

Reassessment of the Lineage Fusion Hypothesis for the Origin of Double Membrane Bacteria

Kristen S. Swithers¹, Gregory P. Fournier³, Anna G. Green¹, J. Peter Gogarten¹, Pascal Lapierre^{2*}

1 Department of Molecular and Cell Biology, University of Connecticut, Storrs, Connecticut, United States of America, **2** University of Connecticut Biotechnology Center, University of Connecticut, Storrs, Connecticut, United States of America, **3** Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America

Abstract

In 2009, James Lake introduced a new hypothesis in which reticulate phylogeny reconstruction is used to elucidate the origin of Gram-negative bacteria (Nature 460: 967–971). The presented data supported the Gram-negative bacteria originating from an ancient endosymbiosis between the Actinobacteria and Clostridia. His conclusion was based on a presence-absence analysis of protein families that divided all prokaryotes into five groups: Actinobacteria, Double Membrane bacteria (DM), Clostridia, Archaea and Bacilli. Of these five groups, the DM are by far the largest and most diverse group compared to the other groupings. While the fusion hypothesis for the origin of double membrane bacteria is enticing, we show that the signal supporting an ancient symbiosis is lost when the DM group is broken down into smaller subgroups. We conclude that the signal detected in James Lake's analysis in part results from a systematic artifact due to group size and diversity combined with low levels of horizontal gene transfer.

Citation: Swithers KS, Fournier GP, Green AG, Gogarten JP, Lapierre P (2011) Reassessment of the Lineage Fusion Hypothesis for the Origin of Double Membrane Bacteria. PLoS ONE 6(8): e23774. doi:10.1371/journal.pone.0023774

Editor: Jonathan H. Badger, J. Craig Venter Institute, United States of America

Received: May 17, 2011; **Accepted:** July 25, 2011; **Published:** August 18, 2011

Copyright: © 2011 Swithers et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported through grants from the National Aeronautic and Space Administration Exobiology Programs (NNX08AQ10G) and the National Science Foundation Assembling the Tree of Life Programs (DEB 0830024). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: pascal.lap@gmail.com

Introduction

Symbioses and endosymbioses have shaped and continue to shape microbial evolution [1]. As such, it is of little surprise that endosymbiotic events and chimaerism are often considered useful hypotheses for explaining the phylogenetic and gene content complexities of bacterial, archaeal and eukaryotic genomes. James Lake used a reconstruction of reticulate phylogeny to argue that the double membrane bacteria evolved from an ancient symbiosis (endosymbiosis) between Clostridia and Actinobacteria [2]. By applying a parsimony analysis of protein family presence absence data over five distinct groups of prokaryotes [3], he identified sets of proteins present in double membrane bacteria (DM) that originated from either Clostridia or Actinobacteria. Since the highest number of protein families from the presence-absence patterns had better support for a ring structure compared to a single bifurcating tree, he concluded that the most likely explanation for the data was a fusion event between Clostridia and Actinobacteria. If this fusion occurred through an endosymbiosis, it could also explain the origin of the double membrane architecture. This view has been supported by interpreting polarizing indels (insertions or deletions) within several protein families as excluding the bacterial root from within Actinobacteria and DM bacteria [2,4], compatible with a monophyletic fusion origin of DM bacteria with Actinobacteria as a participating lineage. Additionally, it was argued that the photosynthetic machinery would resist being transferred because of its complexity, and thus be a good candidate to study ancient divergences [2]. These assumptions and the results of the aforementioned analyses

have not gone without criticism, on both theoretical and methodological grounds [5].

