Phylogenetic Analyses of Plastid-Originated Proteins Imply Universal Endosymbiosis in Ancestors of Animals and Fungi

Shu Yuan[§], Jian-Hua Guo[§], Jun-Bo Du, and Hong-Hui Lin*

Key Laboratory of Bio-resources and Eco-environment (Ministry of Education), College of Life Science, Sichuan University, Chengdu, Sichuan 610064, P. R. China.
Fax: 86-28-85412571. E-mail: honghuilin@hotmail.com

* Author for correspondence and reprint requests

Z. Naturforsch. **63c**, 903–908 (2008); received April 25/June 19, 2008

We searched and analyzed cyanobacteria-originated metazoa/fungi proteins (COPs) by phylogenetic analyses. Analysis of them showed that for millions of years universal plastid endosymbiosis and gene transfer occurred in ancestors of metazoa/fungi, and some transferred fragments have been reserved until now even in modern mammals. Most eukaryotes contained plastids in the ancient era, and some of them lost them later. Functions of homologues in cyanobacterial genomes and eukaryotic genomes are in consensus, and most are involved in the organic compound metabolism. With emergence of organelles and subcellular structure in eukaryotic cells, the locations of these proteins diversified. Furthermore, some novel functions were endowed for COPs, especially in vertebrates.

Key words: Plastid-Originated Proteins, Divergence Time, Endosymbiosis, Gene Transfer

Introduction

Previously, we analyzed nuclear-localized plastid-like DNA (nupDNA) fragments in protozoa, metazoa and fungi. Most eukaryotes that do not have plastids contain 40–5000 bp nupDNAs in their nuclear genomes (Yuan *et al.*, 2007). Analysis of them showed that through millions of years of universal endosymbiosis and gene transfer they may have occurred in ancient protists before divergence of plants and animals/fungi, and some transferred fragments have been reserved until now even in modern mammals. But further analyses implied that some nupDNAs are contaminations of bacterial genome fragments, and some other nupDNAs may be common sequences in a large number of species, such as most rDNA sequences. Therefore, the real nupDNAs derived from possible symbiosis may be little less than we originally expected. However, the idea behind the paper is plausible. It is very likely that universal endosymbiosis and gene transfers occurred millions of years ago, and some transferred fragments have been reserved until now even in modern mammals (Embley and Martin, 2006). Therefore we further investigated possible endosymbiosis in ancestors of animals and fungi at the protein level.

Methods

Candidate protein filter procedure

To search cyanobacteria-originated metazoa/fungi proteins (COPs), a hypothetical phylogenetic tree relating to the acceptance of COPs was proposed (Fig. 1). The branching arrangement of the resulting phylogenies is a critical component of a COP assessment.

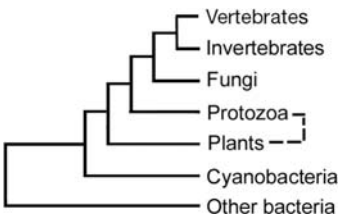

Fig. 1. Hypothetical phylogenetic tree relating to the acceptance of COP. In case of a gene transfer from cyanobacteria to eukaryotes, the necessary phylogeny would show eukaryote paraphyly with protozoa/metazoa/fungi separated from plants by a paraphyletic assemblage of bacterial lineages, where cyanobacteria were more closely related to eukaryotes than other bacteria. The position of some protozoa and plants may be cross (indicated by a broken line), because some protozoa once contained plastids.

[§] These authors contributed equally to this work.

0939–5075/2008/1100–0903 $ 06.00

Four kinds of protein pools were detected to search for COPs:

1) Nuclear-localized plastid-like DNA fragments in protozoa, metazoa and fungi that we analyzed previously (Yuan *et al*., 2007). But all these proteins were rejected as potential COPs by phylogenetic analyses.

