CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Nature Precedings

A Study on the Origin of Peroxisomes: Possibility of Actinobacteria Symbiosis

Duhita Narendra[1], Saruhashi Satoshi[1], Hamada Kazuo[2], Miyata Daisuke[3], Horiike Tokumasa [4], Shinozawa Takao[1]

[1]Department of Integrative Bioscience and Biomedical Engineering, Graduate School of Science and Engineering, Waseda University, Waseda Research Park, Honjo, Saitama, 367-0035, Japan
[2]Radiance Ware Co., Waseda Research Park, Honjo Saitama, 367-0035, Japan
[3]Faculty of Commerce and Economics, Chiba University of Commerce, Chiba, 272-8512, Japan
[4]Department of Integrated Genetics, National Institute of Genetics Center, Mishima, Shizuoka, 411-8540, Japan

The origin of peroxisomes as having developed from the endoplasmic reticulum (ER) was proposed on the basis of the similarity between some peroxisomal proteins and ER proteins, and the localization of some peroxisomal proteins on the ER. To study the evolutionary distance between peroxisomes and ER and Prokaryotes, we carried out a phylogenetic analysis of CDC48 (cell division control 48) and its homologs, including ER-localized CDC48, CDC48 homologs in Prokaryotes and peroxisome-localized PEX1 and PEX6. A similarity search analysis of peroxisomal protein sequences to prokaryotic protein sequences using BLAST at several thresholds (E-values) was also done. We propose Actinobacteria symbiosis for the origin of peroxisomes based on the following evidence: (1) PEX1 and PEX6 are close in distance to CDC48 homologs in Actinobacteria, and these distances are closer than to ER-localized CDC48. (2) Actinobacteria proteins show the highest degree of similarity to peroxisomal proteins compared with other prokaryotes.

The ability of peroxisomes to self-divide[1] and the fact that they possess their own protein import machinery for translated proteins[2] suggest the peroxisome endosymbiosis theory. However, recently, Pex3 (an integral peroxisomal membrane protein) was also observed in the ER and to be transferred to peroxisomes in a PEX19-dependent manner[3]. Others have shown that some essential peroxisomal proteins, including PEX1 and PEX6, are similar to proteins localized in the ER[4][5]. Thus, peroxisomes have been considered to have developed from the ER.

Other discoveries, however, are contradict these views. Proteins localized in mitochondria, an an organelle of alpha proteobacterial origin, have also been observed in the ER[6]. In *Euglena*, some proteins are transported from the ER to the Golgi apparatus prior to import across the three chloroplast membranes[7]. These observations reveal that in organelles originating from endosymbiosis, such as mitochondria and chloroplasts, the localization of post-translated proteins in other organelles prior to entering the target organelle is a common phenomenon. Others have also shown that peroxisomal membrane proteins are properly targeted to peroxisomes in the absence of COPI and COPII, inhibitors of vesicle transport in the early secretory pathway[8].

This study took two approaches to obtain information about peroxisomal evolution. The first approach was a distance comparison of ER-localized CDC48, CDC48 homologs in Prokaryotes and peroxisome-localized PEX1 and PEX6 by phylogenetic tree analysis. In this analysis, the distance of chloroplast-localized Ftsh ATPase to homologues of Cyanobacteria was used as a positive control. The second approach was a similarity search analysis of the whole open reading frames (ORFs) of peroxisomal proteins and their counterparts in each prokaryote. Using this analysis, *Horiike* et. al. proposed symbiosis of an Archaea as the origin of the eukaryotic nucleus[9]. These results show good correlations with the phylogenetic analysis of most ribosomal proteins showing the sisterhood of Archaea and eukaryotes[10].

Our phylogenetic trees were built using parsimony and the Bayesian analysis. In the parsimony analysis, the best tree was obtained using a sequence random addition and Deltran as the character optimization criterion with an approximately unbiased (AU) test value of 0.881 and a bootstrap probability (BP) value of 0.873. In the Bayesian analysis, the highest AU test value (0.824) (BP = 0.584) was obtained using the Hasegawa-Kishino-Yano substitution model with gamma rate (Table 1).

Bayesian and Parsimony phylogenetic trees are very similar in some points (Fig.1). Both Bayesian and Parsimony phylogenetic trees show the distances from peroxisome-localized PEX1 and PEX6 to CDC48 homologs in Actinobacteria are closer than to ER-localized CDC48. ER-localized CDC48 itself has Archea as its closest neighbor, the same as resulting the findings of *Horiike* et. al. [(9)]. Chloroplast-localized Ftsh proteins, as positive controls, also show a close distance to Cyanobacteria as compared to other prokaryotes. In the Bayesian tree, high probability (≥99%) values were obtained for the branching of species inside each group that is, in the Proteobacteria family (with the exception of epsilon-Proteobacteria), between Bacillales and Lactobacillales, between chloroplasts and Cyanobacteria, and among ER, Archea, nucleus, Actinobacteria, PEX1 and PEX6. These branchings were also conserved in the parsimony tree.

The close distance with a high probability score between chloroplasts and Cyanobacteria in our Bayesian tree shows that endosymbiotic evidence can be obtained using CDC48 homologs. In our Bayesian and Parsimony trees, all trees show closer relationships between Actinobacterial homologs and PEX1 and PEX6 than between ER-localized CDC48 with high probability scores. Thus, PEX1 and PEX6 are considered to be phylogenetically closer to Actinobacteria homologs than ER-localized CDC48. This finding calls into question the possibility that PEX1 and PEX6 arose from ER-localized CDC48.

