MICROBIAL GENOMICS

RESEARCH ARTICLE

Quiñonero-Coronel *et al.*, *Microbial Genomics* 2025;11:001409

DOI 10.1099/mgen.0.001409

MICROBIOLOGY SOCIETY

OPEN DATA OPEN ACCESS

The type IV secretion system of *Patescibacteria* is homologous to the bacterial monoderm conjugation machinery

María del Mar Quiñonero-Coronel[1], Pedro J. Cabello-Yeves[2,3], Jose M. Haro-Moreno[3], Francisco Rodriguez-Valera[3] and M. Pilar Garcillán-Barcia[1,*]

Abstract

The Candidate Phyla Radiation, also known as *Patescibacteria*, represents a vast and diverse division of bacteria that has come to light via culture-independent 'omics' technologies. Their limited biosynthetic capacity, along with evidence of their growth as obligate epibionts on other bacteria, suggests a broad reliance on host organisms for their survival. Nevertheless, our understanding of the molecular mechanisms governing their metabolism and lifestyle remains limited. The type IV secretion system (T4SS) represents a superfamily of translocation systems with a wide range of functional roles. T4SS genes have been identified in the *Patescibacteria* class *Saccharimonadia* as essential for their epibiotic growth. In this study, we used a comprehensive bioinformatics approach to investigate the diversity and distribution of T4SS within *Patescibacteria*. The phylogenetic analysis of the T4SS signature protein VirB4 suggests that most of these proteins cluster into a distinct monophyletic group with a shared ancestry to the MPF_{FATA} class of T4SS. This class is found in the conjugative elements of *Firmicutes*, *Actinobacteria*, *Tenericutes* and *Archaea*, indicating a possible horizontal gene transfer from these monoderm micro-organisms to *Patescibacteria*. We identified additional T4SS components near *virB4*, particularly those associated with the MPF_{FATA} class, as well as homologues of other T4SS classes, such as VirB2-like pilins, and observed their varied arrangements across different *Patescibacteria* classes. The absence of a relaxase in most of these T4SS clusters suggests that the system has been co-opted for other functions in *Patescibacteria*. The proximity of T4SS components to the origin of replication (gene *dnaA*) in some *Patescibacteria* suggests a potential mechanism for increased expression. The broad ubiquity of a phylogenetically distinct T4SS, combined with its chromosomal location, underscores the significance of T4SS in the biology of *Patescibacteria*.

Impact Statement

The Candidate Phyla Radiation, or *Patescibacteria*, represents a highly diverse bacterial group that constitutes a significant fraction of the microbial dark matter. Known for their minimal biosynthetic capabilities and dependence on host bacteria, this phylum exemplifies fascinating yet poorly understood lifestyles. Our research sheds light on the molecular strategies underlying their survival by focusing on the type IV secretion system (T4SS), a versatile bacterial machinery typically involved in DNA transfer and effector molecule delivery. We revealed that T4SS is not only widespread in *Patescibacteria* but also forms a distinct evolutionary lineage, likely acquired via horizontal gene transfer from Gram-positive bacteria, and we illustrated its gene organization across different *Patescibacteria* classes. Unlike canonical systems, *Patescibacteria* T4SS lacks key compo-

[*Continued on next page*]

Received 03 January 2025; Accepted 01 April 2025; Published 23 May 2025

Author affiliations: [1]Instituto de Biomedicina y Biotecnología de Cantabria (IBBTEC), Consejo Superior de Investigaciones Científicas-Universidad de Cantabria, Santander, Spain; [2]School of Life Sciences, University of Warwick, Coventry, UK; [3]Evolutionary Genomics Group, Departamento Producción Vegetal y Microbiología, Universidad Miguel Hernández, Alicante, Spain.

***Correspondence:** M. Pilar Garcillán-Barcia, maria.garcillan@csic.es; garcilmp@unican.es

Keywords: *Patescibacteria*; Candidate Phyla Radiation; horizontal gene transfer; type IV secretion system.

Abbreviations: CPR, Candidate Phyla Radiation; GTDB, Genome Taxonomy Database; HMM, Hidden Markov model; ICEs, integrative and conjugative elements; MPF, mating pair formation; SH-aLRT, SH-like approximate likelihood ratio test; T4SS, type IV secretion system; UFBoot, ultrafast bootstrap.

nents for conjugation, hinting at a novel functional adaptation. Remarkably, T4SS genes in *Patescibacteria* are often located near the replication origin, suggesting a regulatory role linked to their expression. This research suggests the repurposing of T4SS in this phylum to thrive under extreme reliance on other organisms. By unravelling the genomic and evolutionary peculiarities of T4SS in *Patescibacteria*, our work provides valuable insights into their unique biology, broadens our understanding of microbial diversity and innovation and highlights this system as a strong candidate to sustain their parasitic episymbiotic lifestyle.

DATA SUMMARY

The authors confirm all supporting data and protocols have been provided either within the article, in supplementary data files, or through supporting data files available on GitHub (https://github.com/mdmqc/Patescibacteria_T4SS).

INTRODUCTION

Type IV secretion systems (T4SSs) are multiprotein nanomachines present in both Gram-negative and Gram-positive bacteria. They span the entire bacterial cell envelope and function as a one-step mating pair formation (MPF) system, capable of delivering effector molecules directly to the cytosol of a target cell, typically requiring direct cell-to-cell contact [1]. T4SSs are mainly involved in bacterial conjugation and effector secretion from bacteria to eukaryotic cells [2]. Bacterial conjugation is one of the most prominent mechanisms of horizontal gene transfer. It involves the transport of a nucleoprotein effector from donor to recipient bacteria [3]. Meanwhile, T4SSs devoted to protein delivery are mainly deployed by intracellular bacterial pathogens to hijack eukaryotic host cell processes, gain access to nutrients and regulate specific virulence programmes to ensure their survival in the host [4]. While T4SSs known to be involved specifically in protein delivery have been mostly found in diderms, only a few reports show the presence of T4SS protein translocators in monoderms, more specifically in streptococci [5–7]. More recent studies, however, highlight a broader functional diversification of the T4SS. T4SS-mediated toxin export to other bacteria pointed to an additional role of T4SS in killing other bacteria [8–11]. Moreover, there are a few examples of T4SSs capable of importing or exporting DNA from or to the environment [12–14]. Furthermore, some reduced T4SS variants have been described in *Archaea,* the Ced system in *Sulfolobales* and the Ted system in *Thermoproteales*, mediating the unidirectional import of DNA between archaeal cells, which is then used as a template for genome repair by homologous recombination [15–17].

Understanding the role of T4SS in different bacterial lineages is crucial for uncovering the main aspects of their physiology and cellular biology. However, in other bacterial groups outside of *Proteobacteria* and *Firmicutes*, T4SSs are poorly characterized. One of these underexplored groups is the Candidate Phyla Radiation (CPR), which encompasses a strikingly diverse collection of ultra-small bacteria present in a variety of habitats with streamlined genomes [18]. This bacterial division was placed as a member of the *Terrabacteria* clade and a sister lineage to the *Chloroflexota* and *Dormibacterota* branches, suggesting that CPR evolved by reductive genome evolution from free-living ancestors [19]. Initially considered a superphylum based on 16S rRNA gene phylogeny [18], CPR has been reclassified as a single phylum, *Candidatus Patescibacteria*, based on a phylogeny inferred from the concatenation of 120 ubiquitous single-copy proteins [20].

Members of this phylum are typically characterized by extremely limited biosynthetic capacities, including the inability to synthesize amino acids, lipids, vitamins and, in many cases, nucleotides [21–23]. Additionally, they lack genes for the citric acid cycle and oxidative phosphorylation, except for the ATPase complex. As a result, a symbiotic dependence for cellular growth has been assumed, although free-living and particle-associated groups have been detected [24]. Most knowledge on *Patescibacteria* has been obtained through DNA-based, culture-independent methods because they are generally recalcitrant to traditional cultivation techniques. However, a few strains, primarily from the class *Saccharimonadia* [25–31], as well as a few from classes *Gracilibacteria* [32, 33], *JAEDAM01* [34] and *Paceibacteria* [35], have been successfully cocultured with a compatible bacterial or archaeal host. Additionally, *Patescibacteria* growing attached to host bacterial surfaces in an epibiotic association has been reported, such as the parasitic relationship between *Nanosynbacter* species and various *Actinobacteria* [25, 31], as well as between *Candidatus Vampirococcus lugosii*, a member of class *JAEDAM01*, and an anoxygenic photosynthetic gammaproteobacterium [33].

The presence of a putative T4SS in *Patescibacteria* was first reported in the parasitic *Saccharimonadia Ca. Nanosynbacter lyticus* strain TM7x [36] and later in *Ca. Southlakia epibionticum* strain ML1 [37]. In this latter strain, a transposon-insertion sequencing genome-wide screen revealed that the T4SS cluster was essential for its epibiotic growth [37]. These T4SSs have also been identified in *Patescibacteria* metagenome-assembled genomes [38] and complete genomes from Lake Baikal [39], the largest and deepest freshwater lake on Earth. This study was particularly motivated by the enigmatic and still poorly understood biology and ecological significance of this lineage, which colonizes most habitats on Earth. Here, we conducted a phylum-wide analysis of the presence and diversity of T4SS in *Patescibacteria* spanning different habitats and identified the main peculiarities derived from this secretion system.

