

RESEARCH ARTICLE

Open Access



Evolutionary gradient of predicted nuclear localization signals (NLS)-bearing proteins in genomes of family Planctomycetaceae

Min Guo¹, Ruifu Yang², Chen Huang¹, Qiwen Liao¹, Guangyi Fan¹, Chenghang Sun³ and Simon Ming-Yuen Lee^{1*}

Abstract

Background: The nuclear envelope is considered a key classification marker that distinguishes prokaryotes from eukaryotes. However, this marker does not apply to the family Planctomycetaceae, which has intracellular spaces divided by lipidic intracytoplasmic membranes (ICMs). Nuclear localization signal (NLS), a short stretch of amino acid sequence, destined to transport proteins from cytoplasm into nucleus, and is also associated with the development of nuclear envelope. We attempted to investigate the NLS motifs in Planctomycetaceae genomes to demonstrate the potential molecular transition in the development of intracellular membrane system.

Results: In this study, we identified NLS-like motifs that have the same amino acid compositions as experimentally identified NLSs in genomes of 11 representative species of family Planctomycetaceae. A total of 15 NLS types and 170 NLS-bearing proteins were detected in the 11 strains. To determine the molecular transformation, we compared NLS-bearing protein abundances in the 11 representative Planctomycetaceae genomes with them in genomes of 16 taxonomically varied microorganisms: nine bacteria, two archaea and five fungi. In the 27 strains, 29 NLS types and 1101 NLS-bearing proteins were identified, principal component analysis showed a significant transitional gradient from bacteria to Planctomycetaceae to fungi on their NLS-bearing protein abundance profiles. Then, we clustered the 993 non-redundant NLS-bearing proteins into 181 families and annotated their involved metabolic pathways. Afterwards, we aligned the ten types of NLS motifs from the 13 families containing NLS-bearing proteins among bacteria, Planctomycetaceae or fungi, considering their diversity, length and origin. A transition towards increased complexity from non-planctomycete bacteria to Planctomycetaceae to archaea and fungi was detected based on the complexity of the 10 types of NLS-like motifs in the 13 NLS-bearing proteins families.

Conclusion: The results of this study reveal that Planctomycetaceae separates slightly from the members of non-planctomycete bacteria but still has substantial differences from fungi, based on the NLS-like motifs and NLS-bearing protein analysis.

Keywords: Planctomycetaceae, Comparative genomics, Nuclear localization signal, Signal peptide transformation

Background

Species in the Planctomycetaceae family are ecologically widespread, including even human gut and blood but ubiquitous in water and soil [1–8]. Although Planctomycetaceae are taxonomically affiliated with bacteria, they have been in past studies reported to possess a number of characteristics that are closer to eukaryotes, especially the absence of peptidoglycan in their cell envelope, synthesis of

membrane sterols and the presence of membrane-coat proteins [1, 9–11]. Another intriguing characteristic of Planctomycetaceae is their cellular compartmentalization due to development of internal lipid intracytoplasmic membranes (ICMs) [1, 12], which is uncommon in prokaryotes. In the Planctomycetaceae family, only *Gemmata obscuriglobus* has double-layer ICMs [12]; the other species of the family contain single-layer ICMs. Nevertheless, recently, species in this family have been experimentally confirmed to contain peptidoglycan in their cell wall [13, 14]. Moreover, nearly all of the “unique” characteristics beyond non-planctomycete bacteria in Planctomycetaceae

* Correspondence: simonlee@umac.mo

¹State Key Laboratory of Quality Research of Chinese Medicine and Institute of Chinese Medical Sciences, University of Macau, Macao, China
Full list of author information is available at the end of the article



have been argued not relevant to homology with eukaryotic characteristics, with many of them proposed to result from convergent evolution or lateral gene transfer [15]. Arguing on the other hand in favor of potential homology is the finding that ICMs divide cells of all examined planctomycete species into two compartments, the paryphoplasm and pirellulosome [16, 17], and consequently may make transcription and translation independent, allowing for the development of eukaryotic cellular complexity [18]. The exact nature and topology of planctomycete cell compartments has been subject to controversy, and the question of a closed nucleoid-associated membrane envelope is especially subject to debate [19, 20] - compartments completely closed by membranes may however imply some form of transport system similar to that used by eukaryotes for nucleocytoplasmic transport. A study of the cellular compartmentalization of *G. obscuriglobus* using an immunogold approach found a substantial difference from *Escherichia coli* in the distribution of FtsK protein, which may give insights into the origin of the eukaryotic endomembrane system [20, 21]. Thus, exploration of unusual molecular features that may contribute to or be a consequence of the complicated internal features of family Planctomycetaceae is urgent.

A eukaryotic nucleus has complicated structural and functional foundation, particularly the nuclear pore complex (NPC) [22], a component of the nuclear envelope, which is involved in communication of macromolecules over 60 KDa between the nucleoplasm and cytoplasm. Two types of short amino acids stretches are the signals that direct the transport of macromolecules through the NPC: nuclear location signals (NLSs) [23] and nuclear export signals (NESs) [24, 25]. With other potential cellular functions [26], NLSs direct molecular transport from the cytoplasm to nucleoplasm, and NESs direct transport in the opposite direction. NES motifs are leucine (L) rich and NLSs are arginine (R) and lysine (K) rich. NLS motifs are monopartite or bipartite [27] and their location and number in proteins can vary. NLSs and NESs have been widely identified in many organisms as conferring the ability on a protein to shuttle through the nuclear membrane [28, 29]. Only a few cytoplasmic proteins without a typical NLS core peptide enter the nucleus and they do this only via a strong interaction with protein factors with a core NLS motif [30]. The intracellular environment is crucial to the function of NLS and NES motifs [31]. NLS or NES motifs generally need to be exposed at the protein surface to bind to importins or exportins. Thus, the cell needs mechanisms to unmask hidden or cryptic NLS or NES motifs in proteins; these mechanisms include phosphorylation or dephosphorylation, dissociation of an inhibitory subunit that masks the NLS, processing of a larger precursor, and binding of hormones at a certain stage of

development [31]. An NLS database (<http://rostlab.org/services/nlsdb/browse.php>) has 114 experimentally identified NLS motifs to date [23, 32].

Earlier reported experimental studies of bacterial NLS sequences demonstrated in *Thermoplasma* [33, 34], *Streptomyces*, and *Agrobacterium* [35] the functionality of prokaryotic NLS in transporting proteins into a eukaryotic nucleus. However, no genomic or experimental investigation of NLS motif or NLS-bearing proteins has so far been reported in Planctomycetaceae [36]. Considering the complicated cellular membrane structures of Planctomycetaceae species and the critical functional role played by NPCs and the correlated NLS-sequences in proteins destined for transporting into the nucleus, herein we aim to determine the status of NLSs and NLS-bearing proteins in the Planctomycetaceae family and other microorganisms by a comparative genomic approach. The analysis of signals in Planctomycetaceae related in eukaryotes to the existence of a nuclear envelope (and functions of which might be expected to be absent in bacteria) may help in understanding the underlying stages in molecular evolution correlating with the origin of cell structure complexity.

Methods

Data normalization

In order to evaluate the significance of transformation of NLS-like motifs among bacteria, Planctomycetaceae and fungi groups, index Q value was developed, in which the sizes of protein pool and genome, and gene amount were considered for normalization, defined as:

$$Q_i = \frac{M_i}{\lg(N_i * G_i)}$$

Where M_i is the NLS-like motif abundance in i th species, and N_i and G_i are the gene amount and genome size of the i th species respectively. The larger Q value, the more NLS-like motifs harbor in the i th species.