One problem with this analysis is that the group designated DM is comprised of many rather divergent groups of bacteria, such as the Dictyoglomi, Thermotogae, *Deinococcus-Thermus*, Cyanobacteria and the different classes of Proteobacteria (see materials and methods for full listing). The definition of what constitutes a genuine double membrane compared to an external proteolipid or protein layer is unclear, and the constituents of the outer layer are difficult to determine [6]. For this reason, the majority of phyla included as double-membrane organisms are controversial and have possibly introduced an artifactual signal in favor of a fusion. Given the amount of interdomain and interphylum horizontal gene transfer that has been identified (e.g., [7,8,9]), one should expect a larger group of organisms to harbor more different protein families than a smaller group. This alternative explanation for Lake's data is testable; if the reticulate signal detected by Lake were due to many transfers of individual genes and operons, it should diminish if the DM group is replaced in the analysis with any of its biologically cohesive constituent subgroups. In contrast, if the signal were due to a single ancient endosymbiotic event at the root of the DM bacteria, then the signal should not disappear even if only a subgroup of the DM were selected in the analysis.

The claim that DM bacteria evolved from an ancient symbiosis is based on an analysis that aggregates all Bacteria and Archaea into 5 groups (the double membrane prokaryotes (DM), Actinobacteria (A), Bacilli (B), Clostridia (C), Archaea (R)), using the Pfam database [10] to determine the number of protein families that were represented in 3 out of the 5 aggregate groups. A protein

family (Pfam) was considered present in a group, if at least one genome within the group encoded a member of this family. The analysis produces a table of all possible combinations of presence-absence profiles and determines the most parsimonious scenario explaining the data (i.e., if they were generated by a tree-like or ring-like evolutionary process, see supplemental Figure S1). The ring structure proposed by Lake [2] joins the DM group to both the A and C groups given the allowed patterns in rows 5, 7, 8, 9 and 10 (Figure 1). The presence of a higher number of genes in those five rows compared to the tree signal (rows 1, 4, 7, 8, 9 and 10 in Lake 2009, Figure S2) reflects a higher number of genes shared by DM members with Actinobacteria and Clostridia. If the argument for a fusion event were valid, trends observed in the gene presence-absence table should not be affected by the breakup of the DM into sub-groups, as the presence of the protein family would be shared derived characters of all DM members.

Results

We repeated Lake’s analysis exactly using the same version of the Pfam database, and in addition to Lake’s DM group, we also analyzed the datasets that resulted after dividing the DM group into twelve subclasses (Figure 1, column one to twelve). We found that for most of the DM subgroups, tree patterns were more highly supported than the patterns allowed under the ring scheme proposed by Lake. Additionally, the signal supporting the hypothesis of an ancient endosymbiosis between Clostridia and Actinobacteria is completely lost (p-values in favor of a ring of 0.0035 or smaller) when these subgroups are used as representatives for the DM group (Figure 2 and S2). This result is compatible with the hypothesis that the reticulate signal is due to several HGTs of individual genes, operons, and gene clusters and not due to a single ancient fusion between lineages. The ring signal is retained only in one case, when all classes of the Proteobacteria are

combined (p-value of 0.98), possibly because this group contains the largest sampled biodiversity as reflected by the number of protein families in Pfam compared to the other groups included in the analysis. Figure 2 summarizes our results. We conclude that the deduced reticulate phylogeny appears to be due to many individual gene transfer events. The division of prokaryotes into groups of different size and containing different amounts of sampled protein diversity produces a systematic artifact suggestive of a fusion at the base the group comprised of the most diverse members.

Discussion

Lake’s result that the DM group arose via a fusion event implies that this group be monophyletic rather than paraphyletic, since this model is inconsistent with the DM ancestor giving rise to any other groups of bacteria included in the analysis. However, while it has been claimed that a polarizing indel within the HSP70/MreB gene families excludes the root of the “tree of life” from gram-negatives [11] it is likely that this result is impacted by extensive horizontal gene transfer, and is complicated by alignment and sampling artifacts [12,13]. More convincingly, a polarizing indel in the HisA/HisF protein families and the quaternary structure of PyrD homologs have also been used to exclude the root from most gram-negatives and actinobacteria [14]. While the results of these analyses do permit the monophyly of the DM group, they also permit any scenario where each DM subclass is derived, including a paraphyly or even polyphyly incompatible with the assumptions in [2]. For these reasons, indels do not provide support for a monophyletic DM group as described in (Lake 2009). The argument that the photosynthetic machinery is reluctant to gene transfers because of its complexity, thus linking the Clostridia from one side of the ring of life to the DM bacteria, is similarly weak: previous reports have shown that many photosynthesis genes,

