The tangled past of eukaryotic enzymes involved in anaerobic metabolism

Vladimir Hampl,1,* Courtney W. Stairs2 and Andrew J. Roger2
1Charles University in Prague; Faculty of Science; Department of Parasitology; Prague, Czech Republic; 2Centre for Comparative Genomics and Evolutionary Bioinformatics; Department of Biochemistry and Molecular Biology; Dalhousie University; Halifax, NS Canada

There is little doubt that genes can spread across unrelated prokaryotes, eukaryotes and even between these domains. It is expected that organisms inhabiting a common niche may exchange their genes even more often due to their physical proximity and similar demands. One such niche is anaerobic or microaerophilic environments in some sediments and intestines of animals. Indeed, enzymes advantageous for metabolism in these environments often exhibit an evolutionary history incoherent with the history of their hosts indicating potential transfers. The evolutionary paths of some very basic enzymes for energy metabolism of anaerobic eukaryotes (pyruvate formate lyase, pyruvate:ferredoxin oxidoreductase, [FeFe]hydrogenase and arginine deiminase) seems to be particularly intriguing and although their histories are not identical they share several unexpected features in common. Every enzyme mentioned above is present in groups of eukaryotes that are unrelated to each other. Although the enzyme phylogenies are not always robustly supported, they always suggest that the eukaryotic homologues form one or two clades, in which the relationships are not congruent with the eukaryotic phylogeny. Finally, these eukaryotic enzymes are never specifically related to homologues from α-proteobacteria, ancestors of mitochondria. The most plausible explanation for evolution of this pattern expects one or two inter-domain transfers to one or two eukaryotes from prokaryotes, who were not the mitochondrial endosymbiont. Once the genes were introduced into the eukaryotic domain they have spread to other eukaryotic groups exclusively via eukaryote-to-eukaryote transfers. Currently, eukaryote-to-eukaryote gene transfers have been regarded as less common than prokaryote-to-eukaryote transfers. The fact that eukaryotes accepted genes for these enzymes solely from other eukaryotes and not prokaryotes present in the same environment is surprising.

It is generally accepted that a transfer of genes between distantly related organisms plays a role in microbial evolution that cannot be ignored.1 Although this lateral or horizontal gene transfer (LGT or HGT) is probably more rampant in prokaryotic organisms, reports of this process occurring in eukaryotes are steadily mounting, for some examples see references 2–4. Generally, any DNA present in the environment can be incorporated in the genome of the new eukaryotic host. The most common sources of the exogenous genes are prokaryotes that reside as endosymbionts in the cells of eukaryotes5 or prokaryotes that eukaryotes feed on6 (i.e., Doolittle’s “you are what you eat” concept). Some genes appear to be more prone to successful lateral transfer than others. For example, the transfers of ribosomal RNAs are very rare,7,8 and generally the transfers of genes involved in translation and ribosome formation occur at much lower frequency than transfers of enzymes for carbohydrate metabolism or biosynthesis of secondary metabolites.9 The complexity hypothesis9,10 posits that genes whose products are involved in
many interactions with other proteins or molecules (e.g., informational genes) are less prone to transfers than genes with fewer interactions (e.g., metabolic enzymes). It has been shown that LGT might play an important role in the adaptation of organisms to their new ecological niche, e.g., acquisition of enzymes for the energy production under anaerobiosis, for the production of secondary metabolites or for the saprotrophic way of life.\textsuperscript{3,11,12}

Detection of laterally transferred genes is not as trivial as one might expect. While there are several approaches to do so\textsuperscript{13} the most common is assessing the phylogenetic discordance of the transferred genes relative to the organismal phylogeny. Genes whose history differs significantly from the history of organisms (e.g., a bacterial gene in a eukaryote) are suspected to have suffered an LGT in their history. Unfortunately, in many cases LGT is not the only plausible interpretation of the incongruent gene history and other scenarios, such as cases of ancestral gene duplications and differential losses must be also considered. It is often difficult to decide definitively which of these scenarios is correct.