Table 1

Comparison among different phylogenetic reconstructions.

Bayesian analysis

Substitution model	Rate	Rank	AU	BP
Felsenstein 1981	Equal	19	3E-68	0
	Propinv	17	0.002	0.0003
	Gamma	18	2E-42	0
	Invgamma	7	0.004	0.0001
General-time-reversible	Equal	11	0.031	0.012
	Propinv	24	3E-44	0
	Gamma	10	3E-12	0
	Invgamma	9	0.0003	0
Hasegawa-Kishino-Yano	Equal	21	0.003	0
	Propinv	15	0.001	0
	Gamma	1	0.824	0.594

	Invgamma	23	0.0004	0
Jukes-Cantor 1969	Equal	14	0.001	0
	Propinv	22	3E-05	0
	Gamma	3	0.163	0.051
	Invgamma	8	8E-05	0
Kimura 2-parameter	Equal	6	0.026	0.003
	Propinv	13	0.0004	3E-05
	Gamma	5	0.049	0.004
	Invgamma	2	0.418	0.325
Symmetric	Equal	20	3E-48	0
	Propinv	16	0.002	0
	Gamma	12	4E-34	0
	Invgamma	4	0.027	0.01

Parsimony analysis

Character	Sequence addition	Rank	AU	BP
Acctran	Simple	1	0.881	0.873
	Random	2	0.119	0.127
Deltran	Simple	4	0.119	0.127
	Random	3	0.119	0.127

Abbreviations: AU: approximately unbiased test; BP: bootstrap probability for the AU test.

Figure 2 (A) shows the average number of genes in the peroxisome database that are orthologous to each prokaryotic group at several thresholds (-log E = 5-200). Actinobacteria show the highest average hit values in the range of –log E-values at all ranges. Cyanobacteria also show the highest average numbers of genes orthologous to the chloroplast database at all ranges (Fig 2 (B)).

To determine statistically, the similarity of peroxisomes to Actinobacteria and that of chloroplasts to Cyanobacteria, we carried out the chi-square test. As a result, the similarity of the chloroplast proteins in our database to the Cyanobacteria ORFs was clearly distinctive ($p < 0.05$ at all thresholds) . On the other hand, despite the high average hit number of Actinobacteria groups, the similarity of peroxisomes to Actinobacteria at a 13% significance level at the lowest (Fig 2).

Nevertheless, considering the closer relationship of PEX1 and PEX6 to Actinobacteria homologs than to ER-localized CDC48 in the phylogenetic trees (Fig 1), and also the finding that the average hit value for peroxisomes to

Actinobacteria were the highest among prokaryotes, there is a high possibility that peroxisomes originated from Actinobacteria.

Moreover, peroxisomes are surrounded by a single membrane, as is also true of Actinobacteria. The Actinobacterial membrane shows various biochemical similarities to peroxisomes such as the presence of sterols[11] and phosphatidylinositol lipids[12], suggesting them as possible ancestors for the archaeans and eukaryotes. This shows a significant relationship with the contribution of ER to peroxisomal formation. Similarities between Actinobacteria and the ER membrane make it possible for membrane substitutions to occur and there is a possibility that post-translated peroxisomal membrane proteins such as PEX3 mislocalized and docked to the ER.

We considered that the low hit value between peroxisomes and actinobacteria might be due to the loss of some proteins. Peroxisomes may transfer their genes to the nucleus, and some genes may be retargeting proteins from mitochondria. Before the symbiosis of peroxisomes, mithochondria first entered the host cell and transferred some of their genes to the nucleus[5]. Thus, when peroxisomes transferred their genes to the nucleus, genes with functions similar to those of mitochondrial genes were lost, and as a substitute, mitochondrial proteins were retargeted to the peroxisomes. One of the proteins required for the beta-oxidation function (Fox2p) in peroxisomes shows alpha-proteobacterial descent[5], indicating that the presence of long-chain fatty acid beta-oxidation in peroxisomes followed the endosymbiosis of mitochondria. The recruitment of proteins of endosymbiotic origin to peroxisomes is not an exceptional event. Nine proteins in the glycosomes of the kinetoplastida *T. brucei* and *Leishmania mexicana* are derived from chloroplasts, and can be traced back to Cyanobacteria [13]. And since mitochondria and peroxisomes show close similarities in function, large numbers of peroxisomal proteins can be considered to have been replaced by mitochondria proteins.

The Actinobacterium used in this study belongs to the Mycobacterium genus and its phenotypically close Nocardia genus. Mycobacteria have been suggested to have an endosymbiotic relationship with the human host, and Adreno-leukodystrophy (ALD), a peroxisome biogenesis disorder, has been proposed to contribute to host circumstances that are favourable for the endosymbiosis of mycobacteria[14]. Furthermore, one of the *Mycobacterium Tuberculosis* proteins that is considered to contribute to cholesterol catabolism, essential for the survival of macrophages[15], shares intriguing sequence

similarity with the eukaryotic multifunctional 17β-hydroxysteroid dehydrogenase IV (17βHSD4) involved in peroxisome-related disorders [16].

