Opinion TRENDS in Genetics Vol.20 No.7 July 2004

Evolution of cell–cell signaling in animals: did late horizontal gene transfer from bacteria have a role?☆

Lakshminarayan M. Iyer[1], L. Aravind[1], Steven L. Coon[2], David C. Klein[2] and Eugene V. Koonin[1]

[1]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA
[2]Section on Neuroendocrinology, Laboratory of Developmental Neurobiology, National Institute of Child Health and Development, Bethesda, MD 20894, USA

Cell–cell signaling is essential for the functioning of the nervous, neuroendocrine and immune systems. Crucial for these processes are small, rapidly diffusible messengers including catecholamines (epinephrine, norepinephrine and dopamine), indoles (serotonin and melatonin), histamine, acetylcholine and nitric oxide. In this article, we show that the evolutionary history of most genes encoding enzymes involved in the metabolism of these messengers is best described by scenarios that include horizontal gene transfer (HGT) from bacteria, with some transfers occurring after the divergence of animals from fungi. The acquisition of bacterial genes via HGT seems to have had the essential role of extending existing biochemical pathways to yield the messengers. The possible relatively late HGT of some signaling enzymes contrasts with the apparent acquisition of central metabolic pathway enzymes early in eukaryotic evolution from the proto-mitochondrial endosymbiont.

Cell–cell communication in animal neural, neuroendocrine and immune systems is dependent largely on low-molecular weight first-messenger molecules, which include amino acids, amino-acid derivatives [e.g. norepinephrine (NE), epinephrine (EPI), dopamine (DA), serotonin (5-HT) and melatonin (MEL)], nucleosides (e.g. adenosine) and histamine (HA), diverse lipid and fatty-acid derivatives including acetylcholine (Ach) and inorganic molecules [e.g. nitric oxide (NO) and carbon monoxide (CO)] [1,2]. Typically, messenger synthesis pathways are extensions of the biosynthetic pathways of the respective precursors, in particular, amino acids (Figure 1).

We observed previously that arylalkylamine N-acetyltransferase (AANAT), the enzyme that regulates the daily changes in MEL biosynthesis in vertebrates, was encoded in the genomes of vertebrates, yeasts and several bacteria but apparently not in plants, worms or flies [3,4]. This led to speculation that the evolution of AANAT might have involved, among other mechanisms, a relatively late horizontal gene transfer (HGT) from bacteria to an ancestral eukaryote [5].

The unusual phyletic pattern of AANAT prompted a systematic investigation of dedicated enzymes of first-messenger metabolism (Figure 1), in an attempt to determine when these pathways emerged and to assess the relative contribution of vertical inheritance versus HGT. The results are consistent with scenarios involving relatively late HGT along with multiple gene losses.

Distribution of messenger-metabolism enzymes among phylogenetic lineages

The patterns of distribution among phylogenetic lineages (phyletic patterns) for vertebrate enzymes involved in messenger metabolism (Figure 1) were derived by searching the non-redundant protein sequence database (http://www.ncbi.nlm.nih.gov) and specialized, organism-specific databases for apparent orthologs of these enzymes (Table 1 and the supplementary material online for details). The resulting picture was non-trivial in that the phyletic patterns of many enzymes resembled that of AANAT (Table 1). Of 17 major enzymes dedicated to messenger metabolism, only two were present in all three major crown-group lineages of eukaryotes and in bacteria [amino-acid decarboxylase (AADC) and histidine decarboxylase (HDC)] and one [dopamine β-hydroxylase (DBH)] appeared to be animal-specific; the remaining enzymes were shared by only one or two eukaryotic lineages and bacteria. It should be emphasized that some of these genes had homologs in other eukaryotes. However, it was shown that these were paralogs, rather than orthologs because they were orthologous to distinct vertebrate genes (Table 1; Table 1 in supplementary material online). None of these enzymes have orthologs in archaea, which (as indicated by phylogenetic analysis of the genes involved in translation, RNA modification, transcription and replication) share a vertically inherited core gene-set with the eukaryotes [6].

The limited and scattered distribution of messenger-metabolism enzymes contrasts with the distribution of

☆ Supplementary data associated with this article can be found at doi: 10.1016/j.tig.2004.05.007

Corresponding author: Eugene V. Koonin (koonin@ncbi.nlm.nih.gov).

