J. Eukaryot. Microbiol., 57(1), 2010 pp. 11–18

DOI: 10.1111/j.1550-7408.2009.00458.x

Integrating Horizontal Gene Transfer and Common Descent to Depict Evolution and Contrast It with ''Common Design''[1]

GUILLERMO PAZ-Y-MIÑO C.[a] and AVELINA ESPINOSA[b]

[a]*Department of Biology, University of Massachusetts Dartmouth, North Dartmouth, Massachusetts 02747, and*
[b]*Department of Biology, Roger Williams University, Bristol, Rhode Island 02809*

ABSTRACT. Horizontal gene transfer (HGT) and common descent interact in space and time. Because events of HGT co-occur with phylogenetic evolution, it is difficult to depict evolutionary patterns graphically. Tree-like representations of life's diversification are useful, but they ignore the significance of HGT in evolutionary history, particularly of unicellular organisms, ancestors of multicellular life. Here we integrate the reticulated-tree model, ring of life, symbiogenesis whole-organism model, and eliminative pattern pluralism to represent evolution. Using *Entamoeba histolytica* alcohol dehydrogenase 2 (EhADH2), a bifunctional enzyme in the glycolytic pathway of amoeba, we illustrate how EhADH2 could be the product of both horizontally acquired features from ancestral prokaryotes (i.e. aldehyde dehydrogenase [ALDH] and alcohol dehydrogenase [ADH]), and subsequent functional integration of these enzymes into EhADH2, which is now inherited by amoeba via common descent. Natural selection has driven the evolution of EhADH2 active sites, which require specific amino acids (cysteine 252 in the ALDH domain; histidine 754 in the ADH domain), iron- and NAD^+ as cofactors, and the substrates acetyl-CoA for ALDH and acetaldehyde for ADH. Alternative views invoking ''common design'' (i.e. the non-naturalistic emergence of major taxa independent from ancestry) to explain the interaction between horizontal and vertical evolution are unfounded.

Key Words. Design creationism, *Entamoeba*, integrative model, lateral evolution, reticulated patterns of unicellular life.

SCIENTISTS use tree-like diagrams to depict life's history. This approach illustrates common descent, the idea conceptualized by Charles Darwin (1859) in *The Origin of the Species*, by which all organisms originate from ancestral forms traceable in time. However, phylogenies rarely include in their vertical representations of evolutionary history the significance of the lateral acquisition of genetic material by organisms (Gogarten, Gogarten, and Olendzenski 2009). Continuous lateral exchange of genes in bacteria, Archaea, and Eukarya complicate the identification of a single origin and/or dichotomous branching pattern during the evolution of unicellular life (Andersson 2005; Lopez and Bapteste 2009). Instead a ''reticulated tree'' or ''net'' has been suggested (Doolittle 1999) to represent the earliest stages of life's history (Fig. 1). The ''ring of life'' proposal (Rivera and Lake 2004) also takes into consideration the role of horizontal gene transfer (HGT) in genome evolution; a ''ring'' connects the origin of eukaryotes to a dual genetic merging, via endosymbiosis, of eubacterial and archaebacterial genomes (Fig. 2). The ''symbiogenetic whole-organism model'' (Margulis 2009; Margulis et al. 2006) posits that entire genomes of ''partner organisms'' became integrated at topological, temporal, metabolic, gene-product, and genetic levels, and that a serial symbiotic fusion or ''anastomosis'' of ancestral lineages gave rise to eukaryotic cells (Fig. 3). The ''pattern pluralism scheme'' (Bapteste and Boucher 2009; Doolittle and Bapteste 2007) considers that various representations of evolutionary relationships are valid for taxonomic units. To account for taxa resemblance, an overlapping combination of traits (''interactive database'') is used to detect ''evolutionary patterns'' without seeking, a priori, a tree-like depiction. Thus, organisms can appear in various natural and non-exclusive taxonomic units (Fig. 4), making vertical inheritance part of the conceptual understanding of evolution, but not the end.

The taxonomic significance of the interaction between HGT and common descent depends on the frequency (= how often), magnitude (= how much), and quality (= fitness value) of lateral and/or vertical acquisition and integration of genetic material. Depicting stages of prokaryotic and unicellular eukaryotic life as a reticulated pattern of interconnected organisms, clustered in units of congruency, might resemble reality, because the frequency and magnitude of lateral genetic exchange can be high (Bapteste and Boucher 2009). The bifurcated tree-like representation of evolutionary relations seems better suited for multicellular organisms, where vertical inheritance of genomes is of larger magnitude than horizontally acquired genetic traits (Andersson 2008; Keeling and Palmer 2008; Lopez and Bapteste 2009). Here we merge the reticulated pattern and tree phylogenies with the ring of life, symbiogenetic whole-organism model, and pattern pluralism schemes; we highlight a gradual spatio-temporal contribution of HGT to common descent and sustain our model with accounts from the literature. We use *Entamoeba histolytica* alcohol dehydrogenase 2 (EhADH2), a bifunctional glycolytic enzyme, to illustrate how EhADH2 could be the product of both horizontally and vertically acquired genetic features. We discuss how mutation rate coupled with natural selection and HGT coupled with common descent drove the evolution of EhADH2, and perhaps of most alcohol dehydrogenases (ADH), and contrast this analysis with proposals invoking ''common design'' (i.e. the independent emergence of major taxa with no common ancestry) to explain the interaction between horizontal and vertical evolution.

MECHANISMS OF LATERAL GENETIC EXCHANGE

The mechanisms involved in lateral transfer and/or genetic exchange are complex and diverse (for reviews see Andersson 2005; Hensel and Schmidt 2008; Keeling and Palmer 2008; Gogarten et al. 2009). In Table 1, we summarize them as follows: indirect or agent-mediated acquisition of genetic fragments or mobile genetic elements among unicellular organisms includes *transduction* (i.e. DNA or RNA sequences transferred by viruses) and *transformation* (i.e. uptake of genes from the environment). Unicellular organisms can donate/receive genetic material directly from one another through various processes: bacterial *conjugation* can transfer a plasmid through a pilus bridge; ciliates undergo *nuclear exchange* conjugation; mitochondrial and chloroplast genes have originated from genome transfer during evolutionary-ecological associations between cells by *symbiogenesis*, or by uptake via *phagotrophism* (i.e. the ingestion of cells). Viruses are the original source of transposons (i.e. ''jumping'' fragments of DNA within a genome, an agent-mediated mechanism) among multi-

[1]Presentation delivered at the workshop: Horizontal Gene Transfer and Phylogenetic Evolution Debunk Intelligent Design, 1st International Society of Protistologists, North American Section Meeting, June 11–13, 2009, Roger Williams University, Bristol, RI, USA.