It has been observed that Mycobacterial proteins also show high similarity to those of eukaryotes. The first identified histone-like protein of *Mycobacterium Tuberculosis* has been demonstrated to possess unique dual domains showing homology to both bacterial histone-like proteins as well as eukaryotic histone H1[17]. Mycobacterial cyclase, the closest progenitor of the mammalian adenylyl cyclase family to date, is also considered to have been spread in eukaryotes by horizontal gene transfer[18]. These observations are consistent with our findings, suggesting that evolutionary peroxisomes are organelles developed from Actinobacteria.

Methods

For these analyses we constructed databases comprising 183 peroxisomal proteins, 690 chloroplast-localized proteins, 52 prokaryotic genomes, and 4 eukaryote-localized CDC48 proteins *(See supplemental material)*.

Phylogenetic trees

Protein sequences of CDC48 homologs in each prokaryote were obtained from the above organism database using the BLAST program with the sequences of known *Saccaromyces cerevicae* ER-localized CDC48 proteins as queries. The same method was used to detect CDC48 homologs in our peroxisome and chloroplast database. From each category, only proteins with the highest similarity were used. We detected 4 peroxisome-localized homologs (PEX1 and PEX6), 3 chloroplast-localized homologs, and 1 CDC48 homolog in each prokaryote.

All protein sequences of ER-localized CDC48, nucleus-localized CDC48, peroxisome-localized homologs (PEX1 and PEX6), chloroplast-localized homologs (Ftsh Protease) and CDC48 homologs in prokaryotes were aligned in Clustal X 1.83 using default parameters .

The statistical significance in the competing phylogenetic hypotheses was assessed under a likelihood model with the approximately unbiased (AU) test[19].

In order to perform the AU test, a set of alternative tree hypotheses must be available; different phylogenetic hypotheses were obtained for parsimony and Bayesian analysis (see below). For each phylogenetic tree, site-wise log-likelihood values were obtained with ProtML 2.3b3, and the AU test was implemented in CONSEL 0.1f.

Parsimony analysis

Parsimony analyses of CDC48 and its homologs were performed in PAUP* 4.0b10; four different analyses were performed with a bootstrap with 1000 pseudoreplicates: (A) heuristic search with 10 random addition sequences, TBR with 100 random additions, Acctran as a character optimization criterion, gaps treated as a fifth state and zero length branches collapsed, (B) heuristic search with 10 random addition sequences, TBR with 100 random additions, Deltran as a character optimization criterion, gaps treated as a fifth state and zero length branches collapsed, (C) heuristic search with simple addition sequence, TBR with 100 random additions, Acctran as a character optimization criterion, gaps treated as a fifth state and zero length branches collapsed, and (D) heuristic search with simple addition sequence, TBR with 100 random additions, Deltran as a character optimization criterion, gaps treated as a fifth state and zero length branches collapsed. For the most parsimonious trees obtained in each analysis, a 50% majority rule consensus was conducted.

Bayesian analysis

Bayesian analyses for CDC48 and its homologs were performed in MrBayes 3.1 with the number of generations set at 250,000, and a sample frequency of 1000 as parameters. Six models were tested: the Felsenstein 1981 model, the General-time-reversible model, the Hasegawa-Kishino-Yano model, the Jukes-Cantor 1969 model, the Kimura 2-parameter, and the Symmetric model. Four "rate" parameters were used (invgamma, propinv, gamma, and equal) for each model.

The first 40,000 generations were pruned after reaching stationarity, and a 50% majority rule consensus was constructed using the remaining trees (210).

Similarity search analysis

Proteins for which there is experimental evidence for their localization in

peroxisomes and chloroplasts, and ORFs from the complete genomes of the 52 prokaryotic organisms were used.

Using BLAST, prokaryotic ORFs (in each organism) with the higest degree of similarity with peroxisomal proteins were detected. Peroxisomal proteins with the highest degree of similarity to prokaryotic ORFs (in each organism) were also detected. Finally, ORF pairs showing the highest degree of similarity between them were included. This operation was done for the purpose of detecting orthologous gene pairs. In this case, the avoidance of the effect of gene duplication after the deviation of peroxisomal proteins from each bacterium was also carried out[20]. Finally, gene pairs were scored as hit numbers. The threshold (E-value) was set at intervals of 5 in the range of 5–200 as –log E. Hit numbers for ORFs at each E-value (5–200 as –log E) were calculated for each organism. Prokaryotic ORFs (in each organism) with the highest degree of similarity with chloroplast proteins were also detected as positive controls.

The Chi-square test for statistical analysis was carried out for prokaryotic groups showing the highest hit numbers in most ranges. We calculated the chi-square statistic as follows:

$$\chi^2 = \frac{\overline{O} - \sum_{i=1}^{i=n} \overline{E_i} / n}{\sum_{i=1}^{i=n} \overline{E_i} / n}$$

Here, $\overline{O}$, $\overline{E_i}$ and n are the mean hit value of prokaryotic groups showing the highest hit numbers in most ranges, the mean hit value of each prokaryotic group, and the number of prokaryotic groups used in this analysis, respectively. To determine the significance level, fifteen $(=n-1)$ degrees of freedom were used. When $p < 0.05$, it was judged that the hit number of that group at that E-value was larger than those of other groups at the 5% significance level.

Acknowledgements

We wish to thank D. Watabe (Saitama Institute of Technology, Japan) for valuable suggestions and assistance with the analyses.

