



Original article

PVCbase: an integrated web resource for the PVC bacterial proteomes

Nicola Bordin^{1,*}, Juan Carlos González-Sánchez^{2,3} and Damien P. Devos^{1,*}

¹Centro Andaluz de Biología del Desarrollo, CSIC, Universidad Pablo de Olavide, Carretera de Utrera, Km. 1, Seville 41013, Spain, ²CellNetworks, BioQuant, University of Heidelberg, Im Neuenheimer Feld 267, 69120 Heidelberg, Germany and ³Biochemie Zentrum Heidelberg (BZH), Heidelberg University, Im Neuenheimer Feld 328, 69120 Heidelberg, Germany

*Corresponding author: Email: nbor1@upo.es

Correspondence may also be addressed to Damien P. Devos. Email: damienpdevos@gmail.com

Citation details: Bordin, N., González-Sánchez, J.C., and Devos, D.P. PVCbase: an integrated web resource for the PVC bacterial proteomes. *Database* (2018) Vol. 2018: article ID bay042; doi:10.1093/database/bay042

Received 26 January 2018; Revised 15 March 2018; Accepted 5 April 2018

Abstract

Interest in the *Planctomycetes-Verrucomicrobia-Chlamydiae* (PVC) bacterial superphylum is growing within the microbiology community. These organisms do not have a specialized web resource that gathers *in silico* predictions in an integrated fashion. Hence, we are providing the PVC community with PVCbase, a specialized web resource that gathers *in silico* predictions in an integrated fashion. PVCbase integrates protein function annotations obtained through sequence analysis and tertiary structure prediction for 39 representative PVC proteomes (PVCdb), a protein feature visualizer (Foundation) and a custom BLAST webserver (PVCblast) that allows to retrieve the annotation of a hit directly from the DataTables. We display results from various predictors, encompassing most functional aspects, allowing users to have a more comprehensive overview of protein identities. Additionally, we illustrate how the application of PVCdb can be used to address biological questions from raw data.

Database URL: PVCbase is freely accessible at www.pvcbacteria.org/pvcbase

Introduction

The *Planctomycetes-Verrucomicrobia-Chlamydiae* (PVC) bacterial superphylum is composed of the three name-giving phyla and some additional ones, like *Lentisphaerae*, ‘*Candidatus* Omnitrophica’ and *Kirimatiellaeota* (1). Despite this diversity, it is now accepted that they form a monophyletic group (2). This bacterial superphylum draws

interest because some species display characteristics not frequently observed in bacteria. Examples of these are condensed DNA (nucleoids), extensive inner membrane organization (3), the ability to internalize external compounds before degradation (4), the presence of membrane coat-like proteins linked to the extensive membrane organization (5, 6) and that they were thought to lack peptidoglycan

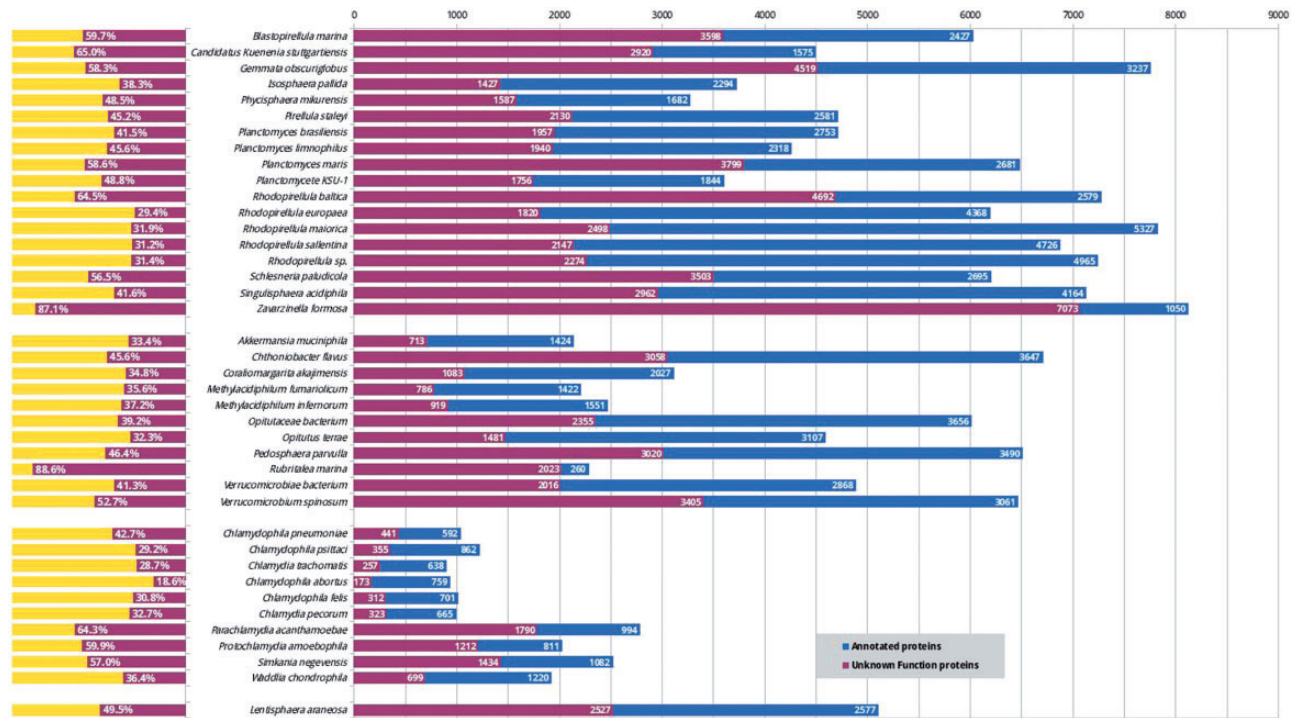


Figure 1. Status of functional annotation of PVC proteins. Total numbers (right) and percentage of total (left) proteins annotated (blue or yellow bars) and unannotated (purple) in each proteomes.