**Several Eukaryotic Enzymes for Anaerobic Metabolism Originated from a Single Prokaryotic Source other than the Mitochondrial Symbiont**

In a recent report\textsuperscript{14} we describe the probable horizontal transfers of the eubacterial enzyme pyruvate formate lyase (PFL) and its activating enzymes (PFLAE) into and within eukaryotes. PFL is one of three enzymes known to catalyze the conversion of pyruvate to acetyl-CoA under anaerobic conditions; the other two are pyruvate:ferredoxin oxidoreductase (PFO) and pyruvate:NADP oxidoreductase (PNO). Although they differ in details of their reactions, these enzymes have been shown to functionally substitute for the pyruvate dehydrogenase complex (PDH) under anoxic conditions. Unlike PFO, PNO and PDH, a free-radical must be added to PFL by PFLAE for catalysis.

We identified a number of PFLs in diverse eukaryote representatives of the supergroups Archaeplastida (chlorophyte, glaucophyte, prasinophyte and rhodophyte algae), Opisthokonta (Amoebobium parasiticum, chytrid fungi), Amoebozoa (Mastigamoeba halamuthi) and Chromista (Thalassiosira pseudonana, Prymnesium parvum). Interestingly, some of these organisms are not regarded as typical anaerobes (diatom Thalassiosira, the opisthokont protist Amoebobium and the primary algae). The well-studied alga, Chlamydomonas reinhardtii, has been shown to harbor not only PFL but also PDH and PFO.\textsuperscript{15} Thorough phylogenetic analysis of both PFL and PFLAE revealed that the eukaryotic homologues of the two enzymes form a clade and have similar evolutionary histories. When the two genes were concatenated, the support for eukaryotic monophyly expressed in corrected bootstrap values\textsuperscript{16} was greater than 0.95 (statistically significant p < 0.05). This suggests that despite the diverse and patchy distribution of PFL and PFLAE across distantly related eukaryote lineages, eukaryote PFLs and PFLAEs are likely ultimately derived from a single prokaryotic source. Upon initial inspection, this finding is consistent with several popular hypotheses on the origin of eukaryotes and mitochondria, namely the Hydrogen hypothesis and the Synthrophic hypothesis.\textsuperscript{17,18} These hypotheses suggest that the mitochondrial endosymbiont was a facultative anaerobic $\alpha$-proteobacterium that could generate ATP in aerobic environments (using PDH, the citric acid cycle and the respiratory chain) as well as under anoxic conditions (using anaerobic enzymes like PFO, PFL or PNO, [FeFe]hydrogenase and hydG).

These hypotheses suggest that the endosymbionts that gave rise to these putativory, mitochondria harboured genes for both anaerobic and aerobic mitochondrial functions. It is well known that most enzymes of aerobic energy metabolism are indeed derived from the $\alpha$-proteobacterial mitochondrial endosymbiont, e.g., pyruvate dehydrogenase complex, cytochrome B, cytochrome c oxidase subunit, $\alpha$ and $\beta$ subunits of ATP synthase and many others.\textsuperscript{19,21} For the enzymes of anaerobic metabolism this phylogenetic affinity does not hold. For PFL, the $\alpha$-proteobacterial origin was strongly rejected using statistical tests. Instead, the phylogeny strongly suggests that the eukaryotic PFLs are derived from within a firmicute-bacteroides clade. The phylogenetic relationships of other enzymes specific for the energy metabolism under anaerobic conditions, PFO/PNO, [FeFe]hydrogenase, arginine deiminase (ADI), show similar patterns, however, for some of these cases, poor resolution of the trees makes it difficult to reach strong conclusions.

Arginine deiminase (ADI) is a part of arginine dehydroylase pathway that is used by several anaerobes instead of the urea cycle for energy generation and deamination of amino acids. The pathway has been more studied in Giardia intestinalis\textsuperscript{22,23} and Trichomonas vaginalis,\textsuperscript{24} where it localizes to the cytosol and hydrogenosomes (mitochondrion-related organelles) respectively. Like PFL, ADIs were found in distantly related eukaryotes and atypical anaerobes (such as the green algae Chlamydomonas and Chlorella). All eukaryotic ADIs form a clade with medium bootstrap support (61%) sister to a clade of Bacteroides and Archaea (bootstrap support 84%).\textsuperscript{23}

