Update TRENDS in Microbiology Vol.15 No.10

Research Focus

Supertrees and symbiosis in eukaryote genome evolution

Christian Esser and William Martin

Institute of Botany III, University of Düsseldorf, D-40225 Düsseldorf, Germany

If we took all of the single copy genes in all sequenced genomes, made phylogenetic trees from them individually, and then made the supertree of those trees, what would we get? Recently, David Pisani and colleagues did that experiment and their results are likely to spark much discussion. Their prokaryote tree looks very familiar, but the genome history of eukaryotes appears dominated by genes of cyanobacterial (plastid) and α-proteobacterial (mitochondrial) origin, while the host component branches within the archaebacteria.

The origin of eukaryotes; the debate

Flames of debate now rage amidst the issues of deep microbial phylogeny and the position of eukaryotes therein. One camp is saying that the tree of life is a tree [1] and that the tree looks like the tree of rRNA sequences [2]. In this scenario, the eukaryotes are a deeply-branching primordial lineage that is the sister of the archaebacteria [3] and eukaryote specific proteins and introns in eukaryote genes provide evidence in favour of that view [4]. A second camp is saying: that the tree of life is a hypothesis that we should be testing [5]; that there has been substantial transfer of genes among lineages during microbial evolution [6]; that eukaryote genomes are chimaeras containing eubacterial and archaebacterial components [7,8]; and that the endosymbiotic origin of mitochondria was possibly causal to that chimaerism, in the wake of which introns and eukaryote specific genes could have arisen [9]. Still others are saying that the prokaryote–to–eukaryote transition requires even more chimaerism and endosymbiosis than the second camp suggests [10–12].

In principle, these ideas can all be tested with molecular data from genomes. But there are some hefty caveats, among them the uncertainty associated with making trees [13] and alignments [14] from deeply diverged sequences. But even if those problems were solved, very few genes are shared by all genomes, especially among prokaryotes. Hence the strategy of concatenating hundreds of sequences into a grandmaster alignment, as is common practice for the study of eukaryote phylogeny [15], is not an option if many genomes from all walks of microbial life are involved. This is where supertrees, which Pisani *et al.* [16] used to explore the issue of deep phylogeny, open new avenues of pursuit.

Enter supertrees

Supertree methods make trees from trees. They take phylogenetic trees as their input. Those input trees stem, in turn, from alignments of different genes. As their output, supertree methods summarise the various input trees as a single tree. Consensus methods do much the same thing, but in consensus methods the trees must all contain exactly the same taxa. In supertree methods, the taxa sets of different trees need not be the same, they just need to overlap [17–19]. Given some overlap among taxon sets, supertree methods puzzle all of the input trees together so as to display the branches that are compatible among the input trees. In other words, supertrees combine many smaller, overlapping phylogenetic trees into a single, more comprehensive tree [18]. Supertree approaches are still comparatively young, and they hold promise for datasets harbouring incompletely or sparsely distributed patterns of gene presence and absence, as with real microbial genome data.

Pisani *et al.* [16] made phylogenetic trees for 5741 genes that are present as single copy genes among 165 sequenced genomes, including more than a dozen eukaryotes. The use of single copy genes only was designed to help avoid problems associated with comparing possible paralogs within large gene families. They then pruned this dataset by looking only at the most conserved regions of the alignments, then only at the alignments that had clearly nonrandom phylogenetic signals, and then only at those in which the amino acid composition of the sequences was not significantly skewed. The supertree of prokaryotic trees that passed those hurdles harboured no major surprises: almost all of the major groups that modern prokaryotic systematics recognises were recovered.

Surprise!

The surprise was the position of the eukaryotes, which came to branch with, of all things, the cyanobacteria. This surprise, however, is hardly nonsense because it can be understood with the help of endosymbiotic theory: the plastids of plants descend from cyanobacteria and plants have acquired many of their nuclear genes from the cyanobacterial ancestor of plastids [16]. Because supertree methods greedily recover the strongest phylogenetic signals in the data, and because there was no strong signal that placed the eukaryotes anywhere else in the supertree, the cyanobacterial signal in the plants predominated. That signal was only one of several conflicting signals in the data, and in the absence of stronger signals speaking to the contrary, it was strong enough to place the eukaryotes within the cyanobacterial group.