until recently (7, 8). Interesting characteristics relevant for wastewater treatment are shown by some planctomycetes having an anammoxosome, possibly the first true prokaryotic organelle, which allows the bacteria to degrade ammonium anaerobically (9). The exceptional diversity of cell plans displayed by some of the phyla, showcased by *Gemmata obscuriglobus* and some verrucomicrobia (10), has been the subject of controversial interpretations (11, 12).

Despite the considerable interest in these bacteria in many fields, including cell biology, evolution (13) and biotechnology (14), these organisms lack a centralized resource for their analysis. While new PVCs are being sequenced, the mean percentage of unannotated proteins constitutes approximately 46% (Figure 1). This issue does not exclusively affect the PVC bacteria since complete genome sequencing rarely translates into a complete characterization of the organism (in UniProt 31% of non-PVC proteomes are uncharacterized) (15). Protein function is not limited to a single feature or description; therefore we developed a pipeline for the simultaneous consideration of many different sequence descriptors. Since our aim is to provide these results to the community of PVC experimentalists, the results were collected in a resource built taking user-friendliness into consideration. Users are able to easily query the results of all functional predictors and download the predictions in bulk for large-scale interrogations.

Methods

Unless specified otherwise, all tools were used with default parameters.

Proteomes collection

The proteomes of 39 PVC representative species (17 planctomycetes, 11 verrucomicrobia, 10 chlamydia and 1 leptospirae) comprising a total of 173 664 protein sequences, were obtained from the UniProtKB and the NCBI-protein databases. The complete list of PVC species is given in supplementary (See online [supplementary material for Table S1](#)).

Homology-based inference

For every sequence, a homology search was performed using PSI-BLAST (16) with three iterations and default parameters, against the UniProtKB/Swiss-Prot database (release 2015_02) (17). The raw output was parsed to extract all the hits showing an *E*-value below $1E-3$ and a minimum coverage of 75% of the query sequence. The first of these matches was selected as the best match and information regarding function under the form of GO description (18), keywords and enzymatic activity was assigned to the query protein. From the remaining matches, GO terms were extracted and counted, and they were reported only if they appeared in at least 10% of the hits.

Domain analysis

The tool InterProScan (v5.16-55.0) (19) was used to search for protein signatures by scanning their sequences against all its member databases: Pfam (release 28.0), TIGRFAM (release 15.0), PANTHER, ProSite (release 20.113), HAMAP (release 2015_11), PIRSF (release 3.01), Gene3D (release 3.5.0), SUPERFAMILY (release 1.75), PRINTS (release 42.0), SMART (release 6.2) and InterPro (20), using the parameters *-goterms* and *-pa*. Thanks to these options, entries from the InterPro database also provided functional information in the form of terms from GO and KEGG-pathway entries (21).

Tertiary structure prediction

A series of programs and utilities included in the HHsuite package (v2.0.16) were used. For each sequence, first HHblits (22) was used to construct a high-quality multiple sequence alignment (MSA) by comparing it against the UniProt20 database of template Hidden Markov models (HMMs) (release 2013_03), with the option *-addss* which adds secondary structure information predicted with PSIPRED v3.5 (23) to the resulting MSA. It was then converted to a HMM (.hmm format) with the *hmmake* function. Finally, HHsearch (24) was used to compare it against the HMM template database pdb70 (release 16May15), which is based on the protein data bank (PDB) (25). Every tool was run with default parameters. The results for both comparisons, against UniProt20 and pdb70, were parsed with an *E*-value cut-off of 1E-3. For the latter, functional information in the form of GO terms and EC codes was gathered from SIFTS mapping (26).

Prediction of signal peptides and transmembrane helices

Signal peptides were predicted with SignalP4.1 (27) using the *gram-* option. Transmembrane helices (TMHs) were predicted with TMHMM (28). The content (%) of integral membrane or transmembrane proteins of the proteome was defined as the fraction of proteins for which at least one TMH was predicted (transmembrane proteins/total number of proteins*100).

Prediction of protein intrinsic disorder

The IUPred tool (29) was used to predict intrinsically disordered regions and globular domains. The default threshold of 0.5 was used to determine whether a residue was considered as structured or disordered. Three metrics were computed to describe disorder within the proteome (30):

(i) the disorder content (%) which was calculated as the fraction of disordered residues in the proteome (total predicted disordered residues/total number of residues*100); (ii) the content (%) of long disordered regions (LDRs), which are defined as those regions where at least 30 disordered residues are predicted continuously along the sequence, calculated as the fraction of residues in those LDRs, (residues in LDRs/total number of residues*100); and (iii) the fraction (%) of highly disordered proteins (HDPs) which are defined as those with >50% of predicted disordered residues in their sequences (number of HDP/total number of proteins*100).

