrote the manuscript; C.G. and M.S. contributed to the conceptualization of the study, supervised investigation and interpretation of results, edited and reviewed the manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL. This work was supported by the Deutsche Forschungsgemeinschaft (priority program SPP2125, project no. 401867691, awarded to MS and CG). CG was also supported by core funding by the Max Planck Society.

Data availability

Original code is publicly available at https://github.com/rsiani/phylogenomic_flagella as of the date of publication. The datasets supporting the conclusions of this article are available in the Zenodo Repository <https://doi.org/10.5281/zenodo.17679973>. RefSeq assembly accession numbers for all genomes analysed in this study are listed in Supplementary Data 1 and can be accessed via the NCBI Genome database (<https://www.ncbi.nlm.nih.gov/datasets/genome/>).

Declarations

Ethics approval and consent to participate

This study did not involve human participants, human data, or animal subjects. Ethics approval was therefore not required.

Competing interests

The authors declare no competing interests.

Received: 6 December 2025 / Accepted: 1 April 2026

Published online: 09 April 2026

References

- Sheppard SK, Guttman DS, Fitzgerald JR. Population genomics of bacterial host adaptation. *Nat Rev Genet.* 2018;19:549–65. <https://doi.org/10.1038/s41576-018-0032-z>.
- Didelot X, Walker AS, Peto TE, Crook DW, Wilson DJ. Within-host evolution of bacterial pathogens. *Nat Rev Microbiol.* 2016;14:150–62. <https://doi.org/10.1038/nrmicro.2015.13>.
- Toft C, Andersson SGE. Evolutionary microbial genomics: insights into bacterial host adaptation. *Nat Rev Genet.* 2010;11:465–75. <https://doi.org/10.1038/nrg2798>.
- Trevelline BK, Fontaine SS, Hartup BK, Kohl KD. Conservation biology needs a microbial renaissance: A call for the consideration of host-associated microbiota in wildlife management practices. *Proc Royal Soc B: Biol Sci.* 2019;286. <https://doi.org/10.1098/RSPB.2018.2448/79078>.
- Moran NA. Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. *Proceedings of the National Academy of Sciences.* 1996;93:2873–8. <https://doi.org/10.1073/pnas.93.7.2873>.
- Albalat R, Cañestro C. Evolution by gene loss. *Nat Rev Genet.* 2016;17:379–91. <https://doi.org/10.1038/nrg.2016.39>.
- Moran NA. Tracing the evolution of gene loss in obligate bacterial symbionts. *Curr Opin Microbiol.* 2003;6:512–8. <https://doi.org/10.1016/j.mib.2003.08.001>.
- Sharp C, Foster KR. Host control and the evolution of cooperation in host microbiomes. *Nat Commun.* 2022;13:3567. <https://doi.org/10.1038/s41467-022-30971-8>.
- Raina JB, Fernandez V, Lambert B, Stocker R, Seymour JR. The role of microbial motility and chemotaxis in symbiosis. *Nat Reviews Microbiol.* 2019. 2019;17:5. <https://doi.org/10.1038/s41579-019-0182-9>.