PFOs and PNOs are again present in unrelated eukaryotes including some atypical anaerobes (green algae and Thalassiosira).\textsuperscript{24} The eukaryotic PFOs/PNOs form a clade, with one exception. However, the phylogeny of PFO/PNO suffers from low resolution and the overall monophyly of eukaryotic enzymes cannot be excluded. The prokaryotic sister group could not be determined with confidence but the $\alpha$-proteobacterial origin could be statistically excluded.\textsuperscript{24}

The history of [FeFe]hydrogenase is the most difficult to interpret.\textsuperscript{25} As in previous cases, it is distributed across unrelated eukaryotic anaerobes as well as some green algae, Thalassiosira and Amoebozium. Although it should be noted that all eukaryotes contain non-hydrogen producing NARF’s homologous to [FeFe]hydrogenase.\textsuperscript{25} In the case of [FeFe]hydrogenase, it is clear that the eukaryotic enzymes are not monophyletic because Monocercomonoides, Trimatix and Entamoeba branch separately from each other eukaryotes emerging within separate clades of bacterial homologs.\textsuperscript{24} The maturase enzymes hydE, hydF and hydG are necessary for maturation of functional [FeFe]hydrogenase. In their phylogenetic
trees, the eukaryotic homologues always form a strongly supported clade. While the prokaryotic source of [FeFe]hydrogenase itself could not be determined, the maturases consistently grouped with the δ-proteobacteria and firmicutes.

In summary, while the resolution of protein trees of the enzymes discussed above—ADI, PFO/PNO and [FeFe]hydrogenase plus its maturases—is not as robust as the PFL/PFLAE phylogeny, all of these analyses demonstrate these anaerobic enzymes (1) are present in disparate eukaryotes emerging from within lineages that predominantly lack these enzymes, (2) may have been derived from a single bacterial source (two sources for [FeFe]hydrogenases) and (3) likely did not originate from the mitochondrial endosymbiont, because they are not related to enzymes of contemporary α-proteobacteria.

Were these Enzymes Transferred to the Eukaryotic Common Ancestor or have these Genes Undergone Eukaryote-to-Eukaryote Lateral Gene Transfers?

Phylogenetic patterns that we observe in the enzymes in question, i.e., patchy representation in distantly related eukaryote lineages and, at the same time, an apparent single origin for all of the eukaryote homologs, could arise in by several means. Firstly, the transfer of the gene may have happened very early in eukaryotic evolution and the gene was present in the common ancestor of all eukaryotes that possess it today. This scenario requires that the vast majority of the descendants of this common ancestor have lost the gene. Secondly, the gene transfer may have happened more recently to one lineage of eukaryotes and subsequently the gene has undergone several eukaryote-to-eukaryote lateral transfers to other unrelated eukaryotes.

One may distinguish between these scenarios by inspecting the relationships in the eukaryotic subtree in the enzyme trees. If the former scenario were correct, we would expect that the relationships among eukaryotes would be consistent with the eukaryotic phylogeny as it is understood today. If the latter scenario were instead correct we would expect the opposite. The relationships among eukaryotes in PFL, PFO/PNO, [FeFe]hydrogenase and ADI trees are broadly inconsistent with well-accepted eukaryotic relationships. The problem, however, is that the statistical support for the relative branching order within the eukaryote subtree is often low and for PFO/PNO and ADI datasets the topology that is consistent with known
eukaryotic phylogenetic relationships cannot be statistically excluded. In case of [FeFe]hydrogenase the relationships are poorly supported yet there are strongly supported nodes that contradicts eukaryotic phylogeny—e.g., there is a robust clad that includes one *Entamoeba* copy as well as oxynomas plus *Trimastix* and a robust clad of another *Entamoeba* copy grouping with diplomonads plus retortamonads. Finally, in the case of PFL, the contradicting relationships are strongly supported and the conflict with eukaryotic phylogeny is statistically significant. In Figure 1 we present an updated tree of eukaryotic PFLs and, as in our previous analyses, the enzymes of green algae, red algae and glaucophytes clearly do not form a clad. The green algae are located in three different positions in the tree—clades of chlorophytes, streptophytes and prasinophytes. This result argues in favour of the aforementioned eukaryote-to-eukaryote LGT scenario. The frequent occurrence of PFL among green algae indicates that ancestor of green algae or the ancestor of Archaeplastida might be the first eukaryotic host for PFL and PFLAE (Figure 1B). From its descendants, the genes could then have been transferred to other PFL-bearing eukaryotes. Alongside with PFL, its activation enzyme (PFLAE) must have been transferred; otherwise PFL would not be functional. Such simultaneous transfers would be facilitated by the fact the PFL and PFLAE are encoded on a single operon in prokaryotes and sometimes transfers would be facilitated by the fact they evolve in anaerobic niche, where unrelated organisms became physically close and may exchange the tools (i.e., enzymes) advantageous for that particular environment. In this view, it is difficult to explain why the transfer from prokaryotes to eukaryotes should have happened only once and why other eukaryotes should preferentially accept the gene from eukaryotes despite the fact that these genes are also present in many surrounding prokaryotes, which are more abundant and more often become their prey. Moreover, gene transfer from prokaryotes to eukaryotes is not uncommon and multiple independent transfers of one particular enzyme from various prokaryotic sources to eukaryotes have been described, for example, in references 26 and 27. Why PFL/PFLAE and maybe some other enzymes of anaerobic metabolism in eukaryotes (such as PFO, PNO, [FeFe]hydrogenase, ADI), have followed this particularly circuitous evolutionary trajectory remains an open question.