Principal component analysis

Covariance analysis used software CanoDraw for Windows 4.0 (<http://www.canodraw.com/>) with diagrams processed in Adobe Illustrator CS6 [37].

Ortholog retrieval

Orthologs were determined using software OrthoMCL [38]. At first, this program conducts an all-against-all BLASTp search in BLAST 2.2.25. OrthoMCL then converts the reciprocal BLAST p-values to a normalized similarity matrix that is analyzed using a Markov Cluster algorithm (MCL). This yields many clusters, each containing a set of orthologs and/or recent paralogs. The BLAST e-value cut-off was $\leq 1e^{-5}$; other parameters were defaults.

Evaluation of the complexity of NLSs-like motifs

We generated a score matrix considering diversity, length and origin of NLS-like motifs (Table 1). We measured the complexity of the NLS motif from two aspects: the length and diversity of the motif (in structure), and the evolutionary origin of the motif (in evolution). We calculated scores with simple conversions or formula based on the appearance/abundance of the motif in the 27 genomes, and the methods (conversion and formula) were also described there (Table 1).

NLS-bearing proteins abundance in the 27 strains

We obtained all 114 experimentally identified NLS motifs from NLS database (<http://roslab.org/services/nlsdb/browse.php>). After searching the 114 NLS motifs in 27 genomes, we obtained 1101 NLS-proteins (Additional file 1: Table S1, A), and generated a heat-map with R software (version 2.13.0).

Function annotation and metabolic pathway analysis

Functions of NLS-bearing protein families were assigned using the best match of the alignments using BLASTp (E-value $\leq 10^{-5}$) searching against the SwissProt (Release 15.10) [39] and KEGG databases (Release 48.2) [40]. If the best hit of the proteins with any of these processes was “function unknown,” or “putative,” second-best hits were used to assign function until no additional hits met the

Table 1 Score matrix of the 10 NLS-motifs in the 13 NLS-bearing protein families, considering length, diversity and origin of NLS-like motifs

NLS motif	length	diversity	origin
PRRRK	5	3	5.66
RKRKK	5	3	5.98
RPRRK	5	3	5.38
KRPRP	5	3	5.36
RKRRR	5	2	7.96
P.KKKRK	7	4	5.3
TKRS...M	8	11	5.14
KR.{10}KKKL	16	23	5.12
K[RK]{3,5}.{11,18}[RK]K.{2,3}K	25	38	8.54
[KR]{4}.{20,24}K{1,4}.K	31	58	11.4

Note: ‘length’ means total amino acid number in NLS motif; ‘diversity’ refer to total amino acid types in a NLS motif, symbols as ‘.’ or ‘[]’ in NLS motif will be given an extra reward score (+1, + 0.5 respectively); ‘origin’ was used here to judge the potential evolutionary dominance of NLS motif through classification of the species accommodating the NLS motif. Taking Planctomycetaceae as control (score = 0), if one NLS motif is only found in non-planctomycete bacteria but not found in eukaryotic or Planctomycetaceae species, this NLS motif will be regarded as having low evolutionary dominance, and its scored will be punished (− 5) regardless of NLS motif abundance in these species. In contrast, if one NLS motif was detected in a eukaryotic species, it will obtain a reward score (+5), meanwhile, if the average abundance of the NLS motif in a eukaryotic species is high, that means, this NLS motif will have much higher evolutionary dominance, and it will be given an extra reward score (reward score = average abundance of the motif in eukaryotes/10)

alignment criteria. Analysis of metabolic pathways was performed by ipath 2.0 (<http://pathways.embl.de/>) using the assigned KO numbers in KEGG Orthology system.

Results

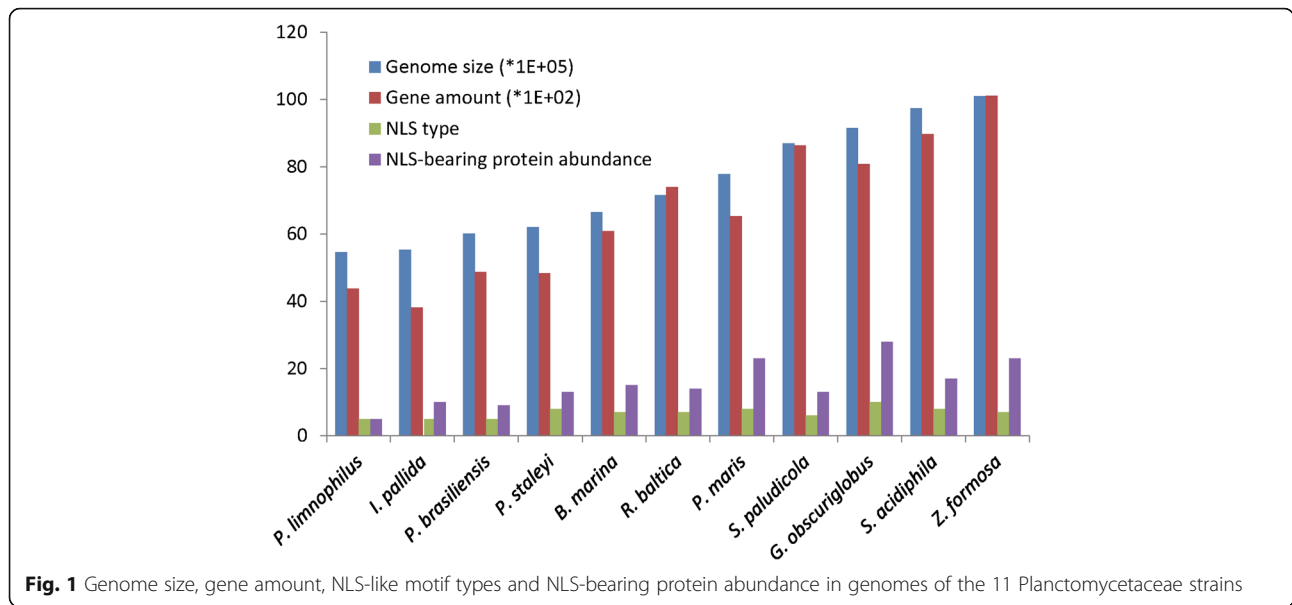
NLS-like motifs in the family Planctomycetaceae

To date, 114 experimentally identified NLS motifs are in the NLS database (<http://roslab.org/services/nlsdb/browse.php>). After searching protein pools of the 11 Planctomycetaceae species using amino acid sequences of the 114 NLS motifs, a total of 15 NLS types and 170 NLS-bearing proteins were detected in the family Planctomycetaceae. We arranged the order of the 11 species in Planctomycetaceae on the basis of genome size (Fig. 1). Multiple regression analysis indicated that NLS type or NLS-bearing proteins abundance express insignificant correlations with genome size or gene amount ($P > 0.05$). However, the double-layer ICMS strain *G. obscuriglobus* had the most abundant NLS-bearing proteins (28) and the most NLS types (10) in the family Planctomycetaceae. Both the NLSs KR.{10}KKKL (the dot means any amino acid; the number in brace means copy number) and KAKRQR were seen and the highest frequency of RKRRR was observed in *G. obscuriglobus* compared to other strains in the family.