Available online 28 May 2004

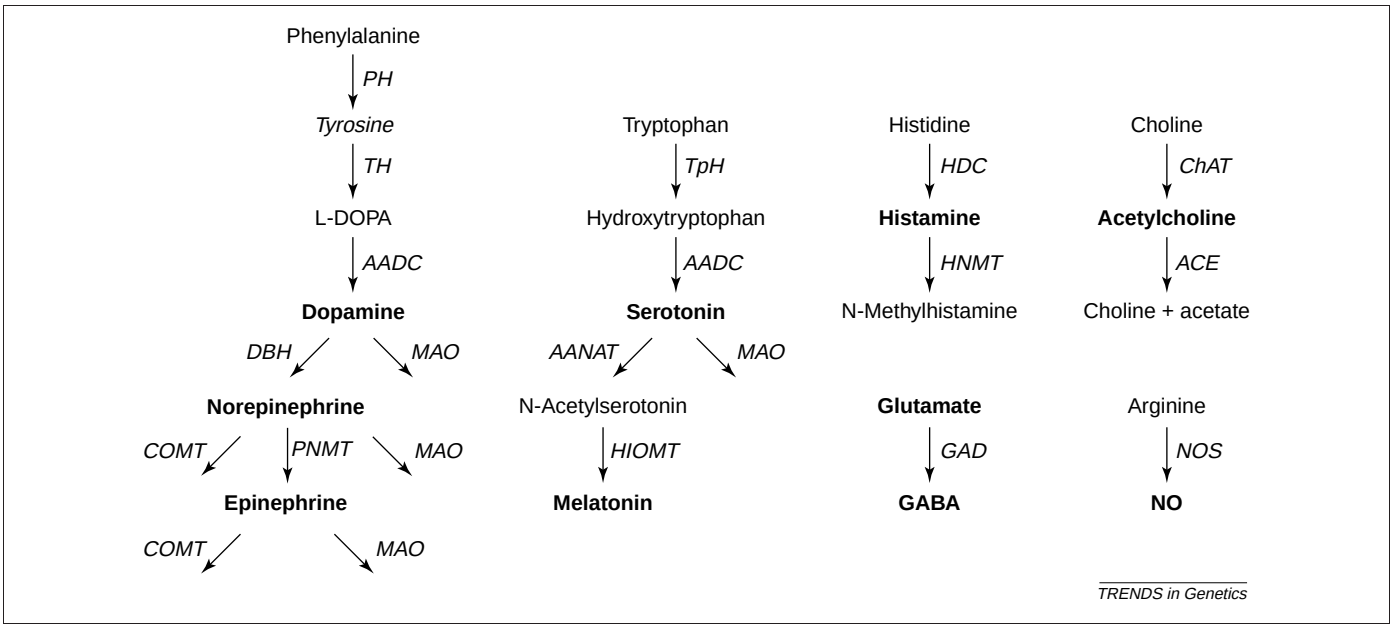


Figure 1. A scheme of metabolic pathways for the principal amino-acid-derived messengers. Compounds in bold are messengers. Abbreviations: AADC, amino-acid decarboxylase; AANAT, arylalkylamine N-acetyltransferase; ACE, acetylcholine esterase; Ach, acetylcholine; COMT, catechol methyltransferase; DA, dopamine; DBH, dopamine β-hydroxylase; EPI, epinephrine; GABA, γ-aminobutyric acid; GAD, glutamate decarboxylase; 5-HT, 5-hydroxytryptamine; HIOMT, hydroxyindole-O-methyltransferase; HDC, histidine decarboxylase; HNMT, histamine N-methyltransferase; *MAO*, monoamine oxidase; Mel, melatonin; NE, norepinephrine; NO, nitric oxide; *NOS*, nitric oxide synthase; TpH, tryptophan hydroxylase.

central metabolic genes, in particular, those encoding amino-acid-metabolizing enzymes, which are typically represented in all major eukaryotic lineages, bacteria and, in many cases, archaea (Table 2 in the supplementary material online). There are two non-mutually exclusive explanations for sporadic phyletic distributions: (i) vertical gene transmission followed by multiple, lineage-specific gene losses; and (ii) HGT.

Table 1. Phyletic patterns of enzymes involved in chemical messenger metabolism[a,b,c]

Enzyme	Gene symbol	Chemical messenger	Bacteria	Archaea	Plants	Fungi	Animals				Other eukaryotes
							Nematodes	Insects	Urochordates	Vertebrates	
Aromatic amino acid decarboxylases	*AADC*	NE, EPI, DA, 5-HT and Mel	+	−	+	+	+	+	+	+	+ (Tepy)
Histidine decarboxylase	*HDC*	Histamine					+	+	+	+	
Glutamate decarboxylase	*GAD*	GABA	+	−	−	+	+	+	+	+	−
Phenylalanine hydroxylase	*PH*	NE, EPI and DA	+	−	−	−	+	+	+	+	(Tgo?)
Tyrosine hydroxylase	*TH*	NE, EPI and DA	+	−	−	−	+	+	+	+	(Tgo?)
Tryptophan hydroxylase	*TpH*	5-HT, Mel	+	−	−	−	+	+	+	+	(Tgo?)
NOS oxygenase domain	*NOS*	NO	+	−	−	−	−	+	+	+	+ (Phpo)
NOS NADPH-P450 and sulfite reductase-like module			+	−	+	+	+	+	+	+	+ (Phpo, Tbr)
Catechol O-methyltransferase	*COMT*	NE, EPI and DA	+	−	−	+	−	−	+	+	−
Serotonin N- acetyltransferase	*AANAT*	Mel	+	−	−	+	−	−	−	+	−
Histamine N-methyltransferase	*HNMT*	Histamine	(+)[d]	−	−	−	−	−	−	+	−
Phenylethanolamine N-methyltransferase	*PNMT*	NE	(+)[e]	−	−	−	+	−	−	+	−
Hydroxyindole O-methyltransferase	*HIOMT*	Mel	+	−	−	−	−	−	−	+	−
Dopamine β hydroxylase	*DBH*	NE, EPI and DA	−	−	−	−	+	+	+	ϖ	−
Monoamine oxidase	*MAO*	NE, EPI, DA and 5-HT	+	−	−	−	−	−	+	+	+ (Ddi)
Choline acetyltransferase	*ChAT*	ACh	(+)[f]	−	−	+	+	+	+	+	−
Acetylcholine esterase	*ACE*	ACh	+	−	−	+	+	+	+ (Cisa)	+	+ (Ddi)