2) Probable bacteria-to-vertebrate horizontally transferred genes in the human genome with best hits in cyanobacteria (International Human Genome Sequencing Consortium, 2001). They are gi6912516, gi8922122, P45381, P21397 (P27338), gi8922946, IGI_M1_ctg13419_35, IGI_M1_ctg14420_10 (IGI_M1_ctg14420_109), IGI_M1_ctg19053_31, gi8923844, and IGI_M1_ctg13238_61. However, by phylogenetic analyses, we found that only four proteins may actually originate from cyanobacteria: NAD/FAD-dependent oxidoreductase (Renalase, gi8922946), zinc phosphodiesterase (ELAC, gi8922122), aspartoacylase (ASPA, P45381), and sodium-dependent multi-vitamin transporter (SMVT, IGI_M1_ctg14420_10).

3) Homologues of proteins found in both cyanobacteria (*Nostoc* sp. PCC 7120) and unicellular eukaryotes (*Saccharomyces cerevisiae*, *Schizosaccharomyces pombe* or *Encephalitozoon cuniculi*), from the website of Clusters of Orthologous Groups (COG, http://www.ncbi.nlm.nih.gov/COG/grace/cogenome.cgi?g=103690) (Tatusov *et al*., 2003). Most of these proteins are ambiguous to define their origins. Only some homologues found in both cyanobacteria and unicellular eukaryotes, but almost not in other bacteria, could be used for COP phylogenetic analyses. They are COG5031, COG5542, COG5398, COG3751, COG4240, COG4294, COG4301, COG4360, COG4725, COG5017, COG5119, COG5126, COG5135, COG5524, COG5548, and COG0659 (for *Synechococcus*). Phylogenetic analyses supposed that heme oxygenase (HMOX, COG5398), small integral membrane protein (SIMP, COG5548), coenzyme Q4 (COQ4, COG5031), adenosine methyltransferase (MA-T, COG4725), glucosyltransferase (GTF, COG5017) and bicarbonate transporter (BT, COG0659, only fungi BT originated from cyanobacteria, therefore not listed in Table I) were putative COPs. According to Obornik and Green's (2005) analyses, the other two proteins porphobilinogen deaminase (PBGD) and phosphoadenosine phosphosulfate (PAPS) reductase (only fungi PAPS reductase originated from cyanobacteria, therefore not listed in Table I) related to HMOX should also be COPs.

4) Horizontally transferred gene from cyanobacteria or apicoplasts to *Leishmania*, *Trypanosoma*, *Toxoplasma* or *Cryptosporidium*. The corresponding proteins are: plant-like APX (Adak and Datta, 2005), glucose-6-phosphate dehydrogenase and other 16 proteins (Hannaert *et al*., 2003), glucose-6-phosphate isomerase and enolase (Dzierszinski *et al*., 1999), leucine aminopeptidase, biopteridine transporter, uridine kinase/uracil phosphoribosyltransferase and calcium-dependent protein kinases (Huang *et al*., 2004). Except for fructose bisphosphate aldolase (FBA), all other proteins were rejected as potential COPs by phylogenetic analyses.

Phylogenetic analysis methods

All sequences were retrieved from the NCBI database (www.ncbi.nlm.nih.gov) and analyzed at the amino acid level. Candidate sequences from the above four resources were used as queries for gene search using BLASTP. Sequences were aligned using CLUSTAL X (Thompson *et al*., 1997) and refined by eye. Gaps resulting from the alignment were treated as missing data. Ambiguous alignments in highly variable regions were excluded from the phylogenetic analyses. The reliability of interior branches was assessed with 1,000 bootstrap resamplings using "pairwise deletion option" of amino acid sequences with gamma parameters. Neighbour-joining (NJ) phylogenetic analysis (Saitou and Nei, 1987) was performed by using PAUP V.4.0β. Maximum likelihood (ML) (Strimmer and von Haeseler, 1996) was performed by using the JTT model (Jones *et al*., 1992). ML analyses were performed by using PHYLIP V. 3.62. NJ phylogenetic analyses usually resulted in lower bootstrap values andare therefore not shown in Results. Hypothetical phylogenetic trees relating to accepted cyanobacterial proteins are available directly form the authors (therefore not shown in the manuscript).

Because the sequences should not show constant evolutionary rates across paralogs, we used the nonparametric rate-smoothing method (Sanderson, 1997) to estimate divergence times by the software r8s. The divergence of amniotes and amphibians 360 million years ago (Kumar and Hedges, 1998) was used as the calibration point.

Table I. Locations, functions and divergence times of potential COPs.