PVCbase

PVCbase is a webserver developed to distribute the predictors results and statistics on disorder and TMHs distributions for the PVC superphylum. The resource acts as a gateway to PVCdb, the BLAST webserver and our secondary structure descriptor, Foundation. PVCbase is built on top of a Linux-Apache-MySQL-Python stack with WordPress as content management system (Figure 2).

PVCdb

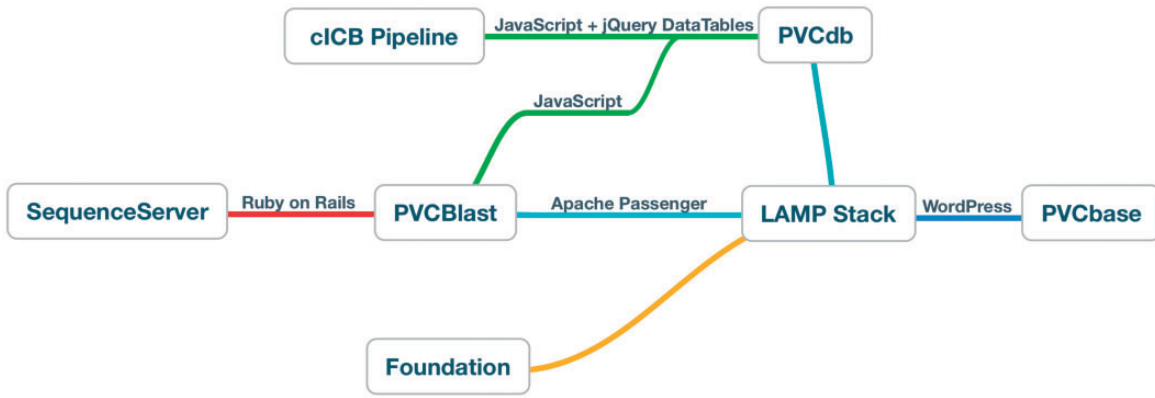
PVCdb data include sequence and structure-based features that were computed using a Python-Perl pipeline that runs, parse and organize the predictors results (Figure 3), generating tabular and HTML web pages for each proteome. The HTML pages include Javascript code that retrieves the query originated from PVCblast from the URI and pre-filters the jQuery DataTable on load. The standard DataTables plugin was modified in order to allow fixed headers, table prefiltering and paging.

PVCBlast

PVCblast is based on the SequenceServer (31) Ruby package running on a RubyOnRails-Passenger-Apache stack. The Ruby was customized to highlight hit significance, links to PVCbase, a sample FASTA file, further sequence checks and a link-out system that connects a hit on a PVC proteome to its annotation in PVCdb. We included some additional controls to the original SequenceServer, such as refresh and back-to-top buttons.

Foundation

For ease of visualization, we provide a protein features visualization tool, Foundation, that combines secondary structure (predicted by PSI-Pred), Transmembrane helices (by TMHMM) and disorder (by IUPred) predictions.



PVCbase Infrastructure

Figure 2. PVCbase organization and relationship between services.