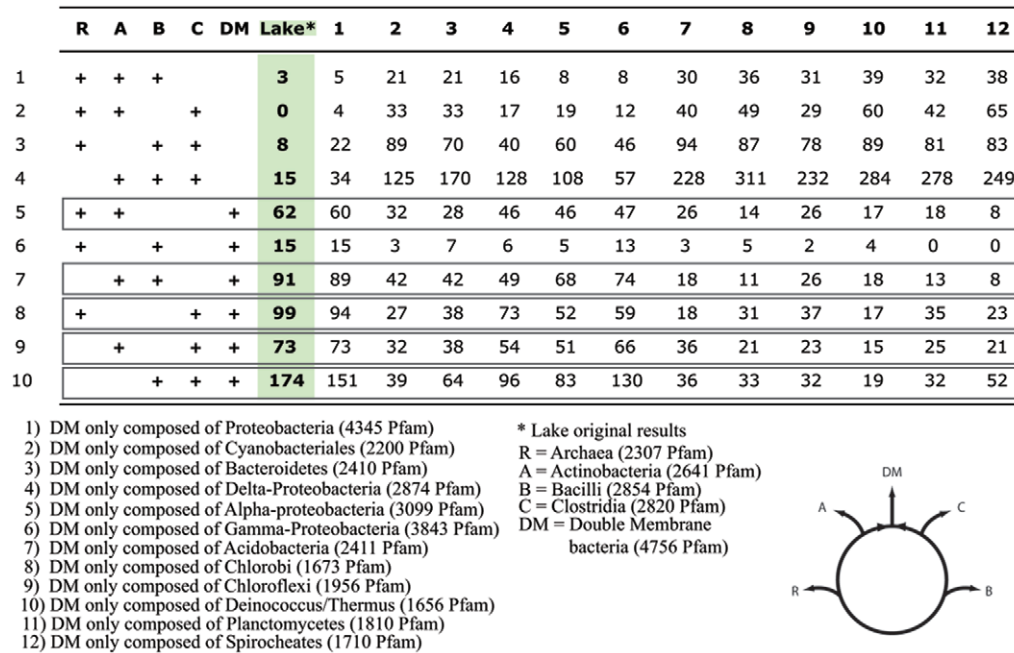


Figure 1. Protein family counts for the ten possible informative profiles. The table was adapted from Lake’s Table 1 [2] to include the Pfam counts that result if different representative classes are chosen for the DM group. Number of Pfam per group is in parentheses the same number as in Lake’s paper was found for all other groups. The circle illustrates Lake’s hypothesis that the double membrane bacteria resulted from a fusion between Clostridia and Actinobacteria. The patterns compatible with this hypothesis are boxed (pattern 5,7,8,9 and 10). doi:10.1371/journal.pone.0023774.g001

	Lake Groups	Proteobact. 1	Cyanobact. 2	Bacteroides 3	Delta- 4	Alpha- 5	Gamma- 6	Acidobact. 7	Chlorobi 8	Chloroflexi 9	Deino/Thermus 10	Plancto. 11	Spiro. 12
Ring Supports :	1	0.9812	0.0001	0.002	0.001	0	0.0001	0	0.0035	0.0002	0	0.0116	0.0023
Tree Supports :	0	0.0148	0.9998	0.9973	0.9983	1	0.9999	0.9999	0.9942	0.9991	1	0.9838	0.9971
Ties :	0	0.004	0.0001	0.0007	0.0007	0	0	0.0001	0.0023	0.0007	0	0.0046	0.0006

