University of South Bohemia
Faculty of Science
Department of Molecular Biology

Master thesis

2010 Petr Růžička

University of South Bohemia

Faculty of Science

Department of Molecular Biology

and

Institute of Parasitology, Academy of Science of the Czech Republic

Laboratory of Molecular Biology of Protists

Mgr. thesis

Experimental examination of EFL and MATX eukaryotic horizontal gene transfers: co-existence of mutually exclusive transcripts predates functional rescue

Bc. Petr Růžička

Supervisor: Prof. RNDr. Julius Lukeš, CSc

České Budějovice, 2010

Ruzicka, P., 2010: Experimental examination of EFL and MATX eukaryotic horizontal gene transfers: co-existence of mutually exclusive transcripts predates functional rescue. Mgr. Thesis, in English. – p., Faculty of Science, The University of South Bohemia, České Budějovice, Czech Republic

Anotation

Aim of this work was testing whether EFL and MATX mRNAs will be compatible with the *T. brucei* transcriptome, which contains the EF-1α and MAT mRNAs. Furthetmore, these two gene couples were examined for their ability to take over the functions of EF1-α and MAT genes.

I declare that I did this work, summarized in this thesis, on my own or in collaboration with co-authors of the presented manuscript, and only using the cited literature.

Prohlašuji, že v souladu s § 47b zákona č. 111/1998 Sb. v platném znění souhlasím se zveřejněním své diplomové disertační práce, a to v nezkrácené podobě – v úpravě vzniklé vypuštěním vyznačených částí archivovaných Přírodovědeckou fakultou - elektronickou cestou ve veřejně přístupné části databáze STAG provozované Jihočeskou univerzitou v Českých Budějovicích na jejích internetových stránkách.

České Budějovice, April 30, 2010 Petr Růžička

CONTENTS

Introduction

Trypanosoma brucei

Trypanosoma brucei belongs to class Kinetoplastida, which includes free-living species, as well as important parasites of human, animals and plants. Based on morfology and life cycles, kinetoplastida were traditionaly subdivided into two suborders – Trypanosomatina (*T. brucei, T. cruzi, T. gambiense, Leishmania donovani, Phytomonas serpens,* etc.) and Bodonina (*Trypanoplasma, Cryptobia, Bodo,* etc.) (Simpson *et al*., 2006). Bodonids, which are composed of families Bodonidae and Cryptobiidae, contains free-living species, parasites and commensals. Their characteristic features are two flagella and a huge kinetoplast DNA. The suborder Trypanosomatina contains only parasitic flagellates including the most studied parasites. In taxonomic system, based on the SSU rRNA genes and protein phylogenies (Simpson and Roger, 2004, Moreira *et al*., 2004), kinetoplastids are devided into Prokinetoplastida and Metakinetoplastida, with the latter subdivided into four orders – Eubodonida, Parabodonida, Neobodonida and Trypanosomatida (Moreira *et al.,* 2004).

Kinetoplast DNA, which is flag feature of the class Kinetoplastida, can contain as many as 40% of total DNA (Lukeš et al., 1998). This unique mitochondrial DNA, usually present as network of thousands of concatenated circular DNA molecules of two types. (Klingbeil et al., 2001). One is minicircles, which are present in thousands of 0,5 to 10 kb-long molecules (depending of species) and dozens of maxicircles, the size of which varies from 20 to 40 kb (Liu *et al.,* 2005). Maxicircles carry typical mitochondrial genes, namaly ribosomal 9S and 12S RNA and several subunits of the respiratory chain (Benne, 1985). Minicircles others genes for guide RNA (gRNA) that function in the editing of maxicircles mRNA transcripts (Sloof and Benne, 1993).

RNA editing is a process of uridine insertion or delerion into RNA molecules, causing changes of the reading frame (Benne *et al.,* 1986).

The replication of kDNA includes the duplication of free detached minicircles and catenated maxicircles, and the generation of two progeny kDNA networks. It is catalyzed by an enzymatic machinery, consisting of kDNA replication proteins that are located at defined sites flanking the kDNA disk in the mitochondrial matrix (Fig.1) (Liu *et al*., 2005). Except kDNA and RNA editing, Kinetoplastida possess a collection of other distinctive features, such as the antigenic variation of surface glycoproteins (Bridgen *et al.,* 1976), polycistronic

transcription (Johnson *et al.,* 1987), massive transplicing (Boothroyd *et al.,* 1982), unique organelles named glycosomes (Opperdoes *et al.,* 1977) and others.

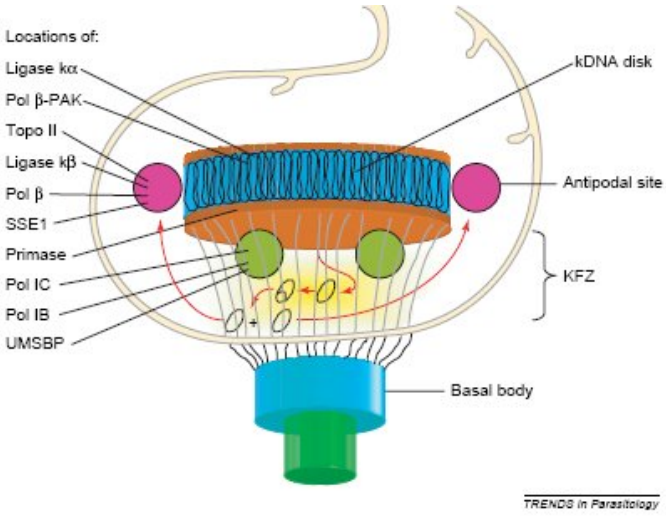

Fig. 1. kDNA replication model (adapter from Liu *et al*., 2005)

Most research is focused on subclass Trypanosomatida, because this taxonomic unit includes genera *Trypanosoma* and *Leishmania.* Into this two species – rich genera belong all medically most imortant parasitic flagellates, that cause tropical and subtropical diseases, such as sleeping sickness, Chagas disease and leishmaniases. *T. brucei brucei* causes African sleeping sickness of wild game and livestock called nagana (Bakker *et al*., 2000), while *T. brucei gambiense* is responsible for chronic form of the disease, afflicting humans that occurs in central and western Africa. People in eastern and southern Africa are faced with *T. brucei rhodesiense*, which causes an acute and, if untreated, invariably fatal form of the disease (Barrett *et al*., 2003). Finally *Trypanosoma cruzi* causes Chagas disease in South America (Fig. 2).

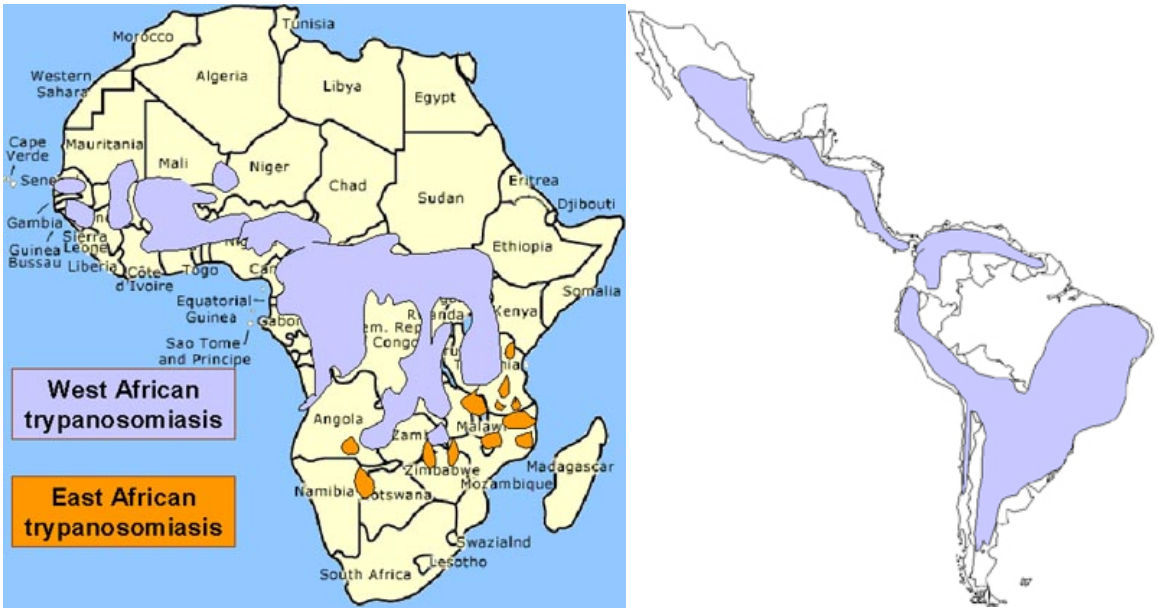

Fig. 2. Distribution of Gambian and Rhodesian sleeping sickness in Africa and Chagas disease in America (adapted from http://pathmicro.med.sc.edu)

Life cycle of *T. brucei*

During its life cycle *T. brucei* switches between two different development stages (Fig. 3). Vehicle of the infection, tse-tse fly (*Glossina*) takes a blood meal containing parasites. Ingested flagellates transform itno procyclic form in the fly's digestive system. Energy metabolism, mainly based on glycolysis, switches to metabolism dependent on a functional mitochonrion (Bowman *et al*., 1972). Procyclic trypanosomes eventually migrate from the hindgut to salivary glands, where they change into epimastigotes. Epimastigotes quickly devide and change once again into metacyclic trypomastigotes, which are ready for release from fly into vertebrate. For their protection from the host's immune system, trypomastigotes produce antigenic variation glycoproteins (Donelson, 2003). This entire process takes about two weeks in *Glossina*. Upon their bite-mediated transmission into the warm-blooded host, metacyclic trypanosomes proliferate in the infection site, which leads to occurrence of a sore. Metacyclic trypanosoma subsequently move in lymphatic system, blood and other tissues and changes into slender bloodstream form, which proliferates profusely. After some time infectious trymastigotes, which are prepared to enter into the tse-tse fly start emerging in the bloodstream (Beinen *et al*., 1991).