[a]Abbreviations: Ach, acetylcholine; Cisa, *Ciona savignyi*; DA, dopamine; Ddi, *Dictyostelium discoideum*; EPI, epinephrine; GABA, γ-aminobutyric acid; 5-HT, 5-hydroxytryptamine; Mel, melatonin; NE, norepinephrine; NO, nitric oxide; Phpo, *Physarum polycephalum*; Tgo, *Toxoplasma gondii;* Tepy, *Tetrahymena pyriformis* (fragment only); Tbr, *Trypanosoma brucei*.

[b]Typically, the taxa that are denoted in Table 1 as not having a particular enzyme of messenger metabolism actually contain homologs of this enzyme. However, reciprocal searches showed that these were not orthologs because they had distinct counterparts in the genomes containing the given enzyme, thus supporting the phyletic patterns shown in the table (see supplementary material online for details). Obviously, conclusions on the absence of orthologs can be made with confidence only for complete genomes. Incomplete sequences that could not be mapped to the available genomic sequences and indicated by a '?'; these are likely to be contaminants.

[c]The plus and minus symbols represent the presence or absence of the enzyme, respectively.

[d]Only detected in two bacteria, *Streptomyces* and *Trichodesmium*, to date.

[e]Only detected in *Streptomyces* to date.

[f]Only detected in *Mycoplasma* to date – probable horizontal gene transfer (HGT) from eukaryotes.

Phylogenies of messenger-metabolism enzymes

To gain further insight into the evolution of the messenger-metabolism pathways, phylogenetic analysis was performed using several independent tree-construction methods and statistical tests to ensure reliable tree topologies. The results are organized according to the two principal phyletic patterns observed (see supplementary material online).

Genes present in animals, other eukaryotes and bacteria

This pattern is most common among decarboxylases, which catalyze the first or second biosynthetic steps for most amino-acid-derived messengers (Figure 1 and Table 1). The decarboxylase phylogenetic tree has two major clades, one of which consists of aromatic amino-acid decarboxylases (AADCs) and HDCs and the second includes glutamate decarboxylases (GADs) and cysteine sulfinate decarboxylases (CSDs) (Figure 1 in the supplementary material online). The topology of the AADC–HDC clade is best compatible with a mitochondrial origin because the eukaryotic cluster, which includes representatives from animals, fungi and plants, is bracketed by a bacterial branch that consists mostly of α-proteobacteria, the progenitors of the mitochondria [7]. In the GAD–CSD clade, there is no plant representative and no clustering with α-proteobacteria, which suggests acquisition from bacteria at a later stage of evolution but before the radiation of animals and fungi from their common ancestor (Figure 1 in the supplementary material online). Under this scenario, gene loss would explain the absence of these enzymes from yeasts.

The phylogenetic trees of AANAT, catechol methyltransferase (COMT), monoamine oxidase (MAO) and nitric oxide synthase (NOS) each have a eukaryotic clade, which is nested within a diverse group of bacterial homologs (Figure 2a, Figures 3–5 and Figure 10 in the supplementary material online). Such patterns could be explained by HGT from a bacterial source to eukaryotes before the separation of animals and fungi or – for MAO and NOS – before the separation of animals and slime molds; the absence of these genes in non-chordate animals would be attributed to multiple gene losses. However, independent HGT (from bacteria to fungi and/or slime mold; from bacteria to chordate) from similar bacterial sources (accounting for the eukaryotic clade) or serial HGT (from bacteria to fungi and/or slime mold to chordate) could also explain the observed phyletic patterns.

The phylogenetic trees of these enzymes suggest that they originated at different times from diverse bacteria. The AANATs appear to have been derived from Gram-positive bacteria, which show the maximum diversity of these enzymes (Figure 2a and Figure 5 in the supplementary material online). Eukaryotic COMTs cluster with homologs from a heterogeneous bacterial assemblage, which includes several α-proteobacteria and actinomycetes (Figure 4 in the supplementary material online). Thus, eukaryotes might have acquired COMT either from the mitochondrial precursor or from a distinct bacterial endosymbiont, probably of the α-proteobacterial lineage. By contrast, the phylogenetic position of the eukaryotic NOS points to the possibility of acquisition of this gene from a bacterium of the *Thermus–Deinococcus* lineage (Figure 3 in the supplementary material online). MAO belongs to a group of oxidases, which seem to be extremely mobile in bacteria and might have been independently transferred to eukaryotes on several occasions (Figure 10 in the supplementary material online). Therefore, the origin of MAO in chordates and slime mold is hard to pinpoint beyond the general indication of a bacterial provenance.