Acknowledgements
V.H. was supported by Ministry of Education, Youth and Sport of the Czech Republic (project MSM0021620828). C.W.S. is supported by scholarships from the Natural Sciences and Engineering Research Council of Canada and Killiam Trusts. A.J.R. was supported by the Canadian Institute for Health Research (grant number MOP 62809).

References
1. Keeling PJ, Palmer JD. Horizontal gene transfer in eukaryotic evolution. Nat Rev Genet 2008; 9:695-8; PMID: 1891983; DOI:10.1038/nrg2386.
2. Andersson JO. Lateral gene transfer in eukaryotes. Cell Mol Life Sci 62:1182-97; PMID: 15761667; DOI:10.1007/s00018-001-0559-9.
3. Richards TA, Dacks JB, Jenkinson JM, Thornton CR, Talbot NJ. Evolution of filamentous plant pathogen genes: gene exchange across eukaryotic kingdoms. Curr Biol 2006; 16:1857-64; PMID: 16979565; DOI:10.1016/j.cub.2006.07.052.
4. Westheide DR. The transfer of metabolic genes explored: analysis of the horizontal transfer of enzyme encoding genes in unicellular eukaryotes. Genome Biol 2009; 10:36; PMID: 19368726; DOI:10.1186/gb-2009-10-4-36.
5. Martin W, Schnarrenberger C. The evolution of the Calvin cycle from prokaryotic to eukaryotic chloroplasts: a case study of functional redundancy in ancient pathways through endosymbiosis. Curr Genet 1997; 32:1-18; PMID: 9309164; DOI:10.1007/s002940050241.
6. Doolittle WF. You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. Trends Genet 1998; 14:507-11; PMID: 9724962; DOI:10.1016/S0168-9525(98)01494-2.
7. Yap WH, Zhang Z, Wang Y. Distinct types of rRNA operons exist in the genome of the actinomycyte *Thermomonospora chromogena* and evidence for horizontal transfer of an entire rRNA operon. J Bacteriol 1998; 14:307-11; PMID: 9724962; DOI:10.1016/S0168-9525(98)01494-2.
8. Miller SR, Augustine S, Olson TL, Blankenship RE, Selker J, Wood AM. Discovery of a free-living chlorophyll d-producing cyanobacterium with a hybrid prochlorocystein/cyanobacterial small-subunit rRNA gene. Proc Natl Acad Sci USA 2002; 102:850-5; PMID: 15673160; DOI:10.1073/pnas.0405672102.
9. Cohen O, Gophna U, Pupko T. The complexity hypothesis revisited: connectivity rather than function constitutes a barrier to horizontal gene transfer. Mol Biol Evol 2011; 28:1481-9; PMID: 2119642; DOI:10.1093/molbev/mss353.
10. Rivera MC, Jain R, Moore JE, Lake JA. Genomic evidence for two functionally distinct gene classes. Proc Natl Acad Sci USA 1998; 95:6239-44; PMID: 960949.