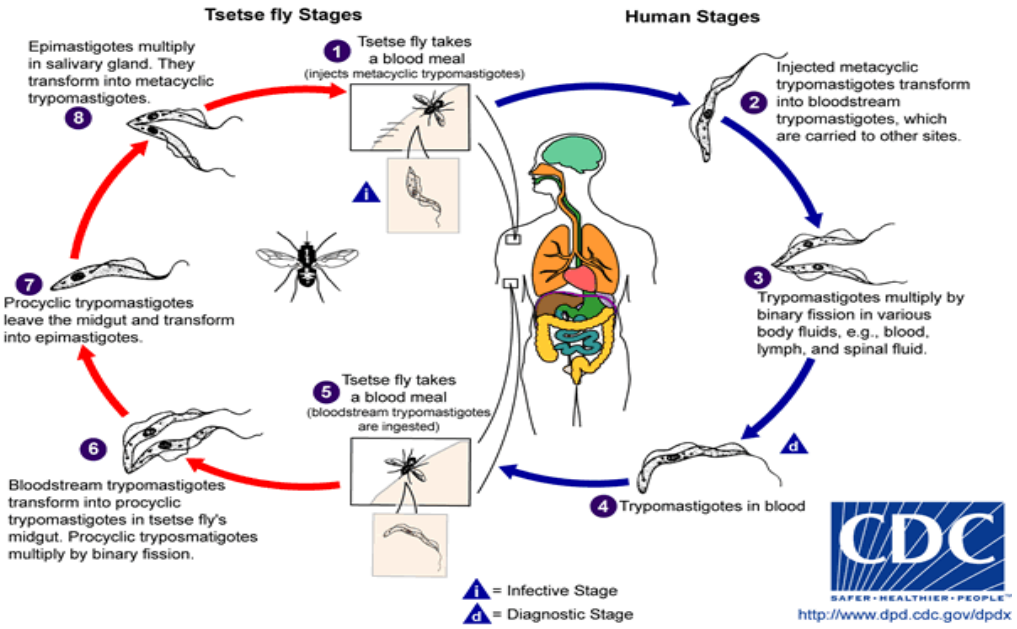

Fig. 3. Life cycle of *T. brucei* (adapted from www.dpd.cdc.gov)

Mat/MatX

Methionine adenosyltransferase (MAT) is an important enzyme that synthetizes S-adenosylmethionine (SAM). SAM is one of the most important metabolites in all known organisms, being involved in key functions such as DNA repair, tRNA modification, DNA precursor biosynthesis and biodegradation pathways (Fontecave *et al*, 2004). However, its most essential function is the addition of methyl group to biological methylation reaction (Cantoni, 1975). SAM is synthesizes in cytosol and is imported to mitochondria and plastids, where it is needed for the methylation of ribulose-1,5-bisphosphate carboxylase/oxygenase (Ying *et al*., 1999). MAT helps to deliver the adenosyl portion of ATP to methionine.

The structure of the MAT complex, which consists of four identical subunits (A, B, C and D) is important for its functions. Two subunits always strongly interact with one another to form a spherical tight dimer, and these dimers associate to form a peanut-shaped tetrameric enzyme. Each dimer has two active sites located between its subunits, with amino acid residues from both subunits contributing to each active site (Takusagawa *et al*., 1996).

The amino acid sequence of MAT is highly conserved among Bacteria and eukaryotes, while the sequence in Archaea is highly divergent (Graham *et al*., 2000). Although MAT is a highly conserve protein, there is evidence, that few eucaryote lineages encode different type of MAT called MATX. Surprisingly the MATX gene is patchily distributed across the

eucaryotic tree (Anderson, 2005). It was also shown that in at least some lineages that possess MATX (f.e. dinoflagellates, euglenids, haptophytes) MAT is absent. Apparently, the gene MAT is mutually exclusive with MATX. However, the diatom *Thalassiosira pseudonana* is the only known organism that possesses genes encoding both MAT and MATX in its nuclear genome (Armbrust *et al.,* 2004). The alignment of MAT and MATX genes contains distinctive features including four insertions and a large number of unique substitutions (Sanchez-Perez *et al*., 2008) (Fig. 4), yet it also shows a high degree of conservation of residues known to play a substantial role in ATP, cation and methionine binding (Gonzales *et al*., 2000).

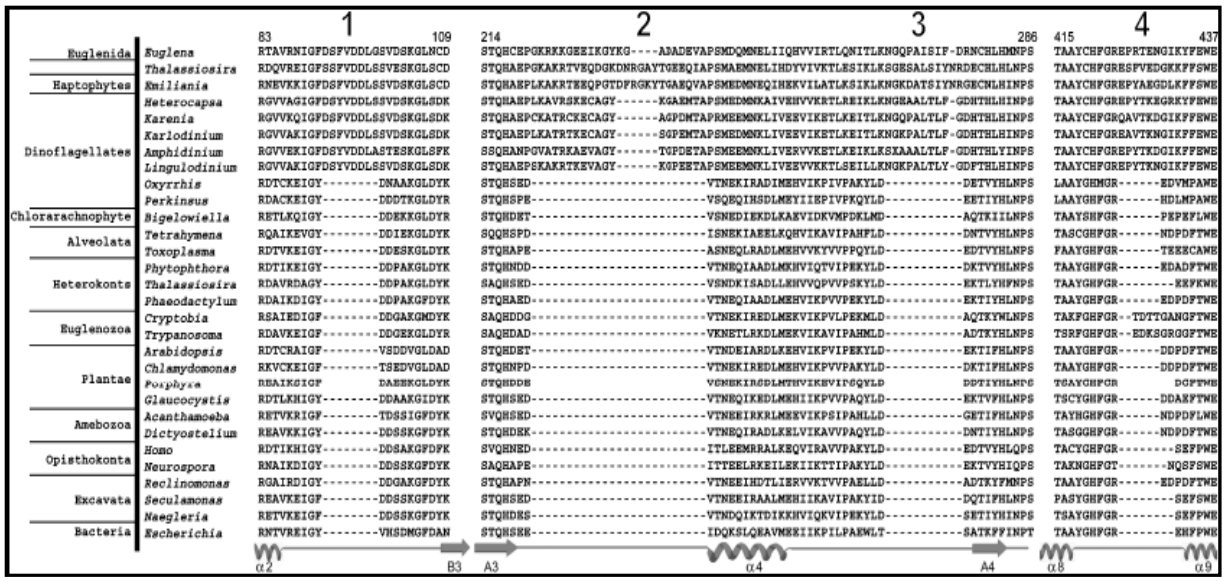

Fig. 4 Regions of the alignment of the methionine adenosyltransferase (MAT) homologs where the four distinctive insertions for MATX are located (adapted from Sanchez-Perez *et al*., 2008)

EF1-α/EFL

Translation elongation factor-1 alpha (EF1-α) is a highly conserved protein in eukaryotes, as obvious from the alignment of amino acids sequences. In particular some domains, such as the putative GTP-binding regions, tRNA-binding regions and the site of actin interaction are very highly conserved (Kaur and Ruben, 1994) (Fig. 5).

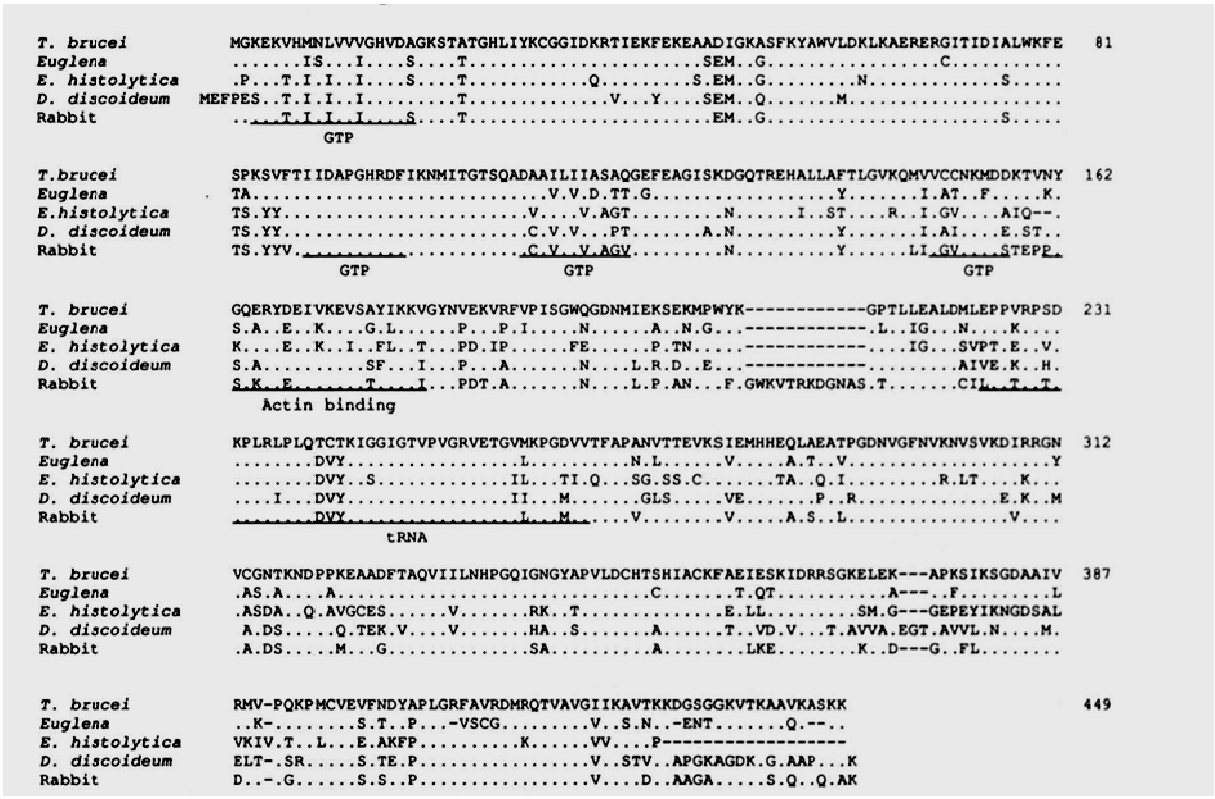

Fig. 5 Alignment of EF-lα from various species (adapted from Kaur and Ruben, 1994)

In Bacteria, was found that the homologue of EF1-α called EF-Tu is encoded by two nearly identical but unlinked genes – tufA and tufB (Jaskunas *et al*., 1975). Both EF1-α and EF-Tu have conserved primary, secondary and tertiary structures. In eukaryotes EF1-α delivers aminoacyl-tRNA (aa-tRNA) to the A site of the ribosome, followed by hydrolyzis of GTP coupled with the release of aa-tRNA. Subsequently, GDP-bound EF-1α interacts with EF-1β (or EF-Ts in eubacteria) to recharge GTP before participating in the second round of peptide elongation process (Anderson *et al*., 2003). In *T. brucei,* EF1-α is involved in Ca^{2+} signal pathway and there is evidence that EF1-α interacts with calmodulin. Furthermore, overexpression of EF1-α was proposed to be important for apoptosis (Lamberti *et al*, 2004). In the vertebrate cells, EF1-α serves as an actin binding protein. In plants EF1-α appears to be implicated in microtubule bundling (Durso and Cyr, 1994), interacting with cytoskeletal proteins, and the ubiquitin-dependent proteolytic system (Chaung *et al*., 2005). Interaction with EF1-α is a prerequisite for tRNA targeting into mitochondria, and it was proposed that this binding determines the specificity of the process (Bouzaidi-Tiali *et al*., 2007).

Elongation factor-like protein is member of the GTPase superfamily involved in translation initiation, elongation and termination, as well as in several other cellular functions.

EFL is exclusively present in eukaryotic groups scattered across the phylogenetic tree. Having a pattern of distribution highly reminiscent of MAT and MATX, most of lineages tend to encode either EF1-α or EFL. Some groups show a tendency to encode EF-1α, (f.e. Euglenozoa) (Gile *et al*., 2009) (Fig. 6) while others (f.e. chlorophytes) rather carry EFL (Noble *et al*., 2007). Such a mutually exclusive distribution of EFL and EF1-α suggests horizontal gene transfer and multiple replacements of EF1-α.