- Chaban B, Hughes HV, Beeby M. The flagellum in bacterial pathogens: For motility and a whole lot more. *Semin Cell Dev Biol.* 2015;46:91–103. <https://doi.org/10.1016/j.semcdb.2015.10.032>.
- Colin R, Ni B, Laganenka L, Sourjik V. Multiple functions of flagellar motility and chemotaxis in bacterial physiology. *FEMS Microbiol Rev.* 2021;45:fuab038. <https://doi.org/10.1093/femsre/fuab038>.
- Grognot M, Taute KM. More than propellers: how flagella shape bacterial motility behaviors. *Curr Opin Microbiol.* 2021;61:73–81. <https://doi.org/10.1016/j.mib.2021.02.005>.
- Liu R, Ochman H. Stepwise formation of the bacterial flagellar system. *Proc Natl Acad Sci.* 2007;104:7116–21. <https://doi.org/10.1073/pnas.0700266104>.
- Snyder LAS, Loman NJ, Fütterer K, Pallen MJ. Bacterial flagellar diversity and evolution: seek simplicity and distrust it? *Trends Microbiol.* 2009;17:1–5. <https://doi.org/10.1016/j.tim.2008.10.002>.
- KEGG PATHWAY. ko02040. <https://www.kegg.jp/entry/ko02040>. Accessed 20 Sep 2024.
- McCarter LL. Regulation of flagella. *Curr Opin Microbiol.* 2006;9:180–6. <https://doi.org/10.1016/j.mib.2006.02.001>.
- Schavemaker PE, Lynch M. Flagellar energy costs across the tree of life. *Elife.* 2022;11:e77266. <https://doi.org/10.7554/eLife.77266>.
- Shaping of microbial phenotypes by trade-offs. *Nat Commun.* <https://doi.org/10.1038/s41467-024-48591-9>.
- Kazmierczak BI, Schniederberend M, Jain R. Cross-regulation of Pseudomonas motility systems: the intimate relationship between flagella, pili and virulence. *Curr Opin Microbiol.* 2015;28:78–82. <https://doi.org/10.1016/j.mib.2015.07.017>.
- Yao J, Allen C. Chemotaxis Is Required for Virulence and Competitive Fitness of the Bacterial Wilt Pathogen *Ralstonia solanacearum*. *J Bacteriol.* 2006;188:3697–708. <https://doi.org/10.1128/JB.188.10.3697-3708.2006>.
- Sule P, Horne SM, Logue CM, Prüb BM. Regulation of Cell Division, Biofilm Formation, and Virulence by FlhC in *Escherichia coli* O157:H7 Grown on Meat. *Appl Environ Microbiol.* 2011;77:3653. <https://doi.org/10.1128/AEM.00069-11>.
- Liao S, Sun A, Ojcius DM, Wu S, Zhao J, Yan J. Inactivation of the flhY gene encoding a flagellar motor switch protein attenuates mobility and virulence of *Leptospira interrogans* strain Lai. *BMC Microbiol.* 2009;9:1–10.
- Guan W, Wang T, Huang Q, Zhao M, Tian E, Liu Y et al. Transcriptomic and Functional Analyses Reveal Roles of AcrR, a LuxR-type Global Regulator, in Regulating Motility and Virulence of *Acidovorax citrulli*. <https://doi.org/10.1094/MPMI-01-21-0020-R>. 2021. <https://doi.org/10.1094/MPMI-01-21-0020-R>.
- Guan W, Wang T, Huang Q, Tian E, Liu B, Yang Y, et al. A LuxR-type regulator, AcrR, regulates flagellar assembly and contributes to virulence, motility, biofilm formation, and growth ability of *Acidovorax citrulli*. *Mol Plant Pathol.* 2020;21:489–501. <https://doi.org/10.1111/mp.12910>.

