ACS OMEGA

Open Access

http://pubs.acs.org/journal/acsodf

Article

T9GPred: A Comprehensive Computational Tool for the Prediction of Type 9 Secretion System, Gliding Motility, and the Associated Secreted Proteins

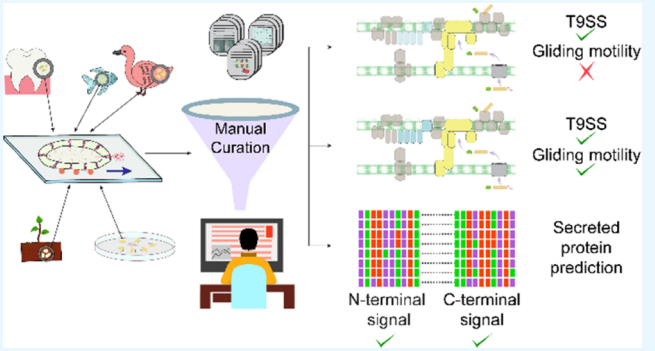
Ajaya Kumar Sahoo, R. P. Vivek-Ananth, Nikhil Chivukula, Shri Vishalini Rajaram, Karthikeyan Mohanraj, Devanshi Khare, Celin Acharya, and Areejit Samal*

Cite This: *ACS Omega* 2023, 8, 34091−34102

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Type 9 secretion system (T9SS) is one of the least characterized secretion systems exclusively found in the *Bacteroidetes* phylum, which comprises various environmental and economically relevant bacteria. While T9SS plays a central role in bacterial movement termed gliding motility, survival, and pathogenicity, there is an unmet need for a comprehensive tool that predicts T9SS, gliding motility, and proteins secreted via T9SS. In this study, we develop such a computational tool, **T**ype **9** secretion system and **G**liding motility **Pred**iction (T9GPred). To build this tool, we manually curated published experimental evidence and identified mandatory components for T9SS and gliding motility prediction. We also compiled experimentally characterized proteins secreted via T9SS and determined the presence of three unique types of C-terminal domain signals, and these insights were leveraged to predict proteins secreted via T9SS. Notably, using recently published experimental evidence, we show that T9GPred has high predictive power. Thus, we used T9GPred to predict the presence of T9SS, gliding motility, and associated secreted proteins across 693 completely sequenced *Bacteroidetes* strains. T9GPred predicted 402 strains to have T9SS, of which 327 strains are also predicted to exhibit gliding motility. Further, T9GPred also predicted putative secreted proteins for the 402 strains. In a nutshell, T9GPred is a novel computational tool for systems-level prediction of T9SS and streamlining future experimentation. The source code of the computational tool is available in our GitHub repository: https://github.com/asamallab/T9GPred. The tool and its predicted results are compiled in a web server available at: https://cb.imsc.res.in/t9gpred/.



INTRODUCTION

Type 9 secretion system (T9SS) is one of the recently characterized secretion systems found particularly in members of the *Bacteroidetes* phylum.[1−4] *Bacteroidetes* is a diverse phylum of bacteria found in different environmental niches, including animal gut microbiota, where they are either beneficial or pathogenic based on their secretory factors.[5,6] In some of these *Bacteroidetes* strains, T9SS has been experimentally identified to play a central role in movement of the bacteria,[1,7−9] assimilation of ions,[10] and secretion of various polysaccharide degrading enzymes,[10,11] adhesins,[12] and virulent proteins.[2,13] Additionally, T9SS is found to be essential for the growth and survival of some species.[13−15] Thus, identification of T9SS is imperative to understand the underlying microbial activity. The identification of T9SS is also of industrial relevance as it can aid in an efficient high-throughput production of recombinant proteins.[16]

T9SS was initially characterized in two different *Bacteroidetes* species, *Porphyromonas gingivalis* (human periodontal pathogen)[1,17] and *Flavobacterium johnsoniae*.[1] In particular, McBride and colleagues had extensively characterized the protein components associated with gliding motility of *F. johnsoniae*[18,19] and later found a large overlap with the T9SS components.[3] Additional experiments proved that the gliding motility is facilitated by the secretion of adhesins through T9SS.[20,21] Further, several groups have identified functionalities of the core protein components of T9SS, namely, GldK, GldL, GldM, GldN, and SprA.[22−25] James et al.[26] have shown that the GldLM complex is conserved across members of the *Bacteroidetes* phylum and is mediated through proton flow to drive the secretion via T9SS.[23,27] There are other such independent studies focused on T9SS or gliding motility in

Received: July 17, 2023
Accepted: August 24, 2023
Published: September 8, 2023

members of the *Bacteroidetes* phylum,[7,8,28] but there is a scarce attempt toward systems-level understanding of the components associated with T9SS or gliding motility.

Secretion via T9SS is a two-step process: (i) transport of protein from cytoplasm to periplasm via the Sec export pathway mediated by N-terminal domain and (ii) transport of protein from periplasm to extracellular space via T9SS mediated by C-terminal domain (CTD).[13,29] Kulkarni et al.[30] have experimentally proved that the N-terminal domain and the CTD signal are necessary and sufficient for the secretion of a protein via T9SS. Additionally, they have identified different types of CTDs. While efforts have been made to characterize the secreted proteins, there have been limited attempts to identify the CTD types and predict such secreted proteins.

Earlier, Abby et al.[4,31] had developed a computational tool, TXSScan, to identify different secretion systems, including T9SS across bacteria. However, TXSScan is neither specific to T9SS nor does it predict the associated gliding motility or secreted proteins. Therefore, in this study, we built a prediction tool, which is dedicated to T9SS, gliding motility, and the proteins secreted via T9SS. Initially, we compiled the proteins associated with T9SS or gliding motility from the published literature and generated hidden Markov model (HMM) profiles for each protein. By leveraging the compiled experimental knowledge, we identified the key protein components and built a computational tool for prediction of T9SS or gliding motility across members of the *Bacteroidetes* phylum. Note that, we have included gliding motility prediction along with that of T9SS as some *Bacteroidetes* strains having T9SS do not exhibit gliding motility.[1,13] Previously, Veith et al.[32] have used HMM-based approaches to predict proteins secreted via T9SS. However, their prediction model did not account for different types of CTD. Here, we leveraged the different CTD types and their motifs from the published literature to generate HMM profiles and built a computational pipeline for the prediction of proteins secreted via T9SS. Finally, we developed a computational tool, **T**ype **9** secretion system and **G**liding motility **Pred**iction (T9GPred), which is accessible at: https://github.com/asamallab/T9GPred. We also compiled the tool and its predictions into a web server, namely, T9GPred, which is accessible at: https://cb.imsc.res.in/t9gpred/.

RESULTS AND DISCUSSION

Compilation of 28 Protein Components Associated with T9SS or Gliding Motility among Members of the *Bacteroidetes* Phylum. We compiled 23 protein components associated with T9SS (Table S1), 22 protein components associated with gliding motility (Table S2), and 102 proteins secreted via T9SS (secreted proteins) (Table S3) in *Bacteroidetes* strains from the published literature (see Methods Section; Figure 1). We observed that there is a large overlap of 17 proteins between the proteins associated with T9SS and gliding motility, which is in line with the previous studies reporting conclusive evidence of a link between T9SS and gliding motility in *Bacteroidetes* strains.[1,3] In total, we identified 28 proteins associated with either T9SS or gliding motility among members of the *Bacteroidetes* phylum.

Key Protein Components of T9SS and Gliding Motility. During the compilation of associated protein components (Figure 1), we observed that not all of the components are present in every bacterium of the *Bacteroidetes* phylum that has experimental evidence for T9SS or gliding motility (Figure 2). Therefore, we were interested in identifying the key components, which are mandatory for a functional T9SS or exhibiting gliding motility. Hidden Markov model (HMM) profiles of protein sequences have been previously used in TXSScan,[4,31] SecReT6,[34] SecReT4,[35] and T346Hunter[36] to make robust predictions of the presence of protein components associated with different secretion systems. Therefore, we generated HMM profiles for the 28 identified protein components (see Methods Section; Supporting Information, Section 1.1) and checked their presence in 30 *Bacteroidetes* strains that have experimental evidence for the presence of T9SS or gliding motility (Table S4) to infer a minimal set of key protein components that are necessary for their functionality.[36] We observed that proteins GldL, GldM, GldK, GldN, SprE, SprA, SprT, PorV, PorP (with one exception), FtsX, and Omp17 are always present in *Bacteroidetes* strains with experimental evidence on the presence of T9SS or gliding motility (Figure 2). The proteins GldL and GldM form an inner membrane proton-driven motor that drives protein secretion via T9SS (Figure 3).[23,27] The GldK−GldN ring and SprE facilitate energy transduction from the inner membrane GldLM motor to the outer membrane translocon SprA for the secretion of proteins to the extracellular space (Figure 3).[22,23,25,37] SprT aids in the maturation of the secreted protein but its role in the secretion mechanism is not yet known.[17] PorV and PorP aid in the delivery of certain secreted substrates and are not necessary for all secreted proteins.[38] FtsX and Omp17 are found to be ubiquitous among all 30 *Bacteroidetes* strains, and their function in T9SS is not known. Therefore, we designated the minimal set of GldK, GldL, GldM, GldN, SprA, and SprE (6 proteins) as mandatory protein components to predict the presence of T9SS. Notably, we observed that these T9SS mandatory components are always present in *Bacteroidetes* strains that lack experimental evidence of the presence of T9SS but have experimentally characterized T9SS-associated gliding motility (Figure 2).