NLS-bearing protein abundance of 27 strains

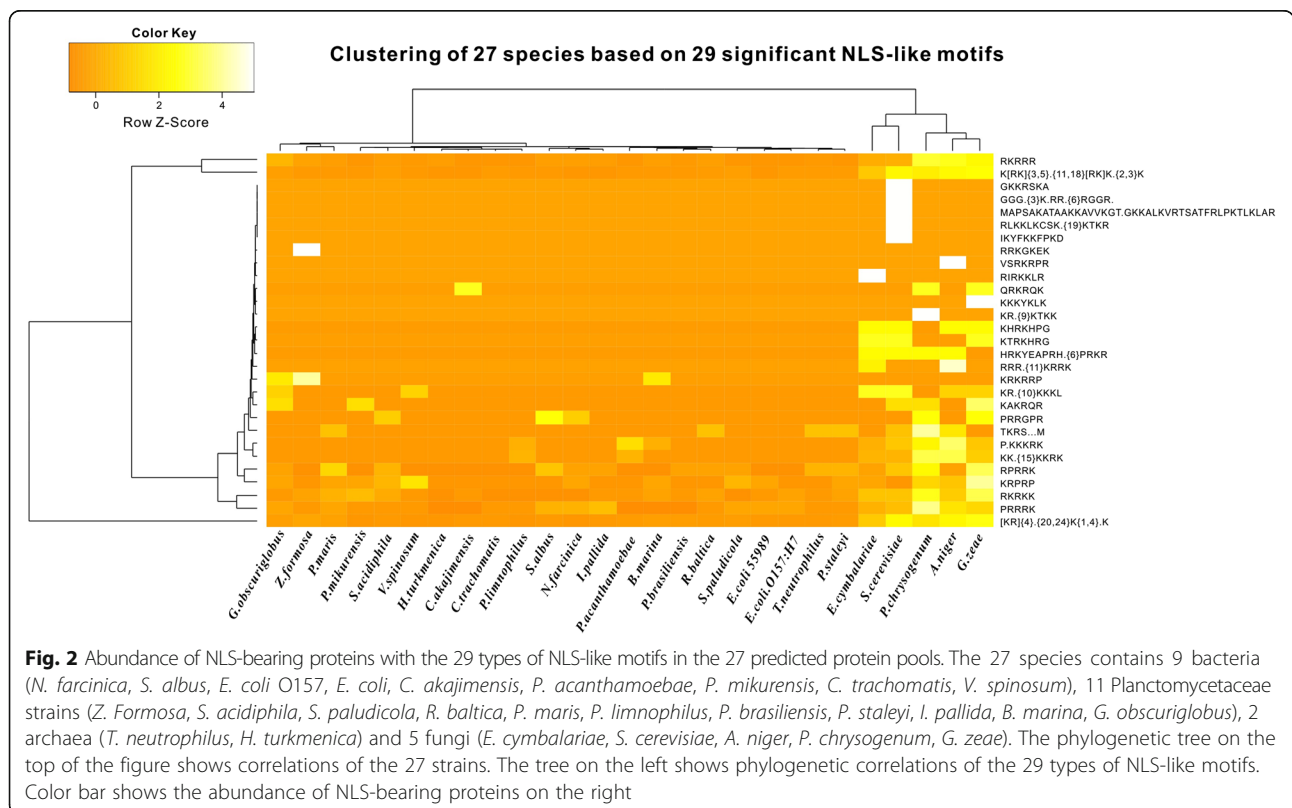
To better illustrate the relative distribution and abundance of NLS and NLS-bearing proteins in Planctomycetaceae relative to other groups of bacteria and eukaryotes, as judged by comparative genomics, to the analysis of 11 strains of Planctomycetaceae we added 16 extra representative microbes from different microbial taxonomical communities and retrieved their genomes from NCBI database (<ftp://ftp.ncbi.nlm.nih.gov/genomes/>). Phylogenetic relatives of Planctomycetaceae [41], especially two members of the Planctomycetaceae-Verrucomicrobia-Chlamydiae (PVC) superfamily [1], were included in the analysis for comprehensive phylogenetic representation. Through searching the 27 predicted protein pools (the 11 Planctomycetaceae strains and the other 16 microorganisms) using the 114 identified NLS motifs, we discovered 29 NLS types and 1101 NLS-bearing proteins (Additional file 1: Table S1, A). For the 29 NLS motifs, 15 NLS types were detected in the family Planctomycetaceae, and the rest of the NLS types were discovered in eukaryotes. ‘QRKRQK’ was only found in non-planctomycete bacteria and eukaryotes; ‘RRKGKEK’ and ‘KRKRPP’ were only found in Planctomycetaceae.

Correlations between the 27 strains were shown by the occurrence frequencies of NLS-bearing proteins with the 29 types of NLS-like motifs in their predicted protein pools (Fig. 2). Phylogenetically, the 27 strains were divided approximately into two branches. The first branch includes bacteria, Planctomycetaceae, and archaea; the



second contains only fungi. In Fig. 2, eukaryotic organisms possessed more NLS-bearing proteins and frequently had longer and more diverse NLS-like motifs than bacteria, and prokaryotes tended to have simple and short NLS-like motifs. However, many short, simple NLS-like motifs were still widely found in fungi (Additional file 1: Table S1, A). We hypothesized

that some short and simple NLS-like motifs were inherited from an evolutionary ancestor, before activation of their function. Afterward, these motifs were first activated and extensively used in NLS-bearing proteins from the perspective of evolutionary economics of energy consumption. Some longer and more complicated NLS-like motifs then appeared in eukaryotic species to meet higher



or special demands of intracellular molecular communication. Our results were consistent with this hypothesis, as shown in Fig. 2, the PRRRK, RKRKK, KRPRP and RPRRK NLS-like motifs, which appeared in bacteria (including non-planctomycete bacteria and planctomycete bacteria), were dramatically increased in fungi. By contrast, the GKRSKA, IKYFKKFPKD, and K[RK]{3,5}x{11,18}[RK]Kx{2,3}K (where x is any amino acid; the characters in bracket means alternative) motifs only occurred in fungi. Curiously, the bipartite NLS-like motif [KR]{4}x{20,24}K{1,4}xK was found in all the 27 strains. We attributed the emergence of long bipartite NLS-like motif in non-planctomycete bacteria to the high plasticity of this NLS-like motif regardless of its length.

After normalizing the data of NLS-bearing proteins abundance in the 27 genomes considering genome size and protein quantity (Additional file 1: Table S1, B), we detected a significant correlation between the 27 strains. Principal component analysis showed a significant transitional gradient (revealed by euclidean distance: planctomycete groups displayed a shorter euclidean distance to eukaryotic groups than non-planctomycete bacteria, Fig. 3) from bacteria to Planctomycetaceae to fungi in NLS-bearing protein abundances. Planctomycetaceae species separated slightly from bacteria, but were substantially distinguished from fungi. Remarkably, two Planctomycetaceae species, *Z. formosa* and *G. obscuriglobus* stand closest to eukaryotes (Fig. 3, in red up-triangles).

Clustering and annotation of NLS-bearing proteins

We used the 993 non-redundant NLS-bearing proteins instead of all the 1101 NLS-like proteins for clustering and functional annotation. Shared Protein families of all 993 nonredundant NLS-like proteins are showed by venn diagram (Fig. 4), excluding orphan proteins. Fungi possessed the most NLS-bearing protein families and NLS-bearing proteins, but shared a very small number of them with Planctomycetaceae (four families) or bacteria (three families). Planctomycetaceae and bacteria shared more NLS-bearing protein families (nine families) [42]. The five fungal strains have as many as 144 unique NLS-bearing protein families. By contrast, bacteria have only eight unique NLS-bearing protein families, and Planctomycetaceae have 12 unique NLS-bearing protein families.