Genes found only in animals and bacteria

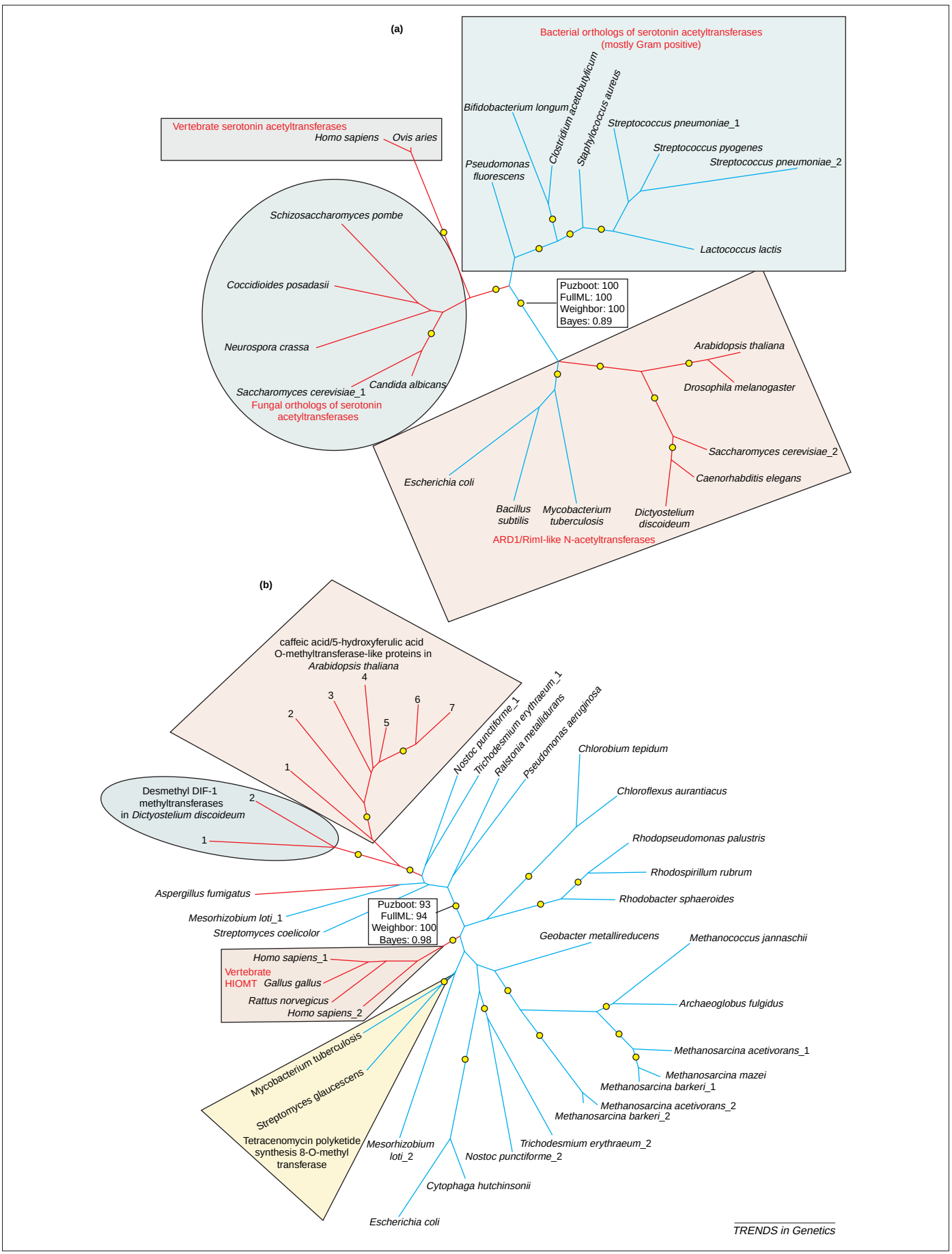
Hydroxyindole-O-methyltransferase (HIOMT) and the three aromatic amino-acid hydroxylases (AAH) are found in animals (HIOMT is found only in vertebrates) and bacteria exclusively (Table 1). HIOMT and AAH form tight clades nested within highly diverse bacterial trees (Figure 2b and Figures 2,8 in the supplementary material online). This pattern is consistent with a late gene transfer from a bacterial source into eukaryotes subsequent to the separation of the animal and fungal lineages. The phylogenetic tree of HIOMT and related methyltransferases suggests extensive horizontal mobility of genes in this family among bacteria, as discussed previously for MAO; this mobility is consistent with several independent events of HGT from bacteria to eukaryotes (Figure 2b and Figure 8 in the supplementary material online).