Corresponding Author: A. Espinosa, Department of Biology, Roger Williams University, One Old Ferry Road, Bristol, Rhode Island 02809—Telephone number: +401 254 3137; FAX number: +401 254 3310; e-mail: aespinosa@rwu.edu

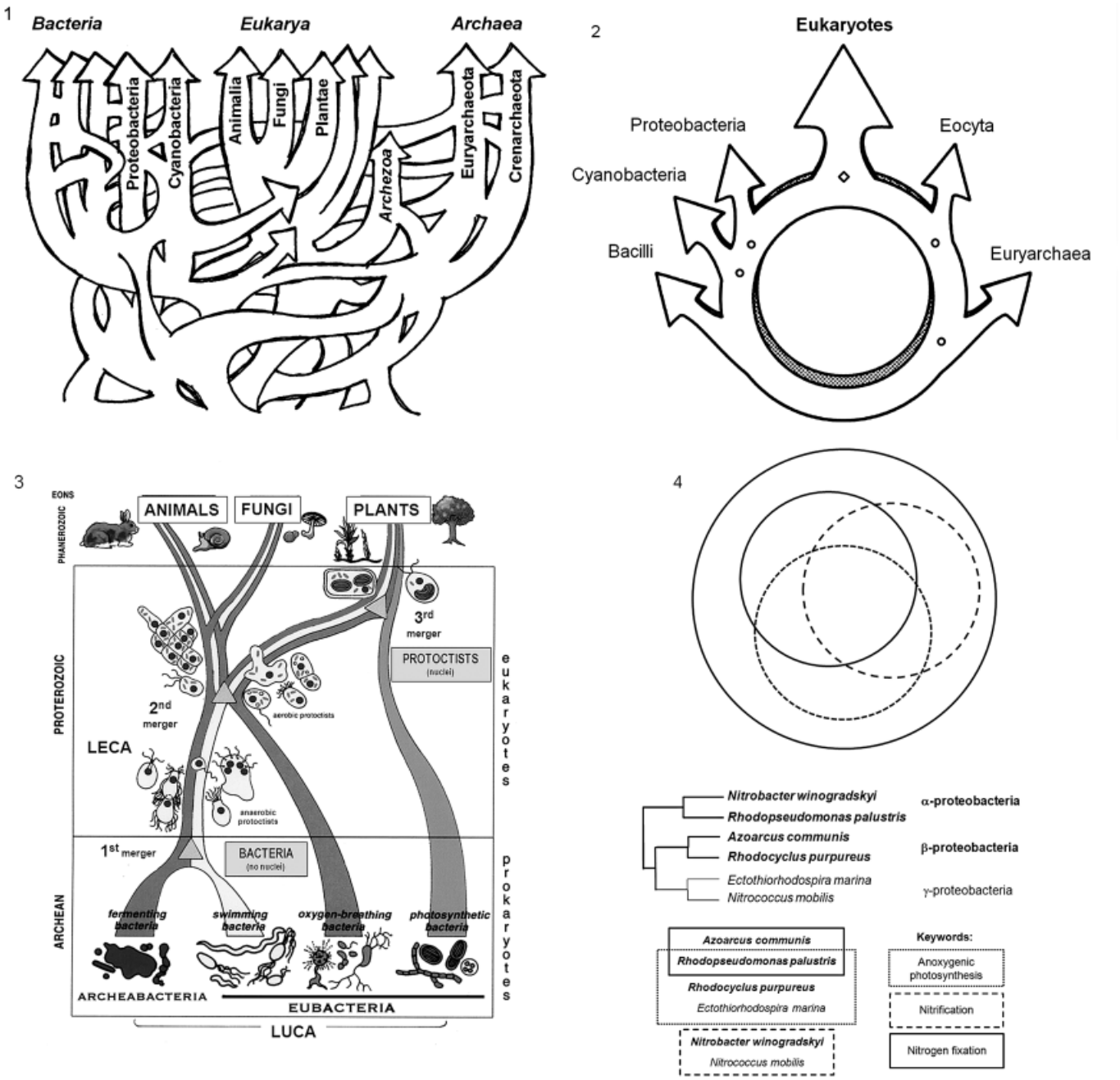

Fig. **1–4.** Significant depictions of life's evolutionary history. **1.** ''Reticulated tree'' or net representing frequent lateral transfer of genetic material among the ancestors of bacteria, Eukarya, and Archaea (after Doolittle 1999; with permission from *Science* Magazine). **2.** The ''ring of life'' connects all unicellular life by horizontal gene transfer and depicts major ancestral groups by tiny circles on the ring (Rivera and Lake 2004); bacteria (three left arrows) generate an ''operational eukaryotic ancestor,'' while archaea (two right arrows) generate an ''informational eukaryotic ancestor,'' both merge at the root of the eukaryotes (with permission from Nature Publishing Group). **3.** The ''symbiogenetic whole-organism reconstruction'' of the evolution of early life (Margulis 2009; illustration modified from original drawing by K. Delisle) depicts life emerging from a last universal common ancestor (LUCA), the origin of eukaryotes after a series of symbiotic fusions of ancestral archaebacteria and eubacteria, which led to the emergence of the last eukaryotic common ancestor (LECA); three significant genome ''anastomoses'' are represented by the merging of archaebacteria and a ''swimming'' bacteria, eukaryotes and ''oxygen-breathing'' bacteria, and eukaryotes and ''photosynthetic'' bacteria; the ulterior origin of animals, fungi, and plants is also depicted (with permission from Springer Science and Business Media). **4.** The ''pattern pluralism scheme'' (Bapteste and Boucher 2009; Doolittle and Bapteste 2007) is depicted as three overlapping circles, within a larger circle, representing ''family resemblance'' among taxa, a well-known concept among taxonomists; the tree in the middle shows a classic phylogeny, where six taxa belong to three identifiable groups (α-, β-, and γ-proteobacteria); the ''interactive data base'' (bottom) is structured by keywords that highlight overlapping features of the six taxa, which can appear grouped in different taxonomic units because they share common properties in specific dimensions, i.e. anoxygenic photosynthesis, nitrification, nitrogen fixation (with permission from Springer Science and Business Media).

Table **1.** Mechanisms of lateral genetic transfer and/or exchange.

Organisms involved (mechanism)	MGE[a] and/or agent involved	Frequency of occurrence	Magnitude of genetic transfer/ exchange relative to host's genome	Accumulation and integration of genetic material into host's genome over time	Detectability of source of origin (donor) after eons
Virus-to-, or unicellular-to-unicellular					
Virus-mediated (transduction)					
Virus-to-prokaryote	Bacteriophages[b]	High	Low to medium	Probably high	Probably low
Virus-to-eukaryote	Transposons	High	Low	Moderate	Probably low
Prokaryote-to-prokaryote					
(transformation)	Gene fragments	High	Low to medium	Moderate	Low to moderate
(conjugation)	Plasmids	High	Low to high	Moderate to high	Moderate
Prokaryote-to-eukaryote					
(symbiogenesis)	Cells	Low to high	Low to medium	Low to moderate	Low
(phagotrophism)	Cells	Probably high	Low to medium	Low to moderate	Low
Eukaryote-to-eukaryote					
(phagotrophism)	Cells	Low to high	Low to medium	Low to moderate	Low
(nuclear exchange)	Micro-nuclei	Probably high	Medium	High	Moderate to high
Virus-to-, unicellular-to-, or multicellular-to-multicellular					
Virus-to-plants/animals/ fungi (transposition)	Transposons[c]	High	Low	Low to moderate	Low
Bacteria-to-plants/animals/ fungi (transfer)[d]	Plasmids or gene fragments	Low to high	Low to medium	Low to moderate	Low to moderate
Plants-to-plants (hybridization)	Gametes	Low to high	High	High	Moderate to high
Animals-to-animals (hybridization)	Gametes	Low to medium	High	High	High
Multicellular-to-virus, or unicellular					
Plants/animals/fungi-to-virus (transfer)	Gene fragments	High	High	High	Low
Plants/animals/fungi-to-prokaryote (transformation)[e]	Gene fragments	High	High	Moderate	Low
Plants/animals/fungi-to-eukaryote (transfer)[e]	Gene fragments	Low to high	Low to medium	Low to moderate	Low