METHODS

Dataset

Assemblies from the *Patescibacteria* group (4,699) were downloaded from the NCBI RefSeq database (version 212, March 2023), and only those confirmed as members of *Patescibacteria* in the Genome Taxonomy Database (GTDB) (release 214) [40] (3,024 assemblies) were kept. To be included in GTDB, genomes must meet a CheckM completeness estimate of >50%, a contamination estimate of <10% and a quality score of >50 (defined as completeness – 5*contamination), ensuring the use of draft-quality genomes. Additionally, *Ca. S. epibionticum* strain ML1 (NZ_CP124550.1) and *Ca. N. lyticus* strain TM7x (NZ_CP007496.1), for which the presence of a T4SS was reported [36, 37], were included. Furthermore, nine complete *Patescibacteria* genomes isolated from Lake Baikal [39] were also included (PRJNA924152). The final dataset contained 3,035 *Patescibacteria* genomes (Table S1 and Fig. S1, available in the online Supplementary Material).

Detection of T4SS components

Hidden Markov model (HMM) profiles of the protein components of the eight phylogenetic classes of T4SS, as described by [41] and available at https://github.com/macsy-models/CONJScan/tree/main/profiles, as well as the Pfam HMM profiles TrbC/VirB2 (PF04956.16) and T4SS_pilin (PF18895.3), were used to inspect the *Patescibacteria* protein dataset. Searches were conducted using the *hmmscan* function of HMMER 3.1b2 [42], with an *i*-Evalue < 0.001 and HMM profile alignment coverage >50% as inclusion criteria. The presence of relaxases associated with T4SS was checked using MOBscan [43].

VirB4 phylogeny

The VirB4 homologues recovered from the *Patescibacteria* dataset were used for the phylogenetic reconstruction. Additionally, 262 VirB4-like proteins from bacterial conjugative plasmids and integrative and conjugative elements (ICEs) encoding a T4SS of the MPF_{FATA} (227), MPF_{FA} (5 proteins), MPF_{T} (5), MPF_{F} (5), MPF_{G} (5), MPF_{C} (5) and MPF_{B} (5) classes were included (Table S2). Furthermore, 12 VirB4-like proteins from *Archaea* were included (5 from the Ced system, 4 from the Ted system and 3 from conjugative archaeal plasmids belonging to MPF_{FATA}) (Table S2). Proteins were aligned using MAFFT v7.310 with the L-INS-i method [44], and the alignment was curated with Trimal 1.2rev59 (option -automated1) [45]. A maximum likelihood phylogenetic reconstruction was generated using IQ-TREE version 1.6.1 [46], based on the substitution model LG+F+R10, which was selected as the best fit according to the Bayesian information criterion provided by ModelFinder [47]. Branch support was estimated with ultrafast bootstrap (UFBoot) approximation [48] and SH-like approximate likelihood ratio test (SH-aLRT) [49], using 1,000 replicates in both cases. The resulting tree was midpoint rooted. It was visualized and annotated using the online tool iTOL [50]. We computed the patristic distance matrix from the tree using the cophenetic.phylo function in the *ape* package (v5.3) for R [51].

Analysis of the *virB4* gene neighbourhood

Coding sequences (CDSs) surrounding the *virB4* gene (20 CDSs upstream and 20 downstream) in the *Patescibacteria* assemblies were extracted whenever a second MPF component was identified. Entrez Protein Family Models, including 12,592 NCBIFAM and 4,486 TIGRFAM profiles (available at https://ftp.ncbi.nlm.nih.gov/hmm/16.0/ [52]), as well as 19,632 protein family models of Pfam-A (available at http://ftp.ebi.ac.uk/pub/databases/Pfam/releases/Pfam35.0/), were used to query the *Patescibacteria* proteins with the HMMER 3.1b2 *hmmscan* function (parameters -E 0.001 --domE 0.001 --incE 0.001 --incdomE 0.001) [42]). Proteins were also clustered with the *mmseqs* cluster module included in the MMseqs2 suite (version 13.45111 [53]) with a minimum sequence identity of 30 and 70% coverage (parameters --cov-mode 0 --cluster-mode 0 --cluster-reassign). For each cluster of more than five members, a multiple alignment was built with MAFFT v7.310 using default parameters [44] and trimmed with Trimal 1.2rev59 (option -automated1) [45]). These multiple sequence alignments were used to construct an HMM profile for each cluster, with the function *hmmbuild* of HMMER 3.1b2 [42]. These HMM profiles were compared with the MPF_{FATA} HMM profiles using the *HHsearch* function of HH-suite v3 [54]. Clinker [55] was used to visualize the *virB4* gene vicinity of representative genomes.

RESULTS AND DISCUSSION

The phylogenetic distribution of VirB4 in *Patescibacteria*

VirB4 is the only protein with detectable homologues present in all known T4SSs and is, therefore, a hallmark of T4SS presence [56, 57]. It has a strong phylogenetic signal, making it detectable through HMM searches. It functions as an ATPase involved in the assembly of the translocation machinery and pilus biogenesis [2]. Its phylogenetic distribution is the basis for the eight classes of T4SS described in bacteria, namely MPF_{T}, MPF_{F}, MPF_{I}, MPF_{G}, MPF_{C}, MPF_{B}, MPF_{FA} and MPF_{FATA} [41, 56]. The first six are distributed among Gram-negative bacteria, while the last two are exclusive to monoderms, with the final one also present in the conjugative systems of *Archaea*. VirB4 was previously reported to be enriched in CPR compared to non-CPR bacteria [58]. We searched a *Patescibacteria* dataset containing 3,035 genomes (Table S1 and Fig. S1), using a VirB4 HMM profile. We retrieved 2,474 homologues in 2,417 out of 3,035 *Patescibacteria* assemblies (Table S3). They were present in 18 out of the 20 taxonomic classes contained in the dataset (Fig. 1a), and their presence was evenly distributed, regardless of the assembly size (Fig. 1b). These

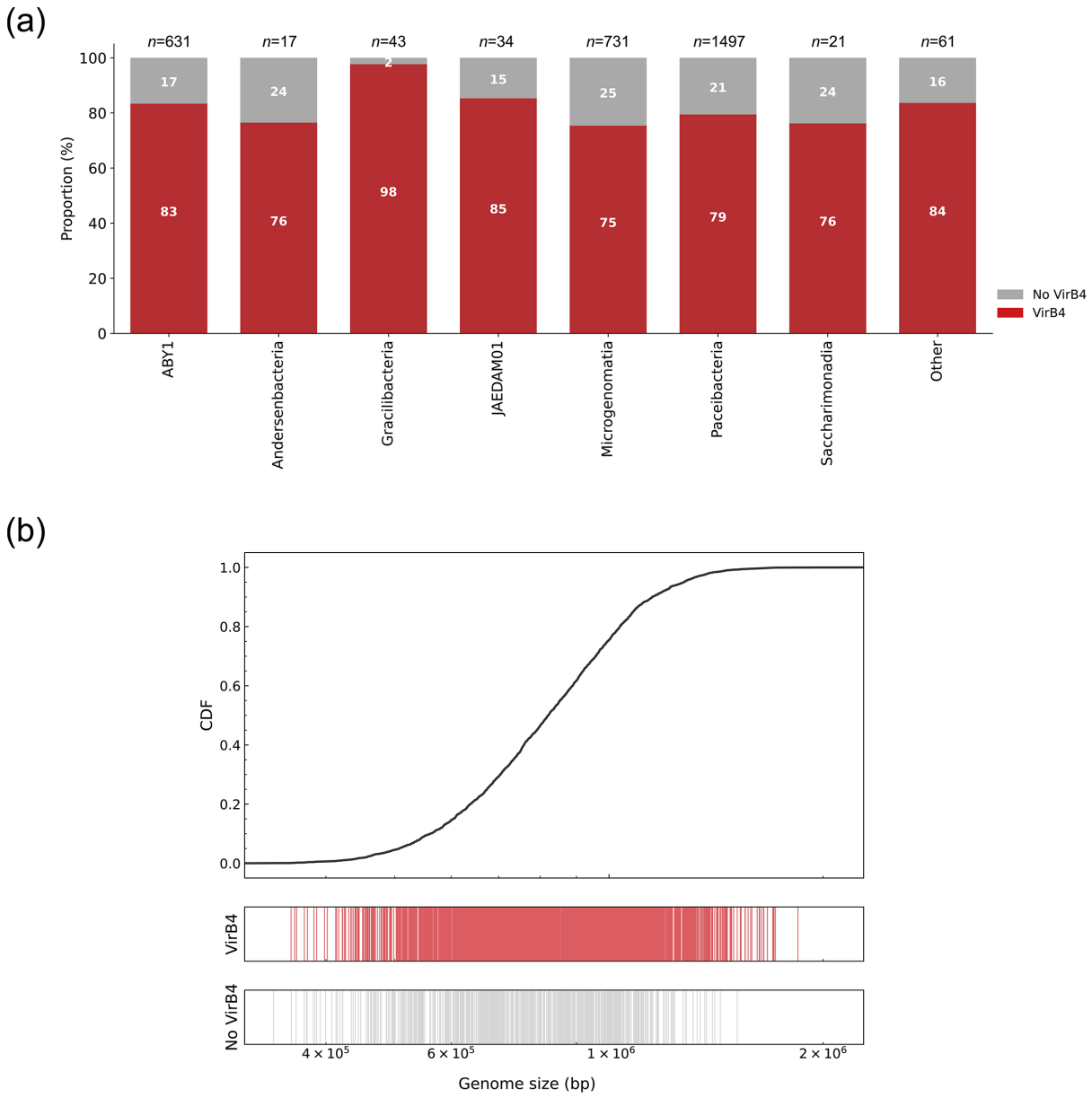

Fig. 1. Detection of VirB4 homologues in *Patescibacteria*. (a) VirB4 abundance in the CPR dataset. Stacked bar plot showing the proportion of genomes with (red) and without (grey) VirB4 homologues for each class in the dataset. The total number of genomes per class (*n*) is displayed at the top of each bar. (b) Distribution of VirB4 homologues according to genome size. The 3,026 assemblies were ranked by size (307,478–2,280,175 bp), shown on the *x*-axis. The top panel displays a cumulative distribution function (CDF) plot of genome size, while the lower panels use vertical lines to indicate the presence (red) or absence (grey) of a VirB4 homologue in each corresponding genome.