Origin and evolution of the messenger-metabolism pathways

The above analysis enables us to propose a parsimonious evolutionary scenario for the enzymes of first-chemical-messenger metabolism in chordates (Figure 3a). The enzymes that catalyze the decarboxylation steps of messenger biosynthesis have the broadest phyletic distribution and might have been acquired by eukaryotes as part of the basic amino-acid-catabolism pathways, possibly, from the proto-mitochondrion. Thus, simple amino-acid derivatives, such as γ-aminobutyric acid (GABA) and HA, which were generated by these catabolic decarboxylases, could have been recruited as messengers in different branches of the eukaryotic tree. The glycine decarboxylase complex component (GcvP), which functions as a NOS in plants [8] also belongs in this category (Table 1).

The advent of the pan-metazoan neurotransmitters 5-HT and DA appears to be linked to the acquisition of the common ancestor of the three AAHs. Some bacterial versions of these enzymes are involved in the synthesis of tyrosine from phenylalanine, and their original function after the acquisition by early eukaryotes could have been the same. Subsequent duplications in animals produced enzymes with slightly different specificities. These newly derived hydroxylases combined with the pre-existing decarboxylases yielding 5-HT and DA. Similarly, the evolution of the lipid-derived neurotransmitter Ach apparently involved the recruitment of pre-existing generic lipid-metabolism enzymes, with subsequent duplication and diversification.

The key enzyme of the NO pathway in animals, NOS, emerged through a more complex set of events. The principal animal NOS contains two enzymatic domains,


(a)
Bacterial orthologs of serotonin acetyltransferases
(mostly Gram positive)
Bifidobacterium longum
Clostridium acetobutylicum
Staphylococcus aureus
Streptococcus pneumoniae_1
Streptococcus pyogenes
Streptococcus pneumoniae_2
Pseudomonas fluorescens
Lactococcus lactis
Vertebrate serotonin acetyltransferases
Homo sapiens
Ovis aries
Schizosaccharomyces pombe
Coccidioides posadasii
Neurospora crassa
Candida albicans
Saccharomyces cerevisiae_1
Fungal orthologs of serotonin acetyltransferases
Puzboot: 100
FullML: 100
Weighbor: 100
Bayes: 0.89
Arabidopsis thaliana
Drosophila melanogaster
Saccharomyces cerevisiae_2
Caenorhabditis elegans
Escherichia coli
Bacillus subtilis
Mycobacterium tuberculosis
Dictyostelium discoideum
ARD1/Riml-like N-acetyltransferases
(b)
caffeic acid/5-hydroxyferulic acid
O-methyltransferase-like proteins in
Arabidopsis thaliana
1
2
3
4
5
6
7
Desmethyl DIF-1
methyltransferases
in Dictyostelium discoideum
1
2
Nostoc punctiforme_1
Trichodesmium erythraeum_1
Ralstonia metallidurans
Pseudomonas aeruginosa
Chlorobium tepidum
Chloroflexus aurantiacus
Rhodopseudomonas palustris
Rhodospirillum rubrum
Rhodobacter sphaeroides
Aspergillus fumigatus
Mesorhizobium loti_1
Streptomyces coelicolor
Puzboot: 93
FullML: 94
Weighbor: 100
Bayes: 0.98
Geobacter metallireducens
Methanococcus jannaschii
Homo sapiens_1
Vertebrate HIOMT
Gallus gallus
Rattus norvegicus
Homo sapiens_2
Archaeoglobus fulgidus
Methanosarcina acetivorans_1
Methanosarcina mazei
Methanosarcina barkeri_1
Methanosarcina acetivorans_2
Methanosarcina barkeri_2
Mycobacterium tuberculosis
Streptomyces glaucescens
Tetracenomycin polyketide synthesis 8-O-methyl transferase
Mesorhizobium loti_2
Trichodesmium erythraeum_2
Nostoc punctiforme_2
Cytophaga hutchinsonii
Escherichia coli
TRENDS in Genetics

the N-terminal oxygenase domain and the C-terminal module related to the NADPH-cytochrome P450 and the sulfite reductases. The oxygenase domain was most likely to have been acquired via HGT from bacteria before the divergence of animals and slime molds (Figure 3 in the supplementary material online). The C-terminal module is widespread in eukaryotes (Table 1) and appears to have evolved from a bacterial sulfite reductase (CysJ)-like enzyme, probably acquired from the proto-mitochondrial endosymbiont. The fusion of these modules into a single functional enzyme probably occurred in the animal–slime mold clade. This scenario is compatible with the bacterial phylogenetic affinity of NO and/or CO receptors in animals [9]. Thus, the ancestral eumetazoan apparently had the biosynthetic capabilities for GABA, DA, 5-HT, HA and NO (Figure 3a) and might have employed at least some of these molecules as messengers.