Corresponding author: Martin, W. (w.martin@uni-duesseldorf.de).
Available online 19 September 2007.

To remove the cyanobacterial signal, Pisani *et al.* [16] simply removed the trees that supported that grouping. When they did that, the eukaryotes then came to branch within the α-proteobacteria. That result would be altogether inexplicable were it not for endosymbiotic theory, some versions of which predict a strong genetic contribution from the genome of the α-proteobacterial ancestor of mitochondria to the complement of eukaryote nuclear genes. The α-proteobacterial signal too was only one of several conflicting signals in the data, and in the absence of stronger signals speaking to the contrary, it was again strong enough to place the eukaryotes among the α-proteobacteria as a group. That result will probably come as a surprise to many, but for those who view eukaryotes from the perspective of endosymbiosis, it is readily interpretable: mitochondria were quantitatively important in shaping the contours of eukaryotic genomes.

Of course, the α-proteobacterial signal was just the strongest of several competing eubacterial signals in the decyanobacterialised eukaryotic data. At face value, the evidence from the individual trees might suggest the existence of gene contributions to the eukaryote common ancestor from all major groups of eubacteria sampled [16], including spirochaetes and δ-proteobacteria, as some current symbiotic views would predict [10,11]. But no current view predicts both signals, and that puts a bit of a strain on theories making either prediction. So there is ample room for more debate on such issues. There also looms the issue that prokaryotes have been exchanging their genes over time [5–7], so that it is hard to say exactly what collection of genes the ancestor of mitochondria possessed at the time that it became an endosymbiont [20].

When the eubacterial genes were stripped out of the supertree data, then the eukaryotes came to rest with the Thermoplasmatales, a group of archaebacteria belonging to the euryarchaeotes. Dennis Searcy [21] has written about *Thermoplasma* as a candidate lineage for the host that acquired mitochondria for a couple of decades. The *Thermoplasma* genome sequence gave no direct hints to support that view [22], but also did not consider supertree analyses, so the new result is of interest [16] because it suggests that the host lineage of eukaryotes arose from within the archaebacteria, rather than as a sister to them. That is a topic of current debate [23], as is the issue of how much the supertree data favour the grouping of eukaryotes with the *Thermoplasma* lineage, because other recent whole genome analyses have found evidence to suggest that eukaryotes group with the crenarchaeotes (a grouping known as the eocyte tree [7]). Other analyses have found whole genome signals that link eukaryotes to members of the euryarchaeotes, but not to the *Thermoplasma* lineage [12].

Geologists are also telling us something important

Finally, Pisani *et al.* [16] discuss their findings in light of models for the origin of mitochondria that involve anaerobic or sulfur-based metabolic symbioses at mitochondrial origin. Is there any connection between supertree inferences and ancient environments? It is a possibility. While biologists have been debating the shape of microbial evolution, geologists have been developing a fundamentally new model of ocean geochemistry [24–27]. Briefly, it suggests that ~2.3 billion years ago, the O_2 that was accumulating in the atmosphere started to oxidise and erode continental sulfide deposits, thus large amounts of sulfate were carried into the oceans, so that the substrate required for sulfate-reducing prokaryotes was provided [27]. Marine sulfate reducers became significant on a global scale and they produced sulfide. That means that the photic zone (marine surface water) produced oxygen during that period, but below the photic zone, the oceans were anoxic and sulfidic [24,25]. Two recent reports provide evidence that indicate the end of this period occurred only ~580 million years ago [26,27]. With geologists telling us that the oceans were anoxic and sulfidic during the time 2.3–0.58 billion years ago, we should perhaps keep ancient environments in mind when considering microbial evolution and the place of eukaryotes in it. After all, if mitochondria arose more than 1.4 billion years ago [23,24], it means that the ancestor of mitochondria, if it was an average α-proteobacterium of its day, was probably well suited to anoxia and sulfide by virtue of the collection of genes in its genome.