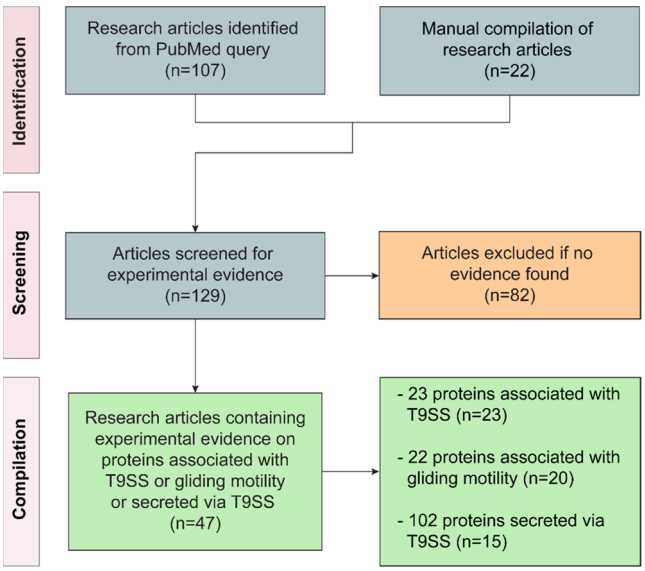


Figure 1. Workflow to identify published research articles containing experimental evidence on the proteins associated with T9SS or gliding motility or proteins secreted via T9SS. This workflow is presented according to the PRISMA statement.[33]

https://doi.org/10.1021/acsomega.3c05155
ACS Omega 2023, 8, 34091−34102

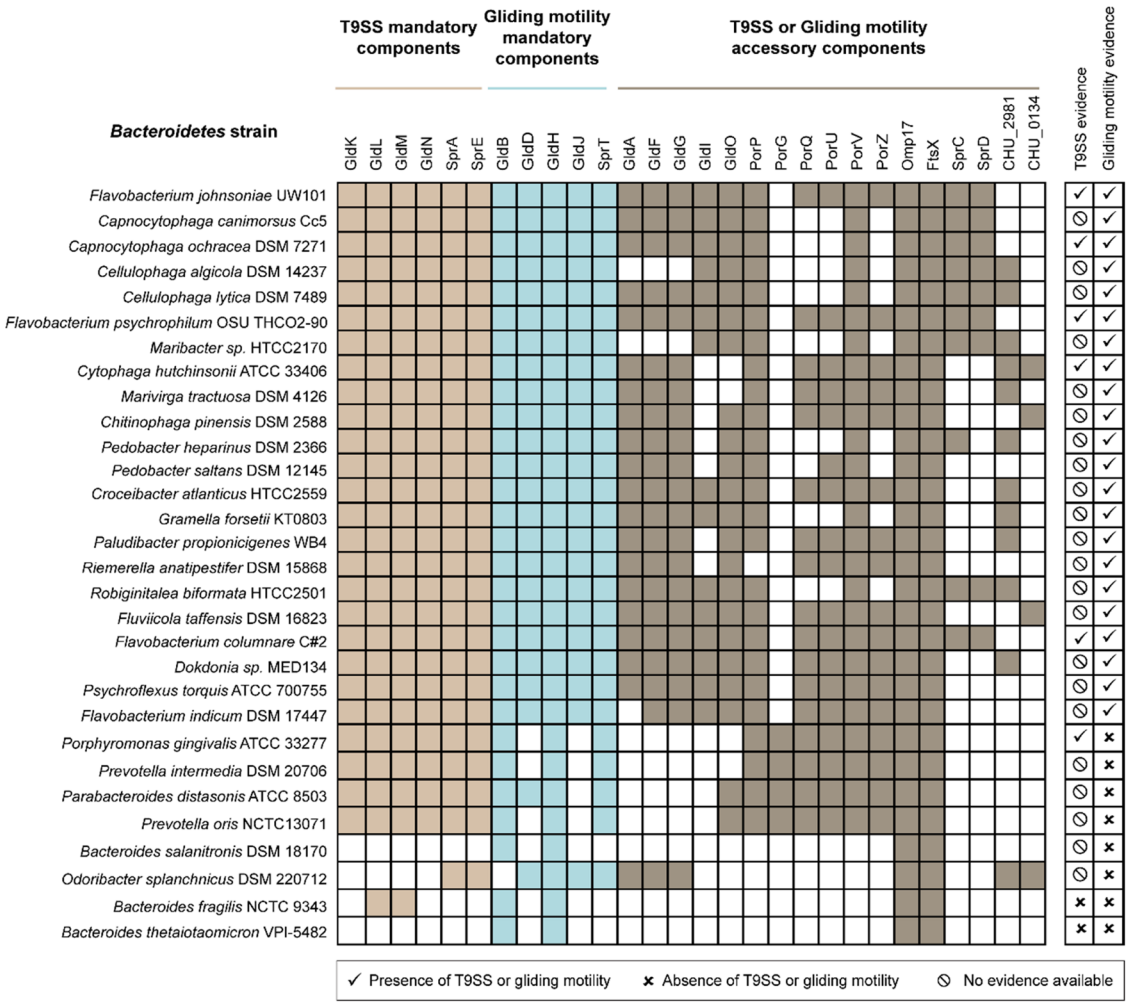


Figure 2. Occurrence of 28 identified protein components associated with T9SS or gliding motility checked across 30 experimentally characterized *Bacteroidetes* strains using the generated HMM profiles. The presence of T9SS mandatory components is denoted by a light brown colored box, gliding motility mandatory components are denoted by a cyan colored box, and T9SS or gliding motility accessory components are denoted by a dark brown colored box. The experimental evidence (presence, absence, or no evidence) for T9SS or gliding motility is denoted alongside each *Bacteroidetes* strain.

In addition to the 6 T9SS mandatory components, we observed that proteins GldB, GldD, GldH, GldJ, and SprT are present in the *Bacteroidetes* strains exhibiting gliding motility in comparison with those which lack gliding motility (Figure 2). Importantly, McBride and Zhu[3] had earlier characterized the presence of these proteins in *Bacteroidetes* strains exhibiting gliding motility. Therefore, we designated the 6 T9SS mandatory components along with GldB, GldD, GldH, GldJ, and SprT (a total of 11 proteins) as mandatory protein components to predict the presence of gliding motility. We designated the remaining 17 proteins as accessory components.

Computational Tool for Prediction of T9SS or Gliding Motility across *Bacteroidetes* Strains. We leveraged the generated HMM profiles to check for the presence of the 28 proteins associated with T9SS or gliding motility in the proteome of 693 completely sequenced *Bacteroidetes* strains (see Methods Section; Table S5). As discussed in the preceding section, if a bacterium contains all 6 T9SS mandatory components in its proteome, we classify it as having T9SS. In addition, if a bacterium contains all 11 gliding motility mandatory components, we classify it as having gliding motility. Using these criteria, we developed a tool in python programming language that employs the generated HMM profiles for 28 proteins to predict the presence of T9SS or gliding motility across 693 *Bacteroidetes* strains. Our tool identified 402 *Bacteroidetes* strains to have T9SS, of which 327 were found to also have gliding motility. Typically, we observed that 19 of the 28 proteins are present in the 402 *Bacteroidetes* strains predicted to have T9SS (Figure S1a), while 22 of the 28 proteins are present in the 327 *Bacteroidetes* strains predicted to have gliding motility (Figure S1b). In summary, we developed a computational tool that can predict the presence of both T9SS and gliding motility in members of the *Bacteroidetes* phylum, including mandatory and accessory protein components.

Validation of T9GPred Predictions of T9SS and Gliding Motility. We validated the T9GPred predictions of T9SS and gliding motility using the recently published experimental evidence (after August 2020) on T9SS and gliding motility that had not been considered while building T9GPred (Supporting Information, Figure S2). Note that we identified only 6 experimentally characterized *Bacteroidetes* strains to have T9SS (Table S4), and therefore, we could not split the compiled evidence into train and test sets for cross-

https://doi.org/10.1021/acsomega.3c05155
ACS Omega 2023, 8, 34091−34102

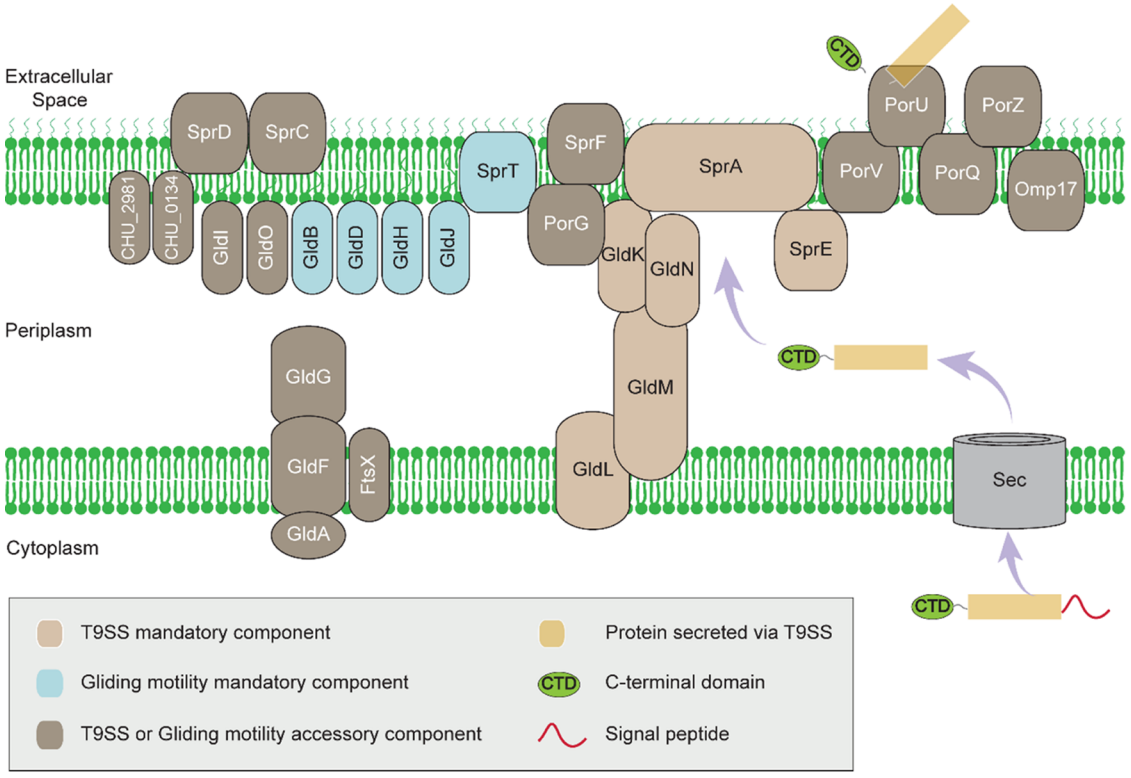


Figure 3. Schematic diagram encapsulating the available information on localization of 28 protein components on the bacterial membrane based on the published literature.

validation of T9GPred. In the published literature subsequent to August 2020, we identified 4 new *Bacteroidetes* strains to have T9SS (Table S6), in which all of the strains are predicted to have T9SS by T9GPred. Moreover, 2 of the 4 new strains are new species that were not part of compiled evidence used to build T9GPred. Similarly, we also identified 7 new *Bacteroidetes* strains to have gliding motility (Table S7) of which all of the strains are predicted to have gliding motility by T9GPred. Four of 7 new strains having gliding motility are new species, which were not considered while building the tool for gliding motility prediction. We noted that though the majority of the new evidence contained draft sequences, it did not limit T9GPred's ability for the prediction of T9SS or gliding motility (Tables S6 and S7). Further, we did not find any recently characterized protein components associated with T9SS or gliding motility in the published literature after August 2020.