25. Gómez-Gómez L, Boller T. Flagellin perception: a paradigm for innate immunity. *Trends Plant Sci.* 2002;7:251–6. [https://doi.org/10.1016/S1360-1385\(02\)2261-6](https://doi.org/10.1016/S1360-1385(02)2261-6).
26. Fulde M, Sommer F, Chassaing B, van Vorst K, Dupont A, Hensel M, et al. Neonatal selection by Toll-like receptor 5 influences long-term gut microbiota composition. *Nat*. 2018;560:7719. <https://doi.org/10.1038/s41586-018-0395-5>.
27. Cullender TC, Chassaing B, Janzon A, Kumar K, Muller CE, Werner JJ, et al. Innate and Adaptive Immunity Interact to Quench Microbiome Flagellar Motility in the Gut. *Cell Host Microbe.* 2013;14:571–81. <https://doi.org/10.1016/J.CHOM.2013.10.009>.
28. Rossez Y, Wolfson EB, Holmes A, Gally DL, Holden NJ. Bacterial Flagella: Twist and Stick, or Dodge across the Kingdoms. *PLoS Pathog.* 2015;11:e1004483. <https://doi.org/10.1371/JOURNAL.PPAT.1004483>.
29. De Maayer P, Pillay T, Coutinho TA. Flagella by numbers: Comparative genomic analysis of the supernumerary flagellar systems among the Enterobacteriales. *BMC Genomics.* 2020;21:1–16. <https://doi.org/10.1186/S12864-020-07085-W/FIGURES/7>.
30. Colaianni NR, Parys K, Lee HS, Conway JM, Kim NH, Edelbacher N, et al. A complex immune response to flagellin epitope variation in commensal communities. *Cell Host Microbe.* 2021;29:635–e6499. <https://doi.org/10.1016/J.CHO.2021.02.006/ATTACHMENT/51272F0C-7896-4386-9416-F3E0341DA070/MMC7.PDF>.
31. Parys K, Colaianni NR, Lee HS, Hohmann U, Edelbacher N, Trgovcevic A, et al. Signatures of antagonistic pleiotropy in a bacterial flagellin epitope. *Cell Host Microbe.* 2021;29:620–e6349. <https://doi.org/10.1016/J.CHOM.2021.02.008/ATTACHMENT/BC301CFC-1EC4-4024-AE8D-C55397E8FE8B/MMC2.XLSX>.
32. Hu D, Reeves PR. The Remarkable Dual-Level Diversity of Prokaryotic Flagellins. *mSystems.* 2020. <https://doi.org/10.1128/MSYSTEMS.00705-19>.
33. Schön ME, Martijn J, Vosseberg J, Köstlbacher S, Ettema TJG. The evolutionary origin of host association in the Rickettsiales. *Nat Microbiol.* 2022. <https://doi.org/10.1038/s41564-022-01169-x>.
34. George EE, Husnik F, Tashyreva D, Prokopchuk G, Horák A, Kwong WK, et al. Highly Reduced Genomes of Protist Endosymbionts Show Evolutionary Convergence. *Curr Biol.* 2020;30:925–e9333. <https://doi.org/10.1016/J.CUB.2019.12.070/ATTACHMENT/458B075A-76EF-4C77-9A21-A09968463DCC/MMC2.PDF>.
35. Zhu S, Sun X, Li Y, Feng X, Gao B. The common origin and degenerative evolution of flagella in Actinobacteria. *mBio.* 2023;14:e02526–23. <https://doi.org/10.1128/mbio.02526-23>.
36. Gatt YE, Margalit H. Common Adaptive Strategies Underlie Within-Host Evolution of Bacterial Pathogens. *Mol Biol Evol.* 2020;38:1101. <https://doi.org/10.1093/MOLBEV/MSAA278>.
37. Ni B, Ghosh B, Paldy FS, Colin R, Heimerl T, Sourjik V. Evolutionary Remodeling of Bacterial Motility Checkpoint Control. *Cell Rep.* 2017;18:866–77. <https://doi.org/10.1016/j.celrep.2016.12.088>.
38. Toft C, Fares MA. The Evolution of the Flagellar Assembly Pathway in Endosymbiotic Bacterial Genomes. *Mol Biol Evol.* 2008;25:2069–76. <https://doi.org/10.1093/molbev/msn153>.
39. Maezawa K, Shigenobu S, Taniguchi H, Kubo T, Aizawa S, Morioka M. Hundreds of Flagellar Basal Bodies Cover the Cell Surface of the Endosymbiotic Bacterium *Buchnera aphidicola* sp. Strain APS. *J Bacteriol.* 2006;188:6539. <https://doi.org/10.1128/JB.00561-06>.
40. Abby SS, Rocha EPC. The Non-Flagellar Type III Secretion System Evolved from the Bacterial Flagellum and Diversified into Host-Cell Adapted Systems. *PLoS Genet.* 2012;8:e1002983. <https://doi.org/10.1371/journal.pgen.1002983>.
41. Pruitt KD, Tatusova T, Maglott DR. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* 2005;33(suppl1):D501–4. <https://doi.org/10.1093/nar/gki025>.
42. O'Leary NA, Cox E, Holmes JB, et al. Exploring and retrieving sequence and metadata for species across the tree of life with NCBI Datasets. *Sci Data.* 2024;11:732. <https://doi.org/10.1038/s41597-024-03571-y>.
43. Reimer LC, Sardà Carbasse J, Koblitz J, Ebeling C, Podstawa A, Overmann J. BacDive in 2022: the knowledge base for standardized bacterial and archaeal data. *Nucleic Acids Res.* 2022;50:D741–6. <https://doi.org/10.1093/nar/gkab961>.