Figure 2. Posterior bootstrap support values (p-values) for a ring model, tree model or equal probabilities for each of the sampled groups. The p-values were calculated from 10,000 re-samplings with replacement and extracting the total number of times the tree model, ring model or when both were equally supported from the parsimony counts. Only in the case where all the double membrane prokaryotes as defined by Lake [2], or when all the proteobacteria were included, did a ring model better explain the data. doi:10.1371/journal.pone.0023774.g002

including the chlorophyll biosynthetic pathway were transferred between bacterial classes and phyla [15,16]. Sharon *et al.* have even reported the discovery of a complete photosystem I operon in a marine phage [17], and analysis by Igarashi *et al.* [18] suggested that a photosynthetic gene super-cluster in the β -Proteobacteria was acquired through transfer from the α -Proteobacteria.

Endosymbiosis between two single-membrane organisms as proposed by Lake [2] offers a possible scenario for the evolution of double membranes as a derived character possessed by descendants of the fusion event. Our analysis using subclasses of DM bacteria shows that the parsimony approach of Lake [2] only found an origin by fusion for groups of organisms containing a large amount of protein diversity, as evidenced by our recovery of a predicted fusion event only for the Proteobacteria subclass, and not for any other constituent subclass of DM bacteria. Therefore, presence-absence analysis of protein families does not provide evidence of reticulate evolution between Actinobacteria and Clostridia.

Materials and Methods

The complete Pfam database v.22.0 was downloaded from ftp.sanger.ac.uk/pub/databases/Pfam/releases/Pfam22.0/ and was locally searched for presence of protein families across different groups. We divided the Pfam database into five groups according to Lakes specifications as described in [2]. Group 1 was composed of all the Archaeal protein families; group 2 was composed of the Actinobacteria; Group 3 are the Bacilli which includes the Lactobacillales and the Bacillales; Group 4 was represented by the Clostridia and Mollicutes which also included the *Symbiobacterium*, Coriobacteriales and the Rubrobacteridae; and finally group five which represents all the double membrane prokaryotes (Acidobacteria, Aquificae, Bacteroidetes, Chrysiogenetes, Chloroflexi, Chlorobi, Chlamydiae, Cyanobacteria, Deferribacteres, Deinococcus/Thermus, Dictyoglomi, Fibrobacteres, Fusobacteria, Nitrospirae, Proteobacteria, Planctomycetes, Spirochaetes, Thermodesulfobacteria, Thermotogales and Verucomicrobia). Figure 1 shows the ten parsimonious informative character states for the five group comparisons. Following Lake’s methods, a protein family was deemed present if at least one member of the three subject groups contained that protein family and was absent in all members of both query groups. Using the

original group classification, we recovered the exact numbers of protein families for the ten-character state as described by Lake. We then compiled the number of protein families present when the double membrane group was broken up into twelve subgroups: Proteobacteria, Cyanobacteria, Bacteroidetes, δ -Proteobacteria, α -Proteobacteria, γ -Proteobacteria, Acidobacteria, Chlorobi, Chloroflexi, Deinococcus/Thermus, Planctomycetes and Spirochaetes. The posterior bootstrap support values (p-values) for all possible ring and tree models were calculated from 10,000 re-samplings with replacement and extracting the total number of times the tree model, ring model or both were equally supported. For each bootstrap replicate, the best supported model was determined by finding the tree or ring with the lowest minimum parsimony count. The minimum parsimony counts were calculated by weighting the number of Pfams supporting a particular tree or ring twice that of the number of Pfams that do not support the model [3].

Supporting Information

Figure S1 List of all possible trees and rings for five taxa sampling. Each possible tree and ring is listed with the compatible presence-absence pattern of gene families (Pfam) given in Figure 1. For example, the tree and ring corresponding to ABCDR are shown at the left of each table. A corresponds to Actinobacteria, B to Bacilli, C to Clostridia, D for double membrane prokaryotes and R for Archaea. (TIF)

Figure S2 Minimum parsimony counts supporting each of the possible trees (A) and rings (B). The lowest count is used to determine if the data supports a tree or a ring [3]. In the original analyses by Lake [2], the best ring had a minimum parsimony count of 581 versus 625 for the best supported tree (first column). Best supported trees or rings for each tested case are highlighted. (TIF)

Author Contributions

Conceived and designed the experiments: PL KSS JPG GPF. Performed the experiments: PL. Analyzed the data: PL KSS JPG GPF AGG. Wrote the paper: PL KSS JPG GPF.