Comparison of Predictions between T9GPred and TXSScan. We compared the features, predictions, and performance of T9GPred with an already existing computational tool TXSScan.[4,31] TXSScan predicts the presence of different secretion systems (T1SS−T9SS), including T9SS, whereas T9GPred is a tool specific to T9SS, which not only predicts the presence of T9SS but also predicts the associated gliding motility and secreted proteins. TXSScan's prediction of T9SS is based on the identification of mandatory protein components, such as GldK, GldL, GldM, GldN, SprE, SprA, SprT, and PorV (any 7 of 8 designated mandatory proteins) and their relative gene location on the genome. Comparatively, T9GPred has a simpler rule, which only considers the presence of 6 mandatory proteins (GldK, GldL, GldM, GldN, SprE, and SprA) for the prediction of T9SS. Note that SprT and PorV have not been considered as mandatory proteins for the prediction of T9SS by T9GPred as their function is not clearly elucidated or they are not essential for the secretion of all proteins.

We checked the TXSScan's prediction against the compiled list of 30 experimentally characterized *Bacteroidetes* strains used to build T9GPred (Table S8). In gliding *Bacteroidetes* strains, such as *Pedobacter heparinus* DSM 2366 and *Pedobacter saltans* DSM 12145, TXSScan predicted the presence of mandatory protein components but failed to report the presence of T9SS due to the gene location constraints. In the gliding *Bacteroidetes* strain *Fluviicola taffensis* DSM 16823, TXSScan failed to predict the presence of T9SS due to its inability to predict the presence of mandatory proteins. Though TXSScan predicts *P. gingivalis* ATCC 33277 to have T9SS, it fails to predict the presence of GldK, which has been experimentally characterized.[1,25] Based on these observations, we conclude that the T9SS prediction rule in TXSScan is more restrictive in comparison to T9GPred, and the incorporated HMM profiles in TXSScan seem to be missing known protein sequences.

Finally, we compared the performance of T9GPred and TXSScan on the validation set consisting of new experimentally characterized *Bacteroidetes* strains having T9SS or associated gliding motility. TXSScan predicted the presence of T9SS in 8 of the 10 identified strains (sensitivity is 0.8), whereas T9GPred predicted the presence of T9SS in all 10 strains (sensitivity is 1). Overall, T9GPred is more efficient and provides better predictions than TXSScan (Table 1).

Classification of the Secreted Proteins into Three C-Terminal Domain (CTD) Types. To classify the proteins secreted via T9SS, we relied on the CTD of the compiled list of 102 secreted proteins with published experimental evidence (see Methods Section). Earlier studies on T9SS have reported the motif sequences for the CTD of the secreted proteins,[32,39] while some have additionally reported the motif classifica-

https://doi.org/10.1021/acsomega.3c05155
ACS Omega 2023, 8, 34091−34102

Table 1. Comparative Analysis of the Predictive Tools T9GPred and TXSScan

	T9GPred	TXSScan
prediction of		
T9SS	√	√
gliding motility	√	×
secreted proteins	√	×
protein components		
total number of proteins	28	11
mandatory proteins	6 for T9SS, 11 for gliding motility prediction	8 for T9SS prediction
accessory proteins	17	3
prediction rule		
for T9SS	presence of all 6 mandatory proteins	presence of at least 7 mandatory proteins with gene location constraints
for gliding motility	presence of all 11 mandatory proteins	×
for secreted proteins	presence of 3 CTD types	×
performance		
predicted strains	10 of 10 strains are predicted to have T9SS	8 of 10 strains are predicted to have T9SS
sensitivity	1	0.8

tion.[8,30] We compiled a list of 18 CTD sequence motifs (see Methods Section; Table S9) from the published literature and generated a motif similarity network (MSN) (Figure S3). From the MSN, we observed that the unclassified motifs clustered with the motifs classified as type A, whereas the motifs classified as type B are distinct (Figure S3). In addition, Kharade and McBride[11] reported a third type of CTD motif (type C) for the secreted proteins. Therefore, we accessed the HMM profiles from TIGRFAM[40] and NCBI protein family models[41] for all three types of CTD (type A, type B, and type C) to classify the compiled list of 102 secreted proteins. We observed that the HMM profiles and the motif-based classification of the 102 secreted proteins yielded similar results with 68 proteins as type A, 18 proteins as type B, 1 protein as type C, and 15 proteins as unclassified (Figure 4). From the classified CTD sequences, we generated HMM profiles (Supporting Information, Section 1.2). Additionally, we report CTD motif sequences specific to the classified proteins (Table 2). Importantly, these motif sequences are better at classifying the different types of CTD in comparison to the existing sequence motifs (Figure 4).

Computational Pipeline to Predict Proteins Secreted via T9SS. We present the first computational pipeline to predict proteins secreted via T9SS across members of the *Bacteroidetes* phylum (Figure 5). The pipeline integrates several tools that predict important features in protein sequences. We used SignalP-5.0[46] and Phobius[47] to detect the N-terminal Sec signal, SignalP-5.0 to detect N-terminal Tat signal, Phobius and TMHMM 2.0[48] to detect the transmembrane (TM) domain and used the generated CTD HMM profiles to detect the CTD types in protein sequences. We used "OR rule" in the cases, where the feature is predicted by multiple tools. Based on the features of 102 experimentally identified proteins

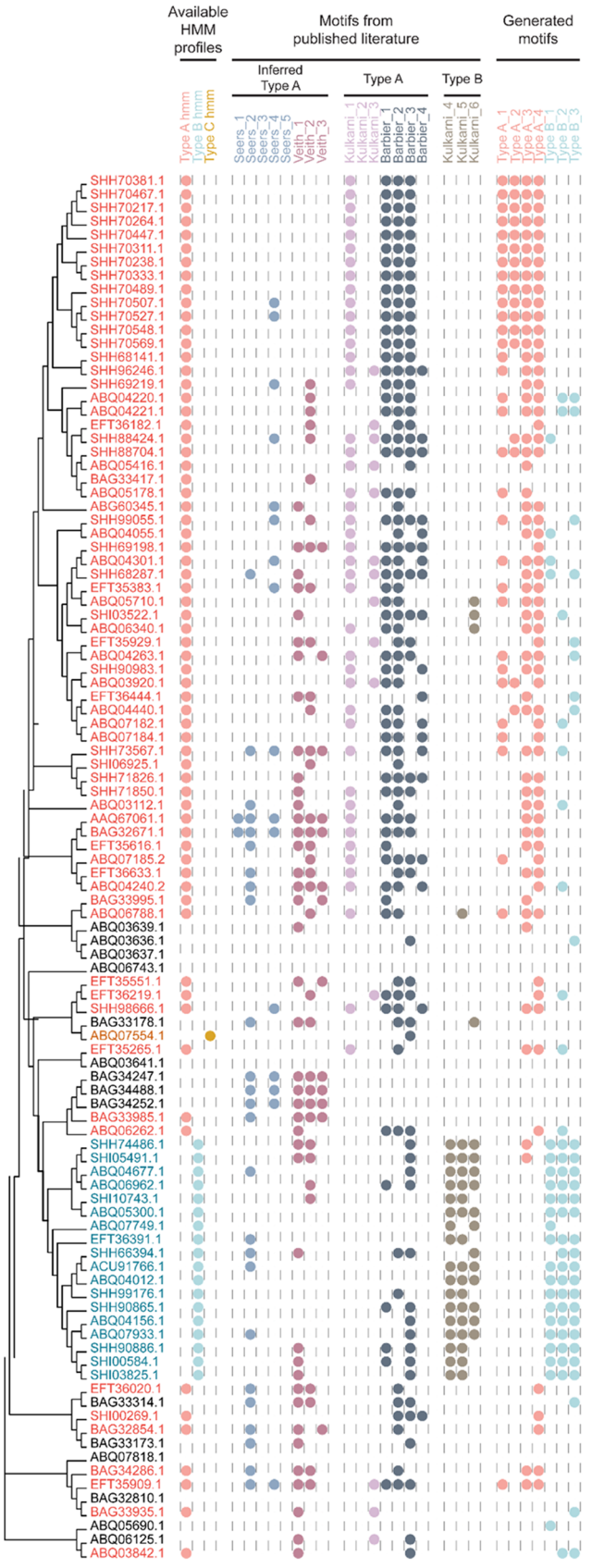


Figure 4. Identification and classification of the C-terminal domain (CTD; last 120 aa) of 102 experimentally characterized secreted proteins into three CTD types, namely, type A (red), type B (cyan), and type C (yellow), based on HMM profiles and sequence motifs from the published literature. The CTD sequences that could not be

https://doi.org/10.1021/acsomega.3c05155
ACS Omega 2023, 8, 34091−34102

Figure 4. continued

classified are designated as unclassified and shown in black. The motifs from Seers et al.[39] and Veith et al.[32] are classified as type A CTD based on the motif similarity network (Figure S3). The CTDs of the secreted proteins are clustered using ClustalW[42] and the colored dots represent the presence of the CTD types. Based on the classification of the secreted proteins, we generated 4 motifs for type A and 3 motifs for type B. The figure was generated using Evolview[43,44] web server.