VirB4-like proteins were aligned with those from bacterial conjugative plasmids and ICEs of the eight T4SS classes, as well as with VirB4-like proteins representative of archaeal conjugative and DNA-importing systems. Their relationships were assessed using a maximum likelihood phylogenetic reconstruction.

The resulting phylogeny (Fig. 2) showed that a minority of VirB4 homologues present in different classes (49 out of 2,474) clustered together with representative members of the MFF_{FATA} (7 with bacterial VirB4 and 37 with archaeal homologues), MPF_{F} (2), MPF_{T} (2) and MPF_{FA} (1) clades, suggesting that a few *Patescibacteria* have acquired different T4SS from various sources in independent transfer events. Notably, most VirB4 proteins (2,425 present in 2,407 *Patescibacteria* assemblies) clustered into a monophyletic group without VirB4 homologues from other origins. This *Patescibacteria* clade, supported by 97% SH-aLRT and 98% UFBoot values, is nested within the MPF_{FATA} clade. This phylogenetic positioning, sharing a common ancestor with VirB4 proteins from other monoderms (*Firmicutes*, *Actinobacteria*, *Tenericutes* and *Archaea*), aligns with the proposed absence of an outer membrane

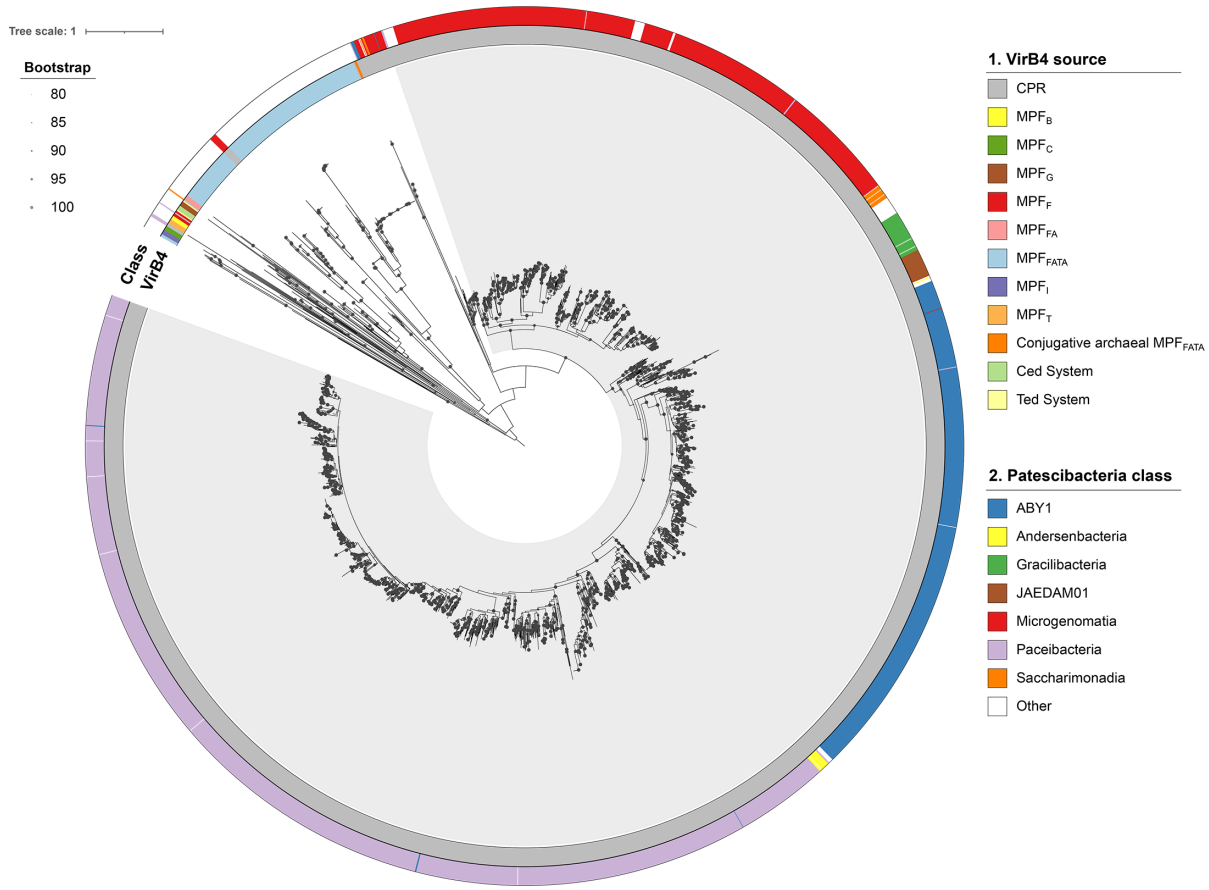

Fig. 2. Phylogenetic analysis of VirB4 proteins. Maximum likelihood tree of the VirB4 homologues of the *Patescibacteria* dataset and members from different T4SS classes rooted at the midpoint. Nodes with UFBoot support values ≥80% are indicated by a grey circle. Rings from inside to outside indicate: (1) T4SS type and (2) *Patescibacteria* class. The *Patescibacteria* clade is shadowed in grey.

in *Patescibacteria* [22, 59]. It also points to the possibility of a single horizontal transfer event of T4SS from monoderms, likely a basibiont, to a *Patescibacteria* ancestor. However, we cannot exclude the possibility that MPF$_{FATA}$ diversification began before the CPR lineage diverged from other monoderm bacterial phyla.

Within the *Patescibacteria* clade, well-defined monophyletic subgroups were observed, each consisting predominantly of VirB4 proteins from a specific taxonomic class. We computed the shortest patristic distance separating each VirB4 protein from the closest homologue in the tree belonging to a different or the same class, as a proxy for estimating T4SS exchange in this phylum (Fig. 3). The cumulative distribution function of patristic distances showed a rapid initial increase for intraclass comparisons, whereas the increase was slower for VirB4 homologues from different classes. In fact, 803 proteins have an identical homolog (patristic distance=0) within the same class, while no identical homologues were found in different classes. Additionally, 99.33% of patescibacterial VirB4 proteins have their closest homologues within the same class, whereas only 0.67% are closest to a homologue from a different class. This suggests that, at short evolutionary distances, it is uncommon to find a VirB4 homologue from a different class, indicating that T4SS transfer events between *Patescibacteria* classes are rare.

T4SS gene repertoire in *Patescibacteria*

Given the phylogenetic proximity of VirB4 proteins in the *Patescibacteria* dataset to those of the MPF$_{FATA}$ type, it is expected that other T4SS components homologous to that system could potentially be found. The MPF$_{FATA}$ class is highly diverse, as are the cell envelopes of the various monoderms included in this class. Moreover, for the MPF$_{FATA}$ class, the exact number of T4SS components, their functions and their essentiality are still uncertain. A set of HMM profiles covers four MPF$_{FATA}$ subclasses, each based on a different prototype: the plasmids pGO1 (*Staphylococcus aureus*) and pCF10 (*Enterococcus faecalis*) and the ICEs CTn2 (*Clostridium difficile*) and ICESaNEM316 (*Streptococcus agalactiae*) [41]. We used HMM profile-based searches to detect components of MPF$_{FATA}$ and the other seven T4SS classes. T4SS components are generally encoded near the *virB4* gene (within 20 CDS) [41], albeit exceptions have been noted, such as the dispersed T4SS islands

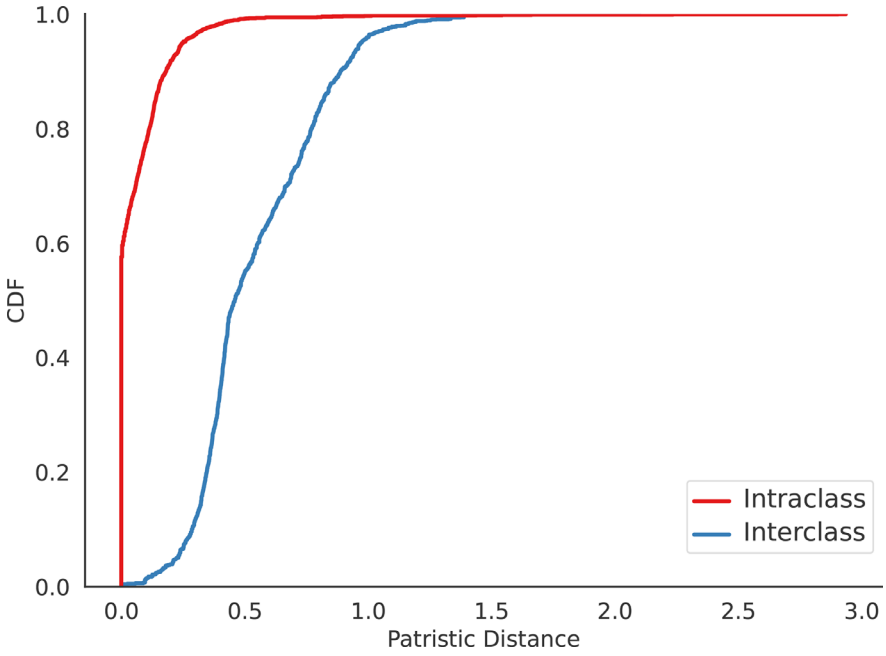

Fig. 3. Cumulative distribution function of the minimal patristic distances in the VirB4 tree. The curves show the sum of the lengths of the branches linking two nodes in the VirB4 phylogenetic tree, belonging to the closest homologues within the same class (red) or different classes (blue).

throughout *Rickettsia* genomes [60] and the coupling protein gene, which is close to the relaxase gene in many conjugative systems [41]. We calculated the distance between the *virB4* genes and the ends of the contigs in our dataset as an estimator of the probability of losing T4SS components in non-closed genomes. Only 5.2% (129 out of 2,474) of the *virB4* genes detected in the *Patescibacteria* dataset were located within 20 CDS of a contig's end. Therefore, the likelihood of missing T4SS components due to genome incompleteness is low. Furthermore, we verified that the detected T4SS components were not skewed toward a specific genome size range (Figs S2–S10).