[a]Mobile Genetic Elements are fragments of genetic material that ''move'' within and among genomes.
[b]Bacteriophages can be considered both organism-participants in HGT and ''agent'' mediators of lateral genetic transfer.
[c]Transposons in multicellular organisms include ''transposable elements'' or ''jumping genes''.
[d]Gene transfer from bacteria to plants (e.g. conjugative plasmids for nitrogen fixation) and animals (e.g. nematode parasitic genes obtained from prokaryotes; rotifers' acquisition of pro and eukaryotic genes) has been documented (Keeling and Palmer 2008; Mitreva, Smant, and Helder 2009).
[e]Metabolic and structural gene or gene-fragment transfer from multi to unicellular organisms is increasingly being reported (Keeling and Palmer 2008).

cellular organisms. Gene travel from bacteria to plants (e.g. conjugative plasmids involved in symbiotic nitrogen fixation) and animals (e.g. parasitic genes obtained from bacteria) is also possible. Nevertheless, the major lateral transfer and integration of genes in plants or animals occurs during *hybridization* when the genomes of two discrete species or between genetically structured populations merge (Schwenk, Brede, and Streit 2008; Soltis and Soltis 2009). It is generally accepted that, on average, vascular plants hybridize more frequently (25%) than animals (0.1–3%), although some vertebrates (e.g. birds) can have hybridization rates comparable to those of vascular plants (Schwenk et al. 2008). The transfer of genetic material from multicellular organisms to unicellular life is increasingly being reported (Keeling and Palmer 2008).

Genetic material can be transferred from and/or exchanged among all groups of organisms (Gogarten et al. 2009), with varying frequency (= how often the transfer and/or exchange occurs), magnitude (= how much genetic material is transferred and/or exchanged), and quality (= fitness value of the transferred and/or exchanged genes). Table 1 can help us make further generalizations and inferences concerning HGT in the context of graphic depictions of evolution. (1) In unicellular organisms HGT is frequent (i.e. it occurs often during an individual's life time), it involves low to high transfer/exchange of genetic material relative to the host's genome (i.e. from few to many genes), and genes acquired horizontally can be integrated rapidly into the population (i.e. depending on their fitness value) due to usually short generation times; therefore, when attempting to reconstruct a phylogenetic tree of unicellular taxa, HGT can obscure the detection of phylogenetic relations if genes acquired horizontally and vertically have become integrated. (2) A reticulated spatio-temporal pattern of genetic interconnection, both horizontally and phylogenetically acquired, among clusters of related unicellular organisms, or ''taxonomic units,'' is plausible (Bapteste and Boucher 2009), and it should be depicted and incorporated into graphic representations of evolutionary history (Bapteste and Boucher 2009; Lopez and Bapteste 2009). (3) Although viruses indeed transport genetic material into multicellular organisms, and at high frequency, the integration of such material into eukaryotic germ line or gametes is lower than in prokaryotes (Siefert 2009); the division of eukaryotic cells into somatic cells and gametes limits the number of vectors that can facilitate gene movement from soma to gametes (Siefert 2009). Note that transposable elements are common and detectable in multicellular genomes, and their presence does not blur significantly phylogenetic reconstructions (i.e. tree-like topologies) (Andersson 2008; Siefert 2009). (4) Most vascular plants and a few animal species (above) can integrate genetic material of lateral origin via hybridization, where the magnitude of the genetic share among taxonomic units can be high, but not as high as in prokaryotes (Schwenk et al. 2008; Soltis

and Soltis 2009). And (5) because spatio-temporal clusters or populations of multicellular organisms are conspicuous to human sampling, scientists can detect more effectively the identity of hybrid genes in plant or animal taxa than in prokaryotes (Grant and Grant 2008; Schwenk et al. 2008; Soltis and Soltis 2009); therefore, tree-like depictions of evolutionary relationships among many plants and animals can illustrate the larger vertical inheritance of genes in respect to those acquired laterally (Keeling and Palmer 2008; Koonin 2009; Lopez and Bapteste 2009).

THE TREE OF LIFE COALESCES TO A RETICULATED PATTERN OF UNICELLULAR HGT

Horizontal gene transfer allowed early unicellular organisms to increase genetic variability faster than would have been the case using only the vertical acquisition and transfer of favorable mutations (Andersson 2008; Keeling and Palmer 2008; Lopez and Bapteste 2009). It is plausible that constant reproductive isolation during colonization of new environments, or simply separation by distance, could have led to reticulated patterns of unicellular life (RPUL) that evolved as clusters of archaeal taxonomic units (CATUs) or bacterial taxonomic units (CBTUs) (Fig. 5). These CATUs and CBTUs are in accordance with the ''pattern pluralism scheme'' (Bapteste and Boucher 2009; Doolittle and Bapteste 2007) depicted as a cluster (= taxonomic unit) of three archaeal or bacterial varieties, represented by the distinct ecological dimensions in which they evolved (Fig. 5: black, gray, and white circles for Archaea, and squares for bacteria). Because HGT interacts closely with vertical evolution, varieties can be grouped by *shape* (see Fig. 5) or *color* (not depicted), thus appearing in various natural and non-exclusive taxonomic units. Environmental contingencies coupled with natural selection could drive these varieties and taxonomic units along different paths of specialization, giving rise later to the Archaea and Eubacteria. Indeed, the last eukaryotic common ancestor (LECA; Fig. 5) could have originated from the symbiogenetic junction of ''archaeal'' and ''bacterial'' cells (Margulis 2009; Margulis et al. 2006); later clusters of eukaryotic taxonomic units (CETUs) interconnected by HGT are, therefore, also conceivable. Our integrative model (Fig. 5) takes its background as the ''ring of life'' (Rivera and Lake 2004). This emphasizes the ongoing HGT during all stages of this evolutionary process. Symbiogenesis explains ''anastomosis'' (Margulis 2009) between genomes: for example, eukaryotes *with* oxygen-respiring bacteria, triggered the evolution of mitochondria; and the later association between oxygen-respiring eukaryotes *with* photosynthetic bacteria led to the origin of chloroplasted cells (Fig. 6).