The origin of the biosynthetic pathways for MEL and EPI is more enigmatic. The synthesis of NE and EPI involves two apparent animal-specific innovations [DBH and phenylethanolamine N-methyltransferase (PNMT)] and might be a true evolutionary novelty that emerged concomitantly with the eumetazoan nervous system. The MEL pathway might have emerged *de novo* in chordates following the acquisition of HIOMT from a bacterial source (the identity of which remains unclear), with AANAT having been acquired before the divergence of animals and fungi (Figure 3a). Alternatively, as discussed previously, AANAT – similar to HIOMT – might have been transferred to a primitive chordate directly from bacteria. MAOs, which are important for the enzymatic inactivation of EPI, NE, DA and 5-HT, have been characterized only in chordates to date. However, given the clustering of the slime mold MAO with the chordate orthologs in the phylogenetic tree (Figure 10 in the supplementary material online) [10], the loss of these pathways in invertebrates should be accepted as the conservative scenario. Nevertheless, more unexpected HGT events, such as the secondary acquisition of these enzymes from fungi or slime molds by the chordates, can not be discounted entirely.

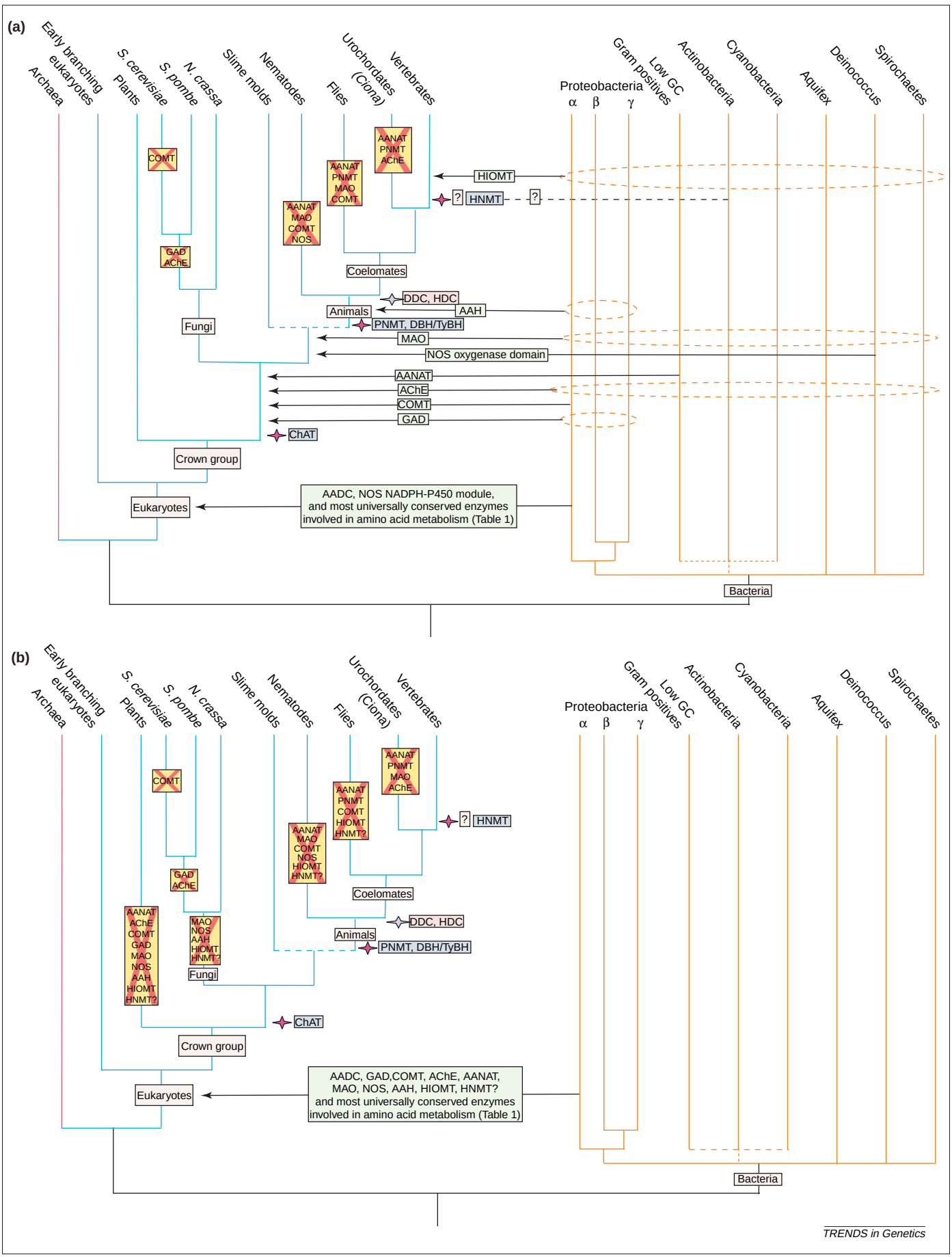
How strong is the case for late HGT in animal first-messenger metabolism?