We, thus, focused primarily on hits associated with VirB4, that is, cases where genes encoding T4SS components are located near the *virB4* gene. T4SS components, primarily associated with MPF_{FATA} subclasses, though not exclusively, were systematically identified in the *Patescibacteria* dataset (Fig. 4 and Table S3). Notably, many of these components were associated with VirB4. To perform more sensitive searches, CDSs encoded near *virB4* were clustered and an HMM profile (CPR profile) was constructed from the alignment of the members of each cluster. These CPR profiles were compared with the MPF_{FATA} HMM profiles.

All conjugative T4SSs, as well as most T4SSs involved in protein secretion, contain a second ATPase that is ancestrally related to VirB4: a coupling protein (T4CP) [56]. This protein is located in the inner membrane and acts as a connector between the T4SS channel and the translocated substrates [2, 61]. Some exceptions include the *Bordetella pertussis* Ptl system and the *Brucella* sp. VirB system [57, 62], which lacks T4CP homologues. Two T4CP families, homologous to VirD4 or TcpA, encompass the diversity of this protein across the eight T4SS classes [41]. We detected 5,713 T4CP homologues to VirD4 in 2,918 *Patescibacteria* genomes, with only a minority of them (148) encoded near *virB4* (Fig. S3). It is not uncommon to find the *t4cp* encoded far from *virB4* [41], as observed with the *t4cp* homologue in *Ca. S. epibionticum* strain ML1. However, given that this gene was not found to be essential for its epibiontic growth, while other T4SS genes were [37], the possibility of *t4cp* being exapted for other functions cannot be ruled out.

The most abundant T4SS component retrieved from the HMM search analysis was TrsD, a protein that shares remote homology with the N-terminal portion of the VirB4 proteins [41]. TrsD is present in only one of the four MPF_{FATA} subclasses (prototype pGO1), and no function has been assigned to this protein. A total of 2,272 TrsD homologues were identified across 2,221 assemblies, including 26 detected exclusively through profile–profile alignments (Table S3). Most of these homologues (2,209) were found in genomes with VirB4 proteins clustered within the *Patescibacteria* clade. A large proportion of these homologues (1,664) were associated with VirB4 (Fig. 4), with the notable exception of TrsD homologues in the *Gracilibacteria* class (Fig. S3). We detected a TrsD homologue associated with VirB4 in *Ca. S. epibionticum* strain ML1, which is listed as essential for its survival [37].

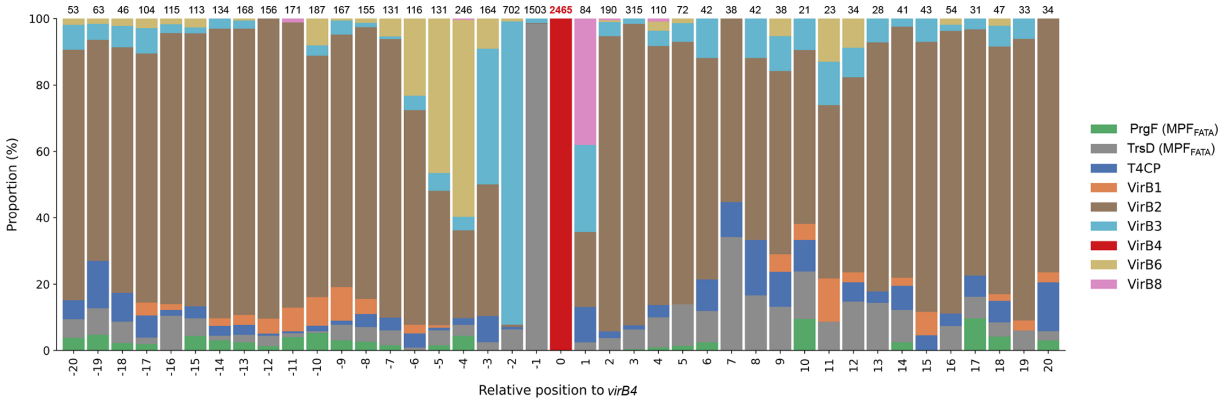

Fig. 4. Abundance of T4SS components near *virB4*. Each column represents the abundance of T4SS components at a specific position relative to the *virB4* gene (within a range of −20 to +20 coding sequences). Only *virB4* genes associated with another T4SS component were considered. The total count of T4SS components at each position is shown above each column and represents 100%. The T4SS components are colour-coded according to the legend. VirB1 comprises homologues retrieved with HMM profiles MPF$_{FATA}$ TrsG and CD419 and MPF$_{T}$ VirB1; VirB2 those retrieved with HMM profiles MPF$_{B}$ TraE, MPF$_{G}$ Tfc9 and Tfc10, MPF$_{I}$ TraQ and TraR, and MPF$_{T}$ VirB2, and Pfam PF04956.16 and PF18895.3; VirB3 with MPF$_{B}$ TraF, MPF$_{C}$ Alr705, MPF$_{F}$ TraL, MPF$_{FATA}$ PrgI and TrsC, MPF$_{G}$ Tfc11, MPF$_{I}$ TraP and MPF$_{T}$ VirB3; VirB6 with MPF$_{FATA}$ PrgH and MPF$_{T}$ VirB6; and VirB8 with MPF$_{F}$ TraE, MPF$_{FATA}$ PrgL and MPF$_{T}$ VirB8.

VirB3, VirB6 and VirB8 are inner membrane proteins that form the cytoplasmic membrane translocon in conjunction with VirB4 and T4CP ATPases [63]. VirB3 is an inner membrane protein that interacts with VirB4, anchoring it to the inner membrane [2, 57, 64]. VirB3 has distinguishable homologues in every bacterial T4SS class, except MPF$_{I}$ and MPF$_{FA}$, and all four MPF$_{FATA}$ prototypes contain VirB3 homologues [41]. VirB3 homologues were recovered in the *Patescibacteria* dataset (1,649 proteins, of which 1,619 in the *Patescibacteria* clade), with a wide variety of HMM profiles from different MPF classes [MPF$_{FATA}$ profiles PrgI (588), and its distant homologue TrsC (97); MPF$_{F}$ profile TraL (142); MPF$_{T}$ profile VirB3 (110); MPF$_{G}$ Tfc11 (53); with MPF$_{B}$ TraF (48) and MPF$_{C}$ Alr1205 (22) (Table S3). HMM–HMM comparisons between CPR and PrgI profiles retrieved 449 additional proteins. VirB3 homologues were mainly associated with VirB4 (880) (Fig. 4), except for the homologues from *Microgenomatia* and *Gracilibacteria* (22.6 and 4.7% of them, respectively) (Fig. S4). PrgI has been previously shown to be more abundant in CPR than in non-CPR bacteria [58] and was found essential for epibiontic growth in *Ca. S. epibionticum* strain ML1 [37]. In some conjugative systems, VirB3 and VirB4 are fused into a single protein [65, 66]. However, we found no instances of VirB3-VirB4 fusion proteins in our dataset.

VirB6 is a polytopic integral membrane protein [67]. VirB6 has recognizable homologues in every MPF class except MPFI [41]. However, in MPF$_{I}$, TraY of MPF$_{I}$ is a strong candidate for a VirB6 analogue. Three of the four MPF$_{FATA}$ subclasses contain VirB6 homologues (PrgH/CD415/GBS1362 HMM profiles), being absent in the pGO1 subclass. However, only 12 out of the 480 homologues detected in the dataset were retrieved with the PrgH HMM profile, while the rest aligned with the VirB6 HMM profile, and most homologues (304) were encoded near *virB4* (Fig. 4 and Table S3). Two VirB6 homologues were additionally retrieved through CPR-PrgH profile–profile alignments. In *Ca. S. epibionticum* strain ML1, the detected *virB6* gene was essential for its growth [37]. It is noticeable that in well-represented classes *Microgenomatia* and *Gracilibacteria*, VirB6 homologues are respectively scarce or absent (Fig. S5), suggesting extreme sequence divergence.

Previous profile–profile comparisons did not identify any MPF$_{FATA}$ HMM profiles with homology to VirB8 [41]. Nevertheless, structural homologues of VirB8 have been identified in the MPF$_{FATA}$ plasmids pIP501 (TraM [68] and TraH [69] and pCF10 (PrgL [70] and PrgD (68)). We detected a few instances of homology to VirB8 HMM profiles from different T4SS classes [MPF$_{T}$ VirB8 (20 proteins), MPF$_{F}$ TraE (8) and MPF$_{FATA}$ PrgL (1)], most associated with VirB4 (Table S3 and Fig. S6). More sensitive searches using CPR-PrgL profile comparisons detected 22 additional homologues. Considering the limited number of VirB8 homologues identified, nearly all of which were found in a single class, *ABY1*, it is likely that this protein has undergone substantial diversification within *Patescibacteria*. A genome-wide structure-based homology analysis of the *S. epibionticum* proteome detected no structural homologue of VirB8 [37]. Alternatively, the role of VirB8 is either redundant due to the structure of the cellular envelope or it is carried out by a completely different protein.

