Protist, Vol. 159, 459—470, July 2008
http://www.elsevier.de/protis
Published online date 26 March 2008

Protist

ORIGINAL PAPER

Evolutionary Analysis of Synteny and Gene Fusion for Pyrimidine Biosynthetic Enzymes in Euglenozoa: An Extraordinary Gap between Kinetoplastids and Diplonemids

Takashi Makiuchi[a], Takeshi Annoura[a], Tetsuo Hashimoto[b], Eri Murata[a], Takashi Aoki[a], and Takeshi Nara[a,1]

[a]Department of Molecular and Cellular Parasitology, Juntendo University School of Medicine, 2-1-1 Hongo, Bunkyo-ku, Tokyo 113-8421, Japan
[b]Institute of Biological Sciences, University of Tsukuba, 1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan

Submitted November 7, 2007; Accepted February 3, 2008
Monitoring Editor: Larry Simpson

A unique feature of the genome architecture in the parasitic trypanosomatid protists is large-scale synteny. We addressed the evolutionary trait of synteny in the eukaryotic group, Euglenozoa, which consists of euglenoids (earliest branching), diplonemids, and kinetoplastids (trypanosomatids and bodonids). Synteny of the pyrimidine biosynthetic (*pyr*) gene cluster, which constitutes part of a large syntenic cluster in trypanosomatids and includes four separate genes (*pyr1—pyr4*) and one fused gene (*pyr6*/*pyr5* fusion), was conserved in the bodonid, *Parabodo caudatus*. In the diplonemid, *Diplonema papillatum*, we identified *pyr4* and *pyr6* genes. Phylogenetic analyses of pyr4 and pyr6 showed the separate origin of each in kinetoplastids and euglenoids/diplonemids and suggested that kinetoplastids have acquired these genes via lateral gene transfer (LGT). Because replacement of genes by non-orthologs within the syntenic cluster is highly unlikely, we concluded that, after separation of the line leading to diplonemids, the syntenic *pyr* gene cluster was established in the common ancestor of kinetoplastids, preceded by their acquisition via LGT. Notably, we found that diplonemid *pyr6* is a stand-alone gene, inconsistent with both euglenoid *pyr5*/*pyr6* and kinetoplastid *pyr6*/*pyr5* fusions. Our findings provide insights into the evolutionary gaps within Euglenozoa and the evolutionary trait of rearrangement of gene fusion in this lineage.

Key words: Euglenozoa; gene cluster; kinetoplastids; lateral gene transfer; pyrimidine biosynthetic enzymes; synteny.

Introduction

Trypanosomatids are flagellated parasitic protists and include medically important pathogens, such as those causing Chagas' disease, African sleeping sickness, and leishmaniasis. The phylogenetic position of trypanosomatids has been extensively studied using molecular phylogeny. Species of trypanosomatids are monophyletic, and this clade is nested in the kinetoplastid clade with bodonids, the sister group of trypanosomatids.

[1]Corresponding author; fax +81 3 5800 0476
e-mail tnara@med.juntendo.ac.jp (T. Nara).

doi:10.1016/j.protis.2008.02.002

Kinetoplastids, together with euglenoids and diplonemids, are assembled into a large monophyletic group, the Euglenozoa, which is characterized by distinctive mitochondria with discoid cristae (Cavalier-Smith 1981). Of the three branches of Euglenozoa, the euglenoids constitute the earliest branch, followed by separation of the diplonemid and kinetoplastid lineages (Simpson and Roger 2004; Simpson et al. 2002).

Synteny, the preserved order of genes, is often observed in the genomes of phylogenetically related eukaryotic species. Conserved synteny is used as an evolutionary marker, which can not only indicate the descent of different species from a common ancestor but also functional and/or evolutionary relationships of the clustered genes (Bennetzen and Freeling 1997; Nadeau 1989).