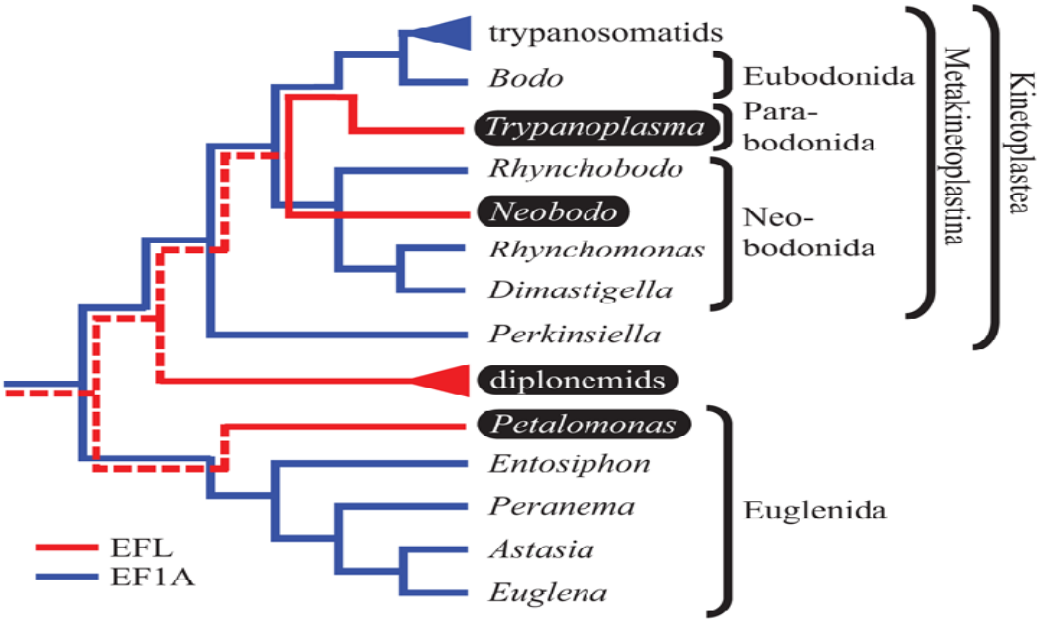

Fig. 6 Schematic tree illustrating currently accepted phylogenetic relationships among euglenozoan taxa (adapted from Gile *et al*., 2009)

Horizontal gene transfer

Horizontal gene transfer (HGT), also known as a latreal gene transfer, is recognized as an important force in eucaryotes evolution. It is movement of genetic information between more or less distantly related organisms. This non-sexual movement of genetic information, can transfer the trait more rapidly than through Darwinian evolution. In opposite to the standard belief in biology that genes are mainly transferred from parental organisms vertically, in the microbial world, this paradigm does not appear to be the best way to explain what occurs. Numerous studies of genes and genomes indicate that multiple HGT events occurred among prokaryotes and less frequently among unicellular and multicellular eukaryotes, however, the importance of HGT in the evolution of multicellular eukaryotes

remains rather obscure (Richardson and Palmer, 2007). Due to the increasing amount of evidence suggesting the importance of these phenomena for evolution, HGT began to be considered as a new paradigm for biology.

HGT is considered as a significant evolutionary driver of eukaryotic evolution. HGT was proposed to have played a considerable role in the origin of the cytoskeleton and endocytosis, which enable eukaryotes to engulf and feed on other cells (Stanier, 1970). According to the endosymbiotic theory mitochondria descended from specialized bacteria (probably purple nonsulfur bacteria) that somehow survived endocytosis by another species of prokaryote or some other cell type and became incorporated into the cytoplasm of a (future) eukaryotic cell. Similar scenario is highly plausible for plastids which arose from either a single, but more likely several ancient endosymbioses between a heterotrophic eukaryote and a photosynthetic cyanobacterium (Martin *et al*., 2002, Curland and Anderson, 2000).

There is evidence that only a handful of genes is encoded by organellar genomes (Gray *et al*., 1999). Genes encoding most organellar proteins are located in the nucleus, which means that an extensive transfer took place among cell's different genomes. Most of these originally organellar genes acquired targeting signals upon their transfer to nucleus, allowing their import into the organelle (Bhattacharya *et al*., 2007).

Generally, there are four categories of techniques for HGT prediction: Codon - based approaches in which HGT is detected by looking for gene that have a G+C content and or codon biases that differ from that of the host (Kaplan and Fine, 1998). BLAST-based approaches allow fast searches of a given protein to a query protein. If the analyzed protein has greater similarity to a homoloque from a distant organism, then it may be a cause of HGT. This approach was used to examine the origin of proteins from *Thalassiossira pseudonana*, which acquired its plastid via a secondary endosymbiosis event (Armbrust *et al*., 2004). Another way of identifying HGT rests on gene–distribution–based approaches that look for genes unevenly distributed between related species. Finally, phylogenetic approaches are based on comparison of the phylogenetic tree (Sicheritz – Pontent and Andersson, 2001). If an organism is found in an unexpected part of the tree, than it is a strong indication of HGT.

References

Simpson L., Aphasizhev R., Gao G., Kang X. 2006. Mitochondrial proteins and complexes in *Leishmania* and *trypanosoma* involved in U-insertion/deletion RNA editing. *RNA*. 10:159 – 170

Simpson A. G. B., Roger A. J. 2004. The real 'kingdoms' of eukaryotes. *Current Biology*. 14: R693-696

Moreira D., Lopez-Garcia P., Vickerman K. 2004.. An updated view of kinetoplastid phylogeny using environmental sequences and a closer outgroup: proposal for a new classification of the class Kinetoplastea. *Int. J. Syst. Evol. Microbiol.* 54: 1861-1875

Lukeš J., Jirků M., Avliyakulov N., Benada O. 1998. Pakinetoplast DNA structure in a primitive bodonid flagellate, *Cryptobia helicis*. *EMBO J*. 17:838 – 846

Klingbeil M. M., Drew M. E., Liu Y., Morris J. C., Motyka S. A., Saxowsky T. T., Wang Z., Englund P. T. 2001. Unlocking the secrets of trypanosome kinetoplast DNA network replication. *Protist*. 152:255 – 262

Liu B., Liu Y., Motyka S., Agbo E., Englund P. 2005. Fellowship of the rings: the replication of kinetoplast DNA. *Trends in Parasitology. 21*:363-369

Benne R. 1985. Mitochondrial genes in trypanosomes. *TIG*. 117 – 121

Sloof P., Bene R. 1993. RNA editing in trypanosome mitochondria: guidelines for models. *FEBS Lett*. 326:146 – 151

Benne R., Van Den Burg J., Brakenhoff J. P. J., Sloof P., Van Boom J. H., Tromp M. C. 1986. Major transcript of the framesshifted coxII gene rfom trypanosome mitochondria contains four nucleotides that are not encoded in the Dna. *Cell* 46:819 – 826

Bridgen P. J., Cross G. A. M., Bridgen J. 1976. N-terminal amino acid sequences of variant-specific surface antigens from *Trypanosoma brucei*. *Nature*. 263:613 – 614

Johnson P. J, Kooter J. M., Borst P. 1987. Inactivation of transcription by UV irradiation of T. brucei provides evidence for a multicistronic transcription unit including a VSG gene. *Cell*. 51(2):273–281

Boothroyd J. C., Paynter C. A., Coleman S. L., Cross G. A.. 1982. Complete nucleotide sequence of complementary DNA coding for a variant surface glycoprotein from *Trypanosoma brucei*. *J. Mo.l Biol*. 157(3):547-56.

Opperdoes F. R., Borst P. 1977. Localization of nine glycolytic enzymes in a microbody-like organelle in Trypanosoma brucei: the glycosome. *FEBS Lett*. 80(2):360–364

Bakker M. B., Westerhoff H. V., Opperdoes F. R., Michels P. A. M. 2000. Metabolic control analysis of glycolysis in trypanosomes as an approach to improve selectivity and effectiveness of drugs. *Mol. Biochem. Parasitol*. 106:1 – 10

Barrett M. P., Burchmore R. J., Stich A., Lazzari J. O., Frasch A.C., Cazzulo J.J., Krishna S. 2003. The trypanosomiases. *Lancet*. 362:1469-80

Bowman I. B. R., Srivastata H. K., Flynn I. W. 1972. Adaptations in oxidative metabolism during the transformation of *trypanosoma rhodesiense* from bloodstream into culture form. *Comparative biochemistry of Parasites.* Academic Press New York NY. 329 – 342

Donelson J. E. 2003. Antigenic variation and the African trypanosome genome. *Acta Trop*. 85(3):391 – 404

Beinen E. J., Saric M., Pollakis G., Grady R. W., Clarkson A. B. Jr. 1991. Mitochondrial development in *Trypanosoma brucei* transitional bloodstream forms. *Mol. Parasitol*. 45(2):185 – 192

Fontecave M., Atta M., Mulliez E. 2004. S-adenosylmethionine: nothing goes to waste. *Trends Biochem. Sci.* 29:243-249

Cantoni G. L. 1975. Biological methylation: selected aspects. *Annu. Rev. Biochem.* 44:435-451

Ying Z., Mulligan R. M., Janney N., Houtz R. L. 1999. Rubisco small and large subunit N-methyltransferases. Bi-and mono-functional methyltransferases that methylate the small and large subunits of Rubisco. *J. Biol. Chem.* 274:36750-36786

Takusagawa F., Kamitori S., Misaki S., Markham G. D. 1996. Crystal structure of S-adenosylmethionine synthetase. *J. Biol. Chem.* 271:136-147

Graham D. E., Bock C. L., Schalk-Hihi C., LU Z. J., Markham G. D. 2000. Identification of a highly divergent class of-S-adenosylmethionine synthetases in the archea. *J. Biol. Chem.* 275:4055-375

Andersson J. O. 2005. Lateral gene transfer in eukaryotes. Cell. Mol. Life. Sci. 62:1182-1197

Armbrust E. V., Berges J. A., Bowler C., Green B. R., Martinez D., Putnam N. H., Zhou S., Allen A. E., Apt K. E., Bechner M. *et al*. 2004. The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science*. 306:79–86

Sanchez-Perez G. F., Hampl V., Simpson A. G. B., Roger A. J. 2008. A new divergent type of eucaryotic methionine adenosyltransferase is present multiple distantly secondary algal lineages. *J. Eucaryot. Microbiol.* 55:374-381

Gonzales B., Pajares M. A., Hermoso J. A., Alvarez L., Garrido F., Sufrin J. R., Sanz-Aparicio J. 2000. The crystal structure of tetrameric methionine adenosyltransferase from rat liver reveals the methionine-binding site. *J. Mol. Biol.* 300:363-375

Kaur K. J., Ruben L. 1994. Protein translation elongation factor-1α trom *Trypanosoma brucei* binds calmodulin. *J. Biol. Chem.* 27:23045-23050

Jaskunas S. R., Lindahl L., Nomura M., Burgess R. R. 1975. Identification of two copies of the gene for the elongation factor EF-Tu in *E. coli*. *Nature*. 257:458-462.

Andersson J. O., Sjogren A. M., Davis L. A., Embley T. M., Roger A. J. 2003. Phylogenetic analyses of diplomonad genes reveal frequent lateral gene transfers affecting eukaryotes. *Curr. Biol.* 13:94–104.