VirB1 is another protein commonly found in T4SSs. This protein is not part of the macromolecular complex but aids in the assembly of the T4SS channel by locally degrading the peptidoglycan [71]. In some monoderms, *virB1* genes are significantly larger than those in diderms [63], as they contain two domains: an N-terminal soluble lytic transglycosylase domain and a C-terminal cysteine-, histidine-dependent amidohydrolase/peptidase domain. Both domains are involved in degrading peptidoglycan [72–76]. VirB1 homologues were retrieved with the HMM profiles MPF$_{T}$ VirB1 (15), MPF$_{FATA}$ TrsG (8) and

CD419 (74) and through HMM-HMM comparison between CPR HMM and CD419 (72) and TrsG (11) (Table S3 and Fig. S7). Of these homologues, 83 contained additional domains, most of which were related to bacterial cell wall metabolism, such as Peptidase_M23 (PF01551.24) and LysM (PF01476.23). In most classes, except for *Paceibacteria*, VirB1 homologues were not highly associated with VirB4. We cannot rule out the possibility that other glycosidases might fulfil the role of VirB1 in *Patescibacteria*. This appears to be the case in *Ca. S. epibionticum* strain ML1, where peptidase and lysozyme homologues encoded immediately downstream of *virB4* have been found essential for its epibiotic growth [37].

In monoderms, T4SS classes lack homologues for the outer membrane core complex proteins (VirB7, VirB9 and VirB10) found in diderm T4SSs [1]. VirB7 is a small, fast-evolving lipoprotein for which no HMM profile is currently available. However, PrgC, which is exclusive to one of the four MPF_{FATA} subclasses, exhibits distant homology to the N-terminal region of VirB9 [41], a protein known to interact with the inner membrane [77]. We detected 56 instances of VirB9 homologues, of which only seven encoded near *virB4* (Table S3 and Fig. S8). In the *Patescibacteria* dataset, no other T4SS components of the outer membrane were associated with VirB4, aside from two VirB10 homologues in assemblies whose VirB4 proteins were located within the MPF_T or MPF_F clade, respectively, and a third homologue present in the *Patescibacteria* clade far from *virB4* (Table S3).

Four VirB4 proteins clustered into clades with homologues present in Gram-negative bacteria: two with MPF_F and two with MPF_T. One genome containing an MPF_F-type VirB4 also encoded homologues of VirB10 (TraB), VirB8 (TraE) and VirB3 (TraL). The second genome with an MPF_F-type VirB4 contained a VirB6 homologue. Both genomes also contained a second VirB4 homologue located within the *Patescibacteria* clade. Among the genomes containing VirB4 within the MPF_T group, one also encoded T4CP, a MOB relaxase and homologues of VirB2, VirB3, VirB5, VirB6, VirB8, VirB9, VirB10 and VirB11, all located in the vicinity of *virB4*. The second genome contained homologues of VirB2, VirB3, VirB6, VirB8 and VirB9, as well as T4CP; however, the latter was not located near *virB4*. Given the low frequency of T4SSs typical of Gram-negative bacteria, it is likely that they have originated from independent transmission events, and we have no data that support (or rule out) the functionality of these Gram-negative-like T4SS in *Patescibacteria*. T4SS classes in monoderms (MPF_{FATA} and MPF_{FA}) encode, near the *virB4* gene, other components specific to each class that are absent in other T4SS types and have no known function. In *Patescibacteria*, a few were detected, including PrgF (627 proteins), TrsJ (47), Gbs1347 (20) and Gbs1350 (37) from MPF_{FATA} and Orf17 (109) from MPF_{FA} (Table S3).

The T4SS classes present in Gram-positive bacteria lack T4SS major pilin genes (*virB2* homologues), relying instead on surface-exposed adhesins or outer membrane proteins to establish and stabilize donor-recipient contacts [14, 78, 79]. The type IV pili (T4P), although not a component of the T4SS, collaborates with it by facilitating MPF rather than directly transporting DNA. T4P encoded in certain conjugative plasmids, such as the IncI plasmid R64, is crucial for conjugative transfer, though only in liquid mating conditions [80]. Similarly, in *Sulfolobales,* it is essential for DNA import by the Ced system [15, 17]. In *Patescibacteria*, T4P was also found to be essential for enabling adhesion to the host during episymbiosis [31]. No adhesins, such as the MPF_{FATA} PrgB, were detected in the *Patescibacteria* dataset. Considering that a VirB2 homologue was identified in the parasitic *Ca. N. lyticus* strain TM7x [36], we searched for VirB2 homologues using various HMM profiles, considering the small size of this protein. Using pilin HMM profiles from different T4SS classes of diderms, we detected VirB2-like pilin proteins in 1,948 assemblies that also contained VirB4, with the pilin gene located near *virB4* in 1135 of these assemblies (Fig. 4, Table S3 and Fig. S9). At this point, we cannot conclude that the original T4SS transferred to *Patescibacteria* contained VirB2-like pilins, but given their absence in MPF_{FATA}, the possibility that the VirB2-like pilin was co-opted from other T4SSs cannot be ruled out. Given that most genomes with VirB2 homologues lack VirB1, it appears that the biogenesis of the putative VirB2 pilus in *Patescibacteria* does not require a VirB1 homologue.

Notably, 441 VirB2 arrays containing 2–11 pilin genes were detected in 383 assemblies. These genes are arranged either in tandem or separated by a non-pilin gene. They include the previously reported essential pilin array in *Ca. S. epibionticum* strain ML1 [37] and what was referred to as the Sec secreted array found in *Ca. N. lyticus* strain TM7x [36]. These VirB2 arrays were especially abundant in *Microgenomatia* and *Saccharimonadia* (Fig. S9). The VirB2 proteins within each array showed high variability, with an average amino acid identity of just 25% and a maximum identity of 54%. Examples of tandem amplification and variation of pilin genes are found in virulence-associated T4SS of intracellular bacteria, such as different species of *Bartonella* [81–83] and *Anaplasma phagocytophilum* [84]. In such cases, the presence of multiple VirB2 paralogs is thought to support a broader immune evasion strategy via antigenic variation or may enhance interactions with a range of host cell surface structures [82]. For *Patescibacteria*, this versatility might aid in adapting to different host strains.

Conjugation-related T4SSs are associated with a relaxase (57), and nine relaxase MOB classes are distinguished (43). We only detected 109 relaxase homologues in 103 *Patescibacteria* genomes, distributed across 8 relaxase classes: MOB_C (58), MOB_M (36), MOB_P (8), MOB_V (2), MOB_Q (2), MOB_T (1), MOB_F (1) and MOB_H (1). In most cases (105 out of 109), the relaxase was identified in a *virB4*-encoding *Patescibacteria* genome, and in only 15 of them were the relaxase and *virB4* genes close, while 39 relaxase genes were close to *t4cp* (Table S3 and Fig. S10). In the *Patescibacteria* clade, we found no distinct characteristics in the T4SSs from the genomes that carry a relaxase gene regarding those that lack relaxases.

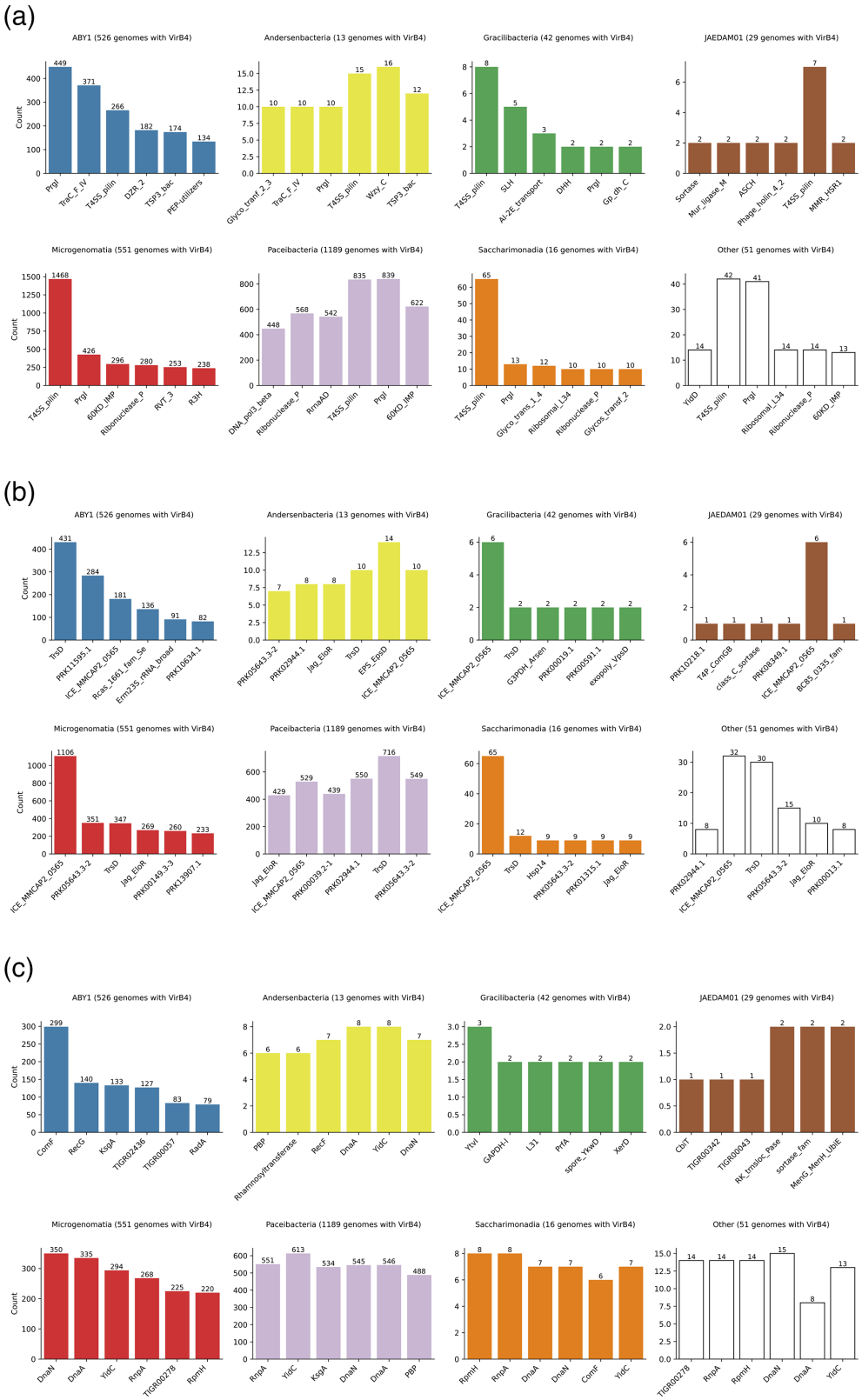

Fig. 5. Proximity of the *virB4* gene to other chromosomal functions. For the 7 *Patescibacteria* classes with 15 or more genomes in the dataset, the 6 most prevalent (a) Pfam, (b) NCBIFAM and (c) TIGRFAM families identified near *virB4* (within a range of −20 to +20 coding sequences) are shown. Data from the remaining classes are grouped under 'Other'. The count of detected members for each protein family is displayed above the respective bar, while the total number of genomes analysed per class is also indicated.