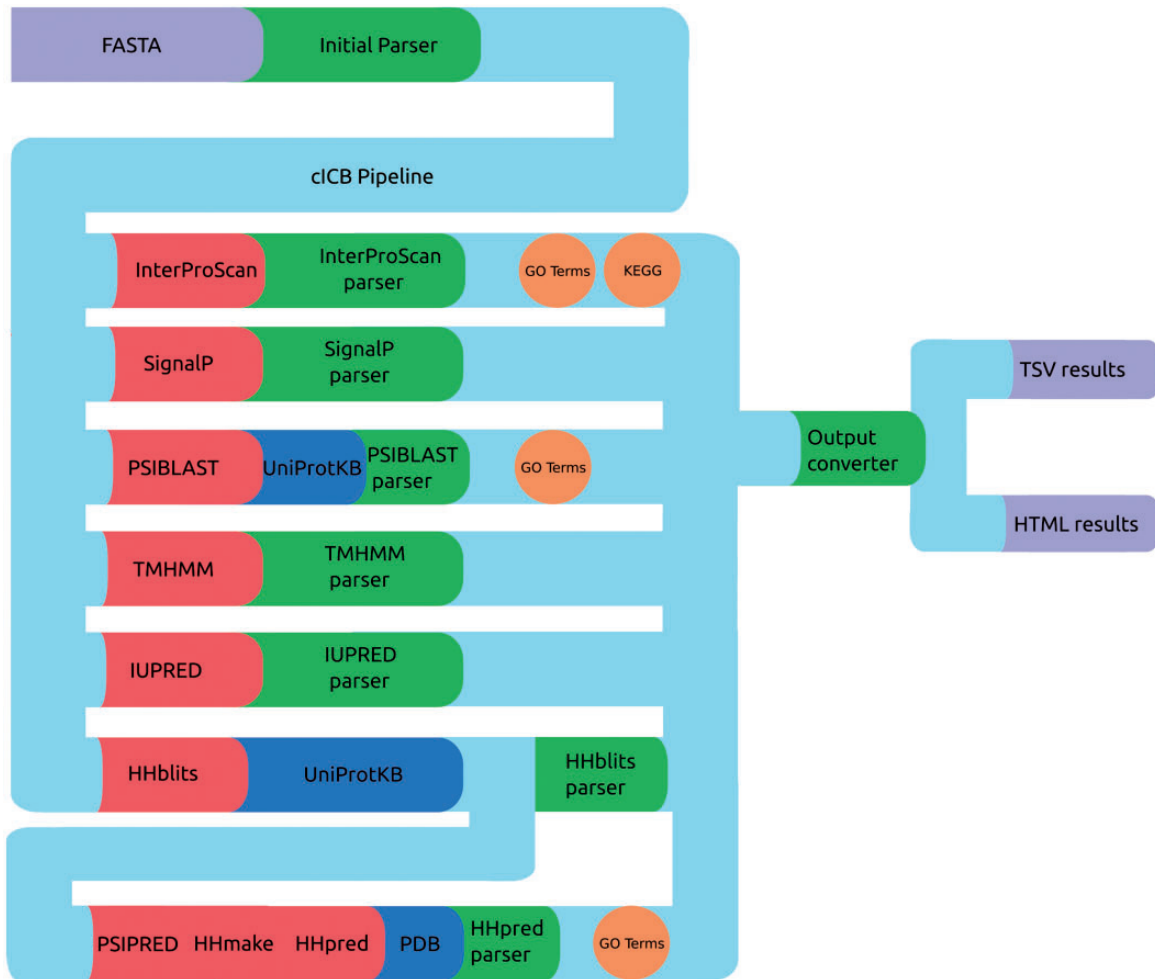


Figure 3. cICB pipeline. Files (purple), Tools (red), databases (blue) and scripts (green) used to generate the annotations for PVCdb. Color spheres indicates additional information obtained through the predictors.

In addition to a quick visualization of sequence feature, an illustration is provided in post-script, allowing the addition of more annotations, such as domains or mutated residues, as well as merging of various images, with limited understanding of postscript scripting. Foundation allows downloading of the raw results from the predictors for further analysis.

Results

PVC proteomes

We gathered the proteomes of 39 PVC species comprising 17 planctomycetes, 11 verrucomicrobia, 10 chlamydia and 1 lentisphaerae (See online [supplementary material](#) for Table S1). The three main phyla, *Planctomycetes*, *Verrucomicrobia* and *Chlamydia*, show an important variance of their proteome sizes. As reference, some of the model bacteria like *Escherichia coli* or *Bacillus subtilis* have 4305 and 4197 different proteins, respectively. Chlamydia has very reduced genomes with low protein numbers (mean/median: 1532–1125 proteins). *Chlamydia trachomatis* possesses one of the tiniest proteome, with only 895 proteins. The *Verrucomicrobia* appears to be intermediary (4307/4588) with some close to the size of reduced chlamydial pathogens with around 2000 proteins. In contrast, the *Planctomycetes* displays much larger proteome sizes (5881/6193), which rank them among the bacteria with the biggest genomes and most protein-coding genes. The largest PVC proteomes belong to the *Planctomycetaceae* family, with *Zavarzinella formosa*, *Rhodopirellula maiorica SM1* and *G.obscuriglobus* encoding 8123, 7825 and 7756 proteins, respectively, almost one order of magnitude bigger than the smallest chlamydial proteome, *C.trachomatis* and bigger than some eukaryotes, like the baker's yeast *Saccharomyces cerevisiae* that encodes 6721 proteins. The biggest planctomycetal genome, *Fimbriiglobus ruber*, has recently been reported to have a size of 12.364 Mbp and it encodes more than 10 000 proteins (32).

For comparison, one of the biggest bacterial genomes is found in *Ktedonobacter racemifer* in the phylum *Chloroflexi*, with 13.7 Mbp and coding for 11 000 proteins (33).

Usage of PVCbase

PVCdb

PVCdb collects protein functional annotations of 39 PVC proteomes. Each proteome can be downloaded as multiFASTA and the corresponding annotation can be either downloaded as a tabular file or easily browsed online. PVCdbs can show a variable amount of entries, based on

user choice. Table rows can be sorted by length or alphabetically, while the search bar filters the table and shows only hits that contain the searched keyword. This is helpful to extract subsets of proteins based on localization, process or related to specific activities (Figure 4).

PVCBlast

PVCBlast allows the user to perform BLAST searches on the PVC proteomes and genomes. The search box supports drag-and-drop and multiple sequences at once. The BLAST search parameters, such as e-values cutoff and number of alignments, can be customized using the 'advanced parameters' bar at the bottom of the page. The results page shows the alignments produced by BLAST, alongside several options for downloading the hits sequences and reports from the tool. The default SequenceServer interface provides a link-out service to NCBI Genbank or UniProt, according to the proteome source. We modified the aligned hits window to indicate the significance of a hit using shades of red. A link-out generator was created to link a hit on a PVC proteome to the corresponding annotation in PVCdb, pre-filtering the DataTable (Figure 5).