Lamberti A., Caraglia M., Longo O., Marra M., Abbruzzese A., Arcari P. 2004. The translation elongation factor 1A in tumorigenesis, signal transduction and apoptosis: review article. *Amino Acids*. 26: 443–448

Durso N. A., Cyr R. J. 1994. A calmodulin-sensitive interaction between microtubules and a higher plant homolog of elongation factor-1 alpha. *Plant Cell*. 6:893-905

Chuang M. S., Madura K. 2005. *Saccharomyces cerevisiae* Ub-Conjugating Enzyme Ubc4 Binds the Proteasome in the Presence of Translationally Damaged Proteins. *Genetics*. 171:1477-1484

Bouzaidi-Tiali N., Aeby E., Charriere F., Pusnik M., Schneider A. 2007. Elongation factor 1a mediates the specificity of mitochondrial tRNA import in *T. brucei. The EMBO Journal.* 26:4302–4312

Gile G. H., Faktorová D., Castlejohn C. A., Burger G., Lang B. F., Farmer M. A., Lukeš J., Keeling P. J. 2009a. Distribution and phylogeny of EFL and EF-1α in Euglenozoa suggest ancestral occurrence followed by differential loss. *PloS One*. 4:e5162

Noble, G. P., Rogers, M. B., Keeling, P. J., 2007. Complex distribution of EFL and EF-1a proteins in the green algal lineage. *BMC Evol. Biol*. 7:82

Richardson A. O., Palmer J. D. 2007. Horizontal gene transfer in plants. *J. Exp. Bot.* 58:1–9

Stanier R. Y. 1970. Some aspects of the biology of cells and their possible evolutionary significance. *Symp. Soc. Gen. Microbiol.* 20:1–38

Martin, W. *et al.* 2002. Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc. Natl Acad. Sci. USA* 99:12246–12251

Kurland C. G., Andersson S. G. 2000. Origin and evolution of the mitochondrial proteome. *Microbiol. Mol. Biol. Rev.* 64:786–820

Gray W. M., Burger G., Lang F. B. 1999. Mitochondrial Evolution. *Science*. 283:1476 - 1481

Bhattacharya D., Archibald J. M., Weber A. P., Reyes-Prieto A. 2007. How do endosymbionts become organelles? Understanding early events in plastid evolution. *Bioessays.* 29:1239–1246

Kaplan J. B., Fine, D. H. 1998. Codon usage in *Actinobacillus actinomycetemcomitans*. *FEMS Microbiol. Lett*. 163:31–36

Armbrust E. V., Berges J. A., Bowler C., Green B. R., Martinez D., Putnam N. H., Zhou S., Allen A. E., Apt K. E., Bechner M. *et al*. 2004. The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science*. 306:79–86

Sicheritz-Ponten T., Andersson S. G. 2001. A phylogenomic approach to microbial evolution. *Nucleic Acids Res.* 29:545–552

Experimental examination of EFL and MATX eukaryotic horizontal gene transfers: co-existence of mutually exclusive transcripts predates functional rescue

Research article

Petr Růžička[1,#], Jana Szabová[2,#], Zdeněk Verner[1], Vladimír Hampl[2] & Julius Lukeš[1,*]

[1]Biology Centre, Institute of Parasitology, Czech Academy of Sciences, and Faculty of Sciences, University of South Bohemia, 370 05 České Budějovice (Budweis), Czech Republic; and [2]Faculty of Science, Charles University in Prague, 128 44 Prague, Czech Republic

[#] Both authors contributed equally to this work

Corresponding author:

Julius Lukeš

Biology Centre, Institute of Parasitology, Czech Academy of Sciences, 370 05 České Budějovice (Budweis), Czech Republic

Phone: +4202387775416

Fax: +420385310388

e-mail: jula@paru.cas.cz

Keywords: EFL, MATX, horizontal gene transfer, functional rescue, RNAi

Running head: Experimental eukaryotic HGT of EFL and MATX

ABSTRACT

Many eukaryotic genes do not follow simple vertical inheritance. Elongation factor 1α (EF-1α) and methionine adenosyl transferase (MAT) belong to enzymes with complicated evolutionary histories and, interestingly, the two cases have several features in common. These essential enzymes come in the form of two relatively divergent paralogs (EF-1α/EFL, MAT/MATX) that have patchy distribution in eukaryotic lineages that is mutually exclusive. To explain such an occurrence, we must invoke either multiple eukaryote-to-eukaryote horizontal gene transfers (HGTs) followed by functional replacement, or presence of both paralogs in the common ancestor followed by long term co-existence and differential losses in various eukaryotic lineages. To model the evolution of these protein couples, we have performed *in vivo* experiments in *Trypanosoma brucei* revealing the consequences of long term co-expression and functional replacement for each gene couple. In the first experiment of its kind, we have demonstrated that EF-1α and MAT can be simultaneously expressed with EFL and MATX, respectively, with the co-expression of EF-1α and EFL being slightly harmful for the trypanosomes. After the endogenous MAT or EF-1α was down-regulated, MATX immediately substituted for its paralog, while EFL was not able to substitute EF-1α leading to flagellate mortality. We conclude that MATX is naturally capable of evolving patchy paralog distribution via HGTs and/or long term co-expression and differential losses. The capability of EFL to pass through such processes is lower and its evolutionary history thus remains rather enigmatic.

INTRODUCTION

The transfer of genetic information among distantly related organisms called horizontal (= lateral) gene transfer (HGT) represents one of major driving forces of evolution (Keeling and Palmer 2008). A pervasive occurrence of HGT among prokaryotic organisms is well reflected in their genomes and can be easily experimentally executed, thus disputing the actual existence of bacterial species (Welch et al. 2002; Doolittle and Papke 2006; Doolittle and Bapteste 2007). Recently, the role of HGT is becoming recognized also as an important force in the evolution of eukaryotes and an increasing number of examples are being reported (Andersson 2005; Watkins and Gray 2006; Richards et al. 2006; Whitaker et al. 2009). Even

though several mechanisms have been proposed (Gogarten 2003), two are believed to prevail: endosymbiotic gene transfer (Martin and Schnarrenberger 1997) and "you are what you eat" (Doolittle 1998). The former occurred upon endosymbiosis of the bacterial ancestors of mitochondria and plastids and represent HGT on the largest scale, since more than 90% of the endosymbiosed genomes were subsequently transferred to the host cell nucleus, and thus form a substantial part of the coding capacity of the nuclei in extant eukaryotes (Esser et al. 2004). Most of these organelle-derived proteins remain functionally associated with the organelle of their evolutionary origin (Kurland and Andersson 2000). However, some of these proteins have eventually found way to other cellular compartments, being responsible for the mosaic pattern of most metabolic pathways in a typical eukaryotic cell (Gabaldón and Huynen 2004; Oborník and Green 2005).

Independently of these massive endosymbiosis-driven HGT events, most eukaryotes were subject to intermittent acquisitions of genomic material from prokaryotes or other eukaryotes. According to the "you are what you eat" concept, the digested prey is the pervasive source of these transferred genes (Doolittle et al. 1998). The extent of the prokaryote-derived process seems to be strongly influenced by the life style and environment of the recipient, as HGTs occurred frequently in the evolutionary history of rumen ciliates (Ricard et al. 2006), anaerobic protists (Andersson et al. 2007) or diatoms (Bowler et al. 2008), whereas very few such events have so far been documented in yeasts (Dujon et al. 2004) or animals (Kondrashov et al. 2006).

Although the eukaryote-to-eukaryote HGTs are considered to be currently underestimated (Keeling and Palmer 2008), it would be daring to consider this process a frequent one. Particularly interesting are highly conserved essential genes that have a surprisingly complex evolutionary history – elongation factor 1α (EF-1α) and methionine adenosyltransferase (MAT) (Keeling and Inagaki 2004; Gile et al. 2006; Gile et al. 2009a; Gile et al. 2009b; Sanchez-Perez et al. 2008; Kamikawa et al. 2009), being the best known examples. The main function of EF-1α is to bring an aminoacyl-tRNA into the A site of the ribosome (Andersen et al. 2003). Furthermore, this extremely abundant protein has been implicated in ubiquitin-dependent protein degradation (Chuang et al. 2005) and localization of selected transcripts via simultaneous binding of EF-1α to actin (Liu et al. 2002). MAT is the only enzyme synthesizing S-adenosyl-L-methionine, which is one of the key metabolites, as it donates the methyl group for most methylation reactions in prokaryotic and eukaryotic cells (Chiang et al. 1996).

While these genes were, same as most other genes, considered to have evolved by vertical descent, the respective phylogenetic trees were inconsistent with such a simple scenario. In both cases, the analyses revealed that subsets of unrelated organisms possess divergent version of the gene - elongation factor like (EFL) and MATX (Keeling and Inagaki 2004; Sanchez-Perez et al. 2008). In general, the patchy distribution of two paralogs can be explained either by i) deep paralogy – presence of both paralogs in the common ancestor – followed by differential retaining of only one variant in individual lineages, ii) more recent origin of one paralog in one lineage of eukaryotes followed by its spread by eukaryote-to-eukaryote HGT or iii) mixture of both. It is difficult to distinguish between these alternatives purely on the basis of phylogenetic analyses of protein sequences and their distributions. Theoretically, since the scenarios differ in the types of events they invoke, we might use the principle of parsimony and choose the most probable path through the tree avoiding the less probable events. While the first scenario minimizes the events of HGT replacements and expects long term co-existence of both paralogs, the second one assumes exactly the opposite. Unfortunately, in the cases of EFL/EF-1α and MATX/MAT we have no idea what the values of these parameters are, i.e. what is the relative probability of long term co-existence of paralogous proteins in one organism without the unfavorable effect on its fitness versus an instant successful HGT replacement.

The co-existence of these paralogs under natural conditions is rare if not totally absent. The distribution of EFL/EF-1α paralogs and MATX/MAT paralogs is almost strictly exclusive, i.e. organisms have either EF-1α or EFL but not both, the same applying to MAT and MATX. Strangely enough, the exceptional group in both cases is the diatoms. The genome of *Thalassiosira pseudonana* harbors both variants of these genes (EFL, EF-1α and MATX), and the dual status of the latter couple was uncovered also in four other diatoms (Kamikawa et al. 2009). The virtually exclusive distribution led to a proposal that the long term co-existence of both variants in one cell and one compartment is deteriorating for the cell, probably due to problems with regulation, competition for substrate or, in the case of MAT/MATX, formation of less functional heteromers (Sanchez-Perez et al. 2008).

The process of HGT replacement of these essential proteins by their paralogs is, however, potentially problematic as well. It is simply difficult to envisage a smooth switch, during which these multifunctional proteins are replaced with a horizontally acquired paralogs that take instantly over orphan functions of the lost proteins. Moreover, the replacement is inevitably preceded by the hazardous period of co-expression of both variants. Any selective

advantage of such a risky acquisition remains to be established, since the given gene and its paralog seem to function equally well in different distantly related lineages.

We have decided to experimentally test the feasibility of long term co-expression and HGT replacement of each gene couple (EFL/EF-1α and MAT/MATX) in the model of *Trypanosoma brucei*. For the first time, we have simulated the whole process under laboratory conditions. We have shown that EFL and MATX can co-exist with EF-1α and MAT, respectively. Moreover, the MATX gene of *Euglena gracilis* was able to rescue the RNA interference knock-down for MAT in this flagellate. Our experiments unveil the outlines of the evolutionary histories of these genes.