These data suggest that *Patescibacteria* have acquired mobile genetic elements from different sources through conjugation, and thus, only a minority of them have conjugative potential. Therefore, the T4SS identified in the *Patescibacteria* clade appears to be functionally specialized for roles other than conjugation, unless very different relaxases are yet to be found in these bacteria.

The chromosomal context of T4SS in *Patescibacteria*

The presence of T4SS components near *virB4* was also identified by homology searches against protein family profiles of other databases (Fig. 5). Besides, we identified several other gene families that are abundant in the vicinity of the *virB4* gene (Table S4). For the seven classes more represented in the dataset, the distribution of the six most abundant protein families from each HMM database is shown in Fig. 5, and the synteny of the *virB4*-containing region in representative complete genomes of the four most abundant classes is illustrated in Fig. 6. These findings suggest that the genomic location of T4SS varies across different *Patescibacteria* classes. Among most classes, including the most represented *Paceibacteria* and *Microgenomatia*, genes encoding proteins related to DNA replication, such as the beta subunit of DNA polymerase III DnaN and the chromosomal replication initiator protein DnaA, are abundantly represented near *virB4*. Ribosomal proteins, such as the ribonuclease P protein component RnpA, the ribosomal RNA small subunit methyltransferase A KsgA and the 50S ribosomal protein L34 RpmH, are also abundant in the *virB4* vicinity. In many bacterial species, the replication genes, *dnaA* and *dnaN* [85], along with ribosomal protein genes [86], are typically located near the origin of replication. Bacteria that undergo overlapping rounds of replication tend to have a higher abundance of genes situated close to the origin than those near the replication terminus [87]. This phenomenon, known as the replication-associated gene dosage effect, results in increased expression of genes located near the origin of replication [88–92]. Therefore, it is plausible to suggest that the proximity of T4SS genes to the putative origin of replication, at least in some *Patescibacteria* classes, may play a role in enhancing the expression of these T4SS components.

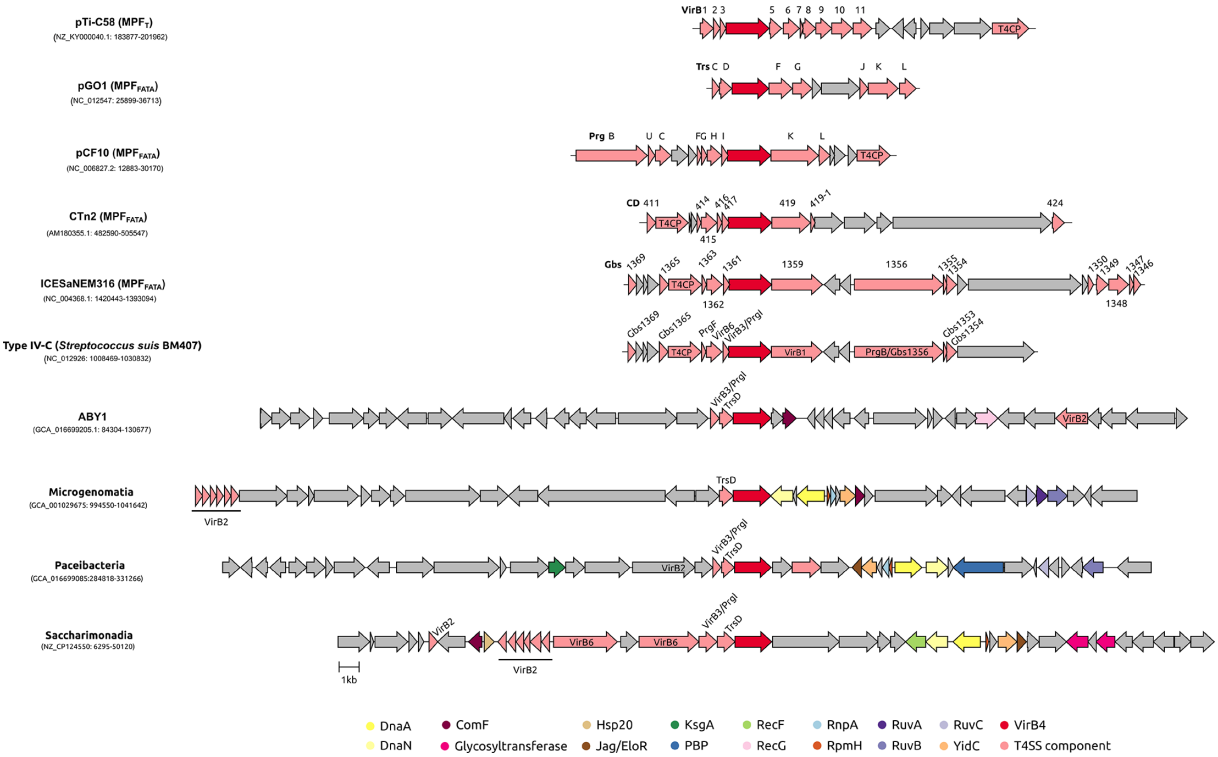

Fig. 6. Genomic context of T4SS in representative genomes. For the four most represented *Patescibacteria* classes in which a T4SS was identified, the synteny of the *virB4* gene neighbourhood is illustrated using a representative from complete *Patescibacteria* genomes. The genetic organization encompasses 20 CDSs upstream and downstream of the *virB4* gene (coloured in red and located at the centre). T4SS genes other than *virB4* are coloured in light red, and non-T4SS genes are depicted in different colours according to the legend. Genes for which no homology to NCBIFAM, Pfam-A or TIGRFAM profiles were found are coloured in grey. T4SS prototypes of the MPF$_T$ and MPF$_{FATA}$ types, as well as an example of the T4SS protein translocators in streptococci, are also included.

On the other hand, in the *ABY1* class, which is also well-represented, genomic loci containing T4SS genes commonly include genes related to natural competence (*comF*) and DNA recombination (*recG* and *radA*). The co-localization of these genes suggests a functional coordination of their activities.

CONCLUDING REMARKS

Patescibacteria was identified as the phylum with the lowest level of protein annotation coverage in GTDB, which may be due to highly divergent gene families that evade detection by standard homology-based annotation methods or indicate the presence of novel protein functions, metabolic activities and biological traits [93]. Moreover, content-based analyses are influenced by genome completeness, and, currently, very few complete genomes are available for this phylum (Fig. S1). These factors underscore the challenges in accurately determining molecular functions in proteins, particularly within highly divergent lineages like *Patescibacteria*. In this study, we identified the presence of a T4SS across most classes of *Patescibacteria* by performing homologue searches using extensively curated and validated HMM profiles. Homologues of VirB4, TrsD, PrgI/VirB3, VirB6 and VirB2-like proteins were found here to be widely distributed in *Patescibacteria*.

Considering the reduced genomes of *Patescibacteria*, it is notable that they have not only retained the T4SS but also often located it at a super-expressed location near the origin of replication. The single experimental analysis of a T4SS carried out in *Patescibacteria* showed that genes encoding these components were essential for the epibiotic host-dependent lifestyle [37].

Patescibacteria, lacking many biosynthetic pathways, presumably grow using molecules derived from active hosts. Notably, although genes encoding homologues of the natural competence *comEC* system mediate DNA uptake from the environment, they were not essential for the epibiotic growth of *Ca. S. epibionticum* strain ML1 [37], suggesting that *Patescibacteria* likely depend on alternative systems to acquire nucleotides needed for growth.

T4SSs act as versatile nanomachines, adapted to transfer large macromolecules across multiple cell membranes. The structural adaptability of T4SS has given rise to a broad range of system variations across bacterial lineages, with T4SS having been co-opted repeatedly throughout evolutionary history to enable the import or export of various substrates [56], including the acquisition of DNA from the environment or other cells. It is, therefore, probable that *Patescibacteria* have repurposed the monoderm conjugative system for functions other than conjugation. The wide distribution and abundance of T4SS in different classes of *Patescibacteria*, along with its probable high expression due to the proximity of T4SS genes to the origin of replication, suggest that T4SS in *Patescibacteria* may function as a mechanism for importing DNA or other macromolecules from their hosts, enabling them to exist as obligate episymbionts on other microbes. Future research should explore this hypothesis.