There were 727 NLS-bearing proteins were annotated in SWISS-PROT database (Additional file 1: Table S2), but only 537 were annotated in Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Additional file 1: Table S3). We aligned the eight homologous NLS-bearing proteins of the only one family shared among bacteria, Planctomycetaceae and fungi (Additional file 2: Dataset S1).

To better demonstrate the functional evolution of real NLS motifs, we investigated core/pan metabolic pathways using the annotated NLS-bearing proteins of Planctomycetaceae (49 NLS-bearing proteins) and fungi (457 NLS-bearing proteins). A total of 66 metabolic pathways were referred, in which fungi occupied 57 metabolic pathways. In Fig. 5, NLS-bearing proteins of Planctomycetaceae

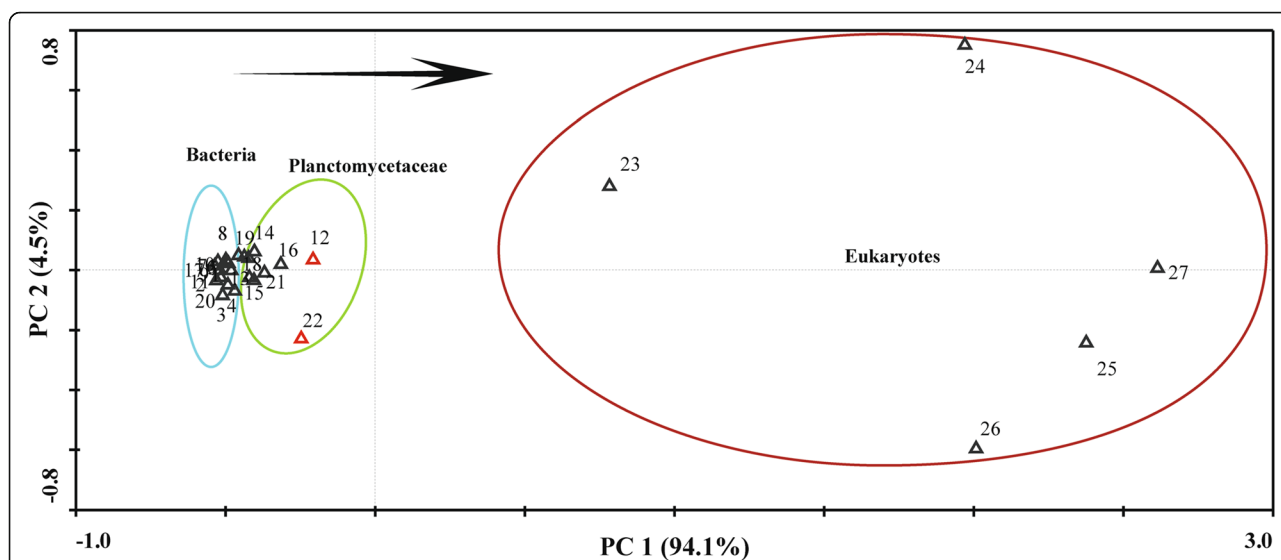
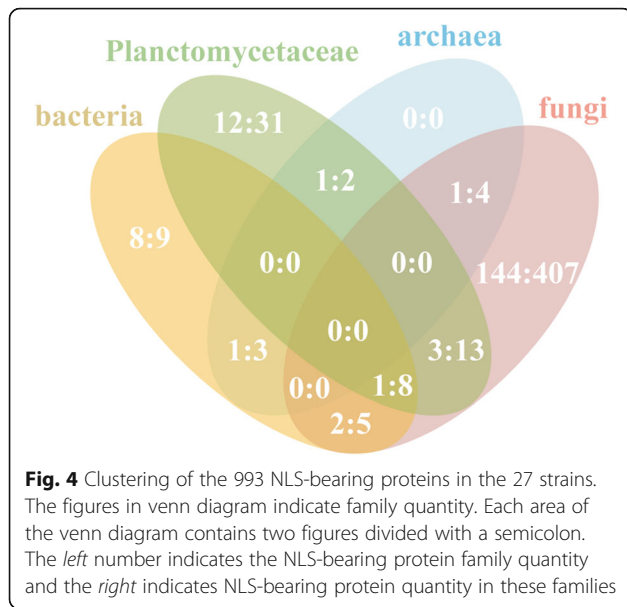


Fig. 3 PCA of the 27 strains. Distance between up-triangles approximates dissimilarity of abundance profiles of the NLS-bearing proteins in the 27 strains, measured by euclidean distance. Red, up-triangles show *G. obscuriglobus* (12) and *Z. formosa* (22). Numbers in the figure indicate: 1: *T. neutrophilus*; 2: *H. turkmenica*; 3: *N. farcinica*; 4: *S. albus*; 5: *E. coli* O157; 6: *E. coli*; 7: *C. akajimensis*; 8: *P. acanthamoebae*; 9: *P. mikurensis*; 10: *C. trachomatis*; 11: *V. spinosum*; 12: *Z. formosa*; 13: *S. acidiphila*; 14: *S. paludicola*; 15: *R. baltica*; 16: *P. maris*; 17: *P. limnophilus*; 18: *P. brasiliensis*; 19: *P. staleyi*; 20: *I. pallida*; 21: *B. marina*; 22: *G. obscuriglobus*; 23: *E. cymbalariae*; 24: *S. cerevisiae*; 25: *A. niger*; 26: *P. chrysogenum*; 27: *G. zeae*