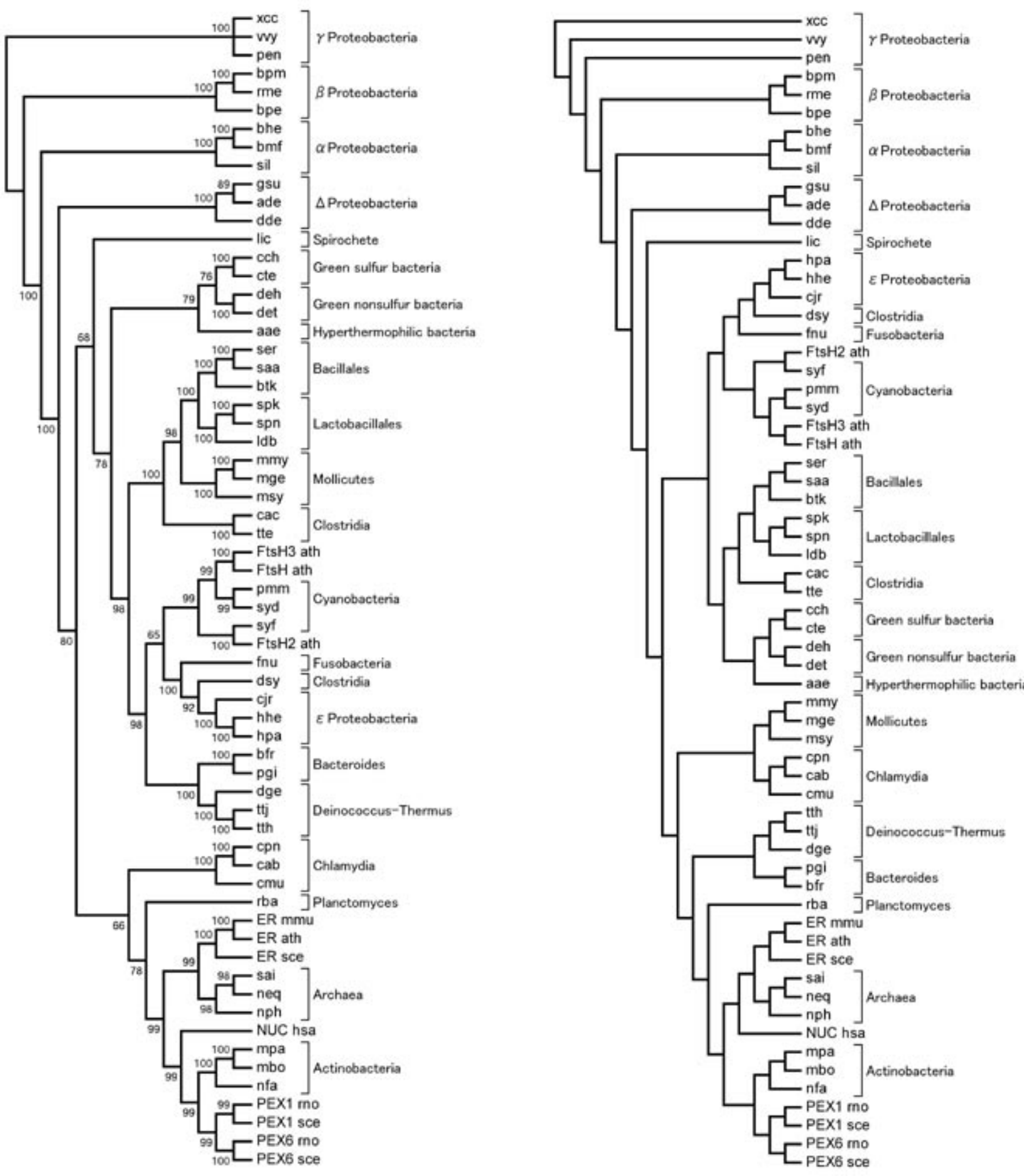
References

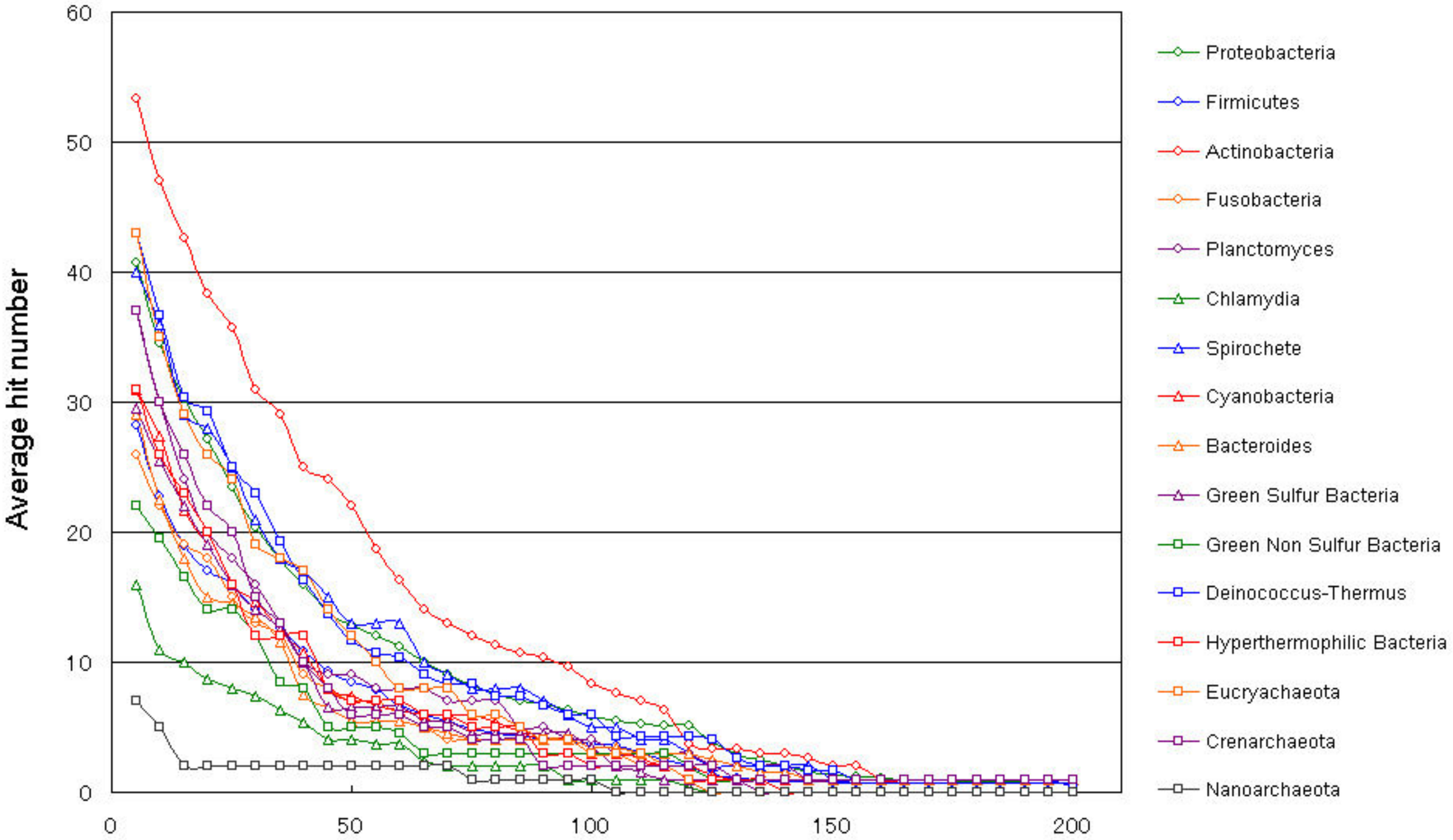
1. P. B. Lazarow, Y. Fujiki. Biogenesis of peroxisomes. Annu Rev Cell Biol 1985, 1:489-530.
2. S. Subramani. Protein Import into Peroxisomes and Biogenesis of the Organelle. Annu Rev Cell Biol 1993, 9:445-478.
3. D. Hoepfner, D. Schildknegt, I. Braakman, P. Philippsen, H.F. Tabak. Contribution of the Endoplasmic Reticulum to peroxisome formation. Cell 2005, 122:85-95.
4. A. Schluter, S. Fourcade, R. Ripp, J.L. Mandel, O. Poch, A. Pujol. The Evolutionary Origin of Peroxisomes: An ER-Peroxisome Connection. Mol. Biol. Evol. 2006, 23:838–845.
5. T. Gabaldón, B. Snel , F. van Zimmeren, W. Hemrika , H. Tabak , M.A. Huynen. Origin and evolution of the peroxisomal proteome. Biol. Direct 2006 1: 8.
6. F. Sun, S. Wei, C. Li, Y. Chang, C. Chao, Y. Lai. Localization of GRP78 to mitochondria under the unfolded protein response. Biochem. J. 2006, 396:31–39.
7. S.D. Schwartzbach, T. Osafune, W. Löffelhardt, Protein import into cyanelles and complex chloroplasts. Plant Mol. Biol. 1998, 38:247-263..
8. T. Voorn-Brouwer, A. Kragt, H.F. Tabak, B. Distel. Peroxisomal membrane proteins are properly targeted to peroxisomes in the absence of COPI- and COPII mediated vesicular transport. J. Cell Sci. 2001,114:2199-2204.