Foundation

Foundation is a tool to quickly visualize the linear and secondary structure features for a provided protein with its secondary structure features. Results can be downloaded as a png or postscript file, as a compressed tar file containing the raw output of the predictors, or can be visualized as an interactive zooming map. Secondary structure features are depicted with different color bars, fuchsia for α helices and cyan for β sheets. TMHs are shown as green boxes and the line underneath the secondary structure shows the disorder level for each amino acid (Figure 6).

Transmembrane proteins, intrinsic disorder and internal membrane: applications of PVCdb

In order to illustrate one of the possible discoveries made possible by PVCbase that would have been difficult to realize with other currently existing databases, we provide the following example. Planctomycetes cells present internal organizations characterized by the presence of developed membrane organizations which are atypical for bacteria (12). These have been extensively studied in *G.obscuriglobus* by three-dimensional tomography reconstructions that reported an extensive network of internal membranes (3, 6). A separate compartment has also been described in anammox planctomycetes (34). Variations of

Figure 4. PVCdb. Example of Web interface of PVCdb. DataTable of Blastopirellula marina. The figure shows portion of the bacterial annotation. Each column can be sorted and browsed using the built-in search bar.

Figure 5. Results page of PVCblast. The results page of PVCblast provides the user with the alignments and the ability to download the results in various formats (top right). Among the linkout systems (center), clicking on 'ciCB annotation' opens the correspondent annotation on PVCdb.

this cellular structure are believed to be shared by other planctomycetes and verrucomicrobia but are still mostly unexplored (10, 12, 35). We investigated a possible relationship between the presence of internal membrane and

the number of transmembrane proteins. We calculated the fraction of proteins with TMHs for each PVC proteome and for each proteome of three reference sets composed of non-PVC bacterial, archaeal and eukaryotic

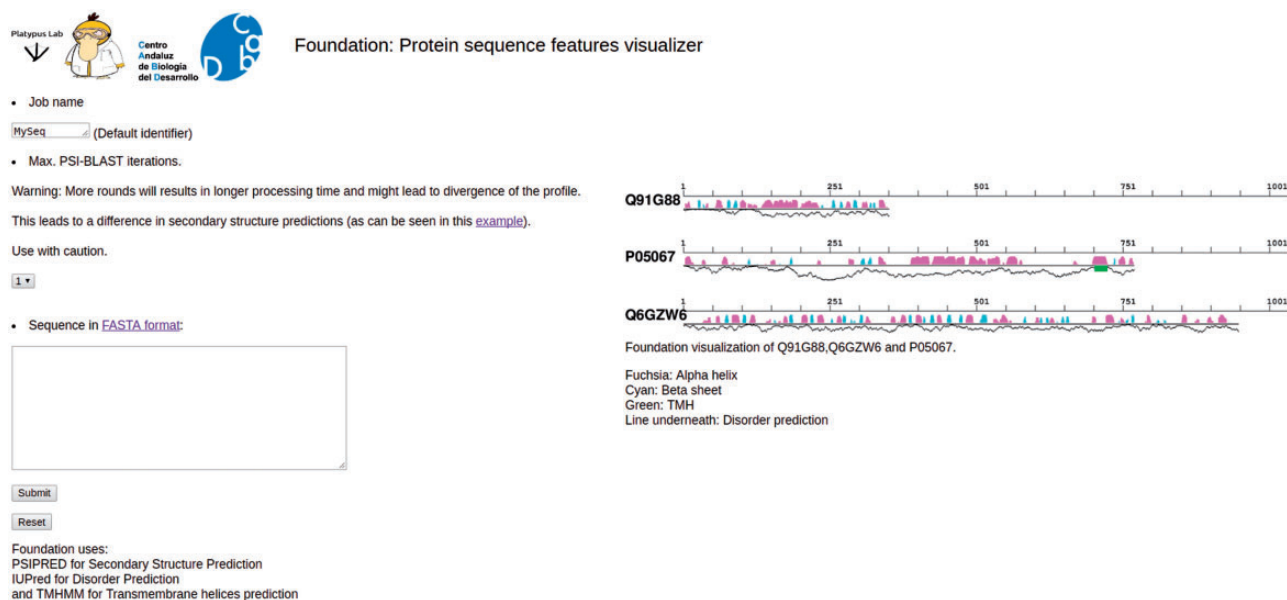


Figure 6. Foundation. Snapshot of Foundation's main page. The input box allows only single protein sequences. The number of PSIBLAST iterations can be modified using the drop-down menu (center-left). On right side of the page there are some examples of the output and the legend.

species (see Methods). These reference proteomes are representative of the three domains of life with diverse cellular plans (See online [supplementary material](#) for Table S2).

We first noticed that bacteria, in general, show a slightly higher content of TMHs than the analyzed species of archaea (P -value = $5.00E-04$) and Eukaryotes (P -value = $6.92E-04$). The content in transmembrane proteins showed however no statistical difference for any of the PVC bacterial groups when compared against all other groups and assessed by the Wilcoxon rank-sum test (See online [supplementary material](#) for Table S5, [Figure 7](#)). Therefore, the PVC genomes possess a smaller fraction of transmembrane protein in comparison to other bacteria (all means and medians between 23 and 24%). Additionally, we compared the fraction of transmembrane proteins to their number of TMHs (See online [supplementary material](#) for Table S3). The results provided further evidence for the previous observation. Thus, transmembrane protein content does not seem to be correlated with membrane complexity. This is illustrated by comparing the human and *C.trachomatis* proteomes, which respectively contain 17.41% and 24.58% of proteins containing at least one TMH.