Enzyme	Function in cyanobacteria[a]	Function in metazoa[a]	Location in metazoa[a]	Divergence time [Myr][a]
NAD/FAD-dependent oxidoreductase (Renalase)	General function	Amine oxidase	C, secreted protein	700
Zinc phosphodiesterase (ELAC)	RNase Z	Endonuclease activity	N (probable)	870
Aspartoacylase (ASPA)	Alanine/aspartate metabolism	Alanine/aspartate metabolism	C	870
Sodium-dependent multi-vitamin transporter (SMVT)	Na+/proline symporter	Na+/proline symporter	CM	680
Heme oxygenase (HMOX)	Heme degradation	Heme degradation	ER	580–720
Porphobilinogen deaminase (PBGD)	Porphyrin synthesis	Porphyrin synthesis	N & C	680
Small integral membrane protein (SIMP)	NA	NA	CM (probable)	690
Coenzyme Q4 (COQ4)	Coenzyme synthesis	Coenzyme synthesis	Mitochondrion	850–1170
Adenosine methyltransferase (MT-A)	DNA methylation	RNA methylation	N	690
Glucosyltransferase (GTF)	Carbohydrate metabolism	Carbohydrate metabolism	C & ER	640
Fructose bisphosphate aldolase (FBA)	Glycolytic pathway	Glycolytic pathway	C	730

[a] C, cytosol; CM, cytoplasmic membrane; ER, endoplasmic reticulum; N, nucleus; NA, not available; Myr, million years.

Results

13 Candidates of cyanobacteria-originated metazoa/fungi proteins (COPs)

The publication of the human genome reported 113 incidents of direct horizontal gene transfer (HGT) between bacteria and vertebrates without any nonvertebrate evolutionary intermediates (International Human Genome Sequencing Consortium, 2001). Later analyses indicated that most of the putative HGT genes can be explained in terms of descent through common ancestry. They are, therefore, unlikely to be examples of direct HGT from bacteria to vertebrates, but may be due to evolutionary rate variation, the small sample of nonvertebrate genomes, or gene loss in the nonvertebrate lineages (Stanhope *et al.*, 2001; Genereux and Logsdon, 2003). We searched for ten possible HGTs in the human genome with best hits in cyanobacteria, and found that six of them are not very similar with cyanobacteria. The left four proteins should be potential COPs and actually originate from cyanobacteria. Our trees also suggested that these genes were not really HGTs from bacteria to vertebrates, which validated the previous conclusion as mentioned above.

Screening homologues found in both cyanobacteria and unicellular eukaryotes would be helpful to find possible COPs transferred from cyanobacteria to unicellular ancestors of metazoa/fungi. By this method, we filtered six potential COPs and two other related proteins.

Horizontally transferred genes from cyanobacteria or apicoplasts to *Leishmania*, *Trypanosoma*, *Toxoplasma* or *Cryptosporidium* were analyzed too. All the corresponding proteins in metazoa

and fungi, except for FBA, were rejected as potential COPs by phylogenetic analyses.

Further analyses of the trees and divergence time estimation

The trees for acceptance of COPs were assayed further. Bootstrap values for the key nodes supporting cyanobacteria origination were usually about 30%, and sometimes up to 78–80% (PBGD and PAPS reductase). The values were relatively low; but by whatever method we constructed the trees (data not shown), the bootstrap values could not be improved, and the most important at all, metazoa/fungi clusters with cyanobacteria could not be changed in all trees. Endosymbiosis and gene transfers occurred 500 million years ago (Embley and Martin, 2006), and thus it is reasonable that the bootstrap values were not high. Actually, much more genes may originate from cyanobacteria. However, due to their very-long-time evolutions, they could not be defined as COPs and the origins became ambiguous. ELAC and GTF did not fit the ideal COP phylogenetic tree very well, but their metazoa/fungi homologues were most probably originated from cyanobacteria (by the BLASTP method), wherefore we considered these two proteins as COPs. Some trees contained two or more subfamilies, but at least one of them potentially originated from cyanobacteria. It is notable that, for ELAC, GTF, FBA and BT, genes seemed to be transferred from cyanobacteria to algae firstly, then to metazoa/fungi, suggesting possibly secondary endosymbiosis in them. Giving that almost all classes of unicellular eukaryotes contain plastid proteins, it can be estimated that most protists once had plastids through primary endosymbiosis or secondary endosymbiosis, including ancestors of metazoa/fungi.