MATERIALS AND METHODS

EF-1α and EFL plasmid constructs

Oligonucleotides for generation of 400-500 bp gene fragments for RNAi knock-down cell lines were designed using the RNAit online tool available on the TrypanoFAN website (http://trypanofan.path.cam.ac.uk/ software/RNAit.html). The 453 bp-long 5' region of the *T. brucei* EF-1α gene was amplified using oligonucleotides EF-1-F (5'-GGATCCTGGAGGCACTAGACATGCTG-3') and EF-1-R (5'-AAGCTTCGATCTTCGACTCGATCTCC-3') (added BamH*I* and Hind*III* restriction sites are underlined) and cloned via the restriction sites into the p2T7-177 RNAi vector carrying phleomycin resistance, creating a construct, which was upon linearization by Not*I* electroporated into the *T. brucei* strain 29-13 as described elsewhere (Hashimi et al. 2008).

For constitutive expression of the EFL gene, the entire open reading frame (NCBI accession number AAV34146) was PCR-amplified from the total DNA of the haptophyte *I. galbana* using oligonucleotides Ig1-F (5'-AAGCTTAT GGCCTCCGAGAAA-3') and Ig1-R (5'GGATCCCTACTTCTTCTTCTT-3') (added Hind*III* and BamH*I* restriction sites are underlined). The amplified fragment was cloned into pCRII TOPO (Invitrogen) and subcloned into the pABPURO vector containing the puromycin resistance (Long et al. 2008). The ensuing construct Ig-EFL or empty pABPURO were electroporated in the *T. brucei* 29-13 cells or in the cell line already containing the EF-1α gene in the inducible p2T7-177 RNAi vector.

MAT and MATX plasmid constructs

To generate the RNAi knock-down cells for the MAT gene, its 438 bp-long 5' fragment was amplified using oligonucleotides IF-F (5'- TCACTCTAGAAC GACGGTGTGTCAAATGAA -3') and IF-R (5'- AGTGAAGCTTGCAGTC GGAAGTTTTTCTGC -3') (added Xba*I* and Hind*III* restriction sites are underlined), cloned into the RNAi vector, and electroporated into trypanosomes as described above for the EF-1α gene. Furthermore, the full-size MATX gene (accession number GU989640) was amplified from the EST clone of the euglenid *E. gracilis* by amplification using oligonucleotides RE-F (5'- TCACATCGATATGGCTGAATCTGCTTC -3') and RE-R (5'- AGTGGCTAGCGTCCACCCACTTCTGCA -3') (added Nhe*I* and Cla*I* restriction sites are underlined). The amplicon was cloned into the pABPURO vector as described above, creating the construct Eg-MATX. Moreover the pABPURO vector contained a HA_3-tag attached to the 3' end of MATX insert. The following clonal cell lines derived from the 29-13 strain were made: i/ RNAi knock-downs containing MAT gene in the p2T7-177 vector; ii/ cells constitutively overexpressing HA_3-tagged MATX from the pABPURO vector; iii/ cells co-transfected with the p2T7-177-MAT and the pABPURO-MATX constructs.

Transfection, cloning and RNAi induction

HA_3-tagged MATX in the pABPURO vector was digested with Mlu*I*. The inducible p2T7-177 RNAi vector containing MAT gene was digested with Not*I* and 10 μg of linearized DNA was transfected into exponentially growing (at 27^{o}C in SDM-79 medium) *T. brucei* 29-13 strain or cell lines derived from thereof, using 2-mm cuvettes and a BTX electroporator with the settings of 1600 V, 25 μfarads and 500 ohms. After transfection, the clones were obtained after about 2-week cultivation by limiting dilution in 24-wells plates at 27 ^{o}C in the presence of 5% CO_2, with 1 μg/ml puromycin used as the selectable agent. Synthesis of double stranded (ds) RNA was induced by the addition of 1 μg/ml tetracycline to six clonal cell lines derived from each electroporation. For further experiments a single clone was selected based on the tightness of tetracycline expression of dsRNA and the robust elimination of target mRNA, as determined by Northern blot analysis using the EF-1α gene fragment as a probe.

Northern blot and growth analysis

Approximately 5 μg of total RNA/lane was loaded on a 1% formaldehyde agarose gel, blotted and cross-linked following standard protocols. After pre-hybridization in NaPi solution

(0.25M Na_2HPO_4 and 0.25M NaH_2PO_4, pH 7.2, 1 mM EDTA, 7% SDS) for 30 min at 50 ^oC to 55 ^oC (depending on the probe; see Results), hybridization was performed overnight in the same solution at 55 ^oC. A wash in 2x SSC + 0.1% SDS at room temperature (RT) for 20 min was followed by three washes in 0.2x SSC + 0.1% SDS for 20 min each at 55 ^oC. Growth curves of selected clones representing the RNAi knock-downs and the other genetically manipulated cells, obtained over a period of 12 days after RNAi induction were established using the Beckman Z2 Cell Counter.

Western blot analysis

Cell lysates corresponding to 2.5 x 10^6 cells/lane were separated on a 15% SDS-PAGE gel, blotted and the membranes were treated with anti-HA_3-tag mouse monoclonal antibody, followed by chicken anti-mouse antibody coupled to horse-radish peroxidase. Western blot bands were quantified with software Bio-Rad quantity one.

RESULTS

EFL and EF-1α mRNAs are not fully compatible

We have first tested whether an EFL mRNA will be compatible with the *T. brucei* transcriptome, which contains the EF-1α mRNA. For that purpose, the full-size EFL gene from distantly related haptophyte *Isochrysis galbana* was cloned into a vector that allows constitutive expression under the procyclin promoter of the inserted gene. The Not*I*-linearized vector was subsequently transfected into the 29-13 *T. brucei* strain where it integrates into the tubulin locus. Total RNA was isolated from a puromycin-resistant clonal cell line and analyzed by Northern blots using the full-size EFL gene as a probe. The analysis showed that in all clones the introduced gene was strongly transcribed, while no signal was detected in the parental 29-13 cells (Fig. 1A). At stringent hybridization temperature (65 ^oC), no cross-hybridization with the EF-1α mRNA, transcribed from three endogenous copies of the EF-1α gene, was observed (Fig. 1A). The morphology of EFL-containing *T. brucei* appeared normal by light microscopy, but their growth was significantly slower as compared with the wild type cells (Fig. 2B, squares) indicating that the co-expression of EF-1αand EFL hinders the growth to some extent.

Silencing of EF-1α inhibits growth

The addition of tetracycline into the medium triggers synthesis of double stranded (ds) RNA in trypanosomes transfected with the EF-1α inducible RNAi construct. The extent of EF-1α mRNA silencing and the tightness of its inducible ablation was determined by Northern blot analysis of total RNA collected from the non-induced cells containing or lacking exogenous EFL (Fig. 1B; lanes 2 and 4) and the same cells 48 hrs after RNAi induction (Fig. 1B; lanes 3 and 5). The hybridization temperature was lowered to 60 ^{o}C in order to detect the EF-1α and EFL transcripts in parallel. Indeed, the target ~ 1.6 kb-long EF-1α mRNA was undetectable 48 hrs after RNAi induction, whereas the level of ~ 1.4 kb-long EFL mRNA originating from *I. galbana* was unaltered in the interfered cells (Fig. 1C; lane 3). Northern analysis thus showed that both transcripts are about equally abundant in the non-induced cells and that RNAi is highly specific, targeting only the EF-1α mRNA (Fig. 1C). Next, we have followed the growth of the cells interfered against EF-1α as well as cells differing from them by the expression of EFL. As expected, the elimination of EF-1α mRNA triggers an almost instant cessation of growth, eventually causing death regardless of the absence (Fig. 2A, triangles) or presence (Fig. 2B, triangles) of EFL that apparently fails to rescue the phenotype.

MAT and MATX mRNAs are compatible

We have used the same strategy for the MAT and MATX system as described above for the EFL and EF-1α genes. First, we have shown that in transfected *T. brucei* the MATX gene from *E. gracilis* is indeed expressed, using Northern blot analysis and the MATX gene as a probe (Fig. 3A). This approach also excluded the unlikely yet possible presence of another MATX gene in the parental 29-13 cells, which contain 9 copies of the MAT gene in their genome (Fig. 3A; lane 1). Growth curve analysis of parental cells and those overexpressing MATX clearly demonstrated that trypanosomes fully tolerate expression of this exogenous gene (Fig. 4B; squares).

In the case of MATX we wanted to address the unlikely possibility that although the gene is transcribed (Fig. 3A), the MATX mRNA is not translated or the protein is instantly degraded. Due to the lack of specific antibodies we have attached an HA_3 tag to the 3' end of the gene, in order to follow its expression. As shown by Western blot analysis using specific

anti-HA_3 tag monoclonal antibody, MATX is not only transcribed but also translated in *T. brucei* transfected with the respective construct (Fig. 3C).

MATX rescues MAT deficiency

Next, we have down-regulated MAT mRNA in *T. brucei* using RNAi. In selected clones, the ablation was very efficient, since after 48 hrs of RNAi induction, virtually no target transcript was detectable by Northern blot analysis (Fig. 3B; lanes 3 and 5). We have then followed the consequences of such depletion on cell growth. As shown in Fig. 4A (triangles), MAT is clearly an essential protein for trypanosomes, as they were unable to propagate in its absence. The situation was, however, strikingly different with cells depleted for MAT but overexpressing MATX. The constitutive expression of the euglenid MATX in the inducible *T. brucei* RNAi MAT knock-down fully rescued the growth, which did differ neither from the 29-13 parental cells, nor from the non-induced knock-downs (Fig. 4B, triangles). This experiment shows that MATX, concurrently expressed with MAT, following the depletion of the latter protein takes quickly over its function(s) and rescues the cells, which would otherwise die.

DISCUSSION

Using *in vivo* experiments we have successfully mimicked the acquisition of an exogenous paralog of an enzyme by HGT followed by a period of simultaneous expression, resulting in functional replacement of the endogenous paralog. Our experiments have several important implications for the evolution of the studied paralogs, as well as for the process of HGT in general.

Using two gene couples (MAT/MATX and EF-1α/EFL) that are virtually never found in nature together we have demonstrated for the first time that their transcripts can co-habitate in *T. brucei* at least for several weeks, which is a substantial period for an organism with eight hours-long generation time. While our results do not attach any selective advantage to this highly risky and cumbersome process, they show that the process can indeed happen under experimentally controlled conditions. The proteins can co-exist for a sufficiently long period, enabling the critical switch from one protein to another or, theoretically, enabling a long term

co-habitation followed by slow and more or less random differential losses of one or the other in various lineages of organisms.

However, trypanosomes co-expressing MAT+MATX or EF-1α+EFL differed in the growth phenotype. While the growth rate of the former was comparable with the wild type cells, the latter trypanosomes grew significantly slower. The co-existence of EF-1α and EFL apparently represents a certain burden. These findings correspond to the observations from nature. While the transcripts of both MAT and MATX were detected in several diatom species indicating the rare yet still possible co-expression (Kamikawa et al. 2009), there is so far no natural example of co-expression of EF-1α and EFL. Although both genes were reported from the genome of *T. pseudonana*, our search of *Thalassiosira* ESTs database (http://avesthagen.sznbowler.com) revealed only EFL transcripts (67 reads) indicating that solely EFL was expressed at the time of mRNA isolation. Still, our results show that the cells can survive the parallel presence of EF-1α and EFL.

Since horizontally transferred genes may be disadvantaged in codon usage, we have compared this parameter for both gene couples. Codon usage of both MATX and EFL departs from the kinetoplastid consensus to about the same extent as the endogenous MAT and EF-1α (Supplementary file 1). We therefore conclude that codon usage bias plays no significant role in establishment of horizontally acquired MATX and EFL.