9. T. Horiike, K. Hamada, S. Kanaya, T. Shinozawa 2001 Origin of eukaryotic cell nuclei by symbiosis of Archaea in Bacteria is revealed by homology-hit analysis Nature Cell Bio. 3:210-214.
10. J.R. Brown, W.F. Doolittle. Archaea and the Prokaryote-to-Eukaryote Transition. Microbiol. Mol. Biol. Rev. 1997, 61:456−502.
11. D.C. Lamba, D.E. Kellya, N.J. Manningb, S.L. Kellya. A sterol biosynthetic pathway in Mycobacterium. FEBS Letters 1998,437:142-144.
12. J. Kordulakova, M. Gilleron, K. Mikusova, G. Puzo, P.J. Brennan, B. Gicquel, M. Jackson. Definition of the First Mannosylation Step in Phosphatidylinositol Mannoside Synthesis. J. Biol. Chem.1992, 277:31335-31344.
13. V. Hannaert, E. Saavedra, F. Duffieux, J.P. Szikora, D.J. Rigden, P.A. Michels, Opperdoes FR: Plant-like traits associated with metabolism of Trypanosoma parasites. *Proc Natl Acad Sci U S A* 2003, 100:1067-1071.
14. G.S. Gericke. Is there an emerging endosymbiotic relationship between mycobacteria and the human host based on horizontal transfer of genetic sequences?. Medical Hypotheses 2006,67:1419–1428.
15. R.V. Geize, K. Yam, T. Heuser, M.H. Wilbrink, H. Hara,M.C. Anderton, E. Sim, L. Dijkhuizen, J.E. Davies, W.W. Mohn, L.D. Eltis. A gene cluster encoding cholesterol