Origins and evolutions of some COPs have been studied before. HMOX is located in the cytosol and should be of either cytosolic or proteobacterial (mitochondrial) origin, but on the phylogenetic tree, it appeared closely related to cyanobacterial sequences. Obornik and Green's (2005) work showed that such clustering of proteins from cyanobacteria and nonphotosynthetic eukaryotes has also been found in PBGD and PAPS reductase. They suggested that these clusters may be an artifact due to the limited sampling of eukaryotes. However, by our expand analyses with more eukaryote samples, their origins were still very clear. Moreover, all representative bacteria (all classes of bacteria, if homologues can be found) were included in the trees, and it is less likely that some class of bacteria was omitted or should be placed between metazoa/fungi and cyanobacteria in the trees, therefore the possibility of artifacts was excluded.

Another instance may weaken our conclusion, that these proteins were derived from bacteria-to-bacteria gene transfers, which means from cyanobacteria to prokaryotic ancestors of metazoa/fungi. Then they may fit the ideal COP tree, but can not be regarded as acquisitions through endosymbiosis. We ruled out this possibility by divergence time estimations. The age estimated for divergence of cyanobacteria and metazoa/fungi was about 640–1170 million years old (most were 680–870 million years, Table I), *i.e.* after eukaryote naissance, but before metazoa appearance (Bengston, 1998), which supported our endosymbiosis hypothesis.

Functional and locational comparisons of the proteins

The functions of homologues in cyanobacterial genomes and eukaryotic genomes are in consensus. With emergence of organelles and subcellular structure in eukaryotic cell, the locations of these proteins diversified. Correspondingly, some novel functions were endowed, especially in vertebrates (Table I). Renalase is a NAD/FAD-dependent oxidoreductase with general functions in cyanobacteria. However, in humans, it is an amine oxidase that is secreted by the kidney, circulates in blood, and modulates cardiac function and systemic blood pressure (Xu *et al*., 2005). ELAC belongs to metallo-β-lactamase superfamily III, that cleaves the 3′ end of tRNAs at the discriminator base, which is similar in both prokaryotes and eukaryotes (Condon and Putzer, 2002). ASPA deficiency leads to increased urinary excretion of *N*-acetyl-aspartic acid (NAA), resulting in human Canavan disease, which is a severe progressive leukodystrophy characterized by swelling and spongy degeneration of the white matter of the brain (Matalon and Michals-Matalon, 2000). HOMX has evolved to carry out the oxidative cleavage of heme, a reaction essential in physiological processes as diverse as iron reutilization and cellular signaling in mammals, synthesis of essential lightharvesting pigments in cyanobacteria and higher plants, and

the acquisition of iron by bacterial pathogens (Wilks, 2002). Free heme is highly toxic to cellular structures. Therefore, humans and other organisms have evolved highly efficient and regulated mechanisms for the synthesis, transport and catabolism of this molecule. HOMX is closely related to cardiovascular and renal disorders, neurologic diseases and malaria (Shibahara *et al.*, 2002). PBGD and PAPS reductase also participate in heme metabolism. Over-expressions of the human PBGD in the glioma cells induced a G1 cell cycle attenuation accompanied by increases in enzyme activity and c6 differentiation toward astrocytes, suggesting a role for PBGD in tumourigenesis (Greenbaum *et al.*, 2002). MT-A is a subunit of mRNA:m^6A methyltransferase, an enzyme that sequence-specifically methylates adenines in pre-mRNAs in eukaryotes, but in cyanobacteria it catalyzes DNA methylation (Bujnicki *et al.*, 2002). GTF found in our analyses belongs to the family 28, which catalyzes sugar chain elongation in biosynthesis. It is located in both cytosol and endoplasmic reticulum (Kikuchi and Narimatsu, 2006).

In summary, COPs can be located in every part of the cell, even in the mitochondrion, or secreted out of the cell. Most of them play key roles in organic compound synthesis and degradation, including nucleic acid modification and metabolism. It is interesting that half of them are related to glycometabolism or pigment metabolism, some of which are regarded as important plastid-to-nucleus signals (Koussevitzky *et al.*, 2007).