44. Field D, Amaral-Zettler L, Cochrane G, Cole JR, Dawyndt P, Garrity GM, et al. The Genomic Standards Consortium. *PLoS Biol.* 2011;9:e1001088. <https://doi.org/10.1371/JOURNAL.PBIO.1001088>.
45. Eloe-Fadrosh EA, Mungall CJ, Miller MA, Smith M, Patil SS, Keliher JM, et al. A Practical Approach to Using the Genomic Standards Consortium MixS Reporting Standard for Comparative Genomics and Metagenomics. *Methods Mol Biol.* 2024;2802:587–609. https://doi.org/10.1007/978-1-0716-3838-5_20/FIGURES/5.
46. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal. Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics.* 2010;11:119. <https://doi.org/10.1186/1471-2105-11-119>.
47. Lee MD. GTOtree: a user-friendly workflow for phylogenomics. *Bioinformatics.* 2019;35:4162–4. <https://doi.org/10.1093/bioinformatics/btz188>.
48. Eddy SR. A new generation of homology search tools based on probabilistic inference. *Genome Inf.* 2009;23:205–11. https://doi.org/10.1142/9781848165632_0019.
49. Edgar RC. Muscle5: High-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. *Nat Commun.* 2022;13:6968. <https://doi.org/10.1038/s41467-022-34630-w>.
50. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics.* 2009;25:1972–3. <https://doi.org/10.1093/bioinformatics/btp348>.
51. Price MN, Dehal PS, Arkin AP. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS ONE.* 2010;5:e9490. <https://doi.org/10.1371/journal.pone.0009490>.
52. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 2016;44:D457–462. <https://doi.org/10.1093/nar/gkv1070>.
53. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, et al. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics.* 2020;36:2251–2. <https://doi.org/10.1093/bioinformatics/btz859>.
54. Wheeler NE, Barquist L, Kingsley RA, Gardner PP. A profile-based method for identifying functional divergence of orthologous genes in bacterial genomes. *Bioinformatics.* 2016;32:3566–74. <https://doi.org/10.1093/bioinformatics/btw518>.
55. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2024.
56. Posit team. RStudio: Integrated Development Environment for R. Boston, MA: Posit Software, PBC; 2023.
57. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, et al. Welcome to the tidyverse. *J Open Source Softw.* 2019;4:1686. <https://doi.org/10.21105/joss.01686>.
58. Revell LJ. phytools 2.0: an updated R ecosystem for phylogenetic comparative methods (and other things). *PeerJ.* 2024;12:e16505. <https://doi.org/10.7717/peerj.16505>.
59. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics.* 2019;35:526–8. <https://doi.org/10.1093/bioinformatics/bty633>.
60. Schliep K, Potts JA, Morrison AD, et al. Intertwining phylogenetic trees and networks. *Methods Ecol Evol.* 2017;8:1212–20.
61. Ho LST, Ane C. A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. *Syst Biol.* 2014;63:397–408.
62. Sommerlad SM, Hendrixson DR. Analysis of the Roles of FlgP and FlgQ in Flagellar Motility of *Campylobacter jejuni*. *J Bacteriol.* 2007;189:179–86. <https://doi.org/10.1128/jb.01199-06>.
63. Philip JS, Grewal S, Scadden J, Puente-Lelievre C, Matzke NJ, McNally L, et al. Mapping the loss of flagellar motility across the tree of life. *ISME J.* 2025;19:111. <https://doi.org/10.1093/ISMEJO/WRAF111>.
64. Madan Babu M, Teichmann SA, Aravind L. Evolutionary Dynamics of Prokaryotic Transcriptional Regulatory Networks. *J Mol Biol.* 2006;358:614–33. <https://doi.org/10.1016/J.JMB.2006.02.019>.
65. Lozada-Chávez I, Janga SC, Collado-Vides J. Bacterial regulatory networks are extremely flexible in evolution. *Nucleic Acids Res.* 2006;34:3434. <https://doi.org/10.1093/NAR/GKL423>.
66. Taylor TB, Mulley G, Dills AH, Alsohim AS, McGuffin LJ, Studholme DJ et al. Evolutionary resurrection of flagellar motility via rewiring of the nitrogen regulation system. *Science (1979).* 2015;347:1014–7. <https://doi.org/10.1126/science.1259145>.
67. Welter DK, Ruaud A, Henseler ZM, De Jong HN, van Coeverden de Groot P, Michaux J, et al. Free-Living, Psychrotrophic Bacteria of the Genus *Psychrobacter* Are Descendants of Pathobionts. *mSystems.* 2021;6. https://doi.org/10.1128/MSYSTEMS.00258-21/SUPPL_FILE/MSYSTEMS.00258-21-ST004.XLSX.
68. Batut B, Knibbe C, Marais G, Daubin V. Reductive genome evolution at both ends of the bacterial population size spectrum. *Nat Rev Microbiol.* 2014;12:841–50. <https://doi.org/10.1038/nrmicro3331>.