Table 2. Table Lists Four Type A and Three Type B CTD Motifs Generated from the Input List of Experimentally Characterized Secreted Proteins Using "Discriminative Mode" in MEME Tool[45]

motif identifier	motif	*E*-value
type A_1	KIYPNPVSEILNIALQE[GN]LQL[EQ] KVNFYNTLGQLIKTTNHSE[IT]N[VI]SSFAKG	7.3e-454
type A_2	VF[GR][ND]V[TN][QK]SNCALNVP[AT]GT[EQ] A[AV]YQAAAVW[KR][DN]FSPISG[SN] LLSNHSFAIES[NA]L	2.6e-398
type A_3	VK[VI]YPN[PQ]GKGT[LK][TN]I[SI]	1.3e-158
type A_4	[LV]SNLA[KS]G[VI]Y[IFL][VL]K[IV]XX	7.4e-115
type B_1	F[TS]PNGDGYNDT[WF]Y[IP]E[NG][IL]EXYP	2.5e-091
type B_2	ELP[SA]G[DT]Y[WY][YF][VI][LVI]K[YF]N[DE] N[KN][NT]XK	6.4e-067
type B_3	[VI]EI[FY][DN]R[YWF]G[KRV]L[IV][KY] ELDG[NYG]DNG[WD]	6.1e-070

secreted via T9SS (Table S3), we designate a protein as secreted protein if (i) it has Sec signal in the N-terminal region; (ii) it has no Tat signal in the N-terminal region; (iii) it has no TM domain; and (iv) it has at least one of the three CTDs, namely, type A, type B, or type C in its C-terminal region. We developed a computational tool implementing the above pipeline (Figure 5) to predict secreted proteins from an input proteome. Note that, we relied only on the 3 classified CTD types to get high-confidence predictions. However, the pipeline can be easily augmented to incorporate more such CTD types in the future. Finally, we used this tool to predict the secreted proteins for the 402 *Bacteroidetes* strains (Table S5) predicted to have T9SS. The predictions are available under the 'Secreted protein prediction' section of respective *Bacteroidetes* strain at: https://cb.imsc.res.in/t9gpred/.

Validation of T9GPred Predictions of Secreted Proteins. We validated the secreted protein prediction from T9GPred using recently published experimental evidence (after August 2020) on secreted proteins (Supporting Information, Figure S2). We identified 22 new secreted proteins that were experimentally characterized from 4 different strains (Table S10). T9GPred predicted 14 of these 22 sequences to be secreted via T9SS. Among the 8 sequences that were not predicted by T9GPred, 2 lacked Sec signal sequences but contained type A CTD signal, while the other 6 sequences did not have any of 3 CTD signals. Additionally, we estimated the precision of our CTD HMM prediction using 2 strategies (see Methods Section). When checked against a total of 16,886 secreted proteins predicted by T9GPred, the first strategy always resulted in zero hits, while the second strategy resulted in 972, 970, and 948 hits (average of 963 hits), yielding the precision of 100% and 94.29%, respectively (see Methods Section).

Comparative Analysis of the Secreted Proteins from Gliding and Nongliding *Bacteroidetes* Strains. Using our computational tool, we predicted 327 *Bacteroidetes* strains to have gliding motility (gliding *Bacteroidetes*) and 75 *Bacteroidetes* strains to lack gliding motility (nongliding *Bacteroidetes*) (see Methods Section). Further, we also predicted the secreted proteins for both gliding and nongliding *Bacteroidetes* using our computational tool. From our predictions, we noted that gliding *Bacteroidetes* secretes more proteins than the nongliding *Bacteroidetes*. To check for the statistical significance of this

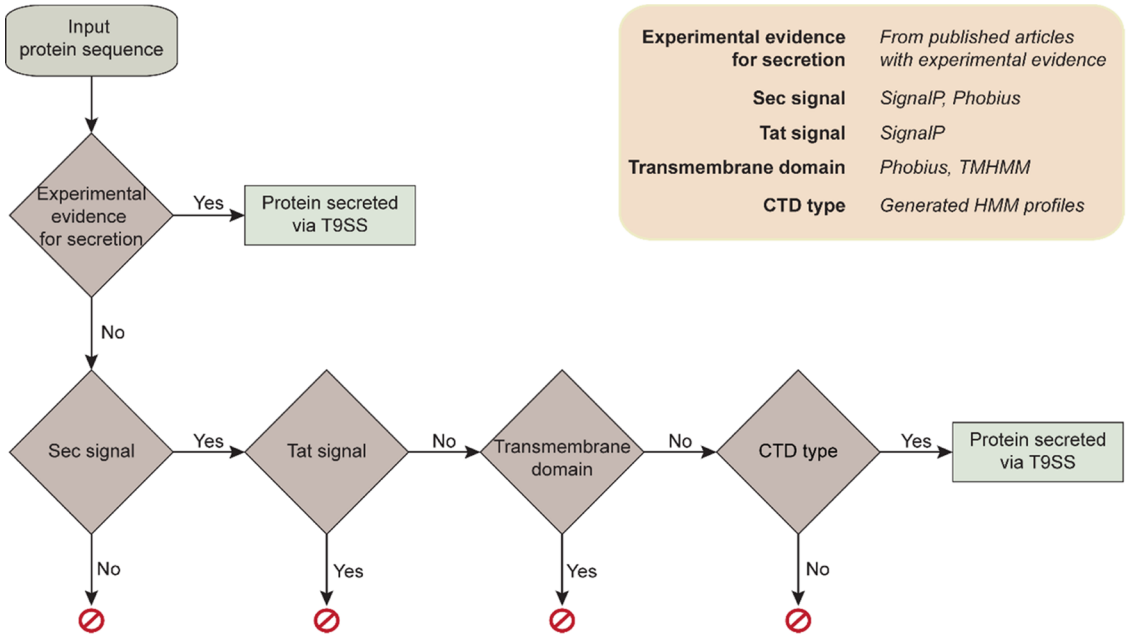


Figure 5. Computational pipeline to predict proteins secreted via T9SS. Computational tools used in this pipeline to predict features in the input protein sequence are mentioned in the top right panel of the figure.

https://doi.org/10.1021/acsomega.3c05155
ACS Omega 2023, 8, 34091−34102

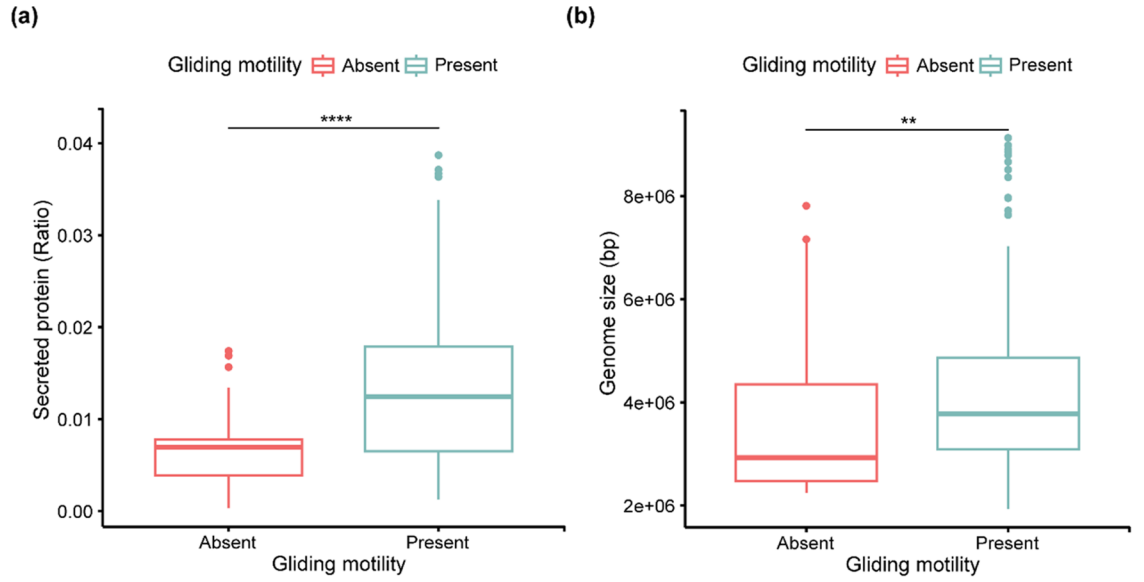


Figure 6. Comparative analysis of the secreted proteins and genome size between gliding and nongliding *Bacteroidetes* strains predicted in this study. (a) Comparison of the predicted secreted proteins normalized with the total number of coding sequences of respective bacteria (ratio) between gliding and nongliding *Bacteroidetes* strains (significantly different at $p < 0.0001$). (b) Comparison of the genome sizes (bp) between gliding and nongliding *Bacteroidetes* strains (significantly different at $p < 0.01$). In each case, the significant difference between two samples is calculated using Bonferroni correction in the rstatix package[49] available in R version 4.2.1. Both plots are made using the ggplot2[50] package in R version 4.2.1.

difference, we first normalized the number of secreted proteins with the total number of proteins (coding sequences) in the respective *Bacteroidetes* strains. We observed that gliding *Bacteroidetes* significantly secretes more proteins than nongliding *Bacteroidetes* ($p < 0.0001$; corrected using the Bonferroni method; Figure 6a). We also observed that the genome of gliding *Bacteroidetes* is significantly larger than nongliding *Bacteroidetes* ($p < 0.01$; corrected using the Bonferroni method; Figure 6b). A similar trend is observed between nongliding and gliding strains predicted from ∼34,000 completely sequenced bacteria (Figure S4).

T9GPred Web Server. We have designed a Findable, Accessible, Interoperable, and Reusable (FAIR)[51] compliant user-friendly web server, **T**ype **9** secretion system and **G**liding motility **Pred**iction (T9GPred) that compiles the predictions and the computational tool developed in this study. The web server is available at: https://cb.imsc.res.in/t9gpred/. T9GPred is an organized web server with a simple navigation that enables easy access to users (Figure S5a).