Funding information
This work was supported by the Spanish Ministry of Science and Innovation (Grant MCIN/AEI/10.13039/501100011033 PID2020-117923GB-I00 to M.P.G.-B.), the Spanish Ministry of Economy, Industry and Competitiveness (FLEX3GEN PID2020-118052GB-I00, cofounded with FEDER funds to F.R-V) and the Spanish Ministry of Universities (predoctoral contract FPU20/04579 to M.d.M.Q.-C.). P.J.C.-Y. work was funded by a Post-Doctoral Fellowship from the Fundación Alfonso Martín Escudero, Spain, and is currently supported by a Marie Sklodowska-Curie postdoctoral fellowship granted by the Horizon Europe programme (1011052332-CYANORUB) and funded by the UKRI (grant ref: EP/Y028384/1).

Conflicts of interest
The authors declare that there are no conflicts of interest.

References

1. Grohmann E, Christie PJ, Waksman G, Backert S. Type IV secretion in Gram-negative and Gram-positive bacteria. *Mol Microbiol* 2018;107:455–471.
2. Cabezón E, Ripoll-Rozada J, Peña A, de la Cruz F, Arechaga I. Towards an integrated model of bacterial conjugation. *FEMS Microbiol Rev* 2015;39:81–95.
3. de la Cruz F, Frost LS, Meyer RJ, Zechner EL. Conjugative DNA metabolism in Gram-negative bacteria. *FEMS Microbiol Rev* 2010;34:18–40.
4. Bienvenu A, Martinez E, Bonazzi M. Undercover agents of infection: the stealth strategies of T4SS-equipped bacterial pathogens. *Toxins* 2021;13:713.
5. Wang J, Feng Y, Wang C, Srinivas S, Chen C, *et al*. Pathogenic *Streptococcus* strains employ novel escape strategy to inhibit bacteriostatic effect mediated by mammalian peptidoglycan recognition protein . *Cell Microbiol* 2017;19:e12724.
6. Jiang X, Yang Y, Zhou J, Zhu L, Gu Y, *et al*. Roles of the putative type IV-like secretion system key component VirD4 and PrsA in pathogenesis of *Streptococcus suis* type 2. *Front Cell Infect Microbiol* 2016;6:172.
7. Zhao Y, Liu G, Li S, Wang M, Song J, *et al*. Role of a type IV–like secretion system of *Streptococcus suis* 2 in the development of streptococcal toxic shock syndrome. *J Infect Dis* 2011;204:274–281.
8. Souza DP, Oka GU, Alvarez-Martinez CE, Bisson-Filho AW, Dunger G, *et al*. Bacterial killing via a type IV secretion system. *Nat Commun* 2015;6:6453.
9. Shen X, Wang B, Yang N, Zhang L, Shen D, *et al*. Lysobacter enzymogenes antagonizes soilborne bacteria using the type IV secretion system. *Environ Microbiol* 2021;23:4673–4688.
10. Bayer-Santos E, Cenens W, Matsuyama BY, Oka GU, Di Sessa G, *et al*. The opportunistic pathogen *Stenotrophomonas maltophilia* utilizes a type IV secretion system for interbacterial killing. *PLoS Pathog* 2019;15:e1007651.

11. Purtschert-Montenegro G, Cárcamo-Oyarce G, Pinto-Carbó M, Agnoli K, Bailly A, *et al*. *Pseudomonas putida* mediates bacterial killing, biofilm invasion and biocontrol with a type IVB secretion system. *Nat Microbiol* 2022;7:1547–1557.
12. Hamilton HL, Domínguez NM, Schwartz KJ, Hackett KT, Dillard JP. *Neisseria gonorrhoeae* secretes chromosomal DNA via a novel type IV secretion system. *Mol Microbiol* 2005;55:1704–1721.
13. Hofreuter D, Odenbreit S, Haas R. Natural transformation competence in Helicobacter pylori is mediated by the basic components of a type IV secretion system. *Mol Microbiol* 2001;41:379–391.
14. Li YG, Hu B, Christie PJ. Biological and structural diversity of type IV secretion systems. *Microbiol Spectr* 2019;7.
15. van Wolferen M, Wagner A, van der Does C, Albers S-V. The archaeal Ced system imports DNA. *Proc Natl Acad Sci USA* 2016;113:2496–2501.
16. Beltran LC, Cvirkaite-Krupovic V, Miller J, Wang F, Kreutzberger MAB, *et al*. Archaeal DNA-import apparatus is homologous to bacterial conjugation machinery. *Nat Commun* 2023;14:666.
17. Ajon M, Fröls S, van Wolferen M, Stoecker K, Teichmann D, *et al*. UV-inducible DNA exchange in hyperthermophilic archaea mediated by type IV pili. *Mol Microbiol* 2011;82:807–817.
18. Brown CT, Hug LA, Thomas BC, Sharon I, Castelle CJ, *et al*. Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* 2015;523:208–211.
19. Coleman GA, Davín AA, Mahendrarajah TA, Szánthó LL, Spang A, *et al*. A rooted phylogeny resolves early bacterial evolution. *Science* 2021;372:eabe0511.
20. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, *et al*. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* 2018;36:996–1004.
21. Wrighton KC, Thomas BC, Sharon I, Miller CS, Castelle CJ, *et al*. Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. *Science* 2012;337:1661–1665.
22. Tian R, Ning D, He Z, Zhang P, Spencer SJ, *et al*. Small and mighty: adaptation of superphylum *Patescibacteria* to groundwater environment drives their genome simplicity. *Microbiome* 2020;8:51.
23. Castelle CJ, Brown CT, Anantharaman K, Probst AJ, Huang RH, *et al*. Biosynthetic capacity, metabolic variety and unusual biology in the CPR and DPANN radiations. *Nat Rev Microbiol* 2018;16:629–645.
24. Chiriac M-C, Bulzu P-A, Andrei A-S, Okazaki Y, Nakano S-I, *et al*. Ecogenomics sheds light on diverse lifestyle strategies in freshwater CPR. *Microbiome* 2022;10:84.
25. He X, McLean JS, Edlund A, Yooseph S, Hall AP, *et al*. Cultivation of a human-associated TM7 phylotype reveals a reduced genome and epibiotic parasitic lifestyle. *Proc Natl Acad Sci USA* 2015;112:244–249.
26. Bor B, Collins AJ, Murugkar PP, Balasubramanian S, To TT, *et al*. Insights obtained by culturing *Saccharibacteria* with their bacterial hosts. *J Dent Res* 2020;99:685–694.
27. Batinovic S, Rose JJA, Ratcliffe J, Seviour RJ, Petrovski S. Cocultivation of an ultrasmall environmental parasitic bacterium with lytic ability against bacteria associated with wastewater foams. *Nat Microbiol* 2021;6:703–711.
28. Ibrahim A, Maatouk M, Rajaonison A, Zgheib R, Haddad G, *et al*. Adapted protocol for *Saccharibacteria* cocultivation: two new members join the club of *Candidate Phyla Radiation*. *Microbiol Spectr* 2021;9:e0106921.
29. Bor B, McLean JS, Foster KR, Cen L, To TT, *et al*. Rapid evolution of decreased host susceptibility drives a stable relationship between ultrasmall parasite TM7x and its bacterial host. *Proc Natl Acad Sci U S A* 2018;115:12277–12282.
30. Nie J, Utter DR, Kerns KA, Lamont EI, Hendrickson EL, *et al*. Strain-level variation and diverse host bacterial responses in episymbiotic *Saccharibacteria*. *mSystems* 2022;7.
31. Xie B, Wang J, Nie Y, Tian J, Wang Z, *et al*. Type IV pili trigger episymbiotic association of *Saccharibacteria* with its bacterial host. *Proc Natl Acad Sci USA* 2022;119:e2215990119.
32. Yakimov MM, Merkel AY, Gaisin VA, Pilhofer M, Messina E, *et al*. Cultivation of a vampire: '*Candidatus Absconditicoccus praedator*'. *Environ Microbiol* 2022;24:30–49.
33. Moreira D, Zivanovic Y, López-Archilla AI, Iniesto M, López-García P. Reductive evolution and unique predatory mode in the CPR bacterium *Vampirococcus lugosii*. *Nat Commun* 2021;12:2454.
34. Cross KL, Campbell JH, Balachandran M, Campbell AG, Cooper CJ, *et al*. Targeted isolation and cultivation of uncultivated bacteria by reverse genomics. *Nat Biotechnol* 2019;37:1314–1321.
35. Kuroda K, Yamamoto K, Nakai R, Hirakata Y, Kubota K, *et al*. Symbiosis between *Candidatus Patescibacteria* and archaea discovered in wastewater-treating bioreactors. *mBio* 2022;13:e0171122.
36. McLean JS, Bor B, Kerns KA, Liu Q, To TT, *et al*. Acquisition and adaptation of ultra-small parasitic reduced genome bacteria to mammalian hosts. *Cell Rep* 2020;32:107939.
37. Wang Y, Gallagher LA, Andrade PA, Liu A, Humphreys IR, *et al*. Genetic manipulation of Patescibacteria provides mechanistic insights into microbial dark matter and the epibiotic lifestyle. *Cell* 2023;186:4803–4817..
38. Cabello-Yeves PJ, Zemskaya TI, Zakharenko AS, Sakirko MV, Ivanov VG, *et al*. Rodriguez-Valera, Microbiome of the deep Lake Baikal, a unique oxic bathypelagic habitat. *Limnol Oceanogr* 2020;65:1471–1488.
39. Haro-Moreno JM, Cabello-Yeves PJ, Garcillán-Barcia MP, Zakharenko A, Zemskaya TI, *et al*. A novel and diverse group of *Candidatus Patescibacteria* from bathypelagic Lake Baikal revealed through long-read metagenomics. *Environ Microbiome* 2023;18:12.
40. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. Gtdb-tk: a toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* 2020;36:1925–1927.
41. Guglielmini J, Néron B, Abby SS, Garcillán-Barcia MP, de la Cruz F, *et al*. Key components of the eight classes of type IV secretion systems involved in bacterial conjugation or protein secretion. *Nucleic Acids Res* 2014;42:5715–5727.
42. Eddy SR. Accelerated profile HMM searches. *PLoS Comput Biol* 2011;7:e1002195.
43. Garcillán-Barcia MP, Redondo-Salvo S, Vielva L, de la Cruz F. MOBscan: automated annotation of MOB relaxases. In: de la Cruz F (eds). *Horizontal Gene Transfer. Methods in Molecular Biology,* vol 2075. New York, NY: Humana; 2020.
44. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013;30:772–780.
45. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 2009;25:1972–1973.
46. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 2015;32:268–274.
47. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 2017;14:587–589.
48. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol* 2018;35:518–522.
49. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, *et al*. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 2010;59:307–321.
50. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* 2021;49:W293–W296.
51. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 2019;35:526–528.
52. Li W, O'Neill KR, Haft DH, DiCuccio M, Chetvernin V, *et al*. RefSeq: expanding the prokaryotic genome annotation pipeline

reach with protein family model curation. *Nucleic Acids Res* 2021;49:D1020–D1028.