MATX and EFL differ in their ability to substitute the endogenous paralog in trypanosomes. After down-regulation of MAT by RNAi, the expressed MATX was able to take over the function of its counterpart, whereas cells with EFL but depleted for EF-1α died. The functional replacement of MAT by MATX happened immediately with no effect on the growth phenotype. This result is in agreement with the work of Ho et al. (Ho et al. 2007), who showed that MATX from the dinoflagellate *Crypthecodinium cohnii* rescued the MAT knock-out of yeast.

Judging by their phylogenies and distribution among eukaryotes, the cases of EF-1α/EFL and MAT/MATX look very similar, yet our laboratory experiments indicate that this similarity may be only superficial. In the case of MAT/MATX both long-term co-existence and functional replacements are plausible. The experimental data indicate that mechanisms able to create the patchy phylogenetic distribution of MAT/MATX are indeed functional for these genes, however do not allow judging which scenario (deep paralogy or recent HGTs) is more probable. In the case of EF-1α/EFL we have demonstrated that the co-existence of both

variants is possible but, at least in our system, represents certain burden for the organism, which would certainly lead to a fast outcompeting of the transformed lineage in the real world. So, it is difficult to imagine how this dual state could be maintained for millions of year in euglenids as the hypothesis of Gile et al. (2009a) postulates, or even for much longer time as the deep-paralogy scenario assumes. However, the functional replacement of EF-1α by EFL was also not successful, so we cannot corroborate the hypothesis that the EFL gene experienced multiple HGT replacements in its evolutionary past. One solution for the corundum of EF-1α/EFL evolutionary history could be the existence of an as yet unknown cofactor essential for the EFL function that must have accompanied EFL during the HGTs. This putative cofactor would likely be located physically close to EFL in the genome. Finally we cannot exclude the possibility that the failure to perform laboratory HGT of EFL may result from unsuitability of *Trypanosoma brucei* as a laboratory host species for such transfer.

In summary, the process of HGT and functional replacement of a paralog by another one was simulated in a step-by-step fashion. For the first time we have demonstrated that two relatively divergent variants of an essential protein can be co-expressed *in vivo*. A trouble-free simultaneous expression facilitates the successful HGT replacement of paralogs and represents a necessary prerequisite for scenarios invoking deep paralogy and differential losses to explain the complex distribution of paralogs in the eukaryotic tree. MAT/MATX exhibit natural capability to generate a patchy distribution, as these two proteins can co-exist without marked phenotypic consequences, with MATX being able to instantly substitute for MAT. The evolutionary history of EF-1α/EFL remains more enigmatic as we have shown that their co-expression is possible, but represents a burden for trypanosomes, and our attempts to functionally replace EF-1α by EFL failed.

ACKNOWLEDGEMENTS

We would like to acknowledge Gabino Sanchez-Perez and Dion G. Durnford for kindly providing the *E. gracilis* MATX clone, and Bryony Williams and Patrick Keeling for sharing the *I. galbana* EFL clone. This work was supported by the Grant Agency of the Czech Republic 204/09/1667 (to J.L.), the Grant Agency of the Charles University 63409 (to J.S.), the Ministry of Education of the Czech Republic (2B06129, 6007665801 and 0021620828)(to J.L. and V.H.), and the Praemium Academiae award (to J.L).

LITERATURE CITED

Andersen GR, Nissen P, Nyborg J. 2003. Elongation factors in protein biosynthesis. *Trends Biochem Sci.* 28:434-441.

Andersson JO. 2005. Lateral gene transfer in eukaryotes. *Cell Mol Life Sci.* 62:1182-1197.

Andersson JO, Sjogren AM, Horner DS, Murphy CA, Dyal PL, Svard SG, Logsdon JM Jr, Ragan MA, Hirt RP, Roger AJ. 2007. A genomic survey of the fish parasite *Spironucleus salmonicida* indicates genomic plasticity among diplomonads and significant lateral gene transfer in eukaryote genome evolution. *BMC Genomics.* 8:51.

Bowler C, Allen AE, Badger JH et al. 2008. The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature.* 456:239-244.

Doolittle WF. 1998. You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. *Trends Genet*. 14:307-311.

Doolittle WF, Bapteste E. 2007. Pattern pluralism and the Tree of Life hypothesis. *Proc Natl Acad Sci USA.* 104:2043-2049.

Doolittle WF, Papke RT. 2006. Genomics and the bacterial species problem. *Genome Biol.* 7:116.

Dujon B, Sherman D, Fischer G et al. 2004. Genome evolution in yeasts. *Nature.* 430:35-44.

Esser C, Ahmadinejad N, Wiegand C et al. 2004. A genome phylogeny for mitochondria among alpha-proteobacteria and a predominantly eubacterial ancestry of yeast nuclear genes. *Mol Biol Evol.* 21:1643-1660.

Gabaldón T, Huynen MA. 2004. Shaping the mitochondrial proteome. *Biochim Biophys Acta.* 1659:212-220.

Gile GH, Faktorová D, Castlejohn CA, Burger G, Lang BF, Farmer MA, Lukeš J, Keeling PJ. 2009a. Distribution and phylogeny of EFL and EF-1α in Euglenozoa suggest ancestral co-occurrence followed by differential loss. *PLoS One.* 4:e5162.

Gile GH, Novis PM, Cragg DS, Zuccarello GC, Keeling PJ. 2009b. The distribution of EF-1α. EFL, and a non-canonical genetic code in the Ulvophyceae: Discrete genetic characters support a consistent phylogenetic framework. *J Euk Microbiol.* 56:367-372.

Gile GH, Patron NJ, Keeling PJ. 2006. EFL GTPase in Cryptomonads and the distribution of EFL and EF-1alpha in Chromalveolates. *Protist.* 157:435-444.

Gogarten JP. 2003. Gene transfer: Gene swapping craze reaches eukaryotes. *Curr Biol.* 13:R53-R54.

Hashimi H, Zíková A, Panigrahi AK, Stuart KD, Lukeš J. 2008. TbRGG1, a component of a novel multiprotein complex involved in kinetoplastid RNA editing. *RNA.* 14:970-980.

Ho P, Kong KF, Tang JSH, Wong JTY. 2007. An unusual S-adenosylmethionine synthetase gene from dinoflagellate is methylated. *BMC Mol Biol.* 8:87.

Chiang PK, Gordon RK, Tal J, Zeng GC, Doctor BP, Pardhasaradhi K, McCann PP. 1996. S-adenosylmethionine and methylation. *FASEB J.* 10: 471-480.

Chuang SM, Chen L, Lambertson D, Nand M, Kinzy TG, Madura K. 2005. Proteasome-mediated degradation of cotranslationally damaged proteins involves translation elongation factor 1A. *Mol Cell Biol.* 25:403-413.

Kamikawa R, Sanchez-Perez GF, Sako Y, Roger AJ, Inagaki Y. 2009. Expanded phylogenies of canonical and non-canonical types of methionine adenosyltransferase reveal a complex history of these gene families in eukaryotes. *Mol Phyl Evol.* 53:565-570.

Keeling PJ, Inagaki Y. 2004. A class of eukaryotic GTPase with a punctuate distribution suggesting multiple functional replacements of translation elongation factor 1alpha. *Proc Natl Acad Sci USA*. 101:15380-15385.

Keeling PJ, Palmer JD. 2008. Horizontal gene transfer in eukaryotic evolution. *Nat Rev Genet.* 9:605-618.

Kurland CG, Andersson SG. 2000. Origin and evolution of the mitochondrial proteome. *Microbiol Mol Biol Rev.* 64:786-820.

Kondrashov FA, Koonin EV, Morgunov IG, Finogenova TV, Kondrashova MN. 2006. Evolution of glyoxylate cycle enzymes in metazoa: evidence of multiple horizontal transfer events and pseudogene formation. *Biol Direct.* 1:31.

Liu G, Grant WM, Persky D, Latham VM Jr, Singer RH, Condeelis J. 2002. Interactions of elongation factor 1alpha with F-actin and beta-actin mRNA: Implications for anchoring mRNA in cell protrusions. *Mol Biol Cell.* 13:579-592.

Long S, Jirků M, Mach J, Ginger ML, Sutak R, Richardson D, Tachezy J, Lukeš J. 2008. Ancestral roles of eukaryotic frataxin: mitochondrial frataxin function and heterologous expression of hydrogenosomal *Trichomonas* homologs in trypanosomes. *Mol Microbiol.* 69:94-109.

Martin W, Schnarrenberger C. 1997. The evolution of the Calvin cycle from prokaryotic to eukaryotic chromosomes: a case study of functional redundancy in ancient pathways through endosymbiosis. *Curr Genet.* 32: 1-18.

Oborník M, Green BR. 2005. Mosaic origin of the heme biosynthesis pathway in photosynthetic eukaryotes. *Mol Biol Evol.* 22:2343-2353.

Ricard G, McEwan NR, Dutilh BE et al. 2006. Horizontal gene transfer from bacteria to rumen ciliates indicates adaptation to their anaerobic, carbohydrate-rich environment. *BMC Genomics.* 7:22.

Richards TA, Dacks JB, Campbell SA, Blanchard JL, Foster PG, McLeod R, Roberts CW. 2006. Evolutionary origins of the eukaryotic shikimate pathway: Gene fusions, horizontal gene transfer, and endosymbiotic replacements. *Eukaryot Cell.* 5:1517-1531.

Sanchez-Perez GF, Hampl V, Simpson AGB, Roger AJ. 2008. A new divergent type of eukaryotic methionine adenosyltransferase is present in multiple distantly related secondary algal lineages. *J Euk Microbiol.* 55: 374-381.

Watkins RF, Gray MW. 2006 The frequency of eubacterium-to-eukaryote lateral gene transfers shows significant cross-taxa variation within amoebozoa. *J Mol Evol.* 63:801-814.

Welch RA, Burland V, Plunkett G 3rd et al. 2002. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc Natl Acad Sci USA*. 99:17020-17024.

Whitaker JW, McConkey GA, Westhead DR. 2009. Prediction of horizontal gene transfers in eukaryotes: approaches and challenges. *Biochem Soc Trans.* 37:792-795.

FIGURE LEGENDS

Fig. 1. Expression of heterologous EFL and (parallel) RNAi silencing of EF-1α in *Trypanosoma brucei*. As a loading control, the gel was stained with ethidium bromide to visualize rRNA bands. (**A**) Level of EFL mRNA was analyzed by blotting 10 μg of total RNA extracted from 29-13 parental cells (lane 1) and cells constitutively expressing EFL from *Isochrysis galbana* containing RNAi vector against EF-1α (lane 2; indicated with an arrowhead). The full-length EFL gene was used as a probe and hybridization was performed at 65 ^{o}C, at which no cross-hybridization with EF-1α occurs. The positions of the EF-1α and EFL mRNAs are indicated with white and black arrowheads, respectively.