Similarly, we explored the differences in protein structural disorder between these groups. We computed three metrics to describe the intrinsic disorder of the proteomes: the total disorder content, the fraction of residues in LDRs and the fraction of highly disordered proteins (HDPs) (see Methods) ([Figure 8](#); See online [supplementary material](#) for Table S4). As demonstrated elsewhere, disorder content is, for both prokaryotes and eukaryotes, generally

independent of the proteome size (36). This observation is also reflected in our data. It is worth noting that the two outliers in the data belonging to archaea, and the largest one from the non-PVC bacteria values, correspond to three extreme halophilic organisms (the archaea *Halobacterium salinarum* NRC1 and *Nanosalina* sp., and the bacterium *Salinibacter ruber*) (See online [supplementary material](#) for [Figures S1](#) and [S2](#)). This observation agrees with the suggestion that intrinsically disordered proteins may help these organisms adapt to the extreme environments they inhabit. Disordered regions have an increased tolerance against mutations which allows for a higher evolutionary rate that results in extraordinary adaptability (37).

We first observed, as previously reported, that eukaryotes have more disordered proteomes (10), which has been related to the importance of disorder for cellular complexity. We then detected that most planctomycetes show a higher content of disordered proteins (mean/median of 12.15–12.39%) than non-PVC bacteria (6.82–5.95%). This trend is not observed in verrucomicrobia (7.43–7.08%) or chlamydia (4.95–5.14%). A statistical evaluation confirmed both observations: the disorder contents of planctomycetes species are significantly superior to the other three bacterial groups (Wilcoxon rank-sum test P -values: P vs V = $7.49E-04$, P vs C = $6.64E-04$, P vs non-PVC = $2.97E-04$) and that the values from the other groups are not statistically different from each other (See online [supplementary material](#) for Table S5).

Low complexity, or intrinsically disordered proteins, are mostly associated with signal transduction, cell-cycle regulation and transcription (38). However, it has been

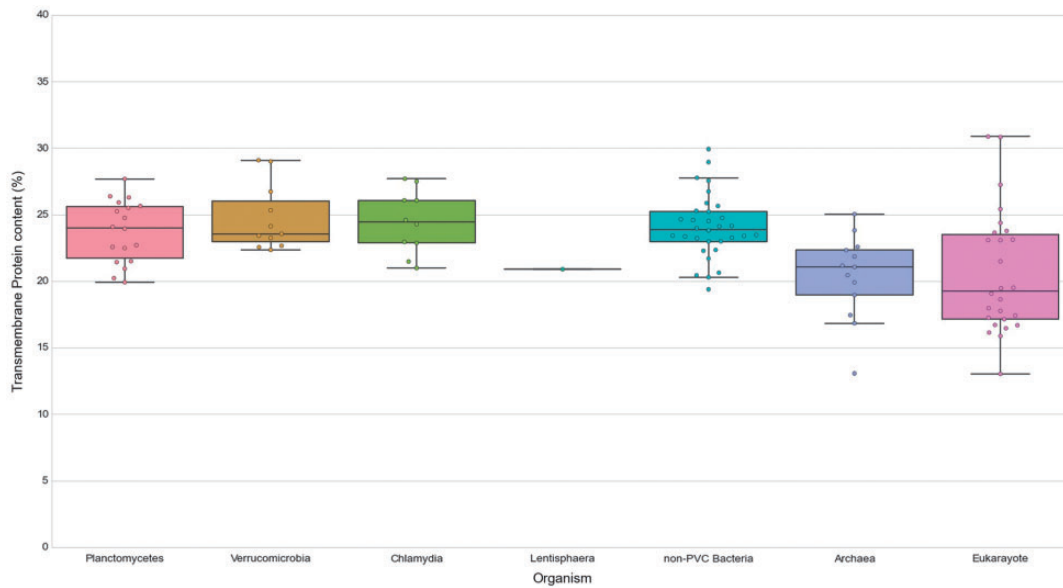


Figure 7. Transmembrane proteins in PVC and representative species from the tree of life. The numbers of TMHs containing proteins, expressed as percentages of the proteomes. Box plots reflect the distribution of the data. The box encloses the quartiles of the dataset, while the whiskers extend to the limits of the distributions. Outliers are determined based on the interquartile range and are not included in the boxes. The middle horizontal line in the box marks the median of the distribution.