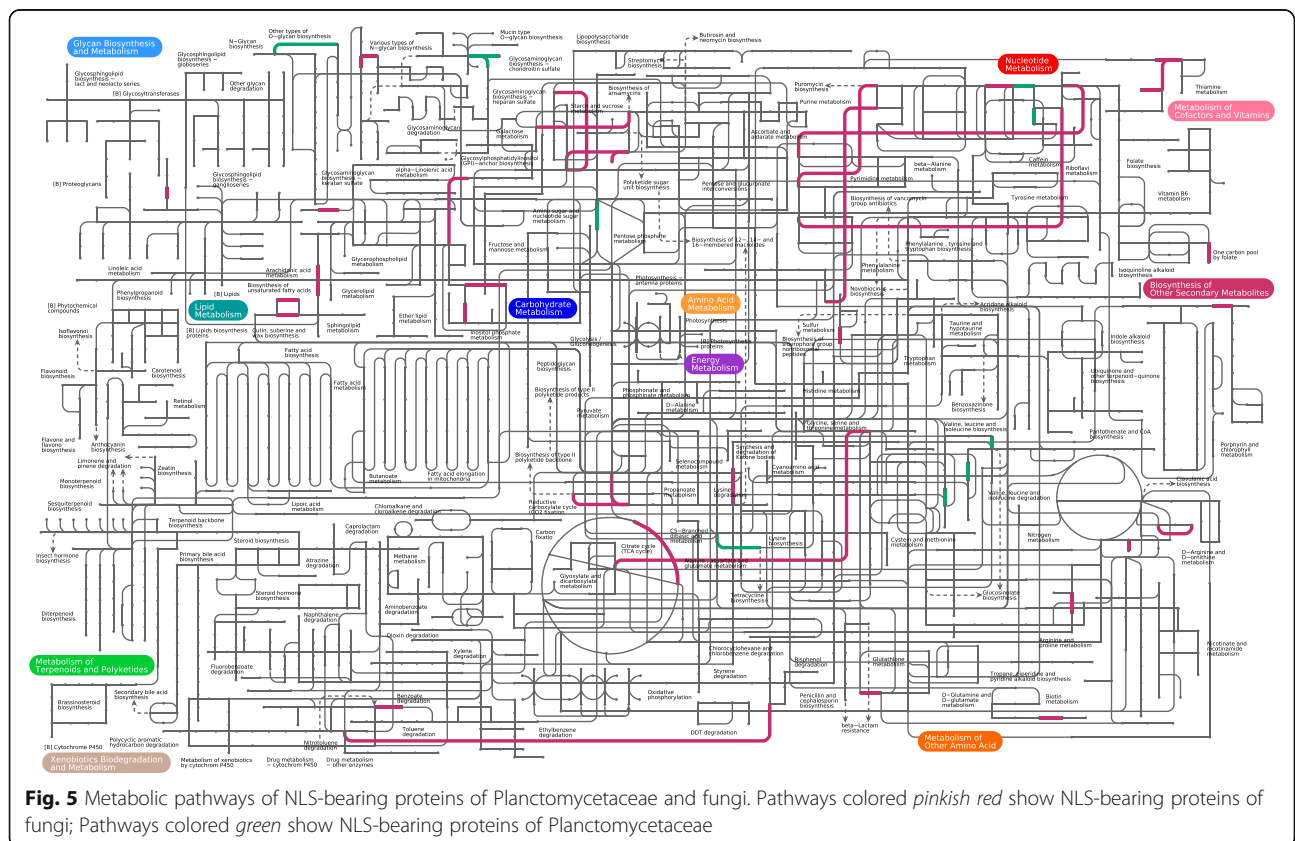


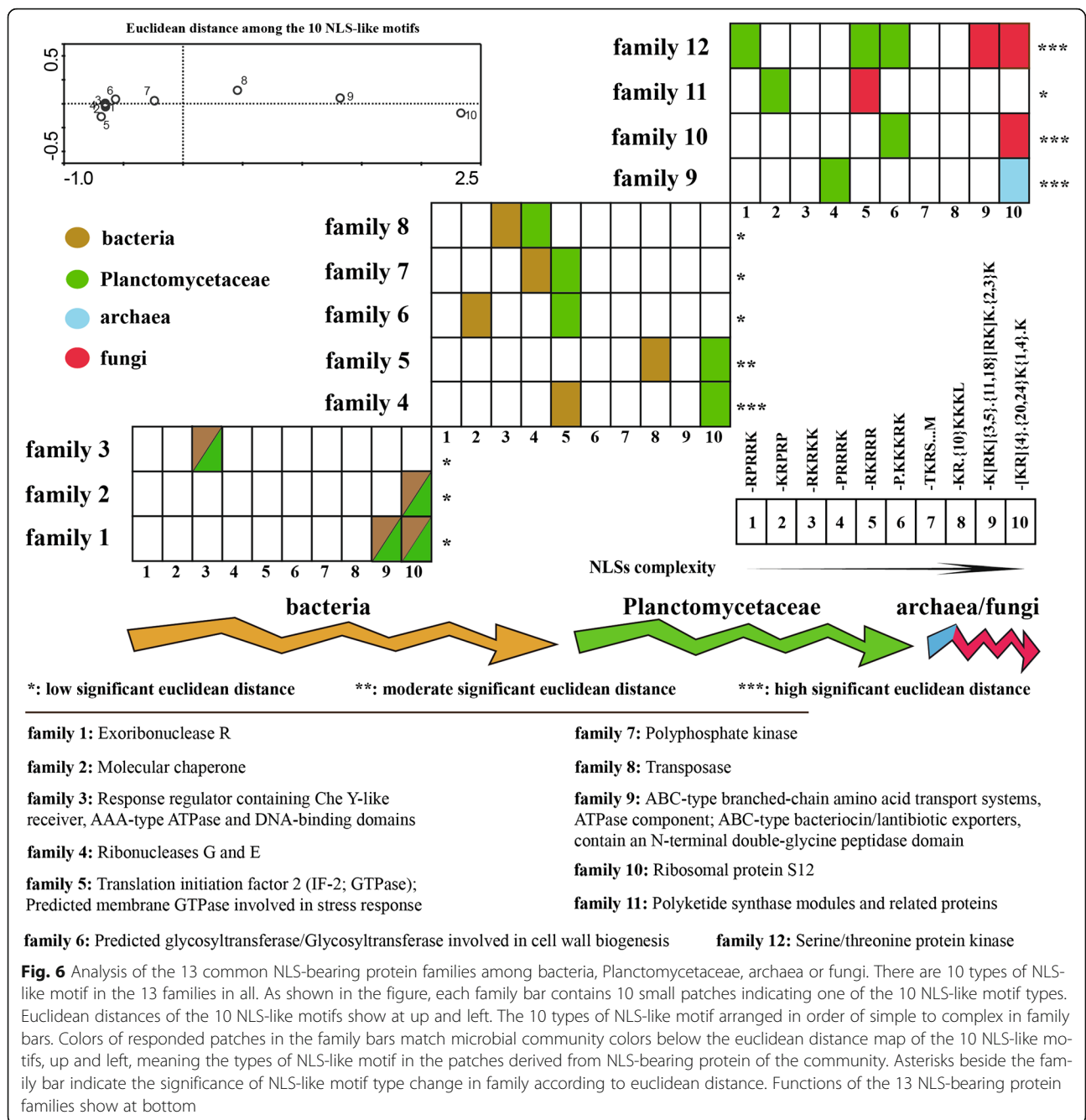
preliminarily stepped in a range of basic material metabolism, such as sulfates [43], O/N-glycan biosynthesis and metabolism, hydrophobic amino acid (valine, leucine and isoleucine) biosynthesis and purine metabolism. NLS-bearing proteins of fungi notably reinforced the pathways Planctomycetaceae’s NLS-bearing proteins referred and

extended the scopes to complex compound metabolism, particularly degradation of benzoate and its derivatives. Likewise, fungi’s NLS-bearing proteins shared with more regulatory pathways than Planctomycetaceae, NLS-bearing proteins of which mainly serve in ribosome and RNA degradation (Additional file 2: Figure S1). Interestingly, we found traces of Planctomycetaceae’s NLS-bearing proteins on protein export.

Transformation of NLS motifs in NLS-bearing protein families

To explore the potential transformations of NLS-like motifs in NLS-bearing protein families, we picked out 13 common NLS-bearing proteins families among bacteria, Planctomycetaceae, archaea or fungi for further analysis (Additional file 2: Dataset S1). The 13 NLS-bearing protein families contained 42 NLS-bearing proteins and ten types of NLS-like motif. We arranged the ten types of NLS-like motifs from simple to complex, considering their diversity, length and origin (Table 1). Consequently, the 13 NLS-bearing protein families were divided into three groups (Fig. 6). The first group contained three NLS-bearing protein families that are common to bacteria and Planctomycetaceae and harbored proteins with the same types of NLS-like motif. The second group contained five NLS-bearing protein families





that are common to bacteria and Planctomycetaceae and harbored proteins with different types of NLS-like motif. The third group contained five NLS-bearing protein families that are common to Planctomycetaceae and archaea or fungi and harbored proteins with different types of NLS-like motif. Interestingly, Planctomycetaceae showed small and large significant changes compared with bacteria and fungi respectively, based on analyzing complexity of the ten NLS-like motifs in the 13 NLS-bearing protein families among bacteria, Planctomycetaceae,

archaea or fungi. This result also point towards to presence of another “a-small-step-forward” genomic change in Planctomycetaceae species along the transformational gradient (Fig. 6). In Fig. 6, the first group accommodating proteins with common NLS-like motif(s) are three families between Planctomycetaceae strains and *Parachlamydia acanthamoebae* (family 2), *Verrucobium pinoum* (family 3) or *Chlamydia trachomatis* (family 1). All of these species belongs to the PVC superfamily [1] [44]. Likewise, *V. pinoum* created low significant changes on NLS-like motif complexity with

Planctomycetaceae members (family 5 and 6). Although *Phyciphaera mikureni* is one of Planctomycetaceae relatives, it revealed significant changes on NLS-like motif complexity with *Z. formosa* (family 4), which was also supported by their significant euclidean distance in Fig. 3.