Users can interactively browse for the 693 *Bacteroidetes* strains in the "Browse" section of the navigation bar. Users can hover over the heatmap to see information on the name of the bacteria, the predicted protein components, and the number of predicted sequences of the protein component in the proteome of the bacteria (Figure S5b). Users can click on the name of the bacteria to see further details. The "Bacteroidetes details" section gives the information on the organism name, genome identifier from NCBI[52] or BV-BRC,[53] isolation source of the bacteria, sequencing center information, predictions on T9SS, gliding motility and the number of secreted proteins, taxonomic information, and proteome of the bacteria as a downloadable file (Figure S5c). Users can view the predicted protein components associated with T9SS on the "T9SS prediction" page (Figure S5d) and the predicted protein components associated with gliding motility on the 'Gliding motility prediction' page (Figure S5e). For bacteria predicted to have T9SS, the predicted components can be visualized on the genome of the bacteria on the "Gene cluster visualization" page (Figure S5f). Finally, the "Secreted protein prediction" page lists the predicted secreted proteins and their annotations including the protein identifier from BV-BRC[53] or NCBI,[54] UniProt[55] identifier, AlphaFold DB[56,57] identifier, type of evidence, type of CTD, and the presence of Carbohydrate-Active enZYmes (CAZy) domain predicted using dbCAN2 tool[58] (Figure S5g).

Users can access the prediction tools in the "Prediction" section of the navigation bar. The 'Prediction of T9SS and gliding motility' page contains the tool for the prediction of T9SS or gliding motility in the input proteome (Figure S6a). Users can upload the proteome in a fasta file format or paste the sequence in the provided text box to run the prediction tool. The tool returns the predicted protein components along with the decision for the presence of T9SS or gliding motility (Figure S6a). The "Prediction of protein secreted via T9SS" page allows users to predict whether the input protein can be secreted via T9SS (Figure S6b). Users can paste the sequence in the provided text box to run the prediction tool. The tool returns the prediction of different features, including the classification of CTD types, and mentions whether the protein can be secreted via T9SS (Figure S6b).

CONCLUSIONS

T9SS is one of the least characterized secretion systems in Gram-negative bacteria (Supporting Information, Section 3; Figure S7). In this study, we leveraged the existing experimental evidence to develop the first comprehensive computational tool, T9GPred, which predicts the presence of T9SS, associated gliding motility, and secreted proteins from bacterial proteome. In particular, we designated 6 key proteins (GldK, GldL, GldM, GldN, SprA, and SprE) as mandatory for a functional T9SS and observed that they are sufficient for T9SS prediction when checked against the proteome of

https://doi.org/10.1021/acsomega.3c05155
ACS Omega 2023, 8, 34091−34102

recently characterized *Bacteroidetes* strains having T9SS (Table S6). Using these 6 T9SS mandatory proteins, we predicted 402 of the 693 completely sequenced *Bacteroidetes* strains to have T9SS. Additionally, we also observed that our tool T9GPred predicted the presence of T9SS only in the *Bacteroidetes* phylum (Supporting Information, Section 4; Table S11), which is in line with the findings by McBride and colleagues.[1,3] Further, we also observed that the T9SS mandatory protein components are clustered in the bacterial genome (Supporting Information, Section 5; Figure S8), suggesting that GldK, GldL, GldM, and GldN form an operon in members of the *Bacteroidetes* phylum having T9SS.[12,25]

In some *Bacteroidetes* strains, T9SS also aids in the movement of the species, which is termed as gliding motility.[1,3] In addition to the 6 T9SS mandatory proteins, we designated 5 key proteins (GldB, GldD, GldH, GldJ, and SprT) as mandatory for exhibiting gliding motility and observed that they are sufficient for gliding motility prediction when checked against the proteome of recently characterized *Bacteroidetes* strains exhibiting gliding motility (Table S7). Using these 11 gliding motility mandatory proteins, we predicted 327 of 402 *Bacteroidetes* strains with T9SS to also have gliding motility. Moreover, we observed that 44 of these 327 strains were predicted to have type IVa pili, and none of them had flagellar motor or type IVb pili (Supporting Information, Section 6). We also noted that there is no overlap between the protein components associated with gliding motility and flagellar motor or type IV pili (Supporting Information, Section 6). Thus, our analysis suggests that gliding motility associated with T9SS provides a unique machinery that is different from other types of bacterial motility.

The proteins secreted via T9SS are involved in multiple processes, including the survival and pathogenicity of the bacteria.[13,29] We identified 3 types of C-terminal domain (CTD) signals (type A, type B, or type C) by leveraging the reported signal sequences[8,30,32,39] in conjunction with published evidence. We leveraged this CTD classification to build a computational pipeline for secreted protein prediction and found that its predictions are highly precise and agree well with recently published experimental evidence (Table S10). A comparative analysis with the proteins secreted via other *Bacteroidetes*-specific secretion systems highlighted the uniqueness of the CTD signal in the proteins secreted via T9SS (Supporting Information, Section 7). We found that members of the *Bacteroidetes* phylum exhibiting gliding motility have significantly larger genome sizes and secrete more proteins in comparison with the members that do not have gliding motility (Figures 6 and S4). This suggests that members of the *Bacteroidetes* phylum exhibiting gliding motility have a larger genome probably encoding more specialized proteins to aid in gliding motility.

T9GPred has its fair share of limitations like most computational tools. T9GPred does not mention the accuracy of the predictions, as it lacks information on negative control. It relies on protein profiles trained on a small set of experimentally characterized *Bacteroidetes* strains to date, which limits its ability to exhaustively capture all variations. Moreover, T9GPred does not comment on the maturation of the secreted proteins, as it does not include the information on the functional characterization of accessory protein components. The secreted protein prediction by T9GPred is limited by the choice of 3 CTD types considered in this study, which is not indicative of all possible secreted proteins.[8] Moreover, T9GPred does not include the structural conservation of the CTD[59] but considers only the protein sequence for the prediction of the secreted proteins. T9GPred predicts putative secreted proteins but does not comment on the variation in the levels of secreted proteins in different environments.

Nevertheless, T9GPred aids in streamlining experiments by providing quick predictions. For pathogens, the predicted T9SS protein components and secreted proteins could be targeted for drug discovery. For nonpathogens, T9GPred predicts secreted proteins that could be the putative subjects for genetic enhancements. Additionally, the secreted protein pipeline in T9GPred can aid in the design of industrially relevant recombinant proteins. Importantly, T9GPred has the potential to provide robust predictions in draft genome sequences. T9GPred has the ability to identify a new uncharacterized bacterium as belonging to the *Bacteroidetes* phylum. In conclusion, we developed T9GPred, a computational tool that not only predicts T9SS but also predicts T9SS-based gliding motility and the proteins secreted via T9SS. We have made the predictions and the tool available on our web server, which is accessible at: https://cb.imsc.res.in/t9gpred/. The source code of the computational tool is available at: https://github.com/asamallab/T9GPred.

METHODS

Workflow to Identify Proteins Associated with T9SS or Gliding Motility among Members of the *Bacteroidetes* Phylum. We mined PubMed[60] to compile an exhaustive list of experimentally characterized proteins associated with T9SS or gliding motility in members of the *Bacteroidetes* phylum from the published literature. We queried PubMed using the following keywords:

("type 9 secretion system"[Title/Abstract] OR "type 9 secretion systems"[Title/Abstract] OR "type IX secretion system"[Title/Abstract] OR "type IX secretion systems"[Title/Abstract] OR "type 9 secretion system"[Title/Abstract] OR "type 9 secretion systems"[Title/Abstract] OR "type IX secretion system"[Title/Abstract] OR "type IX secretion systems"[Title/Abstract] OR "t9ss"[Title/Abstract] OR "T9SS"[Title/Abstract] OR "PoRSS"[Title/Abstract] OR "PoR secretion system"[Title/Abstract] OR "Type 9 secretion"[Title/Abstract] OR "Type IX secretion"[Title/Abstract])

and retrieved 107 published articles on August 26, 2020 (Figure 1). We noted that our PubMed query was unable to retrieve some previous studies focusing on gliding motility and the type 9 based secretion, as such studies were published prior to the use of T9SS nomenclature. Therefore, we manually compiled 22 additional published articles on gliding motility or type 9 based secretion (Figure 1). Thereafter, we excluded the articles that are computational studies or reviews or which did not contain mutation-based experiments. Finally, we shortlisted 47 published articles from which we identified 23 proteins associated with T9SS (Table S1), 22 proteins associated with gliding motility (Table S2), and 102 proteins secreted via T9SS (secreted proteins) in members of the *Bacteroidetes* phylum (Table S3). Figure 1 summarizes the workflow according to the PRISMA statement.[33]

Compilation of Completely Sequenced *Bacteroidetes* Strains. The aim of our study is to predict the presence of T9SS or gliding motility across the *Bacteroidetes* phylum. Thus, we considered *Bacteroidetes* strains for which the complete

https://doi.org/10.1021/acsomega.3c05155
ACS Omega 2023, 8, 34091−34102

genome is available. We queried NCBI assembly[52] using the following keywords:

("Bacteroidetes"[Organism] or "Bacteroidia"[Organism] or Bacteroidetes[All Fields]) and (bacteria[filter] and "latest refseq"[filter] and "complete genome"[filter] and all[filter] not anomalous[filter])

and identified 693 completely sequenced *Bacteroidetes* strains on December 21, 2020 (Table S5). For these 693 *Bacteroidetes* strains, we accessed the corresponding proteome from BV-BRC[53,61] or NCBI assembly.[52] Note that if the proteome for a bacterium was available in both BV-BRC and NCBI assembly, we preferred the proteome from BV-BRC.

Computational Analysis of Secreted Proteins. We collected the sequences of the experimentally identified 102 secreted proteins (Table S3) from NCBI GenBank.[54] We considered the last 120 amino acids of each protein sequence as their C-terminal domain (CTD).[8] Further, we used ClustalW[42] with default parameters to align the CTD sequences of the 102 secreted proteins. Thereafter, we clustered the aligned sequences using MEGA X[62] with the maximum-likelihood method and default parameters.

We relied on the HMM profile-based prediction and motif analysis to classify the 102 CTD sequences. We accessed the HMM profiles for CTD types from TIGRFAM[40] and NCBI protein family models[41] and compiled 18 CTD sequence motifs of secreted proteins from the published literature (Table S9). To find the similarity between the motifs, we created a motif similarity network (MSN) using Tomtom[63,64] by setting the overlap ≥3 and score ≥2[65] and visualized the MSN using Cytoscape[66] (Figure S3). Further, we determined the presence of the motifs in the CTD of the secreted proteins using Find Individual Motif Occurrence (FIMO) tool[67,68] with default parameters. We classified the CTD types of secreted proteins based on the HMM prediction and substantiated with motif-based classification. Using the classified CTD sequences, we created the corresponding HMM profiles (Supporting Information, Section 1.2) and generated motifs using the Multiple Expectation Maximization for Motif Elicitation (MEME)[45,69] tool in the discriminative mode of motif discovery.