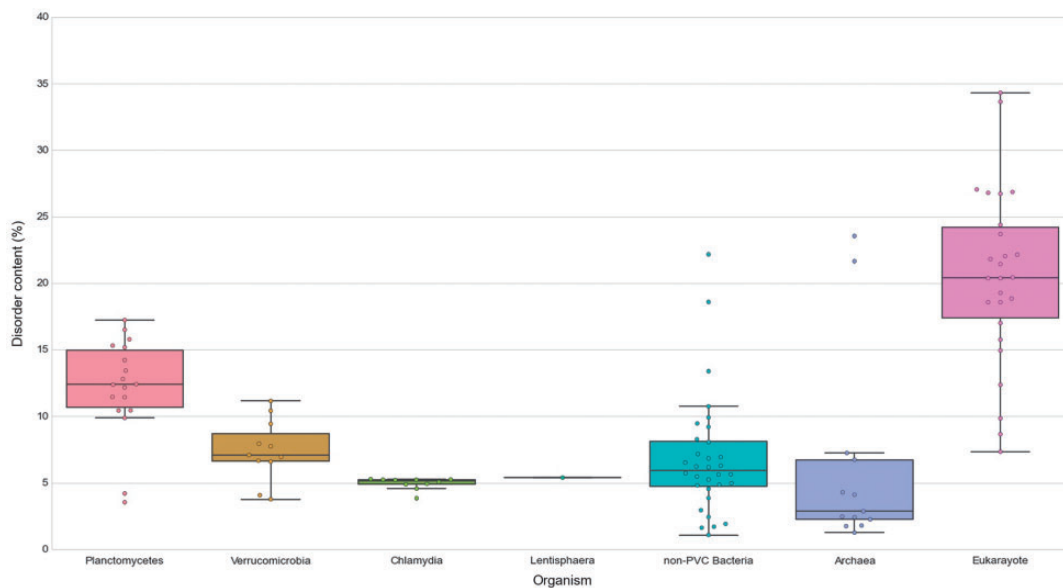


Figure 8. Disorder content in PVC and representative species from the tree of life. The numbers of disordered proteins, expressed as percentages of the proteomes. Box plots reflect the distribution of the data. The box encloses the quartiles of the dataset while the whiskers extend to the limits of the distributions. Outliers are determined based on the interquartile range and are not included in the boxes. The middle horizontal line in the box marks the median of the distribution.

suggested that structural disorder plays a fundamental role in vesicle trafficking pathways (39) and it has been demonstrated that certain unstructured protein domains are highly efficient drivers of membrane curvature. Disordered fragments have a role in membrane coat assembly and vesicle communication, in what has been called the fly-casting

mechanism (40). This is especially the case in the clathrin-coated vesicle system, which mediates endocytosis and the early secretory pathway (35). Thus our observation suggests that the significantly higher ratio of disordered proteins in *Planctomycetes* appears to be correlated with the development of their membranes.

Conclusions

PVCbase offers a convenient one-stop platform for the PVC bacteria community. Its scalability and variety of annotations have already been used in PVC-related publications (41) and newly sequenced organisms will be added regularly. Bulk data collecting also allows users to infer biological discoveries by comparing annotations at the proteome level.

Availability

PVCbase is freely accessible at <http://pvcbacteria.org/pvcbase>.

Acknowledgements

We acknowledge Caitlin Lee Carpenter for her help in proofreading this article.

Supplementary data

Supplementary data are available at Database Online.

Funding

N.B. was funded by Marie Curie ITN FP7-ITN316723-PerFuMe. D.P.D. and J.C.G.S. were funded by the C2A grant EE: 2013/2506 from the Andalusian government. DPD was funded by the Spanish Ministry of Economy and Competitiveness (Grant nos. BFU2013-40866-P and BFU2016-78326-P).

Conflict of interest. None declared.