NESs of the 27 strains

NESs are the functional counterparts to NLSs. NESs are leucine-rich stretches of 8 to 15 amino acids with regularly spaced hydrophobic residues that bind to the export karyopherin *CRM1*. La Cour et al. [25] published a NESbase (version 1.0) database with 75 entries with 80 experimentally determined NESs (<http://www.cbs.dtu.dk/databases/NESbase-1.0/db.html>). Xu et al. compiled an NES database that contained more than 230 experimentally validated leucine-rich NES-bearing *CRM1* cargoes [24, 45]. To investigate the proteins containing NES-like sequences in the 27 predicted protein pools, we collected 279 identified NES motifs that were sufficient to independently export a fused protein out of the nuclear envelope from the NES database constructed by Xu et al. [24]. The search identified only 14 NES-like proteins (Additional file 2: Dataset S2). These NES-bearing proteins were from fungi and were annotated as actin. Furthermore, few proteins in the 27 predicted protein pools perfectly matched the classical NES consensus sequence L-x(2,3)-[LIVFM]-x(2,3)-L-x-[LI] (where x represents any amino acid) [46].

Discussion

Though intracellular compartments, for instance magnetosomes [47], acidocalcisomes [48], chromatophores [49], thylakoids [50] and endospores [51], were reported in specific non-planctomycete bacterial groups, the layout of intracellular compartmentalization of Planctomycetaceae species seem to be more close to eukaryotes in morphology, especially to *G. obscuriglobus* [1, 20] and *Z. formosa* [52]. *Z. formosa* has the largest genome length and coding sequences quantity, and similar to *G. obscuriglobus*, it shows more complicated cellular compartmentalization structures than other species of Planctomycetaceae. Besides, in phylogenetic trees built with conserved positions of ribosomal RNA [53] or feature frequency profiles of whole proteomes [54], Planctomycetales consistently displayed an ancient and independent origin distinct from non-planctomycete bacterial groups, which is topologically in accordance with occurrence of “a-small-step-forward” genomic/complexity change of NLS-like motifs of Planctomycetaceae species when compared with non-planctomycete bacteria.

A number of factors constrained this study. First, more than half of the 11 Planctomycetaceae genomes including *G. obscuriglobus* and *Z. formosa* remain incomplete; second, lots of KEGG Orthology (KO) numbers of NLS-proteins of Planctomycetaceae were excluded from the reconstructed metabolic pathways; third, few experimentally identified NLS/NES motifs deposited in existing databases narrowed genomic searching results of NLS-like motifs. The NLS-like motifs in bacteria may not have the same function as the corresponding eukaryotic NLS motifs. Eubacteria do not have functional NLS-bearing proteins because they do not have a nuclear envelope. The predicted NLS-like motifs in these domains are merely sequence similarities and intended to illustrate the transformational rules of the motif among bacteria, Planctomycetaceae, and fungi. Further studies are required to confirm if these NLS-like components in bacteria are direct functional precursors of the NLS-like motifs in Planctomycetaceae and fungi. In addition, although transcriptomic and proteomic studies of Planctomycetaceae species *Rhodopirellula baltica* (the first Planctomycetaceae species with its genome completely sequenced) have been reported [55–58], however, in perspective of organic evolution, there is still an urgent need transcriptomic and proteomic studies centering on *G. obscuriglobus* or *Z. formosa* in future.

Conclusion

The genomic exploration of NLS-like motifs in species of family Planctomycetaceae provided us with insights into possible genomic changes contributing to the evolution of NLS and nuclear membranes. In the study, we focused on NLS-bearing proteins in 11 strains of the family Planctomycetaceae using comparative genomic approaches. We detected “a-small-step-forward” transitional gradients from non-planctomycete bacteria to Planctomycetaceae to fungi in abundance of NLS-bearing proteins or in complexity of NLS-like motifs evolved in the 13 clustered NLS-bearing protein families (presumable orthologous NLS-bearing proteins) in the 27 strains. The findings expanded our knowledge about the genomic features of family Planctomycetaceae and will facilitate understanding about the impact of NLS motifs in cellular development. The results suggest that a next step might be experimental test of function of NLS-sequences of planctomycetes within a eukaryote cell context (similar to past experiments with *Thermoplasma* and *Streptomyces*) and future experiments aimed at localizing NLS-bearing proteins in relation to cell compartments of *G. obscuriglobus* in particular may be informative.

Additional files

Additional file 1: Table S1 (A). NLS-bearing protein abundance in the 27 strains; **Table S1 (B).** Normalized NLS-bearing protein abundance of the 27 strains. **Table S2.** Annotation of the 993 NLS-bearing proteins with SwissProt database. **Table S3.** Annotation of the 993 NLS-bearing proteins with SwissProt database. (XLSX 107 kb)

Additional file 2: Figure S1. Regulatory pathways of NLS-bearing proteins of Planctomycetaceae and fungi. Pathways colored pinkish red show NLS-bearing proteins of fungi; pathways colored green show NLS-bearing proteins of Planctomycetaceae; pathways colored light blue show the common Regulatory pathways between Planctomycetaceae and fungi.

Dataset S1. The 13 clustered NLS-bearing protein families among non-planctomycete bacteria, Planctomycetaceae or fungi. **Dataset S2.** NES-bearing proteins in the predicted protein pools of the 27 strains. (PDF 408 kb)

Abbreviations

A: Ala; C: Cys; D: Asp; E: Glu/Gln; F: Phe; G: Gly; H: His; I: Ile; K: Lys; KEGG: Kyoto Encyclopedia of Genes and Genomes; L: Leu; M: Met; N: Asn; NCBI: National Center for Biotechnology Information; NLSs: Nuclear localization signals; P: Pro; Q: Gln; R: Arg; S: Ser; T: Thr; V: Val; W: Trp; Y: Tyr

Acknowledgements

We are very grateful to Professor Tianshen Tao and Professor Chengxiang Fang from Wuhan University for their encouragements and thank Professor Inchio Lou from Faculty of Science and Technology, Department of Civil and Environmental Engineering, University of Macau for his kind advice regarding this project.

Funding

This study was supported by grants from the Science and Technology Development Fund (FDC) of Macao SAR (Ref. No. 069/2015/A2 and No. 134/2014/A3) and Research Committee, University of Macau (MYRG2015-00182-ICMS-QRCM, MYRG2015-00214-ICMS-QRCM, MYRG139(Y1-L4)-ICMS12-LMY, and MYRG2016-00129-ICMS-QRCM).