69. Umezawa K, Watanabe T, Miura A, Kojima H, Fukui M. The complete genome sequences of sulfur-oxidizing Gammaproteobacteria *Sulfurifustis variabilis* skN76T and *Sulfuricaulis limicola* HA5T. *Stand Genomic Sci.* 2016;11:1–8. <https://doi.org/10.1186/S40793-016-0196-0/FIGURES/3>.
70. Hallberg KB, Hedrich S, Johnson DB. *Acidiferrobacter thiooxydans*, gen. nov. sp. nov.; an acidophilic, thermo-tolerant, facultatively anaerobic iron- and sulfur-oxidizer of the family Ectothiorhodospiraceae. *Extremophiles.* 2011;15:271–9. <https://doi.org/10.1007/s00792-011-0359-2>.
71. Issotta F, Moya-Beltrán A, Mena C, Covarrubias PC, Thyssen C, Bellenberg S, et al. Insights into the biology of acidophilic members of the Acidiferrobacteraceae family derived from comparative genomic analyses. *Res Microbiol.* 2018;169:608–17. <https://doi.org/10.1016/J.RESMIC.2018.08.001>.
72. An SQ, Potnis N, Dow M, Vorhölter FJ, He YQ, Becker A, et al. Mechanistic insights into host adaptation, virulence and epidemiology of the phyto-pathogen *Xanthomonas*. *FEMS Microbiol Rev.* 2020;44:1–32. <https://doi.org/10.1093/FEMSRE/FUZ024>.
73. Tsois RM, Young GM, Solnick JV, Bäumlér AJ. From bench to bedside: stealth of enteroinvasive pathogens. *Nat Reviews Microbiol* 2008. 2008;6:12. <https://doi.org/10.1038/nrmicro2012>.
74. Minnich SA, Rohde HN. A Rationale for Repression and/or Loss of Motility by Pathogenic *Yersinia* in the Mammalian Host. *Adv Exp Med Biol.* 2007;603:298–311. https://doi.org/10.1007/978-0-387-72124-8_27.
75. Siani R, Si Y, Thaqi SK, Stabl G, Gutjahr C, Schlöter M. Loss of the flagellar regulator FlhC changes the transcriptional response of plant-associated *Acidovorax delafieldii* strains to metabolites from *Rhizophagus irregularis*-colonized *Lotus japonicus* roots. *ISME Commun.* 2026. <https://doi.org/10.1093/ISMECO/YCAF235>.
76. Szurmant H, Muff TJ, Ordal GW. *Bacillus subtilis* CheC and FliY Are Members of a Novel Class of CheY-P-hydrolyzing Proteins in the Chemotactic Signal Transduction Cascade *. *J Biol Chem.* 2004;279:21787–92. <https://doi.org/10.1074/jbc.M311497200>.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.