Conclusions and future perspectives

The new report is hardly the last word in the debate on the origin of eukaryotes and mitochondria, it is merely the first word from the supertree corner, which uses hundreds of genome sequences. Supertrees constructed by Pisani *et al.* [16] placed the position of eukaryotes in the tree of life in unexpected places that do, however, make sense in light of some formulations of endosymbiotic theory. If we want to give the tree of life hypothesis [5] an honest test, we have to listen to what all genes have to say, not just the ones that tell us a familiar story.

Acknowledgements

We thank the Deutsche Forschungsgemeinschaft for financial support.

References

1 Ciccarelli, F.D. *et al.* (2006) Toward automatic reconstruction of a highly resolved tree of life. *Science* 311, 1283–1287
2 Pace, N.R. (2006) Time for a change. *Nature* 441, 289
3 Poole, A.M. and Penny, D. (2007) Response to Dagan and Martin. *Bioessays* 29, 611–614
4 Kurland, C.G. *et al.* (2006) Genomics and the irreducible nature of eukaryote cells. *Science* 312, 1011–1014
5 Doolittle, W.F. and Bapteste, E. (2007) Pattern pluralism and the Tree of Life hypothesis. *Proc. Natl. Acad. Sci. U. S. A.* 104, 2043–2049
6 Gogarten, J.P. and Townsend, J.P. (2005) Horizontal gene transfer, genome innovation and evolution. *Nat. Rev. Microbiol.* 3, 679–687
7 Rivera, M.C. and Lake, J.A. (2004) The ring of life provides evidence for a genome fusion origin of eukaryotes. *Nature* 431, 152–155
8 Lake, J.A. (2007) Disappearing act. *Nature* 446, 983
9 Martin, W. *et al.* (2007) The evolution of eukaryotes. *Science* 316, 542–543
10 Margulis, L. *et al.* (2006) The last eukaryotic common ancestor (LECA): acquisition of cytoskeletal motility from aerotolerant spirochetes in the Proterozoic Eon. *Proc. Natl. Acad. Sci. U. S. A.* 103, 13080–13085
11 Lopez-Garcia, P. and Moreira, D. (2006) Selective forces for the origin of the eukaryotic nucleus. *Bioessays* 28, 525–533
12 Horiike, T. *et al.* (2004) The origin of eukaryotes is suggested as the symbiosis of *Pyrococcus* into γ-proteobacteria by phylogenetic tree based on gene content. *J. Mol. Evol.* 59, 606–619
13 Embley, T.M. and Hirt, R.P. (1998) Early branching eukaryotes? *Curr. Opin. Genet. Dev.* 8, 655–661

14 Landan, G. and Graur, D. (2007) Heads or Tails? A simple reliability check for multiple sequence alignments. *Mol. Biol. Evol.* 24, 1380–1383
15 Bapteste, E. *et al.* (2002) The analysis of 100 genes supports the grouping of three highly divergent amoebae: *Dictyostelium, Entamoeba*, and *Mastigamoeba*. *Proc. Natl. Acad. Sci. U. S. A.* 99, 1414–1419
16 Pisani, D. *et al.* (2007) Supertrees disentangle the chimeric origin of eukaryotic genomes. *Mol. Biol. Evol.* 24, 1752–1760
17 Steel, M. *et al.* (2000) Simple but fundamental limitations on supertree and consensus tree methods. *Syst. Biol.* 49, 363–368
18 Bininda-Emonds, O.R.P. (2004) The evolution of supertrees. *Trends Ecol. Evol.* 19, 315–322
19 Wilkinson, M. *et al.* (2005) The shape of supertrees to come: treeshape related properties of fourteen supertree methods. *Syst. Biol.* 54, 419–431
20 Esser, C. *et al.* (2007) The origin of mitochondria in light of a fluid prokaryotic chromosome model. *Biol. Lett.* 3, 180–184
21 Searcy, D.G. (2006) Rapid hydrogen sulfide consumption by *Tetrahymena pyriformis* and its implications for the origin of mitochondria. *Eur. J. Protistol.* 42, 221–231
22 Cowan, D. (2000) Use your neighbor's genes. *Nature* 407, 466–467
23 Embley, T.M. and Martin, W. (2006) Eukaryote evolution: changes and challenges. *Nature* 440, 623–630
24 Anbar, A.D. and Knoll, A.H. (2002) Proterozoic ocean chemistry and evolution: a bioinorganic bridge. *Science* 297, 1137–1142
25 Poulton, S.W. *et al.* (2004) The transition to a sulphidic ocean ~1.84 billion years ago. *Nature* 431, 173–177
26 Fike, D.A. *et al.* (2006) Oxidation of the Ediacaran ocean. *Nature* 444, 744–747
27 Canfield, D.E. *et al.* (2007) Late-Neoproterozoic deep-ocean oxygenation and the rise of animal life. *Science* 315, 92–95