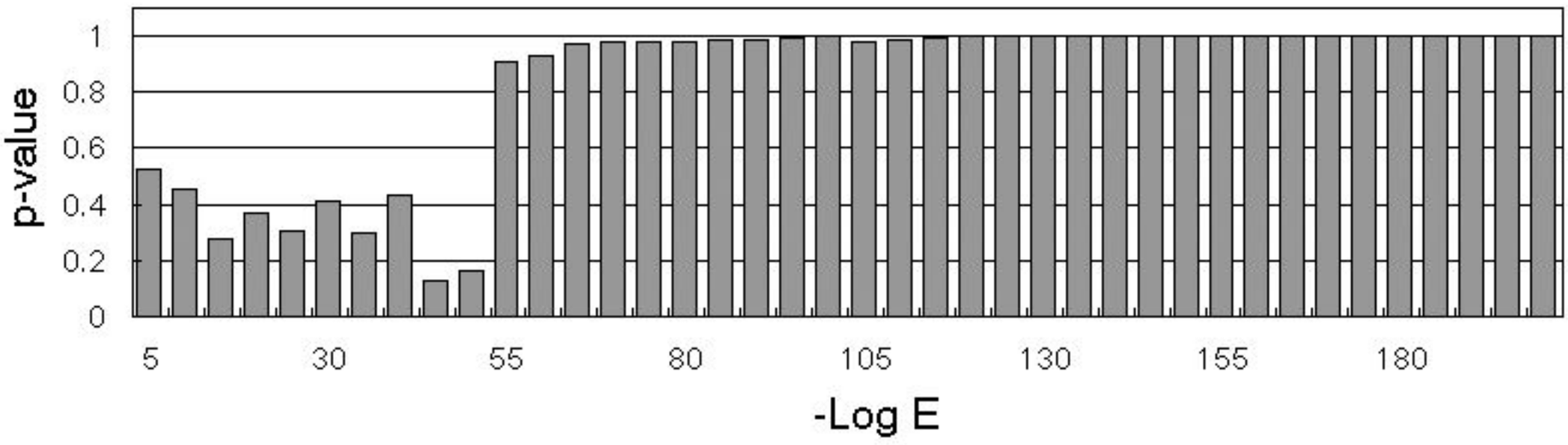
catabolism in a soil actinomycete provides insight into Mycobacterium tuberculosis survival in macrophages. PNAS 2007,104(6):147-152.