Availability of data and materials

Genomes of the 27 strains belonging to non-planctomycete bacteria, Planctomycetaceae, archaea or fungi are available in NCBI database. They are 2 Archaea: *Thermoproteus neutrophilus* (NC_010525) and *Haloterrigena turkmenica* (NC_013743); 9 Bacteria: *Nocardia farcinica* (NC_006361), *Staphylococcus albus* (NZ_ABYC000000000), *Escherichia coli* O157:H7 (NC_002695), *E. coli* 55989 (NC_011748), *Coraliomargarita akajimensis* (NC_014008), *Parachlamydia acanthamoebae* (NC_015702), *Phycisphaera mikurensis* (NC_017080), *Chlamydia trachomatis* (NC_007429), and *Verrucomicrobium spinosum* (NZ_ABI200000000); 2 Yeast species of fungi: *Saccharomyces cerevisiae* (NC_001134) and *Eremothecium cymbalariae* (NC_016449); 3 Other species of filamentous fungi: *Penicillium chrysogenum* (NS_000201), *Aspergillus niger* (NC_007445), and *Gibberella zeae* (NC_009493) and 11 type strains of family Planctomycetaceae: *Planctomyces limnophilus* DSM 3776^T (NC_014148), *Isosphaera pallida* DSM 9630^T (NC_014962), *Planctomyces brasiliensis* DSM 5305^T (NC_015174), *Pirellula staleyi* DSM 6068^T (NC_013720), *Blastopirellula marina* DSM 3645^T (AANZ000000000), *Rhodopirellula baltica* DSM 10527^T (NC_005027), *Planctomyces maris* DSM 8797^T (NZ_ABCE000000000), *Gemmata obscuriglobus* DSM 5831^T (ABGO000000000)[†] and three of the 11 species of family Planctomycetaceae were previously sequenced by us [59]: *Schlesneria paludicola* DSM 18465^T (AHZR000000000), *Singulisphaera acidiphila* DSM 18658^T (AHZQ000000000), *Zavarzinella formosa* DSM 19928^T (AIAB000000000).
[†]: the draft genome of *G. obscuriglobus* was submitted to NCBI databases by J. Craig Venter Institute in 2007, the total length of its genome is about 9.16 Mb, and the genome composed of 922 contigs, GC% = 67.2%. After annotation, a total of 7828 potential genes were detected, including 7060 coding genes, 677 pseudo genes and 91 RNA genes.

Authors' contributions

MG, RY, CS. and SML designed research; MG performed research; MG, QL and CH contributed analytic tools; MG analyzed data; and MG and SML wrote the paper. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Publisher's Note

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

Author details

¹State Key Laboratory of Quality Research of Chinese Medicine and Institute of Chinese Medical Sciences, University of Macau, Macao, China. ²State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology, Beijing, China. ³Department of Microbial Chemistry, Institute of Medicinal Biotechnology, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China.

Received: 13 May 2016 Accepted: 11 March 2017

Published online: 04 April 2017

References

- Fuerst JA, Sagulenko E. Beyond the bacterium: planctomyces challenge our concepts of microbial structure and function. *Nat Rev Microbiol.* 2011;9:403–13.
- Kulichevskaya IS, Ivanova AO, Belova SE, Baulina OI, Bodelier PL, et al. *Schlesneria paludicola* gen. nov., sp. nov., the first acidophilic member of the order Planctomycetales, from Sphagnum-dominated boreal wetlands. *Int J Syst Evol Microbiol.* 2007;57:2680–7.
- Kulichevskaya IS, Ivanova AO, Baulina OI, Bodelier PL, Damste JS, et al. *Singulisphaera acidiphila* gen. nov., sp. nov., a non-filamentous, Isosphaera-like planctomycete from acidic northern wetlands. *Int J Syst Evol Microbiol.* 2008;58:1186–93.
- Jetten MS. The microbial nitrogen cycle. *Environ Microbiol.* 2008;10:2903–9.
- Drancourt M, Prebet T, Aghnatiou R, Edouard S, Cayrou C, et al. Planctomyces DNA in febrile aplastic patients with leukemia, rash, diarrhea, and micronodular pneumonia. *J Clin Microbiol.* 2014;52:3453–5.
- Buckley DH, Huangyutitham V, Nelson TA, Rumberger A, Thies JE. Diversity of Planctomyces in soil in relation to soil history and environmental heterogeneity. *Appl Environ Microbiol.* 2006;72:4522–31.
- Wang J, Jenkins C, Webb RI, Fuerst JA. Isolation of Gemmata-like and Isosphaera-like planctomycete bacteria from soil and freshwater. *Appl Environ Microbiol.* 2002;68:417–22.
- Schlesner H. The Development of Media Suitable for the Microorganisms Morphologically Resembling Planctomyces Spp, Pirellula Spp, and Other Planctomycetales from Various Aquatic Habitats Using Dilute Media. *Syst Appl Microbiol.* 1994;17:135–45.
- Jeske O, Jogler M, Petersen J, Sikorski J, Jogler C. From genome mining to phenotypic microarrays: Planctomyces as source for novel bioactive molecules. *Antonie Van Leeuwenhoek.* 2013;104:551–67.
- Pearson A, Budin M, Brocks JJ. Phylogenetic and biochemical evidence for sterol synthesis in the bacterium Gemmata obscuriglobus. *Proc Natl Acad Sci U S A.* 2003;100:15352–7.
- Santarella-Mellwig R, Franke J, Jaedicke A, Gorjanacz M, Bauer U, et al. The compartmentalized bacteria of the planctomyces-verrucomicrobia-chlamydiae superphylum have membrane coat-like proteins. *PLoS Biol.* 2010;8:e1000281.
- Fuerst JA, Webb RI. Membrane-bounded nucleoid in the eubacterium Gemmata obscuriglobus. *Proc Natl Acad Sci U S A.* 1991;88:184–8.
- van Teeseling MC, Mesman RJ, Kuru E, Espaillet A, Cava F, et al. Anammox Planctomyces have a peptidoglycan cell wall. *Nat Commun.* 2015;6:6878.
- Jeske O, Schuler M, Schumann P, Schneider A, Boedeker C, et al. Planctomyces do possess a peptidoglycan cell wall. *Nat Commun.* 2015;6:7116.
- McInerney JO, Martin WF, Koonin EV, Allen JF, Galperin MY, et al. Planctomyces and eukaryotes: a case of analogy not homology. *Bioessays.* 2011;33:810–7.
- Lindsay MR, Webb RI, Strous M, Jetten MS, Butler MK, et al. Cell compartmentalisation in planctomyces: novel types of structural organisation for the bacterial cell. *Arch Microbiol.* 2001;175:413–29.
- Fuerst JA. Intracellular compartmentation in planctomyces. *Annu Rev Microbiol.* 2005;59:299–328.