The evolutionary evidence shows beyond reasonable doubt that chordate enzymes of messenger metabolism were derived from enzymes of bacterial secondary metabolism at some stage(s) during evolution. The evolutionary history of the genes encoding these enzymes, both in bacteria and in eukaryotes, shows extreme plasticity, involving lineage-specific gene loss and HGT. The acquisition of bacterial enzymes by eukaryotes via HGT would not be surprising if these enzymes came from the proto-mitochondrial endosymbiont, the source of many eukaryotic metabolic enzymes [11]. However, the scenario shown in Figure 3a postulates a different timing, namely, relatively late HGT for most of the enzymes that are essential for small molecule signaling. These events might have occurred at the stage of the common ancestor of animals and fungi or at a later stage (i.e. at the onset of animal evolution or even after the divergence of chordates from other animal phyla) (Figure 3a). Another crucial aspect of the postulated scenario is that various enzymes of messenger metabolism might have originated from different groups of bacteria (Figures 2,3a and Figures 1–12 in the supplementary material online).

Rampant HGT during prokaryotic evolution, often between distantly related lineages, is largely accepted [12–15], albeit not universally [16]. Furthermore, it is becoming increasingly apparent that HGT from bacteria was one of the formative processes during the early, unicellular phase of eukaryotic evolution [15,17,18]. Although the proto-mitochondrial endosymbiont (and the chloroplast endosymbiont in plants) might have been the most important source of acquired bacterial genes, it was by no means the only source [15,17,19].

By contrast, HGT from bacteria to animals has been a subject of intense debate, with opinion leaning heavily

Figure 2. Phylogenetic analysis of enzymes of messenger metabolism. The phylogenetic trees were constructed using several independent methods (see supplementary material online). The figure shows the maximum-likelihood trees constructed using the Proml and Puzzleboot (Puzboot) programs. Eukaryotic branches are colored in blue. Nodes with full maximum-likelihood (FullML) bootstrap-support of >80% are marked with yellow circles. In each tree, the support obtained with the four phylogenetic analysis methods (Puzboot, FullML, Weighbor and Bayes see supplementary material online) for the internal node, which was crucial for the conclusions regarding possible horizontal gene transfer (HGT) from bacteria to eukaryotes, is indicated. This node separates the cluster of probable orthologs of the particular enzyme of vertebrate messenger from paralogs. In particular, note that the acetyltransferase from *Drosophila melanogaster* shown in the tree is not identical to the *Drosophila* arylalkylamine N-acetyltransferase (AANAT) [33], which is even more distantly related to the vertebrate AANATs. **(a)** Phylogenetic tree of AANAT. Gene names and gene identification numbers (GI numbers) for the proteins are as follows: *Arabidopsis thaliana*: *At1 g03150*, 18379062; *Bacillus subtilis*: *YdiD*, 16077660; *Bifidobacterium longum*: *Blon1705*, 23336665; *Caenorhabditis elegans*: *4J55*, 17541270; *Clostridium acetobutylicum*: *CAP0111*, 15004814; *Dictyostelium discoideum*: *NATA*, 543838; *Drosophila melanogaster*: *CG14222*, 7293615; *Escherichia coli*: *RimI*, 15804944; *Homo sapiens*: *AANAT*, 4501845; *Lactococcus lactis*: *yveC*, 15674038; *Mycobacterium tuberculosis*: *RimI*, 15610556; *Neurospora crassa*: *NCU05291.1*, 32408333; *Ovis aries*: *AANAT*, 5107555; *Pseudomonas fluorescens*: *Pflu3191*, 23061056; *Saccharomyces cerevisiae_1*: *YDR071C*, 6320276; *Saccharomyces cerevisiae_2*: *ARD1*, 806323; *Schizosaccharomyces pombe*: *SPAC9.02*, 7492407; *Staphylococcus aureus*: *SAV2661*, 15925651; *Streptococcus pneumoniae_1*: *SP1519*, 15901365; *Streptococcus pneumoniae_2*: *spr1372*, 15903415; *Streptococcus pyogenes*: *SPy0346*, 15674502. **(b)** Phylogenetic tree of hydroxyindole-O-methyltransferase (HIOMT). Gene names and GI numbers for the proteins are as follows: *Arabidopsis thaliana* protein 1: *At4 g35160*, 15236282; *Arabidopsis thaliana* protein 2: *At3 g53140*, 15231756; *Arabidopsis thaliana* protein 3: *At1 g51990*, 15218111; *Arabidopsis thaliana* protein 4: *At1 g33030*, 15223364; *Arabidopsis thaliana* protein 5: caffeic acid/5-hydroxyferulic acid O-methyltransferase (*OMT1*), 2781394; *Arabidopsis thaliana* protein 6: *At1 g21100*, 15218133; *Arabidopsis thaliana* protein 7: *At1 g63140*, 25315953; *Archaeoglobus fulgidus*: *AF0429*, 15668257; *Chloroflexus aurantiacus*: *Chlo2720*, 22972859; *Chlorobium tepidum*: *CT0028*, 21672869; *Cytophaga hutchinsonii*: *Chut0176*, 23135057; *Dictyostelium discoideum* protein 1: desmethyl DIF-1 methyltransferase (*DmtA*), 19702458; *Dictyostelium discoideum* protein 2: dd_02962, 28829805; *Escherichia coli*: *ECs4325*, 15833579; *Gallus gallus*: *HIOMT*, 62926; *Geobacter metallireducens*: *Gmet0093*, 23053176; *Homo sapiens_1*: *HIOMT*, 1170276; *Homo sapiens_2*: *ASMTL*, 30172770; *Mesorhizobium loti_1*: *mll8050*, 13476664; *Mesorhizobium loti_2*: *mlr2960*, 13472613; *Methanosarcina acetivorans_1*: *MA3466*, 23052092; *Methanosarcina acetivorans_2*: *MA2773*, 20091596; *Methanosarcina barkeri_1*: *Meth1819* Mba23050288; *Methanosarcina barkeri_2*: *Meth1515*, 23049911; *Methanococcus jannaschii*: *MJ0086*, 15668257; *Methanosarcina mazei*: *MM2572*, 21228674; *Mycobacterium tuberculosis*: *Rv0567*, 15607707; *Nostoc punctiforme_1*: *Npunp4371*, 23128061; *Nostoc punctiforme_2*: *Npun1841*, 23125511; *Pseudomonas aeruginosa*: *PA4209*, 15599404; *Ralstonia metallidurans*: *Reut2533*, 22977792; *Rattus norvegicus*:*HIOMT*, 21426785; *Rhodopseudomonas palustris*: *Rpal2794*, 22963493; *Rhodospirillum rubrum*: *Rrub1010*, 22966411; *Rhodobacter sphaeroides*: *Rsph2198*, 22958616; *Streptomyces coelicolor*: *SC5C11.09c*, 21225721; *Streptomyces glaucescens*: *TCMO*, 730914; *Trichodesmium erythraeum_1*: *Tery3497*, 23042854; *Trichodesmium erythraeum_2*: *Tery4060*, 230434440. Trees for the remaining enzymes of messenger metabolism listed in Table 1 are shown in the supplementary material online.