Comparative genomics of three trypanosomatids, *Trypanosoma cruzi, T. brucei*, and *Leishmania major*, have highlighted large-scale synteny of polycistronic gene clusters as a feature of their unique genome architecture (El-Sayed et al. 2005). That is, protein-encoding genes, most of which are functionally unrelated, are tandemly arrayed on either strand of DNA as syntenic gene clusters and constitute polycistronic transcription units (Bonen 1993; Liang et al. 2003; Martínez-Calvillo et al. 2004). Due largely to the lack of genomic information on other euglenozoan groups, however, the origin of conserved synteny and gene clustering in trypanosomatids has not yet been determined (Dávila and Lukeš 2003; Jackson et al. 2006).

The de novo pyrimidine biosynthetic pathway is one of the essential catalytic activities in organisms, which produces uridine-5′-monophosphate (UMP) for incorporation into DNA and RNA. This pathway consists of six enzymes: pyr1 (EC 6.3.5.5, carbamoyl-phosphate synthetase II), pyr2 (EC 2.1.3.2, aspartate carbamoyltransferase), pyr3 (EC 3.5.2.3, dihydroorotase), pyr4 (EC 1.3.3.1, dihydroorotate dehydrogenase), pyr5 (EC 2.4.2.10, orotate phosphoribosyltransferase), and pyr6 (EC 4.1.1.23, orotidine-5′-monophosphate decarboxylase), in their order of reaction.

We previously showed that all *pyr* genes are clustered in the genomes of two trypanosomatid species, *T. cruzi* (Gao et al. 1999) and *L. mexicana* (GenBank™ Accession number AB029444), in both of which five genes, *pyr1, pyr3, pyr6/pyr5* fused gene, *pyr2,* and *pyr4* are juxtaposed in this order on chromosomal DNA. Similarly, *T. brucei* possesses the *pyr* gene cluster, although *pyr3* is not annotated in this cluster (Berriman et al. 2005). The *pyr* gene cluster is the only known clustering of genes that encode all enzymes in an essential metabolic pathway in eukaryotes, while there are examples in filamentous fungi of the clustering of genes encoding enzymes catalyzing secondary metabolites (Saikia et al. 2007; Young et al. 2006).

Clustering of genes coding for a metabolic pathway is structurally similar to bacterial operons, which may be advantageous for concerted expression of functionally related enzymes at the appropriate times in these bacteria. However, in trypanosomatids, no regulatory mechanism of transcription initiation is found and regulation of expression seems to occur at the post-transcriptional level. Thus, the biological significance of the occurrence of the *pyr* gene cluster in trypanosomatids is still unknown.

The *pyr* gene cluster appears to constitute part of a large polycistronic gene cluster in trypanosomatids and, importantly, includes genes acquired via lateral gene transfer (LGT). Indeed, trypanosomatid genomes contain large numbers of genes thought to have been acquired via LGT (Opperdoes and Michels 2007). Phylogenetic analyses of pyr4 and pyr6 have shown that both genes have a prokaryotic origin, not only in trypanosomatids but also in bodonids (Annoura et al. 2005; Makiuchi et al. 2007; Nara et al. 2000). Thus, LGT events are likely to have preceded the establishment of the *pyr* gene cluster, as well as contributing to it.

In the present study, we regarded the *pyr* gene cluster as a model synteny and addressed whether this cluster is present in non-trypanosomatid kinetoplastids, i.e. bodonids. We demonstrate clustering of the *pyr* genes in the bodonids, *Parabodo caudatus* (formerly *Bodo caudatus*) and *Neobodo saliens* (formerly *B. saliens*), suggesting that this gene synteny emerged in a common ancestor of kinetoplastids. In addition, we found that *pyr4* and *pyr6* genes in the diplonemid, *Diplonema papillatum*, had a different origin from the kinetoplastid genes. Phylogenetic and gene organization analyses suggested that the standalone *pyr6* in diplonemids might represent a transitional status from the fused *pyr5/pyr6* in euglenoids to the inversely fused *pyr6/pyr5* in kinetoplastids. Our findings provide insights into an evolutionary gap between kinetoplastid and non-kinetoplastid groups in the Euglenozoa.

Results

The *pyr* Gene Cluster in the *P. caudatus* Genome

The *pyr1* and *pyr4* genes are the 5′- and 3′-terminal genes, respectively, of the *pyr* gene