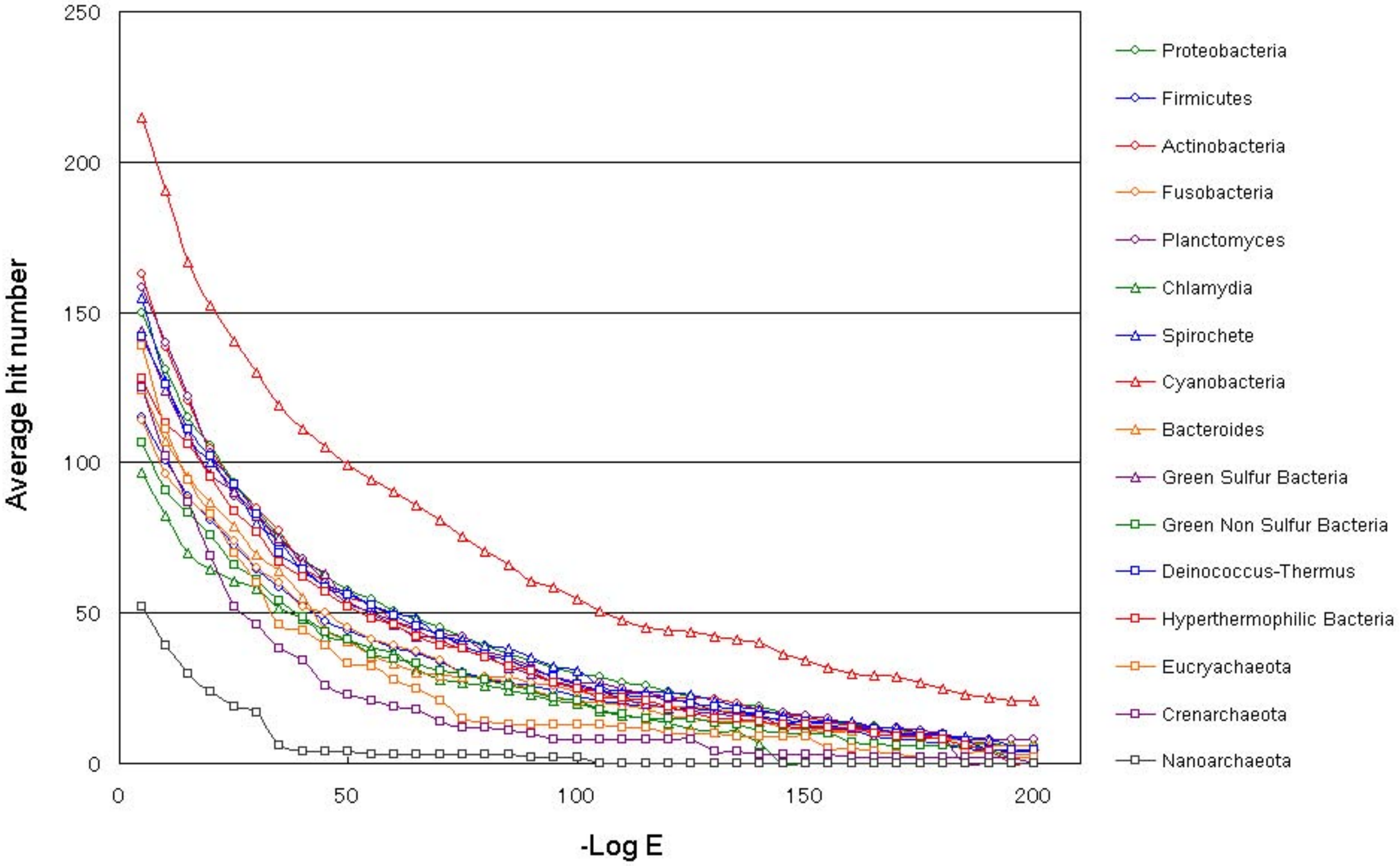
16. R. Mindnich, G. Moller, J. Adamski. The role of 17 beta-hydroxysteroid dehydrogenases. Mol Cell Endocrinol 2004, 218:7–20.
17. S. Prabhakar, P.S. Annapurna, N.K. Jain, A.B. Dey, J.S. Tyagi, H.K. Prasad. Identification of an immunogenic histone-like protein (HLPMt) of Mycobacterium tuberculosis. Tuber Lung Dis 1998, 79(1):43–53.
18. Y.L. Guo, T. Seebacher, U. Kurz, J.U. Linder, J.E. Schultz. Adenylyl cyclase Rv1625c of Mycobacterium tuberculosis: a progenitor of mammalian adenylyl cyclases. Embo J 2001,20(14):3667–3675.
19. H. Shimodaira. An approximately unbiased test of phylogenetic test selection. Syst. Biol. 2002, 51:492–505.
20. A. Zharkikh. Estimation of evolutionary distances between nucleotide sequences. J. Mol. Evol. 1994, 39:315-329.
21. Y. Fukao, M. Hayashi, M. Nishimura. Proteomic analysis of leaf peroxisomes in greening cotyledons of Arabidopsis thaliana. Plant Cell Physiol 2002, 43:689-696.
22. M. Kikuchi, N. Hatano, S. Yokota, N. Shimozawa, T. Imanaka, H. Taniguchi. Proteomic analysis of rat liver peroxisome: presence of peroxisome-specific isozyme of Lon protease. J. Biol. Chem. 2004, 279:421-428.
23. H. Schäfer, K. Nau, A. Sickmann, R. Erdmann, H.E. Meyer. Identification of peroxisomal membrane proteins of Saccharomyces cerevisiae by mass spectrometry. Electrophoresis 2001, 22:2955–2968.
24. M. Marelli, J.J. Smith , S. Jung, E. Yi, A.I. Nesvizhskii , R.H. Christmas, R.A. Saleem, Y.Y Tam, A. Fagarasanu, D.R. Goodlett, R. Aebersold, R.A. Rachubinski, J.D. Aitchison. Quantitative mass spectrometry reveals a role for the GTPase Rho1p in actin organization on the peroxisome membrane. J Cell Biol. 2004, 167:1099-1112.
25. E.C Yi, M. Marelli, H. Lee, S.O. Purvine, R. Aebersold, J.D. Aitchison, D.R. Goodlett. Approaching complete peroxisome characterization by gas-phase fractionation. Electrophoresis 2002, 23:3205-3216.
26. T. Kleffmann, D. Russenberger, A. von Zychlinski, W. Christopher, K. Sjölander, W. Gruissem, S. Baginsky. The Arabidopsis thaliana chloroplast proteome reveals pathway. abundance and novel protein functions. Curr Biol. 2004, 14:354-62.
27. K.U. Frohlich, H.W. Fries, M. Riidiger, R. Erdmann, D. Botstein, D. Mecke. Yeast cell cycle protein CDC48p shows full-length homology to the mammalian protein VCP and is a member of a protein family involved in secretion, peroxisome formation, and gene expression. *J. Cell Biol.* 1991, 114: 443-453.

28. M. Egerton, O.R. Ashe, D. Chen, B.J. Druker, W.H. Burgess, L.E. Samelson. VCP, the mammalian homolog of cdc48, is tyrosine phosphorylated in response to T cell antigen receptor activation. *Embo J.* 1992, 11(10):3533-3540.
29. H.S. Feiler, T. Desprez, V. Santoni, J. Kronenberger, M. Caboche, J. Traas. The higher plant Arabidopsis thaliana encodes a functional CDC48 homologue which is highly expressed in dividing and expanding cells. *Embo J.* 1995, 14(22): 5626-5637.
30. M. Nagahama, T. Yamazoe, Y. Hara, K. Tani, A. Tsuji, M. Tagaya. The AAA-ATPase NVL2 is a component of pre-ribosomal particles that interacts with the DExD/H-box RNA helicase DOB1. *Biochem. Biophys. Res. Commun.* 2006, 346(3):1075-1082.