Checking for Random Hits of CTD HMM Profiles. We followed Veith et al.[32] to check for random hits of the 3 CTD HMM profiles (type A, type B, and type C) on proteome of the 402 *Bacteroidetes* strains using 2 different strategies: (i) randomizing the CTD regions (last 120 amino acids) of each protein while maintaining amino acid composition and (ii) randomly selecting 120 amino acid long sequence from the full sequence of each protein. Since we considered the CTD sequence length as 120 amino acids, we filtered out those predicted protein sequences that were less than or equal to 120 amino acids to be consistent with the randomization strategy. We generated three data sets for each strategy to estimate the number of random hits. We then used the following formula to calculate the precision

$$\text{precision} = \frac{\text{number of true positives}}{\text{number of total predictions}}$$

where we estimated the number of true positives as

$$\text{number of true positives} = \text{number of total predictions} - \text{number of random hits}$$

Web Interface and Web Server Management System. We created a user-friendly web server **T**ype **9** secretion system and **G**liding motility **Pred**iction (T9GPred) to compile and share the predictions and the computational tool generated in this study. T9GPred compiles the proteins associated with T9SS or gliding motility in 693 completely sequenced *Bacteroidetes* strains. For the *Bacteroidetes* strains predicted to have T9SS, T9GPred lists the proteins predicted to be secreted via T9SS. Moreover, T9GPred provides a prediction tool to check for the presence of T9SS or gliding motility and secreted proteins in any sequenced genome. T9GPred is openly available at https://cb.imsc.res.in/t9gpred/.

To create the web server, we used MariaDB[70] to store the compiled information and Structured Query Language (SQL) to retrieve information. We used PHP[71] with custom HTML, CSS, jQuery,[72] Bootstrap 4,[73] and Plotly JavaScript[74] to create the web interface of T9GPred. T9GPred web server is hosted on an Apache[75] server running on the Debian 9.4 Linux Operating System.

■ ASSOCIATED CONTENT

Data Availability Statement

The predictions and the computational tool are accessible via the web server, T9GPred available at: https://cb.imsc.res.in/t9gpred/. The standalone version of the developed computational tool and the codes to generate different plots are available in our GitHub repository: https://github.com/asamallab/T9GPred.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.3c05155.

Distribution of the number of proteins associated with T9SS or gliding motility in each of the 402 *Bacteroidetes* strains predicted to have T9SS (Figure S1); workflow to compile new experimental evidence on T9SS, gliding motility, and associated secreted proteins (Figure S2); motif similarity network (MSN) of the 18 compiled CTD motifs (Figure S3); comparative analysis of the secreted proteins and genome size between gliding and nongliding strains identified from ∼34,000 completely sequenced bacteria (Figure S4); snapshots of the web interface of the T9GPred web server (Figure S5); snapshots of the prediction tool in T9GPred web server (Figure S6); cumulative estimate of published research articles for the nine different secretion systems (T1SS−T9SS) in bacteria (Figure S7); gene cluster visualization of the proteins associated with T9SS or gliding motility in the genome of *F. johnsoniae* UW101 and *P. gingivalis* ATCC 33277 (Figure S8); and flowchart describing the workflow to create the HMM profile from protein sequences (Figure S9) (PDF)

Compiled list of 23 proteins with experimental evidence of being associated with T9SS in the members of *Bacteroidetes* phylum (Table S1); compiled list of 22 proteins with experimental evidence of being associated with gliding motility in the members of *Bacteroidetes* phylum (Table S2); compiled list of 102 proteins with experimental evidence of being secreted via T9SS in the members of *Bacteroidetes* phylum (Table S3); compiled list of 30 *Bacteroidetes* strains with experimental evidence on the presence or absence of T9SS and gliding motility (Table S4); complied list of 693 completely sequenced

https://doi.org/10.1021/acsomega.3c05155
ACS Omega 2023, 8, 34091−34102

Bacteroidetes strains considered in this study (Table S5); compiled list of 4 new *Bacteroidetes* strains with experimental evidence on the presence of T9SS (Table S6); compiled list of 7 new *Bacteroidetes* strains with experimental evidence on the presence of gliding motility (Table S7); prediction of T9SS and associated protein components by TXSScan across 30 experimentally characterized *Bacteroidetes* strains (Table S8); compiled list of 18 sequence motifs for the C-terminal domain (CTD) of the proteins secreted via T9SS (Table S9); compiled list of experimentally characterized 22 new secreted proteins (Table S10); compiled list of completely sequenced bacteria from NCBI assembly (Table S11); compiled list of 43 *Bacteroidetes* strains used to build protein HMM profiles (Table S12); true positive hits for type C CTD (Table S13); and PubMed query for each of the nine different secretion systems in bacteria (Table S14) (XLSX)

AUTHOR INFORMATION

Corresponding Author

Areejit Samal − *The Institute of Mathematical Sciences (IMSc), Chennai 600113, India; Homi Bhabha National Institute (HBNI), Mumbai 400094, India;* orcid.org/0000-0002-6796-9604; Email: asamal@imsc.res.in

Authors

Ajaya Kumar Sahoo − *The Institute of Mathematical Sciences (IMSc), Chennai 600113, India; Homi Bhabha National Institute (HBNI), Mumbai 400094, India;* orcid.org/0000-0003-3543-8021

R. P. Vivek-Ananth − *The Institute of Mathematical Sciences (IMSc), Chennai 600113, India; Homi Bhabha National Institute (HBNI), Mumbai 400094, India;* orcid.org/0000-0002-3232-3299

Nikhil Chivukula − *The Institute of Mathematical Sciences (IMSc), Chennai 600113, India; Homi Bhabha National Institute (HBNI), Mumbai 400094, India;* orcid.org/0000-0002-8335-9427

Shri Vishalini Rajaram − *The Institute of Mathematical Sciences (IMSc), Chennai 600113, India; Centre for Biotechnology, Anna University, Chennai 600025, India;* orcid.org/0000-0002-2446-9642

Karthikeyan Mohanraj − *Institute for Clinical Chemistry and Laboratory Medicine, Technische Universität Dresden, Dresden 01307, Germany;* orcid.org/0000-0003-2792-0627

Devanshi Khare − *Molecular Biology Division, Bhabha Atomic Research Centre, Mumbai 400085, India; Homi Bhabha National Institute (HBNI), Mumbai 400094, India*

Celin Acharya − *Molecular Biology Division, Bhabha Atomic Research Centre, Mumbai 400085, India; Homi Bhabha National Institute (HBNI), Mumbai 400094, India*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.3c05155

Author Contributions

A.K.S.: Conceptualization, data compilation, data curation, formal analysis, software, visualization, and writing; R.P.V.-A.: conceptualization, formal analysis, software, visualization, and writing; N.C.: conceptualization, formal analysis, software, visualization, and writing; S.V.R.: conceptualization, data compilation, data curation, and formal analysis; K.M.: software and visualization; D.K.: formal analysis; C.A.: conceptualization, formal analysis, and writing; A.S.: conceptualization, supervision, formal analysis, software, and writing.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank Gokul Balaji Dhanakoti and B. Raveendra Reddy for computational support and Kishan Kumar for help in generating a figure. They thank Anshu Bhardwaj, Dhiraj Kumar, Mitali Mukerji, Vinay Nandicoori, Hema Rajaram, and Smita Srivastava for discussions. Areejit Samal would like to acknowledge support from the Department of Atomic Energy (DAE), Government of India (GoI), the Science and Engineering Research Board (SERB), GoI [Ramanujan Fellowship SB/S2/RJN-006/2014], and the Max Planck Society, Germany [Max Planck Partner Group in Mathematical Biology]. The funders have no role in the study design, data collection, data analysis, manuscript preparation, or decision to publish.

REFERENCES

(1) Sato, K.; Naito, M.; Yukitake, H.; Hirakawa, H.; Shoji, M.; McBride, M. J.; Rhodes, R. G.; Nakayama, K. A Protein Secretion System Linked to Bacteroidete Gliding Motility and Pathogenesis. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107* (1), 276−281.

(2) Sato, K.; Yukitake, H.; Narita, Y.; Shoji, M.; Naito, M.; Nakayama, K. Identification of Porphyromonas Gingivalis Proteins Secreted by the Por Secretion System. *FEMS Microbiol. Lett.* **2013**, *338* (1), 68−76.

(3) McBride, M. J.; Zhu, Y. Gliding Motility and Por Secretion System Genes Are Widespread among Members of the Phylum Bacteroidetes. *J. Bacteriol.* **2013**, *195* (2), 270−278.

(4) Abby, S. S.; Cury, J.; Guglielmini, J.; Néron, B.; Touchon, M.; Rocha, E. P. C. Identification of Protein Secretion Systems in Bacterial Genomes. *Sci. Rep.* **2016**, *6* (1), No. 23080.

(5) Thomas, F.; Hehemann, J.-H.; Rebuffet, E.; Czjzek, M.; Michel, G. Environmental and Gut Bacteroidetes: The Food Connection. *Front. Microbiol.* **2011**, *2*, No. 93, DOI: 10.3389/fmicb.2011.00093.

(6) Wexler, H. M. Bacteroides: The Good, the Bad, and the Nitty-Gritty. *Clin. Microbiol. Rev.* **2007**, *20* (4), 593−621.

(7) Li, N.; Zhu, Y.; LaFrentz, B. R.; Evenhuis, J. P.; Hunnicutt, D. W.; Conrad, R. A.; Barbier, Paul.; Gullstrand, C. W.; Roets, J. E.; Powers, J. L.; Kulkarni, S. S.; Erbes, D. H.; García, J. C.; Nie, P.; McBride, M. J. The Type IX Secretion System Is Required for Virulence of the Fish Pathogen Flavobacterium Columnare. *Appl. Environ. Microbiol.* **2017**, *83* (23), No. e01769-17, DOI: 10.1128/AEM.01769-17.

(8) Barbier, P.; Rochat, T.; Mohammed, H. H.; Wiens, G. D.; Bernardet, J.-F.; Halpern, D.; Duchaud, E.; McBride, M. J. The Type IX Secretion System Is Required for Virulence of the Fish Pathogen Flavobacterium Psychrophilum. *Appl. Environ. Microbiol.* **2020**, *86* (16), No. e00799-20, DOI: 10.1128/AEM.00799-20.