(**B**) Level of EF-1α mRNA was analyzed in total RNA extracts from the following cell lines: 29-13 cells (lane 1); non-induced cells containing RNAi vector against EF-1α and expressing constitutively EFL (lane 2); same cells as in lane 2, in which RNAi was induced by the addition of 1 μg/ml of tetracycline to the medium (lane 3); non-induced cells containing RNAi vector against EF-1α (lane 4); same cells as in lane 4, in which RNAi was induced (lane 5). The full-length EF-1α gene was used as a probe and hybridization was performed at 65 ^{o}C, at which no cross-hybridization with EFL occurs. The positions of the EF-1α mRNAs and respective double stranded (ds) RNA are indicated with white and grey arrowheads, respectively.

(**C**) Levels of EFL and EF-1α mRNAs were analyzed in parallel in total RNA extracts from cells constitutively expressing exogenous EFL (lane 1), non-induced cells expressing EFL and also containing RNAi vector against EF-1α (lane 2), same cells as in lane 2, in which RNAi was induced (lane 3). The full-length EFL gene was used as a probe and hybridization was performed at 60 ^{o}C, at which cross-hybridization with EF-1α occurs. The positions of the EF-1α and EFL mRNAs are indicated with white and black arrowheads, respectively.

Fig. 2. Lethality of EF-1α RNAi is not rescued by EFL in *T. brucei*. Cell numbers were measured using a Coulter Counter Z2. The *y* axis is labeled by a logarithmic scale and represents the product of cell densities measured and total dilution. Growth curves are one representative set from three independent experiments. (**A**) The growth of cells with ablated EF-1α mRNA is inhibited (triangles), as compared with the non-induced cells (squares), which grow at the same rate as the 29-13 cells (diamonds). (**B**) The growth of cells with ablated EF-1α mRNA that also express exogenous EFL is not rescued (triangles). The non-induced cells co-expressing EF-1α and EFL (squares) grow slower than the 29-13 cells.

Fig. 3. Expression of heterologous MATX and (parallel) RNAi silencing of MAT in *T. brucei*. As a loading control, the gel was stained with ethidium bromide to visualize rRNA bands. (**A**) Level of MATX mRNA was analyzed by blotting 10 μg of total RNA extracted from 29-13 cells (lane 1), cells constitutively expressing MATX from *Euglena gracilis* (lane 2), non-induced cells constitutively expressing heterologous MATX and containing RNAi vector against MAT (lane 3), and the same cells as in lane 3, in which RNAi was induced

(lane 4). The full-size MATX gene was used as a probe and hybridization was performed at 65 ^{o}C. The position of the MATX mRNA is indicated with a black arrowhead.

(**B**) Levels of MAT mRNA and respective dsRNA were analyzed in total RNA extracted from the following cell lines: 29-13 cells (lane 1); non-induced cells containing RNAi vector against MAT and expressing constitutively MATX (lane 2); same cells as in lane 2, in which RNAi was induced by the addition of 1 μg/ml of tetracycline to the medium (lane 3); non-induced cells containing RNAi vector against MAT (lane 4); same cells as in lane 4, in which RNAi was induced (lane 5). 5' region of the MAT gene was used as a probe and hybridization was performed at 65 ^{o}C. The positions of the targeted MAT mRNA and the dsRNA are indicated with white and gray arrowheads, respectively.

(**C**) The MATX protein is expressed in *T. brucei* cells. The levels of the HA_3-tagged exogenous MATX protein and endogenous enolase were followed using mouse monoclonal antibodies and rabbit polyclonal antibodies, respectively, by blotting cell lysates from 5 x 10^6 cells. The levels of both proteins were analyzed in total lysates from 29-13 cells (lane 1), cells constitutively expressing MATX (lane 2), non-induced cells constitutively expressing heterologous MATX and containing RNAi vector against MAT (lane 3), and the same cells as in lane 3, in which RNAi was induced (lane 4). The Coomassie-stained gel and enolase were used as loading controls.

Fig. 4. Lethality of MAT RNAi is rescued by MATX in *T. brucei*. The growth curves were performed as described in Fig. 2. The numbers of 29-13 cells (diamonds), non-induced cells (squares) and those, in which RNAi was induced by the addition of 1 μg/ml tetracycline (triangles). (**A**) The growth of cells with ablated MATXmRNA is inhibited (triangles), as compared with the non-induced cells (squares) and wild type (diamonds). (**B**) The growth of cells with ablated MATXmRNA is rescued by the expression of exogenous MATX (triangles).

Fig. 1

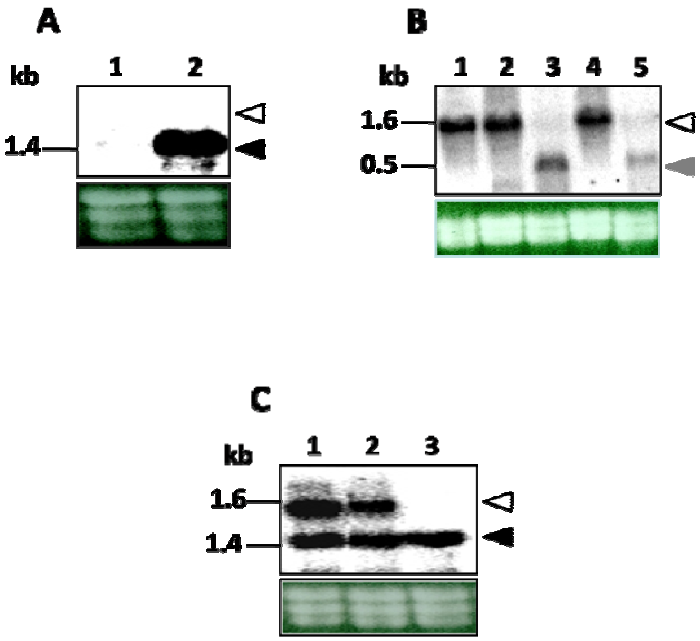

Fig. 1. Expression of heterologous EFL and (parallel) RNAi silencing of EF-1α in Trypanosoma brucei. As a loading control, the gel was stained with ethidium bromide to visualize rRNA bands. (A) Level of EFL mRNA was analyzed by blotting 10 μg of total RNA extracted from 29-13 parental cells (lane 1) and cells constitutively expressing EFL from Isochrysis galbana containing RNAi vector against EF-1α (lane 2; indicated with an arrowhead). The full-length EFL gene was used as a probe and hybridization was performed at 65 oC, at which no cross-hybridization with EF-1α occurs. The positions of the EF-1α and EFL mRNAs are indicated with white and black arrowheads, respectively. (B) Level of EF-1α mRNA was analyzed in total RNA extracts from the following cell lines: 29-13 cells (lane 1); non-induced cells containing RNAi vector against EF-1α and expressing constitutively EFL (lane 2); same cells as in lane 2, in which RNAi was induced by the addition of 1 μg/ml of tetracycline to the medium (lane 3); non-induced cells containing RNAi vector against EF-1α (lane 4); same cells as in lane 4, in which RNAi was induced (lane 5). The full-length EF-1α gene was used as a probe and hybridization was performed at 65 oC, at which no cross-hybridization with EFL occurs. The positions of the EF-1α mRNAs and respective double stranded (ds) RNA are indicated with white and grey arrowheads, respectively. (C) Levels of EFL and EF-1α mRNAs were analyzed in parallel in total RNA extracts from cells constitutively expressing exogenous EFL (lane 1), non-induced cells expressing EFL and also containing RNAi vector against EF-1α (lane 2), same cells as in lane 2, in which RNAi was induced (lane 3). The full-length EFL gene was used as a probe and hybridization was performed at 60 oC, at which cross-hybridization with EF-1α occurs. The positions of the EF-1α and EFL mRNAs are indicated with white and black arrowheads, respectively.

Fig. 2

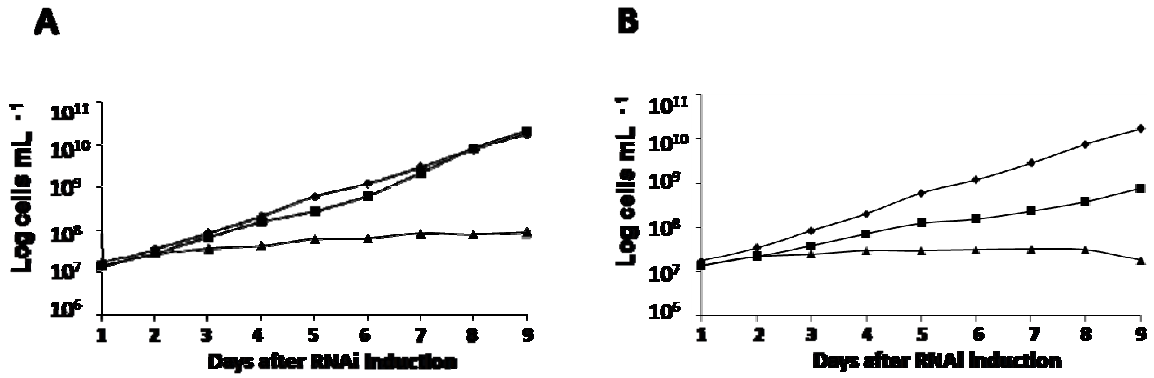

Fig. 2. Lethality of EF-1α RNAi is not rescued by EFL in T. brucei. Cell numbers were measured using a Coulter Counter Z2. The y axis is labeled by a logarithmic scale and represents the product of cell densities measured and total dilution. Growth curves are one representative set from three independent experiments. (A) The growth of cells with ablated EF-1α mRNA is inhibited (triangles), as compared with the non-induced cells (squares), which grow at the same rate as the 29-13 cells (diamonds). (B) The growth of cells with ablated EF-1α mRNA that also express exogenous EFL is not rescued (triangles). The non-induced cells co-expressing EF-1αand EFL (squares) grow slower than the 29-13 cells.