xcc
vvy
pen
γ Proteobacteria
bpm
rme
bpe
β Proteobacteria
bhe
bmf
sil
α Proteobacteria
gsu
ade
dde
Δ Proteobacteria
lic
Spirochete
cch
cte
Green sulfur bacteria
deh
det
Green nonsulfur bacteria
aae
Hyperthermophilic bacteria
ser
saa
btk
Bacillales
spk
spn
ldb
Lactobacillales
mmy
mge
msy
Mollicutes
cac
tte
Clostridia
FtsH3 ath
FtsH ath
pmm
syd
syf
FtsH2 ath
Cyanobacteria
fnu
Fusobacteria
dsy
Clostridia
cjr
hhe
hpa
ε Proteobacteria
bfr
pgi
Bacteroides
dge
ttj
tth
Deinococcus–Thermus
cpn
cab
cmu
Chlamydia
rba
Planctomyces
ER mmu
ER ath
ER sce
sai
neq
nph
Archaea
NUC hsa
mpa
mbo
nfa
Actinobacteria
PEX1 mo
PEX1 sce
PEX6 mo
PEX6 sce


Average hit number
60
50
40
30
20
10
0
0
50
100
150
200
-Log E
Proteobacteria
Firmicutes
Actinobacteria
Fusobacteria
Planctomyces
Chlamydia
Spirochete
Cyanobacteria
Bacteroides
Green Sulfur Bacteria
Green Non Sulfur Bacteria
Deinococcus-Thermus
Hyperthermophilic Bacteria
Eucryachaeota
Crenarchaeota
Nanoarchaeota


p-value
1
0.8
0.6
0.4
0.2
0
5
30
55
80
105
130
155
180
-Log E


Average hit number
250
200
150
100
50
0
0
50
100
150
200
-Log E
Proteobacteria
Firmicutes
Actinobacteria
Fusobacteria
Planctomyces
Chlamydia
Spirochete
Cyanobacteria
Bacteroides
Green Sulfur Bacteria
Green Non Sulfur Bacteria
Deinococcus-Thermus
Hyperthermophilic Bacteria
Eucryachaeota
Crenarchaeota
Nanoarchaeota

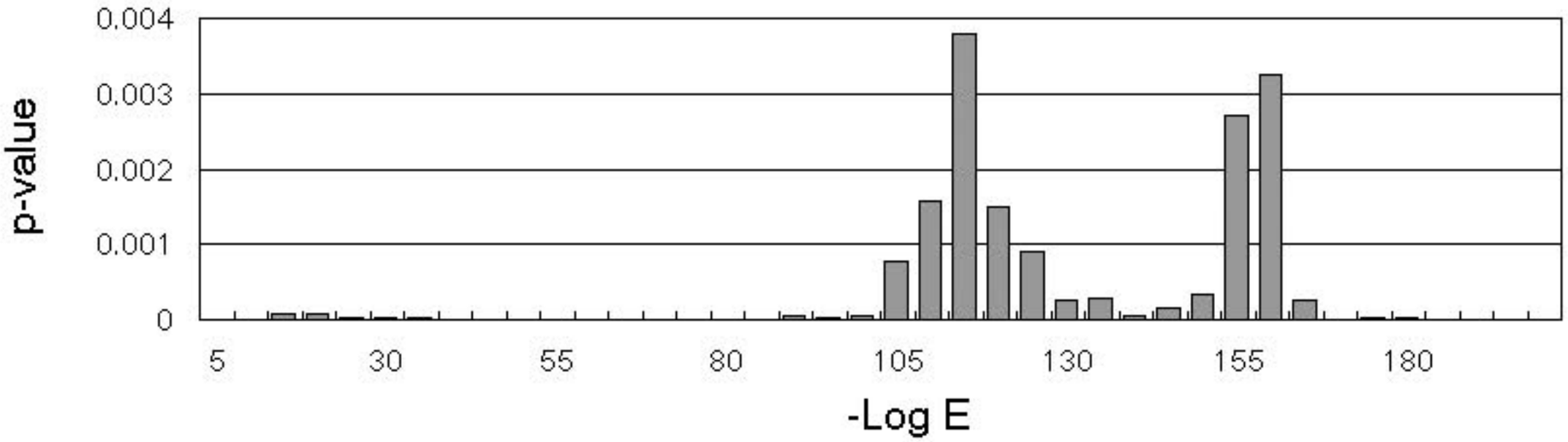

p-value
0.004
0.003
0.002
0.001
0
5
30
55
80
105
130
155
180
-Log E

Fig 1. Bayesian and parsimony phylogenetic trees. A 50% percent majority rule consensus of (Left) the Bayesian phylogenetic analysis when Hasegawa-Kishino-Yano substitution model and gamma rate was used. (Right) the Parsimony phylogenetic analysis when sequence addition was simple and Acctran was used as character optimization strategy. *Abbreviations*: ER: ER-localized CDC48, Nuc: Nucleus-localized CDC48 (*See supplemental material for organisms three letters abbreviation*).

Fig 2. Average hit number of databased peroxisome (A) and chloroplast (B) proteins to various prokaryotes ORFs and p-values of Actinobacteria (A) and Cyanobacteria (B) at each threshold. E-values as –log E scale are shown on the horizontal axes.