0966-842X/$ – see front matter
doi:10.1016/j.tim.2007.09.001

The elusive activity of the *Yersinia* protein kinase A kinase domain is revealed

Michelle A. Laskowski-Arce and Kim Orth

Department of Molecular Biology, UT Southwestern Medical Center, 6000 Harry Hines Blvd, Dallas, TX 75390-9148, USA

***Yersinia* spp. pathogens use their type III secretion system to translocate effectors that manipulate host signaling pathways during infection. Although molecular targets for five of the six known *Yersinia* effectors are known, the target for the serine/threonine kinase domain of *Yersinia* protein kinase A (YpkA) has remained elusive. Recently, Navarro *et al*. (2007) demonstrated that YpkA phosphorylates Gαq, and inhibits Gαq-mediated signaling. Inhibition by YpkA could contribute to one of the most documented symptoms of *Yersinia pestis* infection, extensive bleeding.**

Yersinia and type III secretion

Yersinia pestis was the cause of the Black Death in the Middle Ages, killing over one-third of the population of Europe [1]. Today, this pathogen is becoming an increasing concern, as it represents a potential threat as a biological weapon. Moreover, other *Yersinia* spp. (*Y. enterocolitica* and *Y. pseudotuberculosis*) are responsible for causing severe disease in humans [1]. These pathogens, like many other gram-negative bacteria, encode a type III secretion system (T3SS) and a repertoire of effectors. These secretion systems are macromolecular structures that span the bacterial inner and outer membranes, and allow for direct translocation of effectors from the bacterial cytosol into the host cell [2–4]. Translocation of these effectors function to disrupt and manipulate the host, to create an environment that is conducive to bacterial survival. Consequently, T3SS significantly contributes to virulence [1,4]. The role of each effector and the effect it has on host cells is dependent on both its catalytic activity and host cell target. In this article, we review recent studies from Navarro *et al.* [5] that identify a eukaryotic target for the serine/threonine kinase domain of the *Yersinia* protein kinase A [YpkA (YopO in *Y. enterocolitica*)].

The *Yersinia* spp. pathogens use their repertoire of T3SS effectors to actively block phagocytosis, induce apoptosis and disrupt the host immune response. These phenotypes are mediated by the biochemical activities of the translocated effectors, called Yersinia outer proteins (Yop). Two effectors, YopJ and YopM, are responsible for the ability of *Yersinia* to modulate host cell signaling. YopJ functions as an acetyltransferase and inhibits all mitogen-activated protein kinase (MAPK) pathways and the NFκB pathway. YopJ acetylates crucial serine and threonine residues on the activation loop of the MAPK kinase (MAPKK) superfamily of kinases, thereby preventing their activation, subsequent downstream signaling and promotion of apoptosis [4,6–8]. In addition, YopM contains a leucine rich repeat that scaffolds nuclear kinases that results in inappropriate activation of signaling pathways downstream of the YopJ block [9,10]. *Yersinia* also actively manipulates the actin cytoskeleton by way of four of its T3SS effectors. YopH is a tyrosine phosphatase that dephosphorylates proteins associated with focal adhesions that results in the collapse of these structures [11–13]. YopE is a GTPase activating protein (GAP) for Rho-like G-proteins that induces hydrolysis of GTP to GDP that results in their inactivation and subsequent depolymerization of actin [14]. YopT, although not expressed in all virulent *Yersinia* strains, is a cysteine protease that cleaves off the lipid modification on Rho-like G-proteins that results in their inhibition owing to mislocalization, which further contributes to the inhibition of phagocytosis [15].

Corresponding author: Orth, K. (Kim.Orth@UTSouthwestern.edu).
Available online 24 October 2007.