Multicellular organisms rely mostly on vertical inheritance to acquire genetic material; the combined coding plus non-coding regions of their genomes can be huge ($r = 10^7$–10^{10} bp) in comparison to unicellular life ($r = 10^5$–10^7 bp) (Koonin 2009; Lynch 2006). The transition from uni- to multicellularity was likely paralleled by a gradual decline in frequency and magnitude of HGT within multicellular lineages; their increasingly complex and highly integrated genomes might have become more difficult to transfer horizontally than vertically (Lake 2009). Note that HGT is a crucial adaptive feature of modern Archaea and bacteria, and that their genomes profit from continuous lateral influx of genes, which has allowed them to radiate into ecologically diverse niches (Lawrence 2002; Lopez and Bapteste 2009). The benefits of HGT, largely acquisition of genetic diversity, have been replaced among multicellular organisms by recombination during sexual reproduction. If they are genetically compatible, multicellular individuals from close or distant populations can reproduce and leave fertile offspring. Hybridization usually occurs within relatively close genetic lineages (Fig. 6). Thus, tree-like representations of relatedness among multicellular plants and animals might reflect a natural process of adaptive divergence, which coalesces to a

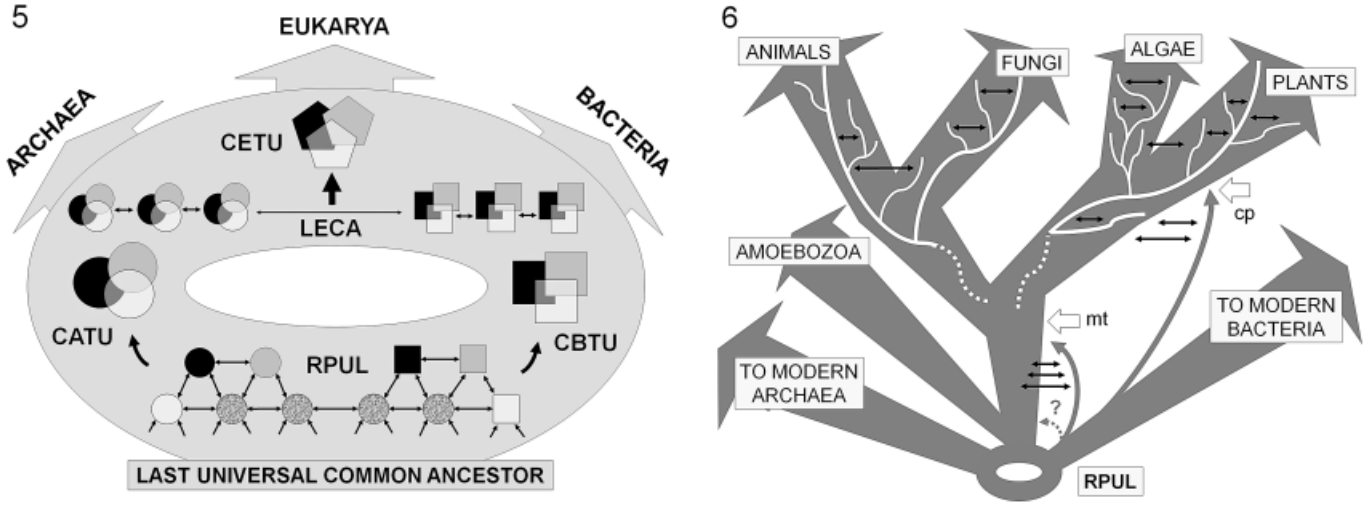

Fig. **5–6.** Integrative model of lateral and vertical evolution. **5.** Reticulated patterns of unicellular life (RPUL), connected by both horizontal gene transfer and common descent (arrows), emerged from a last universal common ancestor (LUCA) and diverged into clusters of archaeal (CATU, overlapping circles) or bacterial (CBTU, overlapping squares) taxonomic units; the last eukaryotic common ancestor (LECA) probably originated from symbiogenetic junction of archaeal and bacterial cells; overlapping pentagons represent clusters of eukaryotic taxonomic units (CETU); the ''ring of life'' depicted in the background emphasizes ongoing HGT. **6.** A tree-like representation of relatedness among multicellular organisms coalesces to a RPUL, note how modern archaea and bacteria also coalesce to RPUL; ''anastomosis'' between eukaryotes with oxygen-breathing bacteria probably gave origin to mitochondria (mt, white arrow; a much earlier event is also possible: gray-dashed arrow with question mark on top), and between oxygen-breathing eukaryotes with photosynthetic bacteria led to chloroplast evolution (cp white arrow); hybridization (lateral exchange of genetic material via recombination) is probably more frequent in plants, algae or fungi than in animals (black horizontal arrows); a bifurcation pattern of diversification (white branching lines) is probably real among multicellular organisms because they acquire genetic material mostly via common descent. Note that only Amoebozoa are shown to account for the discussion on the origin of *Entamoeba*'s alcohol dehydrogenase (see text); however, other unicellular Eukaryotes—Excavata, Chromalveolata, and Rhizaria—(Andersson 2008, 2009) likely emerged from early CETU.

reticulated pattern of unicellular organisms interconnected by HGT at the earliest stages of unicellularity (Fig. 6).

EhADH2 ORIGIN: AN EXAMPLE OF HGT AND VERTICAL EVOLUTION

Entamoeba histolytica alcohol dehydrogenase 2 is a bifunctional enzyme essential for energy production. It evolved from the fusion of aldehyde dehydrogenase (ALDH) and ADH (Espinosa et al. 2001). *Entamoeba* spp. lack mitochondria and use EhADH2 to ferment glucose; the last stages of this process convert acetyl-CoA into acetaldehyde, which is catalyzed by ALDH, and acetaldehyde into ethanol, which is mediated by ADH (Chen, Li, and Stanley 2004; Espinosa et al. 2001, 2009). The catalytic activities of EhADH2 reside in two separate, interacting domains—the N-terminal ALDH and the C-terminal ADH (Fig. 7).

Comparative analysis of both genomic sequences of *E. histolytica*, in respect to other amoebozoans, and amino acid sequences of EhADH2, in respect to 49 other alcohol dehydrogenase E (ADHE) proteins from entamoebids, bacteria, green algae, pelobionts, diplomonads, fungi, and apicomplexans (Andersson et al. 2006), suggest a prokaryote-to-*Entamoeba* origin of the ancestral forms of EhADH2, probably acquired via HGT (Espinosa et al. 2001). This inference relies on the following evidence: (1) Genomic analysis of small subunit rRNA sequences places *E. histolytica* within the eukaryotic amoebozoa, along with *Mastigamoeba*, *Pelomyxa*, *Phreatamoeba*, and others (Lecointre and Le Guyader 2006); *E. histolytica* and *Mastigamoeba balamuthi* cluster together as sister taxa according to a 123-gene analysis of undisputed orthologs (Bapteste et al. 2002). (2) The protein sequences of ADHE from *Entamoeba terrapinae*, *Entamoeba invadens*, *Entamoeba moshkovskii*, and *E. histolytica* branch together next to a cohesive cluster of low-G+C Gram-positive and γ-proteobacteria (i.e. *Streptococcus* spp., *Mannhelmia* sp., *Pasteurella* sp., and *Actinobacillus* sp.; Andersson et al. 2006), suggesting that ancient amoeba most likely ingested, via phagotrophism, prokaryotes capable of glucose fermentation, and later integrated early bacterial *adhE* genes into the ancestral *Entamoeba* genome (Espinosa et al. 2001). (3) This lateral acquisition of proteobacterial forms of *adhE* within the genus *Entamoeba* must have occurred before the differentiation of *Entamoeba* into the four varieties above, because they branch together in the protein sequence analysis (Andersson et al. 2006). (4) The *adhE* sequence of *M. balamuthi*, the close relative of *Entamoeba* spp., branches distant from *Entamoeba*'s spp., in a different region of the phylogenetic tree and proximate to green algae (i.e. *Chlamydomonas reinhardii* and *Polytomella* sp.) and cyanobacteria (*Thermosynechococcus elongatus*; Andersson et al. 2006), corroborating the independent lateral acquisition by HGT of ancestral forms of EhADH2 by the *Entamoeba* cluster. A lateral origin of EhADH2 from a prokaryotic source, coupled with vertical evolution within *Entamoeba* spp. seems the most parsimonious explanation for its origin. Furthermore, about 70 *Entamoeba* genes, seven of which are involved in energy metabolism, have sequence features consistent with HGT origin (Alsmark et al. 2009).