53. Steinegger M, Söding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol* 2017;35:1026–1028.
54. Steinegger M, Meier M, Mirdita M, Vöhringer H, Haunsberger SJ, *et al*. HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics* 2019;20:473.
55. Gilchrist CLM, Chooi Y-H. Clinker & clustermap.js: automatic generation of gene cluster comparison figures. *Bioinformatics* 2021;37:2473–2475.
56. Guglielmini J, de la Cruz F, Rocha EPC. Evolution of conjugation and type IV secretion systems. *Mol Biol Evol* 2013;30:315–331.
57. Alvarez-Martinez CE, Christie PJ. Biological diversity of prokaryotic type IV secretion systems. *Microbiol Mol Biol Rev* 2009;73:775–808.
58. Méheust R, Burstein D, Castelle CJ, Banfield JF. The distinction of CPR bacteria from other bacteria based on protein family content. *Nat Commun* 2019;10:4173.
59. Luef B, Frischkorn KR, Wrighton KC, Holman H-YN, Birarda G, *et al*. Diverse uncultivated ultra-small bacterial cells in groundwater. *Nat Commun* 2015;6:6372.
60. Gillespie JJ, Phan IQH, Driscoll TP, Guillotte ML, Lehman SS, *et al*. The Rickettsia type IV secretion system: unrealized complexity mired by gene family expansion. *Pathog Dis* 2016;74:ftw058.
61. Llosa M, Alkorta I. "Coupling Proteins in Type IV Secretion" in (2017); http://link.springer.com/10.1007/978-3-319-75241-9_6, pp. 143–168.
62. Christie PJ, Whitaker N, González-Rivera C. Mechanism and structure of the bacterial type IV secretion systems. *Biochim Biophys Acta - Mol Cell Res* 2014;1843:1578–1591.
63. Bhatty M, Laverde Gomez JA, Christie PJ. The expanding bacterial type IV secretion lexicon. *Res Microbiol* 2013;164:620–639.
64. Mossey P, Hudacek A, Das A. *Agrobacterium tumefaciens* type IV secretion protein VirB3 is an inner membrane protein and requires VirB4, VirB7, and VirB8 for stabilization. *J Bacteriol* 2010;192:2830–2838.
65. Batchelor RA, Pearson BM, Friis LM, Guerry P, Wells JM. Nucleotide sequences and comparison of two large conjugative plasmids from different *Campylobacter* species. *Microbiology* 2004;150:3507–3517.
66. Strauch E, Goelz G, Knabner D, Konietzny A, Lanka E, *et al*. A cryptic plasmid of *Yersinia enterocolitica* encodes a conjugative transfer system related to the regions of CloDF13 Mob and IncX Pil. *Microbiology* 2003;149:2829–2845.
67. Judd PK, Mahli D, Das A. Molecular characterization of the *Agrobacterium tumefaciens* DNA transfer protein VirB6. *Microbiology* 2005;151:3483–3492.
68. Goessweiner-Mohr N, Grumet L, Arends K, Pavkov-Keller T, Gruber CC, *et al*. The 2.5 Å structure of the enterococcus conjugation protein TraM resembles VirB8 type IV secretion proteins. *J Biol Chem* 2013;288:2018–2028.
69. Fercher C, Probst I, Kohler V, Goessweiner-Mohr N, Arends K, *et al*. VirB8-like protein TraH is crucial for DNA transfer in *Enterococcus faecalis*. *Sci Rep* 2016;6:24643.
70. Jäger F, Lamy A, Sun W-S, Guerini N, Berntsson R-A. Structure of the enterococcal T4SS protein PrgL reveals unique dimerization interface in the VirB8 protein family. *Structure* 2022;30:876–885.
71. Zahrl D, Wagner M, Bischof K, Bayer M, Zavecz B, *et al*. Peptidoglycan degradation by specialized lytic transglycosylases associated with type III and type IV secretion systems. *Microbiology* 2005;151:3455–3467.
72. Arends K, Celik E-K, Probst I, Goessweiner-Mohr N, Fercher C, *et al*. TraG encoded by the pIP501 type IV secretion system is a two-domain peptidoglycan-degrading enzyme essential for conjugative transfer. *J Bacteriol* 2013;195:4436–4444.
73. Guglielmetti S, Balzaretti S, Taverniti V, Miriani M, Milani C, *et al*. TgaA, a VirB1-like component belonging to a putative type IV secretion system of Bifidobacterium bifidum MIMBb75. *Appl Environ Microbiol* 2014;80:5161–5169.
74. Zhong Q, Zhao Y, Chen T, Yin S, Yao X, *et al*. A functional peptidoglycan hydrolase characterized from T4SS in 89K pathogenicity island of epidemic *Streptococcus suis* serotype 2. *BMC Microbiol* 2014;14:73.
75. Laverde Gomez JA, Bhatty M, Christie PJ. PrgK, a multidomain peptidoglycan hydrolase, is essential for conjugative transfer of the pheromone-responsive plasmid pcf10. *J Bacteriol* 2014;196:527–539.
76. Sun W-S, Torrens G, Ter Beek J, Cava F, Berntsson RP-A. Breaking barriers: pCF10 type 4 secretion system relies on a self-regulating muramidase to modulate the cell wall. *mBio* 2024;15:e0048824.
77. Rivera-Calzada A, Fronzes R, Savva CG, Chandran V, Lian PW, *et al*. Structure of a bacterial type IV secretion core complex at subnanometre resolution. *EMBO J* 2013;32:1195–1204.
78. Grohmann E, Muth G, Espinosa M. Conjugative plasmid transfer in gram-positive bacteria. *Microbiol Mol Biol Rev* 2003;67:277–301.
79. Kohler V, Keller W, Grohmann E. Enterococcus adhesin PrgB facilitates type IV secretion by condensation of extracellular DNA. *Mol Microbiol* 2018;109:263–267.
80. Komano T, Yoshida T, Narahara K, Furuya N. The transfer region of IncI1 plasmid R64: similarities between R64 tra and legionella icm/dot genes. *Mol Microbiol* 2000;35:1348–1359.
81. Seubert A, Hiestand R, de la Cruz F, Dehio C. A bacterial conjugation machinery recruited for pathogenesis. *Mol Microbiol* 2003;49:1253–1266.
82. Nystedt B, Frank AC, Thollesson M, Andersson SGE. Diversifying selection and concerted evolution of a type IV secretion system in *Bartonella*. *Mol Biol Evol* 2008;25:287–300.
83. Vayssier-Taussat M, Le Rhun D, Deng HK, Biville F, Cescau S, *et al*. The Trw type IV secretion system of *Bartonella* mediates host-specific adhesion to erythrocytes. *PLoS Pathog* 2010;6:e1000946.
84. Al-Khedery B, Lundgren AM, Stuen S, Granquist EG, Munderloh UG, *et al*. Structure of the type IV secretion system in different strains of *Anaplasma phagocytophilum*. *BMC Genomics* 2012;13:678.
85. Messer W. The bacterial replication initiator DnaA. DnaA and oriC, the bacterial mode to initiate DNA replication. *FEMS Microbiol Rev* 2002;26:355–374.
86. Couturier E, Rocha EPC. Replication-associated gene dosage effects shape the genomes of fast-growing bacteria but only for transcription and translation genes. *Mol Microbiol* 2006;59:1506–1518.
87. Rocha EPC. The organization of the bacterial genome. *Annu Rev Genet* 2008;42:211–233.
88. Schmid MB, Roth JR. Gene location affects expression level in *Salmonella typhimurium*. *J Bacteriol* 1987;169:2872–2875.
89. Sousa C, de Lorenzo V, Cebolla A. Modulation of gene expression through chromosomal positioning in *Escherichia coli*. *Microbiology* 1997;143:2071–2078.
90. Sharp PM, Shields DC, Wolfe KH, Li W-H. Chromosomal location and evolutionary rate variation in enterobacterial genes. *Science* 1989;246:808–810.
91. Lato DF, Golding GB. Spatial patterns of gene expression in bacterial genomes. *J Mol Evol* 2020;88:510–520.
92. Pountain AW, Jiang P, Yao T, Homaee E, Guan Y, *et al*. Transcription-replication interactions reveal bacterial genome regulation. *Nature* 2024;626:661–669.
93. Lobb B, Tremblay BJ-M, Moreno-Hagelsieb G, Doxey AC. An assessment of genome annotation coverage across the bacterial tree of life. *Microbial Genomics* 2020;6.