(9) Khare, D.; Chandwadkar, P.; Acharya, C. Gliding Motility of a Uranium-Tolerant Bacteroidetes Bacterium Chryseobacterium Sp. Strain PMSZPI: Insights into the Architecture of Spreading Colonies. *Environ. Microbiol. Rep.* **2022**, *14* (3), 453−463.

(10) Gao, L.; Guan, Z.; Gao, P.; Zhang, W.; Qi, Q.; Lu, X. Cytophaga Hutchinsonii GldN, Encoding a Core Component of the Type IX Secretion System, Is Essential for Ion Assimilation, Cellulose Degradation, and Cell Motility. *Appl. Environ. Microbiol.* **2020**, *86* (11), No. e00242-20, DOI: 10.1128/AEM.00242-20.

(11) Kharade, S. S.; McBride, M. J. *Flavobacterium Johnsoniae* Chitinase ChiA Is Required for Chitin Utilization and Is Secreted by

the Type IX Secretion System. *J. Bacteriol.* **2014**, *196* (5), 961−970, DOI: 10.1128/JB.01170-13.
(12) Shrivastava, A.; Johnston, J. J.; van Baaren, J. M.; McBride, M. J. *Flavobacterium Johnsoniae* GldK, GldL, GldM, and SprA Are Required for Secretion of the Cell Surface Gliding Motility Adhesins SprB and RemA. *J. Bacteriol.* **2013**, *195* (14), 3201−3212.
(13) Lasica, A. M.; Ksiazek, M.; Madej, M.; Potempa, J. The Type IX Secretion System (T9SS): Highlights and Recent Insights into Its Structure and Function. *Front. Cell. Infect. Microbiol.* **2017**, *7*, No. 215, DOI: 10.3389/fcimb.2017.00215.
(14) Zhu, Y.; McBride, M. J. The Unusual Cellulose Utilization System of the Aerobic Soil Bacterium Cytophaga Hutchinsonii. *Appl. Microbiol. Biotechnol.* **2017**, *101* (19), 7113−7127.
(15) Benedyk, M.; Marczyk, A.; Chruścicka, B. Type IX Secretion System Is Pivotal for Expression of Gingipain-Associated Virulence of *Porphyromonas Gingivalis*. *Mol. Oral Microbiol.* **2019**, *34* (6), 237−244.
(16) Burdette, L. A.; Leach, S. A.; Wong, H. T.; Tullman-Ercek, D. Developing Gram-Negative Bacteria for the Secretion of Heterologous Proteins. *Microb. Cell Fact.* **2018**, *17* (1), No. 196, DOI: 10.1186/s12934-018-1041-5.
(17) Sato, K.; Sakai, E.; Veith, P. D.; Shoji, M.; Kikuchi, Y.; Yukitake, H.; Ohara, N.; Naito, M.; Okamoto, K.; Reynolds, E. C.; Nakayama, K. Identification of a New Membrane-Associated Protein That Influences Transport/Maturation of Gingipains and Adhesins of Porphyromonas Gingivalis. *J. Biol. Chem.* **2005**, *280* (10), 8668−8677.
(18) Agarwal, S.; Hunnicutt, D. W.; McBride, M. J. Cloning and Characterization of the *Flavobacterium Johnsoniae* (Cytophaga Johnsonae) Gliding Motility Gene, GldA. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94* (22), 12139−12144.
(19) Braun, T. F.; Khubbar, M. K.; Saffarini, D. A.; McBride, M. J. *Flavobacterium Johnsoniae* Gliding Motility Genes Identified by Mariner Mutagenesis. *J. Bacteriol.* **2005**, *187* (20), 6943−6952.
(20) Nelson, S. S.; Bollampalli, S.; McBride, M. J. SprB Is a Cell Surface Component of the *Flavobacterium Johnsoniae* Gliding Motility Machinery. *J. Bacteriol.* **2008**, *190* (8), 2851−2857.
(21) Shrivastava, A.; Rhodes, R. G.; Pochiraju, S.; Nakane, D.; McBride, M. J. *Flavobacterium Johnsoniae* RemA Is a Mobile Cell Surface Lectin Involved in Gliding. *J. Bacteriol.* **2012**, *194* (14), 3678−3688.
(22) Lauber, F.; Deme, J. C.; Lea, S. M.; Berks, B. C. Type 9 Secretion System Structures Reveal a New Protein Transport Mechanism. *Nature* **2018**, *564* (7734), 77−82.
(23) James, R. H.; Deme, J. C.; Kjær, A.; Alcock, F.; Silale, A.; Lauber, F.; Johnson, S.; Berks, B. C.; Lea, S. M. Structure and Mechanism of the Proton-Driven Motor That Powers Type 9 Secretion and Gliding Motility. *Nat. Microbiol.* **2021**, *6* (2), 221−233, DOI: 10.1038/s41564-020-00823-6.
(24) Leone, P.; Roche, J.; Vincent, M. S.; Tran, Q. H.; Desmyter, A.; Cascales, E.; Kellenberger, C.; Cambillau, C.; Roussel, A. Type IX Secretion System PorM and Gliding Machinery GldM Form Arches Spanning the Periplasmic Space. *Nat. Commun.* **2018**, *9* (1), No. 429.
(25) Vincent, M. S.; Canestrari, M. J.; Leone, P.; Stathopulos, J.; Ize, B.; Zoued, A.; Cambillau, C.; Kellenberger, C.; Roussel, A.; Cascales, E. Characterization of the Porphyromonas Gingivalis Type IX Secretion Trans-Envelope PorKLMNP Core Complex. *J. Biol. Chem.* **2017**, *292* (8), 3252−3261.
(26) James, R. H.; Deme, J. C.; Hunter, A.; Berks, B. C.; Lea, S. M. Structures of the Type IX Secretion/Gliding Motility Motor from across the Phylum Bacteroidetes. *mBio* **2022**, *13* (3), No. e00267-22, DOI: 10.1128/mbio.00267-22.
(27) Song, L.; Perpich, J. D.; Wu, C.; Doan, T.; Nowakowska, Z.; Potempa, J.; Christie, P. J.; Cascales, E.; Lamont, R. J.; Hu, B. A Unique Bacterial Secretion Machinery with Multiple Secretion Centers. *Proc. Natl. Acad. Sci. U.S.A.* **2022**, *119* (18), No. e2119907119.
(28) Kita, D.; Shibata, S.; Kikuchi, Y.; Kokubu, E.; Nakayama, K.; Saito, A.; Ishihara, K. Involvement of the Type IX Secretion System in Capnocytophaga Ochracea Gliding Motility and Biofilm Formation. *Appl. Environ. Microbiol.* **2016**, *82* (6), 1756−1766.
(29) Veith, P. D.; Glew, M. D.; Gorasia, D. G.; Reynolds, E. C. Type IX Secretion: The Generation of Bacterial Cell Surface Coatings Involved in Virulence, Gliding Motility and the Degradation of Complex Biopolymers. *Mol. Microbiol.* **2017**, *106* (1), 35−53.
(30) Kulkarni, S. S.; Johnston, J. J.; Zhu, Y.; Hying, Z. T.; McBride, M. J. The Carboxy-Terminal Region of *Flavobacterium Johnsoniae* SprB Facilitates Its Secretion by the Type IX Secretion System and Propulsion by the Gliding Motility Machinery. *J. Bacteriol.* **2019**, *201* (19), No. e00218-19, DOI: 10.1128/JB.00218-19.
(31) Abby, S. S.; Denise, R.; Rocha, E. P. Identification of Protein Secretion Systems in Bacterial Genomes Using MacSyFinder Version 2 *bioRxiv* 2023 DOI: 10.1101/2023.01.06.522999.
(32) Veith, P. D.; Nor Muhammad, N. A.; Dashper, S. G.; Likić, V. A.; Gorasia, D. G.; Chen, D.; Byrne, S. J.; Catmull, D. V.; Reynolds, E. C. Protein Substrates of a Novel Secretion System Are Numerous in the Bacteroidetes Phylum and Have in Common a Cleavable C-Terminal Secretion Signal, Extensive Post-Translational Modification, and Cell-Surface Attachment. *J. Proteome Res.* **2013**, *12* (10), 4449−4461, DOI: 10.1021/pr400487b.
(33) Page, M. J.; McKenzie, J. E.; Bossuyt, P. M.; Boutron, I.; Hoffmann, T. C.; Mulrow, C. D.; Shamseer, L.; Tetzlaff, J. M.; Akl, E. A.; Brennan, S. E.; Chou, R.; Glanville, J.; Grimshaw, J. M.; Hróbjartsson, A.; Lalu, M. M.; Li, T.; Loder, E. W.; Mayo-Wilson, E.; McDonald, S.; McGuinness, L. A.; Stewart, L. A.; Thomas, J.; Tricco, A. C.; Welch, V. A.; Whiting, P.; Moher, D. The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews. *BMJ* **2021**, *372*, No. n71, DOI: 10.1136/bmj.n71.
(34) Li, J.; Yao, Y.; Xu, H. H.; Hao, L.; Deng, Z.; Rajakumar, K.; Ou, H.-Y. SecReT6: A Web-Based Resource for Type VI Secretion Systems Found in Bacteria. *Environ. Microbiol.* **2015**, *17* (7), 2196−2202.
(35) Bi, D.; Liu, L.; Tai, C.; Deng, Z.; Rajakumar, K.; Ou, H.-Y. SecReT4: A Web-Based Bacterial Type IV Secretion System Resource. *Nucleic Acids Res.* **2013**, *41* (D1), D660−D665.
(36) Martínez-García, P. M.; Ramos, C.; Rodríguez-Palenzuela, P. T346Hunter: A Novel Web-Based Tool for the Prediction of Type III, Type IV and Type VI Secretion Systems in Bacterial Genomes. *PLoS One* **2015**, *10* (4), No. e0119317.
(37) Gorasia, D. G.; Lunar, S. I.; Butler, C. A.; Chabalier, M.; Doan, T.; Cascales, E.; Veith, P. D.; Reynolds, E. C. Protein Interactome Analysis of the Type IX Secretion System Identifies PorW as the Missing Link between the PorK/N Ring Complex and the Sov Translocon. *Microbiol. Spectrum* **2022**, *10* (1), No. e01602-21, DOI: 10.1128/spectrum.01602-21.
(38) Johnston, J. J.; Shrivastava, A.; McBride, M. J. Untangling *Flavobacterium Johnsoniae* Gliding Motility and Protein Secretion. *J. Bacteriol.* **2017**, *200* (2), No. e00362-17, DOI: 10.1128/jb.00362-17.
(39) Seers, C. A.; Slakeski, N.; Veith, P. D.; Nikolof, T.; Chen, Y.-Y.; Dashper, S. G.; Reynolds, E. C. The RgpB C-Terminal Domain Has a Role in Attachment of RgpB to the Outer Membrane and Belongs to a Novel C-Terminal-Domain Family Found in Porphyromonas Gingivalis. *J. Bacteriol.* **2006**, *188* (17), 6376−6386.
(40) Li, W.; O'Neill, K. R.; Haft, D. H.; DiCuccio, M.; Chetvernin, V.; Badretdin, A.; Coulouris, G.; Chitsaz, F.; Derbyshire, M. K.; Durkin, A. S.; Gonzales, N. R.; Gwadz, M.; Lanczycki, C. J.; Song, J. S.; Thanki, N.; Wang, J.; Yamashita, R. A.; Yang, M.; Zheng, C.; Marchler-Bauer, A.; Thibaud-Nissen, F. RefSeq: Expanding the Prokaryotic Genome Annotation Pipeline Reach with Protein Family Model Curation. *Nucleic Acids Res.* **2021**, *49* (D1), D1020−D1028.
(41) NCBI Protein Family Models. 2021. https://www.ncbi.nlm.nih.gov/protfam. (accessed December 07, 2021).
(42) Thompson, J. D.; Higgins, D. G.; Gibson, T. J. CLUSTAL W: Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice. *Nucleic Acids Res.* **1994**, *22* (22), 4673−4680.
(43) Subramanian, B.; Gao, S.; Lercher, M. J.; Hu, S.; Chen, W.-H. Evolview v3: A Webserver for Visualization, Annotation, and