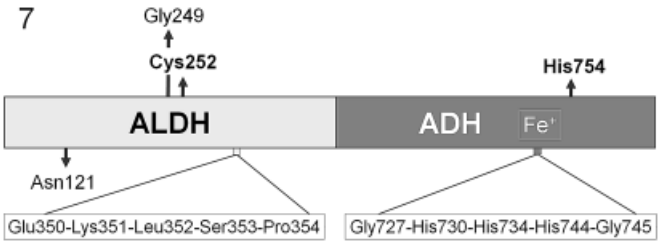

Fig. **7.** Schematic diagram of *Entamoeba histolytica* alcohol dehydrogenase 2 (EhADH2). The enzymatic activities reside in two separate, interacting domains: N-terminal aldehyde dehydrogenase (ALDH) and C-terminal alcohol dehydrogenase (ADH) (catalytic residues: Cys252 and His754, respectively). Conserved amino acids essential for each domain's function are shown; the iron-binding region in ADH is crucial for EhADH2 activity and interaction between domains (Espinosa et al. 2001, 2009).

HGT AND COMMON DESCENT VS. ''COMMON DESIGN''

Entamoeba histolytica alcohol dehydrogenase 2 is an 870 aa ADHE (N-terminal ALDH = 446 aa, C-terminal ADH = 424 aa; Fig. 7), which has ≈ 40% similarity with the ADHE sequences of *Mastigamoeba*, *Spironucleus*, *Giardia*, *Piromyces* (Boxma et al. 2004) and ≈ 60% aa identity with bacteria (e.g. *Streptococcus* spp., above; Field, Rosenthal, and Samuelson 2000). Like most ADHE fusion enzymes, EhADH2 requires NAD^+ and iron as cofactors.

Aldehyde dehydrogenase catalyses the conversion of acetyl-CoA into acetaldehyde; an invariable cysteine in position 252 (Cys252) is the catalytic residue, assisted by a glutamate in position 350 (Glu350), which extracts a proton from Cys252 to initiate catalysis (Fig. 7). Four other conserved amino acids seem crucial to the activity of ALDH: asparagine (Asn121), involved in binding the cofactor NAD^+; glycine (Gly249), which is proximate to Cys252 and presumed to be important to this catalytic residue; and both leucine (Leu352) and proline (Pro354), which are proximate to Glu350 and likely associated with it (Chen et al. 2004). The sequence Glu350-Lys351-Leu352-Ser353-Pro354 is conserved in most members of the ADHE family (Atteia et al. 2003; Boxma et al. 2004; Chen et al. 2004). Alcohol dehydrogenase catalyses the conversion of acetaldehyde, the end product of the reaction above, into ethanol; an invariable histidine in position 754 (His754) is the catalytic residue for this reaction (Espinosa et al. 2001, 2009) (Fig. 7). Three other histidines (His730, 734, and 744), located within an iron-binding region (Gly727–Gly745), facilitate binding to the iron cofactor (Espinosa et al. 2001, 2009; Espinosa, Clark, and Stanley 2004).

Aldehyde dehydrogenase and ADH are interdependent within EhADH2. Although in vitro experiments report 25% activity of recombinant ALDH (Chen et al. 2004) and full activity of recombinant ADH (Espinosa et al. 2001), compared with wild-type EhADH2, site-directed mutagenesis of the ALDH residues Cys252 and Glu350, or the ADH residues His730, 734, 744, and 754, impact significantly each other's catalytic performance, suggesting a synergistic interaction between the two domains, crucial for the survival of *E. histolytica* (Chen et al. 2004; Espinosa et al. 2001, 2009; Espinosa et al. 2004).

How can the origin and evolution of EhADH2 and, moreover, the ADHE enzymes be explained? A random process is improbable for the following reasons. (1) The probability of arriving at random to the correct arrangement of 2,610 nucleotides, excluding introns, in the genetic code for EhADH2 is equal to the allocation of any of the four nucleotides (A, G, C, T) each multiplied by four per nucleotide position, or 4×4 two-thousand six-hundred and ten times ($4^{2,610}$); the probability of generating by chance the correct codon sequence for the 870 aa of EhADH2, plus one stop codon, is equal to 64 (the number of codons in the genetic code) multiplied by 64 eight-hundred and seventy-one times (64^{871}). The entire process would require 2,200 Myr of evolution, assuming a mutation rate of one nucleotide every 850,000 years and 730 generations per year. This didactic estimate is based on an average mutation rate of 1.6 bp every 10^9 nucleotides per generation (Lynch 2006). But mutations are complex, occur in clusters,