References

- Rivas-Marín, E. and Devos, D.P. (2017) The Paradigms They Are a-Changin': past, present and future of PVC bacteria research. *Antonie Leeuwenhoek* doi: 10.1007/s10482-017-0962-z. [Epub ahead of print]
- Wagner, M. and Horn, M. (2006) The Planctomycetes, Verrucomicrobia, Chlamydiae and sister phyla comprise a superphylum with biotechnological and medical relevance. *Curr. Opin. Biotechnol.*, **17**, 241–249.
- Santarella-Mellwig, R., Prugnaller, S., Roos, N. *et al.* (2013) Three-dimensional reconstruction of bacteria with a complex endomembrane system. *PLoS Biol.*, **11**, e1001565.
- Lonhienne, T.G.A., Sagulenko, E., Webb, R.I. *et al.* (2010) Endocytosis-like protein uptake in the bacterium *Gemmata obscuriglobus*. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 12883–12888.
- Santarella-Mellwig, R., Franke, J., Jaedicke, A. *et al.* (2010) The compartmentalized bacteria of the Planctomycetes-Verrucomicrobia-Chlamydiae superphylum have membrane coat-like proteins. *PLoS Biol.*, **8**, e1000281.
- Acehan, D., Santarella-Mellwig, R. and Devos, D.P. (2014) A bacterial tubulovesicular network. *J. Cell. Sci.*, **127**, 277–280.
- Jeske, O., Schüler, M., Schumann, P. *et al.* (2015) Planctomycetes do possess a peptidoglycan cell wall. *Nat. Commun.*, **6**, 7116.
- van Teeseling, M.C.F., Mesman, R.J., Kuru, E. *et al.* (2015) Anammox planctomycetes have a peptidoglycan cell wall. *Nat. Commun.*, **6**, 6878.
- van Niftrik, L. (2013) Cell biology of unique anammox bacteria that contain an energy conserving prokaryotic organelle. *Antonie Leeuwenhoek*, **104**, 489–497.
- Lee, K.-C., Webb, R.I., Janssen, P.H. *et al.* (2009) Phylum verrucomicrobia representatives share a compartmentalized cell plan with members of bacterial phylum planctomycetes. *BMC Microbiol.*, **9**, 5.
- Fuerst, J.A. (2013) The PVC superphylum: exceptions to the bacterial definition? *Antonie Leeuwenhoek*, **104**, 451–466.
- Devos, D.P. (2014) PVC bacteria: variation of, but not exception to, the gram-negative cell plan. *Trends Microbiol.*, **22**, 14–20.
- González-Sánchez, J.C., Costa, R. and Devos, D.P. (2015) A multi-functional tubulovesicular network as the ancestral eukaryotic endomembrane system. *Biology (Basel)*, **4**, 264–281.
- Devos, D.P. and Ward, N.L. (2014) Mind the PVCs. *Environ. Microbiol.*, **16**, 1217–1221.
- Galperin, M.Y. and Koonin, E.V. (2010) From complete genome sequence to “complete” understanding? *Trends Biotechnol.*, **28**, 398–406.
- Altschul, S.F., Madden, T.L., Schäffer, A.A. *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
- Bairoch, A., Boeckmann, B., Ferro, S. *et al.* (2004) Swiss-Prot: juggling between evolution and stability. *Brief. Bioinformatics*, **5**, 39–55.
- Ashburner, M., Ball, C.A., Blake, J.A. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.
- Jones, P., Binns, D., Chang, H.-Y. *et al.* (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics*, **30**, 1236–1240.
- Hunter, S., Apweiler, R., Attwood, T.K. *et al.* (2009) InterPro: the integrative protein signature database. *Nucleic Acids Res.*, **37**, D211–D215.
- Kanehisa, M. and Goto, S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*, **28**, 27–30.
- Remmert, M., Biegert, A., Hauser, A. *et al.* (2012) HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat. Methods*, **9**, 173–175.
- McGuffin, L.J., Bryson, K. and Jones, D.T. (2000) The PSIPRED protein structure prediction server. *Bioinformatics*, **16**, 404–405.
- Söding, J. (2005) Protein homology detection by HMM-HMM comparison. *Bioinformatics*, **21**, 951–960.
- Berman, H.M., Battistuz, T., Bhat, T.N. *et al.* (2002) The protein data bank. *Acta Crystallogr D Biol. Crystallogr.*, **58**, 899–907.
- Velankar, S., Dana, J.M., Jacobsen, J. *et al.* (2012) SIFTS: structure integration with function, taxonomy and sequences resource. *Nucleic Acids Res.*, **41**, D483–D489.
- Petersen, T.N., Brunak, S., von Heijne, G. *et al.* (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods*, **8**, 785–786.
- Krogh, A., Larsson, B., von Heijne, G. *et al.* (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.*, **305**, 567–580.

29. Dosztányi,Z., Csizmok,V., Tompa,P. *et al.* (2005) IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics*, **21**, 3433–3434.
30. Pietroseoli,N., Pancsa,R. and Tompa,P. (2013) Structural disorder provides increased adaptability for vesicle trafficking pathways. *PLoS Comput. Biol.*, **9**, e1003144.
31. Priyam,A., Woodcroft,B.J., Rai,V. *et al.* (2015). SequenceServer: a modern graphical user interface for custom BLAST databases. biorxiv doi: 10.1101/033142.
32. Kulichevskaya,I.S., Ivanova,A.A., Baulina,O.I. *et al.* (2017) *Fimbriglobus ruber* gen. nov., sp. nov., a Gemmata-like planctomycete from Sphagnum peat bog and the proposal of Gemmataceae fam. *Nov. Int. J. Syst. Evol. Microbiol.*, **67**, 218–224. 2017 Feb;
33. Chang,Y-j., Land,M., Hauser,L. *et al.* (2011) Non-contiguous finished genome sequence and contextual data of the filamentous soil bacterium *Ktedonobacter racemifer* type strain (SOSP1-21). *Stand. Genomic Sci.*, **5**, 97–111.
34. Neumann,S., Wessels,H.J.C.T., Rijkstra,W.I.C. *et al.* (2014) Isolation and characterization of a prokaryotic cell organelle from the anammox bacterium *Kuenenia stuttgartiensis*. *Mol. Microbiol.*, **94**, 794–802.
35. Boedeker,C., Schüler,M., Reintjes,G. *et al.* (2017) Determining the bacterial cell biology of Planctomycetes. *Nat. Commun.*, **8**, 14853.
36. Xue,B., Dunker,A.K. and Uversky,V.N. (2012) Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. *J. Biomol. Struct. Dyn.*, **30**, 137–149.
37. Xue,B., Williams,R.W., Oldfield,C.J. *et al.* (2010) Archaic chaos: intrinsically disordered proteins in Archaea. *BMC Syst. Biol.*, **4**, S1.
38. Tantos,A., Han,K.-H. and Tompa,P. (2012) Intrinsic disorder in cell signaling and gene transcription. *Mol. Cell. Endocrinol.*, **348**, 457–465.
39. Busch,D.J., Houser,J.R., Hayden,C.C. *et al.* (2015) Intrinsically disordered proteins drive membrane curvature. *Nat. Commun.*, **6**, 7875.
40. Shoemaker,B.A., Portman,J.J. and Wolynes,P.G. (2000) Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 8868–8873.
41. Faria,M., Bordin,N. *et al.* (2017) Planctomycetes attached to algal surfaces: insight into their genomes. *Genomics*, pii: S0888-7543(17)30135-0.