Discussion

Besides phylogenetic analyses, existent protists also support our hypothesis of ancient universal endosymbiosis. Some protozoa without plastid were considered to have plastid in their evolutionary history, such as *Trypanosoma*, *Leishmania* and Apicomplexa (*Plasmodium*, *Toxoplasma*) (McFadden *et al.*, 1996; Hannaert *et al.*, 2003). Several groups of eukaryotes are believed to have secondarily lost their plastids, of which the oomycetes are probably the most notable example (Cavalier-Smith, 2000). Okamoto and Inouye (2005) also demonstrated that a secondary symbiosis of green algae in a flagellate is in progress at present, indicating that plastid endosymbiosis is a common process in protists even under current natural conditions. Retrospecting to the Proterozoic era when eukaryotes emerged, universal endosymbiosis occurred. Most protozoa may not have evolved effectively kinetic organelles at that time. Consequently, they adopted amphitrophy and temporarily contained some cyanobacteria or algae, and used plastids to fix light energy, especially when organic nutrition was limited. Hundred million years later, some of the ancient protists evolved into protozoa, metazoa and fungi, and discarded plastids finally. However, there were also some protozoa which reserved some plastid relicts, such as apicoplast in Apicomplexa and plate-like chloroplast in *Ochromonas danica* (Semple, 1998). Besides, a few protists still kept their ability to engulf photosynthetic eukaryotes heretofore, such as *Lotharella amoeboformis* (Ishida *et al.*, 2000) and the flagellate "Hatena" (Okamoto and Inouye, 2005). From the first eukaryote naissance (1200 million years ago) to the first metazoa appearance (570 million years ago), ancestors of metazoa and fungi should have much chance to acquire and fix some fragments from cyanobacteria or algae (Embley and Martin, 2006; Yuan *et al.*, 2007), which is rehearsed in this paper by phylogenetic analyses. Endosymbiosis may be a very important impetus for gene origin and evolution.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (30571119 and 30670166) and the New Century Excellent Talents Program (NCET-05-0786).

Adak S. and Datta A. K. (2005), *Leishmania major* encodes an unusual peroxidase that is a close homologue of plant ascorbate peroxidase: a novel role of the transmembrane domain. Biochem. J. **390**, 465–474.

Bengston S. (1998), Animal embryos in deep time. Nature **391**, 529–530.

Bujnicki J. M., Feder M., Radlinska M., and Blumenthal R. M. (2002), Structure prediction and phylogenetic analysis of a functionally diverse family of proteins homologous to the MT-A70 subunit of the human mRNA:m^6A methyltransferase. J. Mol. Evol. **55**, 431–444.

Cavalier-Smith T. (2000), Membrane heredity and early chloroplast evolution. Trends Plant Sci. **5**, 174–182.

Condon C. and Putzer H. (2002), The phylogenetic distribution of bacterial ribonucleases. Nucleic Acids Res. **30**, 5339–5346.

Dzierszinski F., Popescu O., Toursel C., Slomianny C., Yahiaoui B., and Tomavo S. (1999), The protozoan parasite *Toxoplasma gondii* expresses two functional plant-like glycolytic enzymes. J. Biol. Chem. **274**, 24888–24895.

Embley T. M. and Martin W. (2006), Eukaryotic evolution, changes and challenges. Nature **440**, 623–630.

Genereux D. P. and Logsdon J. M. Jr. (2003), Much ado about bacteria-to-vertebrate lateral gene transfer. Trends Genet. **19**, 191–195.

Greenbaum L., Gozlan Y., Schwartz D., Katcoff D. J., and Malik Z. (2002), Nuclear distribution of porphobilinogen deaminase (PBGD) in glioma cells: a regulatory role in cancer transformation? Brit. J. Cancer **86**, 1006–1011.

Hannaert V., Saavedra E., Duffieux F., Szikora J. P., Rigden D. J., Michels P. A. M., and Opperdoes F. R. (2003), Plant-like traits associated with metabolism of *Trypanosoma* parasites. Proc. Natl. Acad. Sci. USA **100**, 1067–1071.