Fig. 3

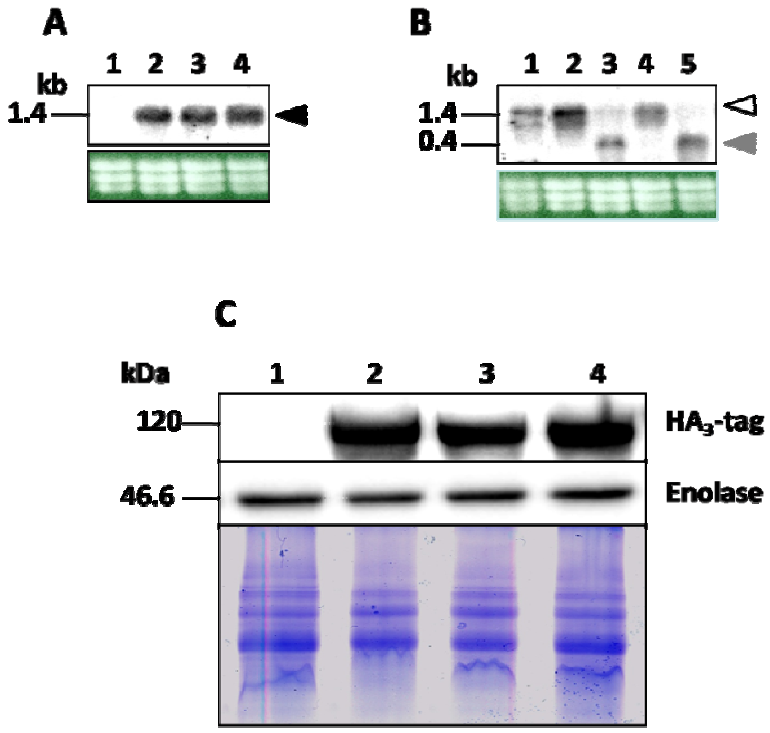

Fig. 3. Expression of heterologous MATX and (parallel) RNAi silencing of MAT in T. brucei. As a loading control, the gel was stained with ethidium bromide to visualize rRNA bands. (A) Level of MATX mRNA was analyzed by blotting 10 µg of total RNA extracted from 29-13 cells (lane 1), cells constitutively expressing MATX from Euglena gracilis (lane 2), non-induced cells constitutively expressing heterologous MATX and containing RNAi vector against MAT (lane 3), and the same cells as in lane 3, in which RNAi was induced (lane 4). The full-size MATX gene was used as a probe and hybridization was performed at 65 oC. The position of the MATX mRNA is indicated with a black arrowhead. (B) Levels of MAT mRNA and respective dsRNA were analyzed in total RNA extracted from the following cell lines: 29-13 cells (lane 1); non-induced cells containing RNAi vector against MAT and expressing constitutively MATX (lane 2); same cells as in lane 2, in which RNAi was induced by the addition of 1 µg/ml of tetracycline to the medium (lane 3); non-induced cells containing RNAi vector against MAT (lane 4); same cells as in lane 4, in which RNAi was induced (lane 5). 5' region of the MAT gene was used as a probe and hybridization was performed at 65 oC. The positions of the targeted MAT mRNA and the dsRNA are indicated with white and gray arrowheads, respectively. (C) The MATX protein is expressed in T. brucei cells. The levels of the HA3-tagged exogenous MATX protein and endogenous enolase were followed using mouse monoclonal antibodies and rabbit polyclonal antibodies, respectively, by blotting cell lysates from 5 x 106 cells. The levels of both proteins were analyzed in total lysates from 29-13 cells (lane 1), cells constitutively expressing MATX (lane 2), non-induced cells constitutively expressing heterologous MATX and containing RNAi vector against MAT (lane 3), and the same cells as in lane 3, in which RNAi was induced (lane 4). The Coomassie-stained gel and enolase were used as loading controls.

Fig. 4

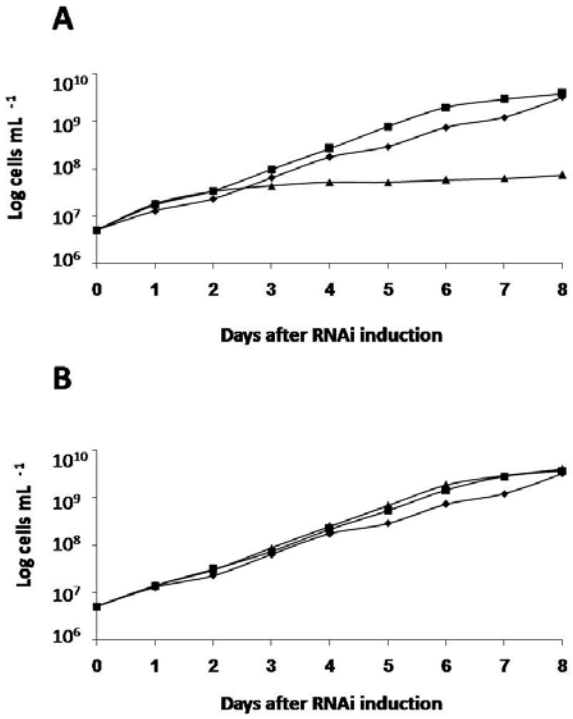

Fig. 4. Lethality of MAT RNAi is rescued by MATX in T. brucei. The growth curves were performed as described in Fig. 2. The numbers of 29-13 cells (diamonds), non-induced cells (squares) and those, in which RNAi was induced by the addition of 1 μg/ml tetracycline (triangles). (A) The growth of cells with ablated MAT mRNA is inhibited (triangles), as compared with the non-induced cells (squares) and wild type (diamonds). (B) The growth of cells with ablated MAT mRNA is rescued by the expression of exogenous MATX (triangles).

In the attached manuscript we have attempted, for the first time, to experimentally test HGT in the model protist *T. Brucei*. The paper was recently rejected from Molecular Biology and Evolution on the grounds of one aditional missing control. The reviews are attached. The editor offered resubmission once this experiment is done indeed, we plan to perform it and resubmit.

Reviewer:1
Comments to the Author
Review of Experimental examination of EFL and MATX eukaryotic horizontal gene transfers: co-existence of mutually exclusive transcripts predates functional rescue by Ruzicka et al.

This paper seeks to investigate two alternative evolutionary scenarios for the extremely mosaic distribution of two ancient paralogues (EFL and MATX) across the eukaryotes. Previously two scenarios have been suggested: 1) Multiple horizontal gene transfer events and, 2) hidden paralogy (i.e. maintenance of two anciently derived paralogues combined with differential loss). In the case of these two gene families both scenarios are extremely complicated so deciding between the two is extremely difficult on grounds of parsimony. Extending work from Ho et al (2007 – see the references in the manuscript), in an evolutionary context, this paper seeks to test the implications of these two scenarios by forcing Trypanosoma brucei, to carry both paralogues and investigate the effect of this extra genetic load on growth and to see if the native paralogues can be complimented by the xenologue. This paper suggests that there is a cost to carrying two paralogues forms of EFL as culture growth rate is reduced but suggests that only the MATX xenologue will complement under their experimental conditions. There is a lot to admire about this paper, however there are several points that require further clarification.

MajorPoints

1) Regarding the amplification the EFL gene from the haptophyte. This sequence was amplified from total DNA - it is therefore unclear from my reading of the methods section if this contained haptophyte introns. If so were these removed or were they also introduced in to the Trypanosome line. The haptophyte EFL gene did not seem to compliment while the MATX gene would – however the MATX gene came from cDNA - it is therefore important

that the authors are clear on this because these experiments are potentially not equivalent if one set of alien genes was containing alien introns while the other was not. Further to this the Acknowledgements section seems to suggest it was a haptophyte EFL clone after all suggesting an EST/cDNA. Please clarify this.

2) Further to the last point the authors should discuss and justify the choice of a Euglena gene in one case and a haptophyte gene in the second experiment. In phylogenetic terms these are hugely different evolutionary distances. Consequently, the complementation of a haptophyte gene is less likely to be achieved because the interactome differences between haptophyte-trypansoma and Euglena-trypanosoma are considerable. Why did the authors choose the haptophyte gene?

3) Usually in most complementation experiments involving growth curves a control is included involving transformation of an empty vector. This is particularly important because the electroporation process can affect growth dynamics. I may have missed this but the authors should consider including this information or justify why this control was not completed. This is especially important because the EFL experiment shows a cost to the introduction of the EFL vector or alternatively a cost to the process of introducing the EFL gene. Currently it is not possible to distinguish between these two alternative explanations?

4) Relating to points 1 and 2 above. I'm concerned that the authors did not confirm that the EFL haptophyte gene introduced to the trypanosome produced a translated protein. This was confirmed for the MATX gene but not for the EFL gene. This is very important given that the MATX protein will compliment and the EFL will not.

Minor Points

1) Abstract. Pg 1 – line 17. Change – 'is mutually exclusive' to 'that appears to be mutually exclusive with one notable exception' – see page 5, second paragraph of the manuscript.

2) Abstract. Pg 1 – line 37. Define 'slightly harmful'

3) Pg 3 – line 44. 'eventually found way to' replace with 'eventually found their way to'

4) Pg 5 line 46. replace 'deteriorating' with 'detrimental'

5) Pg 6 line 59. please give full name of haptophyte at first usage.

6) Pg 12 line 53 – specify if the Euglena and haptophyte gene do not depart from standard Tryp. codon usage – currently this sentence reads like allversions of this gene do not depart from standard Tryp. codon usage.

7) Figure 1 and 3 specify if a, b, etc is – northern blot – western blot? Etc. Makes it easier to follow.

Reviewer: 2

Comments to the Author

Comments on "Experimental examination of EFL and MATX eukaryotic horizontal gene transfers: co-existence of mutually exclusive transcripts predates functional rescue (MBE-10-0172)" by Ruzicka, Verner, Hampl and Lukes.

Elongation factor-like protein (EFL) is a putative translation factor in eukaryotic cells that is probably compatible with EF-1alpha. Since EFL-containing lineages are phylogenetically spread within EF-1alpha-containing lineages, horizontal/lateral gene transfer acts the pivotal role in EFL/EF-1alpha evolution. The distribution of a divergent version of methionine adenosyltransferase (MAT), MATX, was recently identified, and MATX genes appeared to be patchily distributed in eukaryotes. Thus, as postulated for EFL genes, MATX genes have been frequently transferred between two distantly related eukaryotes.

We have assumed that all eukaryotes are originally EF-1alpha-containing, and exogenous EFL genes invaded and subsequently replace the indogenous functional counterparts in the current EFL-containing cells. This scenario requires the period of the cells utilizing both EF-1alpha and EFL, but we have no idea what happened when two functionally equivalent proteins existed in a single cell, and how the endogenous EFL took over the indogenous EF-1alpha. Likewise, it is uncertain how the cells managedthe two methionine adenosyltransferases after putative MATX gene transfer.

The authors prepared two trypanosome cell lines, one expressing the EFL gene of the haptophyte Isochrysis galbana and the other expressing the MATX gene of the euglenid Euglena glacilis, and investigated the impact of long term co-expression of EFL and EF-1alpha/MAT and MATX on cell viability. In addition, by knock-down of the indogeous EF-1alpha or MAT gene, they tested whether the experimentally-introduced (i.e. horizontally transferred) EFL/MATX can suppress the function of the pre-existed EF-1alpha/MAT gene.

The cells expressing both EFL and EF-1alpha grew slower than the wild-type cells, while co-expression of both MATX and MAT did not affect the growth rate. Knock-down experiments

of the exogenous EFL and MATX genes revealed that (1) the haptophyte EFL gene was not capable to replace the indogeous EF-1alpha, and (2) Euglena MATX and the indogenous MAT are functionally compatible.

I'm very interested in the results presented in this manuscript. However, I would like to ask the authors to add extra experiments to address major issues listed below.

[Experiment 1]
I'm not sure why the authors chose the Isochrysis gene to replace the EF-1alpha gene in T. brucei. The authors failed to suppress the function of T. brucei EF-1alpha by expressing Isochrysis EFL gene, but the phylogenetic 'distance' between two genes is likely critical for this type of experiments. Haptophytes and T. brucei are distantly related in the organismal phylogeny, and so the experiments presented in this work should be repeated by using one of euglenozoan EFL genes instead of the gene sampled from Isochrysis.

[Experiment 2]
It would be also interesting to repeat the MAT/MATX replacement using Isochrysis MATX gene. The results presented indicate that Trypanosoma MAT and Euglena MATX are compatible, but might not be compatible in the experiment using a MATX gene from a distantly related lineage (e.g. haptophytes).

[Experiment 3]
The author assumed that a horizontally transferred EFL/MATX gene was fully expressed in the recipient cell. However, the degree of the expression of an exogenous gene right after HGT remains unclear. The expression level may be particularly important for EFL, since the high expression of the exogenous EFL hindered cell growth. I propose to prepare cell lines expressing the EFL gene at various levels and re-examine their growth rates. If the normal cell growth is tolerance to low expression of the exogenous EFL gene, the long-term co-existence of EFL and EF-1alpha genes in a single cell may be possible.