Management of Phylogenetic Trees. *Nucleic Acids Res.* **2019**, *47* (W1), W270−W275.
(44) Evolview. 2022. https://www.evolgenius.info/evolview/. (accessed December 07, 2021).
(45) MEME Multiple Em for Motif Elicitation. 2021 https://meme-suite.org/meme/tools/meme. (accessed December 07, 2021).
(46) Armenteros, J. J. A.; Tsirigos, K. D.; Sønderby, C. K.; Petersen, T. N.; Winther, O.; Brunak, S.; von Heijne, G.; Nielsen, H. SignalP 5.0 Improves Signal Peptide Predictions Using Deep Neural Networks. *Nat. Biotechnol.* **2019**, *37* (4), 420−423, DOI: 10.1038/s41587-019-0036-z.
(47) Käll, L.; Krogh, A.; Sonnhammer, E. L. L. A Combined Transmembrane Topology and Signal Peptide Prediction Method. *J. Mol. Biol.* **2004**, *338* (5), 1027−1036.
(48) Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E. L. L. Predicting Transmembrane Protein Topology with a Hidden Markov Model: Application to Complete Genomes. *J. Mol. Biol.* **2001**, *305* (3), 567−580.
(49) rstatix: Pipe-Friendly Framework for Basic Statistical Tests. 2022 https://cran.r-project.org/web/packages/rstatix/index.html. (accessed December 07, 2021).
(50) Wickham, H. Ggplot2. *Wiley Interdiscip. Rev.: Comput. Stat.* **2011**, *3* (2), 180−185.
(51) Wilkinson, M. D.; Dumontier, M.; Aalbersberg, Ij. J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; da Silva Santos, L. B.; Bourne, P. E.; Bouwman, J.; Brookes, A. J.; Clark, T.; Crosas, M.; Dillo, I.; Dumon, O.; Edmunds, S.; Evelo, C. T.; Finkers, R.; Gonzalez-Beltran, A.; Gray, A. J. G.; Groth, P.; Goble, C.; Grethe, J. S.; Heringa, J.; t'Hoen, P. A. C.; Hooft, R.; Kuhn, T.; Kok, R.; Kok, J.; Lusher, S. J.; Martone, M. E.; Mons, A.; Packer, A. L.; Persson, B.; Rocca-Serra, P.; Roos, M.; van Schaik, R.; Sansone, S.-A.; Schultes, E.; Sengstag, T.; Slater, T.; Strawn, G.; Swertz, M. A.; Thompson, M.; van der Lei, J.; van Mulligen, E.; Velterop, J.; Waagmeester, A.; Wittenburg, P.; Wolstencroft, K.; Zhao, J.; Mons, B. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Sci. Data* **2016**, *3* (1), No. 160018, DOI: 10.1038/sdata.2016.18.
(52) NCBI Assembly. 2020 https://www.ncbi.nlm.nih.gov/assembly. (accessed December 21, 2020).
(53) Bacterial and Viral Bioinformatics Resource Center (BV-BRC). 2020 https://www.bv-brc.org/. (accessed December 21, 2020).
(54) NCBI Genbank. 2020 https://www.ncbi.nlm.nih.gov/genbank/. (accessed August 20, 2020).
(55) The UniProt Consortium. UniProt: The Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.* **2023**, *51* (D1), D523−D531.
(56) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583−589.
(57) Varadi, M.; Anyango, S.; Deshpande, M.; Nair, S.; Natassia, C.; Yordanova, G.; Yuan, D.; Stroe, O.; Wood, G.; Laydon, A.; Žídek, A.; Green, T.; Tunyasuvunakool, K.; Petersen, S.; Jumper, J.; Clancy, E.; Green, R.; Vora, A.; Lutfi, M.; Figurnov, M.; Cowie, A.; Hobbs, N.; Kohli, P.; Kleywegt, G.; Birney, E.; Hassabis, D.; Velankar, S. AlphaFold Protein Structure Database: Massively Expanding the Structural Coverage of Protein-Sequence Space with High-Accuracy Models. *Nucleic Acids Res.* **2022**, *50* (D1), D439−D444.
(58) Zhang, H.; Yohe, T.; Huang, L.; Entwistle, S.; Wu, P.; Yang, Z.; Busk, P. K.; Xu, Y.; Yin, Y. DbCAN2: A Meta Server for Automated Carbohydrate-Active Enzyme Annotation. *Nucleic Acids Res.* **2018**, *46* (W1), W95−W101.
(59) de Diego, I.; Ksiazek, M.; Mizgalska, D.; Koneru, L.; Golik, P.; Szmigielski, B.; Nowak, M.; Nowakowska, Z.; Potempa, B.; Houston, J. A.; Enghild, J. J.; Thøgersen, I. B.; Gao, J.; Kwan, A. H.; Trewhella, J.; Dubin, G.; Gomis-Rüth, F. X.; Nguyen, K.-A.; Potempa, J. The Outer-Membrane Export Signal of Porphyromonas Gingivalis Type IX Secretion System (T9SS) Is a Conserved C-Terminal β-Sandwich Domain. *Sci. Rep.* **2016**, *6* (1), No. 23123.
(60) Pubmed. 2020 https://pubmed.ncbi.nlm.nih.gov/. (accessed August 20, 2020).
(61) Davis, J. J.; Wattam, A. R.; Aziz, R. K.; Brettin, T.; Butler, R.; Butler, R. M.; Chlenski, P.; Conrad, N.; Dickerman, A.; Dietrich, E. M.; Gabbard, J. L.; Gerdes, S.; Guard, A.; Kenyon, R. W.; Machi, D.; Mao, C.; Murphy-Olson, D.; Nguyen, M.; Nordberg, E. K.; Olsen, G. J.; Olson, R. D.; Overbeek, J. C.; Overbeek, R.; Parrello, B.; Pusch, G. D.; Shukla, M.; Thomas, C.; VanOeffelen, M.; Vonstein, V.; Warren, A. S.; Xia, F.; Xie, D.; Yoo, H.; Stevens, R. The PATRIC Bioinformatics Resource Center: Expanding Data and Analysis Capabilities. *Nucleic Acids Res.* **2019**, *48* (D1), D606−D612.
(62) Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **2018**, *35* (6), 1547−1549.
(63) Gupta, S.; Stamatoyannopoulos, J. A.; Bailey, T. L.; Noble, W. S. Quantifying Similarity between Motifs. *Genome Biol.* **2007**, *8* (2), No. R24, DOI: 10.1186/gb-2007-8-2-r24.
(64) Tomtom Motif Comparison Tool. 2021 https://meme-suite.org/meme/tools/tomtom. (accessed December 07, 2021).
(65) Bailey, T. L. STREME: Accurate and Versatile Sequence Motif Discovery. *Bioinformatics* **2021**, *37* (18), 2834−2840.
(66) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **2003**, *13* (11), 2498−2504.
(67) Grant, C. E.; Bailey, T. L.; Noble, W. S. FIMO: Scanning for Occurrences of a given Motif. *Bioinformatics* **2011**, *27* (7), 1017−1018.
(68) FIMO Find Individual Motif Occurences. 2021 https://meme-suite.org/meme/tools/fimo. (accessed December 07, 2021).
(69) Bailey, T. L.; Elkan, C. *Fitting a Mixture Model by Expectation Maximization to Discover Motifs in Bipolymers*; Proceedings/International Conference on Intelligent Systems for Molecular Biology, 1994; pp 28−36.
(70) MariaDB Server: The open source relational database. 2022 https://mariadb.org/. (accessed December 07, 2022).
(71) php. 2022 https://www.php.net. (accessed December 07, 2022).
(72) jQuery. 2022 https://jquery.com/. (accessed December 07, 2022).
(73) Bootstrap 4. 2022 https://getbootstrap.com/docs/4.0/. (accessed December 07, 2022).
(74) Plotly. 2022 https://plotly.com/javascript/. (accessed December 07, 2022).
(75) Apache HTTP server project. 2022 https://httpd.apache.org/. (accessed December 07, 2022).

https://doi.org/10.1021/acsomega.3c05155
ACS Omega 2023, 8, 34091−34102