Huang J., Mullapudi N., Lancto C. A., Scott M., Abrahamsen M. S., and Kissinger J. C. (2004), Phylogenomic evidence supports past endosymbiosis, intracellular and horizontal gene transfer in *Cryptosporidium parvum*. Genome Biol. **5**, R88.

International Human Genome Sequencing Consortium (2001), Initial sequencing and analysis of the human genome. Nature **409**, 860–921.

Ishida K., Ishida N., and Hara Y. (2000), *Lotharella amoeboformis* sp. nov.: A new species of chlorarachniophytes from Japan. Phycol. Res. **48**, 221–230.

Jones D. T., Taylor W. R., and Thornton J. M. (1992), The rapid generation of mutation data matrices from protein sequences. Comput. Appl. Biosci. **8**, 275–282.

Kikuchi N. and Narimatsu H. (2006), Bioinformatics for comprehensive finding and analysis of glycosyltransferases. Biochim. Biophys. Acta **1760**, 578–583.

Koussevitzky S., Nott A., Mockler T. C., Hong F., Sachetto-Martins G., Surpin M., Lim J., Mittler R., and Chory J. (2007), Signals from chloroplasts converge to regulate nuclear gene expression. Science **316**, 715–719.

Kumar S. and Hedges S. B. (1998), A molecular timescale for vertebrate evolution. Nature **392**, 917–920.

Matalon R. and Michals-Matalon K. (2000), Spongy degeneration of the brain, Canavan disease: Biochemical and molecular findings. Front. Biosci. **5**, D307–D311.

McFadden G. I., Reith M. E., Munholland J., and Lang-Unnasch N. (1996), Plastid in human parasites. Nature **381**, 482.

Obornik M. and Green B. R. (2005), Mosaic origin of the heme biosynthesis pathway in photosynthetic eukaryotes. Mol. Biol. Evol. **22**, 2343–2353.

Okamoto N. and Inouye I. A. (2005), A secondary symbiosis in progress? Science **310**, 287.

Saitou N. and Nei M. (1987), The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**, 406–425.

Sanderson M. J. (1997), A nonparametric approach to estimating divergence times in the absence of rate constancy. Mol. Biol. Evol. **14**, 1218–1231.

Semple K. T. (1998), Heterotrophic growth on phenolic mixtures by *Ochromonas danica*. Res. Microbiol. **149**, 65–72.

Shibahara S., Kitamuro T., and Takahashi K. (2002), Heme degradation and human disease: diversity is the soul of life. Antioxid. Redox Signal. **4**, 593–602.

Stanhope M. J., Lupas A., Italia M. J., Koretke K. K., Volker C., and Brown J. R. (2001), Phylogenetic analyses do not support horizontal gene transfers from bacteria to vertebrates. Nature **411**, 940–944.

Strimmer K. and von Haeseler A. (1996), Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. Mol. Biol. Evol. **13**, 964–969.

Tatusov R. L., Fedorova N. D., Jackson J. D., Jacobs A. R., Kiryutin B., Koonin E. V., Krylov D. M., Mazumder R., Mekhedov S. L., Nikolskaya A. N., Rao B. S., Smirnov S., Sverdlov A. V., Vasudevan S., Wolf Y. I., Yin J. J., and Natale D. A. (2003), The COG database: an updated version includes eukaryotes. BMC Bioinformatics **4**, 41–54.

Thompson J. D., Gibson T. J., Plewniak F., Jeanmougin J., and Higgins D. G. (1997), The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25**, 4876–4882.

Wilks A. (2002), Heme oxygenase: Evolution, structure, and mechanism. Antioxid. Redox Signal. **4**, 603–614.

Xu J., Li G., Wang P., Velazquez H., Yao X., Li Y., Wu Y., Peixoto A., Crowley S., and Desir G. V. (2005), Renalase is a novel, soluble monoamine oxidase that regulates cardiac function and blood pressure. J. Clin. Invest. **115**, 1275–1280.

Yuan S., Sun X., Mu L. C., Lei T., Liu W. J., Wang J. H., Du J. B., and Lin H. H. (2007), Nuclear-localized plastid DNA fragments in protozoa, metazoa and fungi. Z. Naturforsch. **62 c**, 123–132.