at different rates within and between genes (= ''hot spots'' in the genome), and networks of genes can coevolve, thus increasing and maintaining informational complexity, decreasing uncertainty, and expediting evolution (for detailed discussions on computational methods and theoretical implications see Durrett and Schmidt 2008; Lynch 2005, 2006; Schneider 2000; Stern and Orgogozo 2009). Indeed, sequence analysis of 96 *Entamoeba* spp. genes suggests only 50.5 ± 13.5 Myr divergence between *E. histolytica* and related genera (Roy, Irimia, and Penny 2006); *E. histolytica* is a relatively recent eukaryote! (2) Because nucleotide transitions (A/G to G/A, or C/T to T/C) are more probable than transversions (purine to/from pyrimidine), due to structural and polar affinity between complementary bases, the sole random arrival at the correct arrangement of the 2,610 nucleotides of EhADH2 would be reduced to 1 in 2, rather than 1 in 4 (above), nucleotides per complementary position of DNA sequence, or $2^{2,610}$ (a much faster process than $4^{2,610}$ but still exceedingly improbable!). Note also that redundancy in codon coding (i.e. 9 aa are coded by two codons each, five by four, three by six, one by three, and two by one) determines differential probability of amino acid site allocation; for example, amino acids coded by two codons each (i.e. Phe, Tyr, His, Gln, Asn, Lys, Asp, Glu, and Cys) have a 2/64 probability of being allocated in a peptide sequence; in contrast, amino acids coded by six codons each (Leu, Ser, Arg) have a 6/64 probability of participating in the protein. This implies that amino acids coded by six codons each would be 3 times more frequent in EhADH2 than those coded by two codons each. However, this is not the case (Table 2): amino acids coded by six codons each occur at an average frequency of 4.9% ($r = 3.3$–6.4), rather than the 9.3% expected by chance, and amino acids coded by two codons each occur at an average frequency of 4.2% ($r = 2.0$–7.6), rather than the 3.1% expected by chance (Table 2). The same discrepancy between observed and expected frequency of occurrence applies to the rest of the amino acids of EhADH2: those coded by four codons each (Gly, Thr, Ala, Val, and Pro) occur at an average frequency of 7.0% ($r = 4.9$–11.7), rather than the 6.2% expected by chance; only Ile is coded by three codons and occurs at a frequency of 7.1%, rather than the 4.6% expected by chance, while Met and Trp are coded by one codon each and occur at frequencies of 4.1% and 0.8%, respectively, rather than the 1.5% expected by chance ($\chi^2 = 275.1$, df = 19, $P \leq 0.001$) (Table 2). (3) The non-random pattern of codon sequence is also evident in the third codon position of EhADH2: 90% AT vs. 10% GC (data generated from nucleotide sequence; NCBI-GenBank U04863.1), rather than the 1:1 ratio expected by chance, a higher AT-bias than the entire *E. histolytica*'s genome (75–84% AT, 25–16% GC; Char and Farthing 1992; Gelderman et al. 1971; Kocik, Sobczak, and Redowicz 2006; Tannich and Horstmann 1992); selection at translation has favored this codon bias composition of EhADH2, which is characteristic of the *Entamoeba* lineage (Romero, Zavala, and Musto 2000) and correlated with its generation times or parasitic/free-living life styles (Dos Reis and Wernisch 2009; Subramanian 2008). (4) Although randomness can be a statistical component of mutation rate, synergistic biological restrictions (e.g. structural compatibility of purine: pyrimidine pairing in DNA; differential codon representation per amino acid; site specificity for cofactor and substrate binding; AT-rich codon bias) impose directionality on molecular assemblage, and EhADH2 is no exception. Natural selection has tinkered molecular improvements in ancestors of EhADH2 by favoring and retaining an adaptive peptide sequence that promotes optimal function, a classical trajectory from simple evolutionary pathways to complex proteins (Lynch 2005).

Table **2.** EhADH2 amino acid (aa) composition[a] ($N = 870$ aa, plus one stop codon).

aa	Number codons in genetic code[b]	Observed[c] number aa (%) in EhADH2	Expected[d] number aa (%) in EhADH2
Gly	4	53 (6.08)	54 (6.25)
Ser	6	44 (5.05)	82 (9.37)
Thr	4	43 (4.93)	54 (6.25)
Cys	2	18 (2.06)	27 (3.12)
Tyr	2	29 (3.32)	27 (3.12)
Asn	2	43 (4.93)	27 (3.12)
Gln	2	22 (2.52)	27 (3.12)
Lys	2	67 (7.69)	27 (3.12)
Arg	6	29 (3.32)	82 (9.37)
His	2	21 (2.41)	27 (3.12)
Asp	2	37 (4.24)	27 (3.12)
Glu	2	59 (6.77)	27 (3.12)
Ala	4	102 (11.71)	54 (6.25)
Val	4	64 (7.34)	54 (6.25)
Leu	6	56 (6.42)	82 (9.37)
Ile	3	62 (7.11)	41 (4.68)
Pro	4	44 (5.05)	54 (6.25)
Met	1	36 (4.13)	14 (1.56)
Phe	2	34 (3.90)	27 (3.12)
Trp	1	7 (0.80)	14 (1.56)
Stop	3	1 (0.11)	41 (4.68)
Total	64	871 (100)	869 (100)

[a]Protein sequence available at NCBI (GenBank AAA81906.1).

[b]Corresponds to the number of codons coding for a specific amino acid.

[c]The observed composition of amino acids in EhADH2 differs from what would be expected by chance ($\chi^2 = 275.1$, df = 19, $P \leq 0.001$; the stop codon was excluded), which suggests that directional selection has favored and retained adaptive peptide sequence for optimal function.

[d]Expected values were estimated from the percentile differential allocation of codons coding for each amino acid or stop signal in the genetic code.

EhADH2, *Entamoeba histolytica* alcohol dehydrogenase 2.

Mutation rate coupled with natural selection and HGT coupled with common descent suffice to explain the origin and diversification of EhADH2 in the *Entamoeba* lineage. But the production of novel protein features that require the participation of specific peptide sequences, via classical evolutionary trajectories, has been challenged by proponents of ''intelligent design'' or ID (Behe 1998, 2001, 2007, 2009; Behe and Snoke 2004, 2005). ID authors attribute prevalent randomness to molecular change, deleterious nature to intermediate mutations in single-copy genes (rather than neutrality or selective advantage), insufficient geological time or population size for molecular improvements to occur, and invoke ''common design'' or ''separate ancestry'' of complex molecular structures (Luskin and Gage 2008) and major taxonomic lineages (Nelson 1996). Proponents of ID diminish the editing role of natural selection on mutation rate and hypothesize supernatural causation to life's essential molecular processes (i.e. Behe 2001). This logic has been dismissed by researchers (Durrett and Schmidt 2008, 2009; Forrest and Gross 2007; Long et al. 2003; Lynch 2005; Pennock 2001; Petto and Godfrey 2007; Schneiderman and Allmon 2009; Young and Edis 2004) and journal editors (Hermodson 2005) who base their criticism of ID on three fundamental premises of molecular evolution. (1) Large variation in mutation rate between and within lineages, and/or protein sites, is susceptible to positive selection (e.g. the complex diversity of the ADHE enzymes across both prokaryotic and eukaryotic taxa, which differ in their natural history and patterns of HGT, has evolved under the selective pressure of anaerobic respiration). (2) Protein-site mutagenesis is associated with mutation and acceptance rates at multiple sites in a genome, called ''compensatory changes'' (e.g. the evolution of EhADH2 seem to have co-depended on the evolution of other alcohol and ALDHs in the *E. histolytica* genome, including EhADH1, EhADH3, and EhALDH1; Espinosa et al. 2001). (3)

New protein functions after domain junction can experience faster evolution (e.g. fused genes of consecutive enzymes in metabolic pathways; see Yanai, Wolf, and Koonin 2002; in our case, this corresponds to the concerted evolution of the two domains of EhADH2, in which ALDH converts acetyl-CoA into acetaldehyde, and ADH converts the latter into ethanol; Espinosa et al. 2001, 2009; Chen et al. 2004).

Our case study of lateral and vertical evolution of EhADH2, a likely exemplar of the ADHE enzymes, is didactic in the context of the Darwinian perspective. Selection has acted continuously and cumulatively on ancestors and intermediates of EhADH2. Therefore, a single or multiple emergence of EhADH2 arising from an ''intelligent design'' followed by adaptive change is improbable.