References

- Margulis L (1995) Symbiosis in Cell Evolution : Microbial Communities in the Archean and Proterozoic Eons. W H Freeman & Co.
- Lake JA (2009) Evidence for an early prokaryotic endosymbiosis. Nature 460: 967–971.
- Lake JA (2008) Reconstructing evolutionary graphs: 3D parsimony. Mol Biol Evol 25: 1677–1682.
- Servin JA, Herbold CW, Skophammer RG, Lake JA (2008) Evidence excluding the root of the tree of life from the actinobacteria. Mol Biol Evol 25: 1–4.
- Pathogens: Genes and Genomes Website. Available : http://pathogenomics.bham.ac.uk/blog/2009/08/evidence-for-an-early-prokaryotic-endosymbiosis-i-dont-believe-it/. Accessed 2011, August 1st.
- Sutcliffe IC (2010) Cell envelope architecture in the Chloroflexi: a shifting frontline in a phylogenetic turf war. Environ Microbiol.
- Zhaxybayeva O, Swithers KS, Lapierre P, Fournier GP, Bickhart DM, et al. (2009) On the chimeric nature, thermophilic origin, and phylogenetic placement of the Thermotogales. Proc Natl Acad Sci U S A 106: 5865–5870.

8. Boussau B, Gueguen L, Gouy M (2008) Accounting for horizontal gene transfers explains conflicting hypotheses regarding the position of aquificales in the phylogeny of Bacteria. *BMC Evol Biol* 8: 272.
9. Beiko RG, Harlow TJ, Ragan MA (2005) Highways of gene sharing in prokaryotes. *Proc Natl Acad Sci U S A* 102: 14332–14337.
10. Finn RD, Mistry J, Tate J, Coghill P, Heger A, et al. (2010) The Pfam protein families database. *Nucleic Acids Res* 38: D211–222.
11. Gupta RS, Singh B (1994) Phylogenetic analysis of 70 kD heat shock protein sequences suggests a chimeric origin for the eukaryotic cell nucleus. *Curr Biol* 4: 1104–1114.
12. Gribaldo S, Lumia V, Creti R, de Macario EC, Sanangelantoni A, et al. (1999) Discontinuous occurrence of the *hsp70* (*dnaK*) gene among Archaea and sequence features of HSP70 suggest a novel outlook on phylogenies inferred from this protein. *J Bacteriol* 181: 434–443.
13. Gogarten JP (1994) Which is the most conserved group of proteins? Homology-orthology, paralogy, xenology, and the fusion of independent lineages. *J Mol Evol* 39: 541–543.
14. Valas RE, Bourne PE (2009) Structural analysis of polarizing indels: an emerging consensus on the root of the tree of life. *Biol Direct* 4: 30.
15. Xiong J, Fischer WM, Inoue K, Nakahara M, Bauer CE (2000) Molecular evidence for the early evolution of photosynthesis. *Science* 289: 1724–1730.
16. Raymond J, Zhaxybayeva O, Gogarten JP, Gerdes SY, Blankenship RE (2002) Whole-genome analysis of photosynthetic prokaryotes. *Science* 298: 1616–1620.
17. Sharon I, Alperovitch A, Rohwer F, Haynes M, Glaser F, et al. (2009) Photosystem I gene cassettes are present in marine virus genomes. *Nature* 461: 258–262.
18. Igarashi N, Harada J, Nagashima S, Matsuura K, Shimada K, et al. (2001) Horizontal transfer of the photosynthesis gene cluster and operon rearrangement in purple bacteria. *J Mol Evol* 52: 333–341.