18. Gottshall EY, Seebart C, Gatlin JC, Ward NL. Spatially segregated transcription and translation in cells of the endomembrane-containing bacterium *Gemmata obscuriglobus*. *Proc Natl Acad Sci U S A*. 2014;111:11067–72.
19. Santarella-Mellwig R, Pruggnaller S, Roos N, Mattaj JW, Devos DP. Three-dimensional reconstruction of bacteria with a complex endomembrane system. *PLoS Biol*. 2013;11:e1001565.
20. Sagulenko E, Morgan GP, Webb RI, Yee B, Lee KC, et al. Structural studies of planctomycete *Gemmata obscuriglobus* support cell compartmentalisation in a bacterium. *PLoS ONE*. 2014;9:e91344.
21. Jogler C, Waldmann J, Huang X, Jogler M, Glockner FO, et al. Identification of proteins likely to be involved in morphogenesis, cell division, and signal transduction in Planctomycetes by comparative genomics. *J Bacteriol*. 2012;194:6419–30.
22. Tran EJ, Wentz SR. Dynamic nuclear pore complexes: life on the edge. *Cell*. 2006;125:1041–53.
23. Nair R, Carter P, Rost B. NLSdb: database of nuclear localization signals. *Nucleic Acids Res*. 2003;31:397–9.
24. Xu D, Grishin NV, Chook YM. NESdb: a database of NES-containing CRM1 cargoes. *Mol Biol Cell*. 2012;23:3673–6.
25. la Cour T, Gupta R, Rapacki K, Skriver K, Poulsen FM, et al. NESbase version 1.0: a database of nuclear export signals. *Nucleic Acids Res*. 2003;31:393–6.
26. Liu X, Erikson RL. The nuclear localization signal of mitotic kinesin-like protein Mklp-1: effect on Mklp-1 function during cytokinesis. *Biochem Biophys Res Commun*. 2007;353:960–64.
27. Lee SMY, Li HY, Ng EKO, Or SMW, Chan KK, et al. Characterization of a brain-specific nuclear LIM domain protein (FHL1B) which is an alternatively spliced variant of FHL1. *Gene*. 1999;237:253–63.
28. Manganaro A, Pizzo F, Lombardo A, Pogliaghi A, Benfenati E. Predicting persistence in the sediment compartment with a new automatic software based on the k-Nearest Neighbor (k-NN) algorithm. *Chemosphere*. 2015;144:1624–30.
29. van der Waal D, den Heeten GJ, Pijnappel RM, Schuur KH, Timmers JM, et al. Comparing Visually Assessed BI-RADS Breast Density and Automated Volumetric Breast Density Software: A Cross-Sectional Study in a Breast Cancer Screening Setting. *PLoS ONE*. 2015;10:e0136667.
30. Kundrat M. When did theropods become feathered?—evidence for pre-Archaeopteryx feathery appendages. *J Exp Zool B Mol Dev Evol*. 2004;302:355–64.
31. Inoue D, Kabata T, Maeda T, Kajino Y, Fujita K, et al. Usefulness of three-dimensional templating software to quantify the contact state between implant and femur in total hip arthroplasty. *Eur J Orthop Surg Traumatol*. 2015;25:1293–300.
32. Zeng Y, Cullen BR. Sequence requirements for micro RNA processing and function in human cells. *RNA*. 2003;9:112–23.
33. Grziwa A, Dahlmann B, Cejka Z, Santarius U, Baumeister W. Localization of a sequence motif complementary to the nuclear localization signal in proteasomes from *Thermoplasma acidophilum* by immunoelectron microscopy. *J Struct Biol*. 1992;109:168–75.
34. Nederlof PM, Wang HR, Baumeister W. Nuclear localization signals of human and *Thermoplasma* proteasomal alpha subunits are functional in vitro. *Proc Natl Acad Sci U S A*. 1995;92:12060–4.
35. Tinland B, Koukolikova-Nicola Z, Hall MN, Hohn B. The T-DNA-linked VirD2 protein contains two distinct functional nuclear localization signals. *Proc Natl Acad Sci U S A*. 1992;89:7442–6.
36. Mans BJ, Anantharaman V, Aravind L, Koonin EV. Comparative genomics, evolution and origins of the nuclear envelope and nuclear pore complex. *Cell Cycle*. 2004;3:1612–37.
37. McLean D. Adobe Photoshop and Illustrator techniques. *J Audio Media Med*. 2001;24:79–82.
38. Li L, Stoeckert Jr CJ, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res*. 2003;13:2178–89.
39. Wu CH, Apweiler R, Bairoch A, Natale DA, Barker WC, et al. The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic Acids Res*. 2006;34:D187–91.
40. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28:27–30.
41. Yarza P, Richter M, Peplies J, Euzéby J, Amann R, et al. The All-Species Living Tree project: a 16S rRNA-based phylogenetic tree of all sequenced type strains. *Syst Appl Microbiol*. 2008;31:241–50.
42. Fuchsman CA, Rocap G. Whole-genome reciprocal BLAST analysis reveals that planctomycetes do not share an unusually large number of genes with Eukarya and Archaea. *Appl Environ Microbiol*. 2006;72:6841–4.
43. Lage OM, Bondoso J. Planctomycetes and macroalgae, a striking association. *Front Microbiol*. 2014;5:267.
44. Speth DR, van Teeseling MC, Jetten MS. Genomic analysis indicates the presence of an asymmetric bilayer outer membrane in planctomycetes and verrucomicrobia. *Front Microbiol*. 2012;3:304.
45. Nam KU, Hong J. Is Three-Dimensional Soft Tissue Prediction by Software Accurate? *J Craniofac Surg*. 2015;26:e729–33.
46. Bogerd HP, Fridell RA, Benson RE, Hua J, Cullen BR. Protein sequence requirements for function of the human T-cell leukemia virus type 1 Rex nuclear export signal delineated by a novel in vivo randomization-selection assay. *Mol Cell Biol*. 1996;16:4207–14.
47. Komeili A, Li Z, Newman DK, Jensen GJ. Magnetosomes are cell membrane invaginations organized by the actin-like protein MamK. *Science*. 2006;311:242–5.
48. Seufferheld M, Lea CR, Vieira M, Oldfield E, Docampo R. The H(+)-pyrophosphatase of *Rhodospirillum rubrum* is predominantly located in polyphosphate-rich acidocalcisomes. *J Biol Chem*. 2004;279:51193–202.
49. Geyer T, Helms V. A spatial model of the chromatophore vesicles of *Rhodobacter sphaeroides* and the position of the Cytochrome bc1 complex. *Biophys J*. 2006;91:921–6.
50. Porta D, Rippka R, Hernandez-Marine M. Unusual ultrastructural features in three strains of Cyanothecae (cyanobacteria). *Arch Microbiol*. 2000;173:154–63.
51. McKenney PT, Eichenberger P. Dynamics of spore coat morphogenesis in *Bacillus subtilis*. *Mol Microbiol*. 2012;83:245–60.
52. Kulichevskaya IS, Baulina OI, Bodelier PL, Rijstpa WI, Damste JS, et al. *Zavarzinella formosa* gen. nov., sp. nov., a novel stalked, Gemmata-like planctomycete from a Siberian peat bog. *Int J Syst Evol Microbiol*. 2009;59:357–64.
53. Brochier C, Philippe H. Phylogeny: a non-hyperthermophilic ancestor for bacteria. *Nature*. 2002;417:244.
54. LaCasse EC, Lochman HA, Walker P, Lefebvre YA. Identification of binding proteins for nuclear localization signals of the glucocorticoid and thyroid hormone receptors. *Endocrinology*. 1993;133:2760.
55. Wecker P, Klockow C, Ellrott A, Quast C, Langhammer P, et al. Transcriptional response of the model planctomycete *Rhodopirellula baltica* SH1(T) to changing environmental conditions. *BMC Genomics*. 2009;10:410.
56. Gade D, Gobom J, Rabus R. Proteomic analysis of carbohydrate catabolism and regulation in the marine bacterium *Rhodopirellula baltica*. *Proteomics*. 2005;5:3672–83.
57. Hieu CX, Voigt B, Albrecht D, Becher D, Lombardot T, et al. Detailed proteome analysis of growing cells of the planctomycete *Rhodopirellula baltica* SH1T. *Proteomics*. 2008;8:1608–23.
58. Voigt B, Hieu CX, Hempel K, Becher D, Schluter R, et al. Cell surface proteome of the marine planctomycete *Rhodopirellula baltica*. *Proteomics*. 2012;12:1781–91.
59. Guo M, Han X, Jin T, Zhou L, Yang J, et al. Genome sequences of three species in the family Planctomycetaceae. *J Bacteriol*. 2012;194:3740–1.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