CONCLUSIONS

Tree-like representations of relatedness among multicellular organisms might reflect a natural process, which coalesces to a reticulated pattern as we retreat back in time to the origin of unicellular life, which may have been highly interconnected by HGT. The frequency and magnitude of HGT probably declined during the evolutionary transition from uni- to multicellularity. Horizontal gene transfer introduced speed into molecular evolution by facilitating genetic exchange among unicellular organisms; this feature was later replaced by recombination of genes during sexual reproduction in multicellular life. The bifunctional enzyme, EhADH2, is an example of ancestral fusion of enzymatic domains (ALDH+ADH), which became interdependent and laterally transferred among unicellular predecessors of *Entamoeba* spp. and later inherited via common descent within the *Entamoeba* lineage. The interaction between mutation rate and natural selection, enhanced by the co-occurrence of HGT and common descent, suffices to explain the origin and evolution of EhADH2 and probably of most ADHE proteins.

ACKNOWLEDGMENTS

We thank the International Society of Protistologists—North American Section for sponsoring the workshop, HGT and Phylogenetic Evolution Debunk Intelligent Design, and Roger Williams University for hosting it. We benefited from discussions with Mark Farmer and Andrea Habura. Two reviewers provided comments that helped us improve the manuscript. G. Paz-y-Miño C. is supported by the Center for Teaching Excellence at the University of Massachusetts, Dartmouth, and A. Espinosa by NIH-NCRR grant #2 P20RR16457-04.

LITERATURE CITED

Alsmark, U. C., Sicheritz-Ponten, T., Goster, P. G., Hirt, R. P. & Embley, T. M. 2009. Horizontal gene transfer in eukaryotic parasites: a case study of *Entamoeba histolytica* and *Trichomonas vaginalis*. *In:* Gogarten, M. B., Gogarten, J. P. & Olendzenski, L. (ed.), Horizontal Gene Transfer: Genomes in Flux. Humana Press, New York. p. 489–500.

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

Andersson, J. O. 2008. Eukaryotic gene transfer: adaptation and replacements. *In*: Hensel, M. & Schmidt, H. (ed.), Horizontal Gene Transfer in the Evolution of Pathogenesis. Cambridge University Press, Cambridge, MA. p. 293–315.

Andersson, J. O. 2009. Horizontal gene transfer between microbial eukaryotes. *In*: Boekels Gogarten, M., Gogarten, J. P. & Olendzenski, L. (ed.), Horizontal Gene Transfer: Genomes in Flux. Humana Press, New York. p. 473–487.

Andersson, J. O., Hirt, R. P., Foster, P. G. & Roger, A. J. 2006. Evolution of four gene families with patchy phylogenetic distributions: influx of genes into protist genomes. *BMC Evol. Biol.*, **6**:27.

Atteia, A., van Lis, R., Mendoza-Hernandez, G., Henze, K., Martin, W., Riveros-Rosas, H. & Gonzalez-Halphen, D. 2003. Bifunctional aldehyde/alcohol dehydrogenase (ADHE) in chlorophyte algal mitochondria. *Plant Mol. Biol.*, **53**:175–188.

Bapteste, E. & Boucher, Y. 2009. Epistemological impacts of horizontal gene transfer on classification in microbiology. *In:* Gogarten, M. B., Gogarten, J. P. & Olendzenski, L. (ed.), Horizontal Gene Transfer: Genomes in Flux. Humana Press, New York. p. 55–72.

Bapteste, E., Brinkmann, H., Lee, J. A., Moore, D. B., Sense, C. W., Gordon, P., Duruflé, L., Gaasterland, T., Lopez, P., Müller, M. & Philippe, H. 2002. The analysis of 100 genes supports the grouping of three highly divergent amoebae: *Dictyostelium*, *Entamoeba*, and *Mastigamoeba*. *Proc. Natl. Acad. Sci. USA*, **99**:1414–1419.

Behe, M. J. 1998. Molecular machines: experimental support for the design inference. *Cos. Pur.*, **1**:27–35.

Behe, M. J. 2001. Reply to my critics: a response to reviews of Darwin's Black Box: the biochemical challenge to evolution. *Biol. Philos.*, **16**:685–709.

Behe, M. J. 2007. The Edge of Evolution. Free Press, New York, 336 p.

Behe, M. J. 2009. Waiting longer for two mutations. *Genetics*, **181**:819–820.

Behe, M. J. & Snoke, D. W. 2004. Simulating evolution by gene duplication of protein features that require multiple amino acid residues. *Prot. Sci.*, **13**:2651–2664.

Behe, M. J. & Snoke, D. W. 2005. A response to Michael Lynch. *Prot. Sci.*, **14**:2226–2227.

Boxma, B., Voncken, F., Jannink, S., van Alen, T., Akhmanova, A., van Weelden, S. W. H., van Hellemond, J. J., Ricard, G., Huynen, M., Tielens, A. G. M. & Hackstein, J. H. P. 2004. The anaerobic chytridiomycete fungus *Piromyces* sp. E2 produces ethanol via pyruvate: formate lyase and an alcohol dehydrogenase E. *Mol. Microb.*, **51**:1389–1399.

Char, S. & Farthing, M. 1992. Codon usage in *Entamoeba histolytica*. *Int. J. Parasitol.*, **22**:381–383.

Chen, M., Li, E. & Stanley, S. L. 2004. Structural analysis of the acetaldehyde dehydrogenase activity of *Entamoeba histolytica* alcohol dehydrogenase 2 (EhADH2), a member of the ADHE enzyme family. *Mol. Biochem. Parasitol.*, **137**:201–205.

Darwin, C. 1859. On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life. 1st ed. John Murray, London, 502 p.

Doolittle, W. F. 1999. Phylogenetic classification and the universal tree. *Science*, **284**:2124–2128.

Doolittle, W. F. & Bapteste, E. 2007. Pattern pluralism and the tree of life hypothesis. *Proc. Natl. Acad. Sci. USA*, **104**:2043–2049.

Dos Reis, M. & Wernisch, L. 2009. Estimating translational selection in eukaryotic genomes. *Mol. Biol. Evol.*, **26**:451–461.

Durrett, R. & Schmidt, D. 2008. Waiting for two mutations: with applications to regulatory sequence evolution and the limits of Darwinian evolution. *Genetics*, **180**:1501–1509.

Durrett, R. & Schmidt, D. 2009. Reply to Michael Behe. *Genetics*, **181**:821–822.

Espinosa, A., Clark, D. & Stanley, S. L. 2004. *Entamoeba histolytica* alcohol dehydrogenase 2 (EhADH2) as a target for anti-amoebic agents. *J. Antimicrob. Chemother.*, **54**:56–59.

Espinosa, A., Perdrizet, G., Paz-y-Miño, C. G., Lanfrachi, R. & Phay, M. 2009. Effects of iron depletion on *Entamoeba histolytica* alcohol dehydrogenase 2 (EhADH2) and trophozoite growth: implications for antiamoebic therapy. *J. Antimicrob. Chemother.*, **63**:675–678.

Espinosa, A., Yan, L., Zhang, Z., Foster, L., Clark, D., Li, E. & Stanley, S. L. 2001. The bifunctional *Entamoeba histolytica* alcohol dehydrogenase 2 (EhADH2) protein is necessary for amebic growth and survival and requires an intact C-terminal domain for both alcohol dehydrogenase and acetaldehyde dehydrogenase activity. *J. Biol. Chem.*, **276**:20136–20143.

Field, J., Rosenthal, B. & Samuelson, J. 2000. Early lateral transfer of genes encoding malic enzyme, acetyl-CoA synthetase and alcohol dehydrogenases from anaerobic prokaryotes to *Entamoeba histolytica*. *Mol. Microbiol.*, **38**:446–455.

Forrest, B. C. & Gross, P. R. 2007. Biochemistry by design. *Trends Biochem. Sci.*, **32**:301–310.

Gelderman, A., Bartgis, I., Keister, D. & Diamond, L. 1971. A comparison of genome sizes and thermal denaturation-derived base composition of

DNAs from several members of *Entamoeba* (*histolytica* group). *J. Parasitol.*, **57**:912–916.
Gogarten, M. B., Gogarten, J. P. & Olendzenski, L. (ed.),. 2009. Horizontal Gene Transfer: Genomes in Flux. Humana Press, New York, 500 p.
Grant, B. R. & Grant, P. R. 2008. Fission and fussion of Darwin's finches populations. *Phil. Trans. R. Soc. B.*, **363**:2821–2829.
Hensel, M. & Schmidt, H. (ed.),. 2008. Horizontal Gene Transfer in the Evolution of Pathogenesis. Cambridge University Press, New York, 342 p.
Hermodson, M. 2005. Editorial and position papers. *Prot. Sci.*, **14**:2215–2216.
Keeling, P. J. & Palmer, J. D. 2008. Horizontal gene transfer in eukaryotic evolution. *Nat. Rev. Gen.*, **9**:605–618.
Kocik, E., Sobczak, M. & Redowicz, M. J. 2006. Codon usage in *Amoeba proteus* significantly differs from *Entamoeba histolytica* and *Acanthamoeba castellanii*. *Act. Protozol.*, **45**:313–316.
Koonin, E. V. 2009. Darwinian evolution in the light of genomics. *Nucleic Acids Res.*, **37**:1011–1034.
Lake, J. A. 2009. Evidence for an early prokaryotic endosymbiosis. *Nature*, **460**:967–971.
Lawrence, J. G. 2002. Gene transfer in bacteria: speciation without species? *Theoret. Popul. Biol.*, **61**:449–460.
Lecointre, G. & Le Guyader, H. 2006. The Tree of Life. The Belknap Press of Harvard University Press, Cambridge, 560 p.
Long, M., Betrán, E., Thornton, K. & Wang, W. 2003. The origin of new genes: glimpses from the young and old. *Nature Rev. Gen.*, **4**:865–875.
Lopez, P. & Bapteste, E. 2009. Molecular phylogeny: reconstructing the forest. *C. R. Biol.*, **332**:171–182.
Luskin, C. & Gage, L. P. 2008. A reply to Francis Collins's Darwinian arguments for common ancestry of apes and humans. *In*: House, H. W. (ed.), Intelligent Design 101. Kregel Publications, Grand Rapids, MI. p. 215–235.
Lynch, M. 2005. Simple evolutionary pathways to complex proteins. *Prot. Sci.*, **14**:2217–2225.
Lynch, M. 2006. The origins of eukaryotic gene structure. *Mol. Biol. Evol.*, **23**:450–468.
Margulis, L. 2009. Genome acquisition in horizontal gene transfer: symbiogenesis and macromolecular sequence analysis. *In:* Gogarten, M. B., Gogarten, J. P. & Olendzenski, L. (ed.), Horizontal Gene Transfer: Genomes in Flux. Humana Press, New York. p. 181–191.
Margulis, L., Chapman, M., Guerrero, R. & Hall, J. 2006. The last eukaryotic common ancestor (LECA): acquisition of cytoskeletal motility from aerotolerant spirochetes in the proterozoic eon. *Proc. Natl. Acad. Sci. USA*, **103**:13080–13085.
Mitreva, M., Smant, G. & Helder, J. 2009. Role of horizontal gene transfer in the evolution of plant parasitism among nematodes. *In:* Gogarten, M. B., Gogarten, J. P. & Olendzenski, L. (ed.), Horizontal Gene Transfer: Genomes in Flux. Humana Press, New York. p. 517–535.
Nelson, P. A. 1996. The role of theology in current evolutionary reasoning. *Biol. Philos.*, **11**:493–517.
Pennock, R. T. (ed.),. 2001. Intelligent Design Creationism and its Critics. Massachusetts Institute of Technology, Cambridge, 825 p.
Petto, A. J. & Godfrey, L. R. (ed.),. 2007. Scientists Confront Intelligent Design and Creationism. W. W. Norton & Company, New York, 416 p.
Rivera, M. C. & Lake, J. A. 2004. The ring of life provides evidence for a genome fusion origin of eukaryotes. *Nature*, **431**:152–155.
Romero, H., Zavala, A. & Musto, H. 2000. Compositional pressure and translational selection determine codon usage in the extremely GC-poor unicellular eukaryote *Entamoeba histolytica*. *Gene*, **242**:307–311.
Roy, S. W., Irimia, M. & Penny, S. 2006. Very little intron gain in *Entamoeba histolytica* genes laterally transferred from prokaryotes. *Mol. Biol. Evol.*, **23**:1824–1827.
Schneider, T. D. 2000. Evolution of biological information. *Nucleic Acids Res.*, **28**:2794–2799.
Schneiderman, J. S. & Allmon, W. D. (ed.),. 2009. For the Rock Record: Geologists on Intelligent Design. University of California Press, Berkeley, 272 p.
Schwenk, K., Brede, N. & Streit, B. 2008. Introduction. Extent, processes and evolutionary impact of interspecific hybridization in animals. *Phil. Trans. R. Soc. B.*, **363**:2805–2811.
Siefert, J. L. 2009. Defining the mobilome. *In:* Gogarten, M. B., Gogarten, J. P. & Olendzenski, L. (ed.), Horizontal Gene Transfer: Genomes in Flux. Humana Press, New York. p. 13–27.
Soltis, P. S. & Soltis, D. E. 2009. The role of hybridization in plant speciation. *Annu. Rev. Plant Biol.*, **60**:561–588.
Stern, D. L. & Orgogozo, V. 2009. Is genetic evolution predictable? *Science*, **323**:746–751.
Subramanian, S. 2008. Nearly neutrality and the evolution of codon usage bias in eukaryotic genomes. *Genetics*, **178**:2429–2432.
Tannich, E. & Horstmann, R. 1992. Codon usage in pathogenic *Entamoeba histolytica*. *J. Mol. Evol.*, **34**:272–273.
Yanai, I., Wolf, Y. I. & Koonin, E. V. 2002. Evolution of gene fusions: horizontal transfer versus independent events. *Gen. Biol.*, **3**:13.
Young, M. & Edis, T. (ed.),. 2004. Why Intelligent Design Fails. Rutgers University Press, New Jersey, 240 p.

Received: 08/13/09, 10/13/09; accepted: 10/14/09