(a)
Early branching eukaryotes
Archaea
S. cerevisiae
Plants
S. pombe
N. crassa
Slime molds
Nematodes
Flies
Urochordates (Ciona)
Vertebrates
Proteobacteria
α
β
γ
Gram positives
Low GC
Actinobacteria
Cyanobacteria
Aquifex
Deinococcus
Spirochaetes
COMT
GAD AChE
AANAT PNMT AChE
AANAT PNMT MAO COMT
AANAT MAO COMT NOS
HIOMT
? HNMT ?
Coelomates
DDC, HDC
Animals
AAH
PNMT, DBH/TyBH
Fungi
MAO
NOS oxygenase domain
AANAT
AChE
COMT
GAD
ChAT
Crown group
Eukaryotes
AADC, NOS NADPH-P450 module, and most universally conserved enzymes involved in amino acid metabolism (Table 1)
Bacteria
(b)
Early branching eukaryotes
Archaea
S. cerevisiae
Plants
S. pombe
N. crassa
Slime molds
Nematodes
Flies
Urochordates (Ciona)
Vertebrates
Proteobacteria
α
β
γ
Gram positives
Low GC
Actinobacteria
Cyanobacteria
Aquifex
Deinococcus
Spirochaetes
COMT
GAD AChE
AANAT AChE COMT GAD MAO NOS AAH HIOMT HNMT?
MAO NOS AAH HIOMT HNMT?
AANAT PNMT MAO AChE
AANAT PNMT COMT HIOMT HNMT?
AANAT MAO COMT NOS HIOMT HNMT?
? HNMT
Coelomates
DDC, HDC
Animals
PNMT, DBH/TyBH
Fungi
ChAT
Crown group
Eukaryotes
AADC, GAD,COMT, AChE, AANAT, MAO, NOS, AAH, HIOMT, HNMT? and most universally conserved enzymes involved in amino acid metabolism (Table 1)
Bacteria
TRENDS in Genetics

towards the extreme rarity, if not impossibility, of such events. During the analysis of the human genome sequence, >100 genes were detected encoding protein products that showed substantially greater sequence similarity to bacterial than to non-chordate eukaryotic orthologs. In the case of these genes, HGT from bacteria to the ancestral chordate was considered a more likely evolutionary scenario than vertical transfer followed by multiple gene losses [20]. However, subsequent re-analyses that employed additional genomic sequences detected apparent orthologs in non-metazoan eukaryotes and thus appeared to refute the case for bacteria to chordate HGT for many of these genes [10,21–25].

The major problem with evolutionary scenarios that involve both vertical transfer followed by gene loss and HGT is that, in principle, any phyletic pattern could be explained by only one type of event (e.g. gene loss [26]). The question is: how many losses are we willing to postulate to avoid accepting a single HGT event? The answer is far from obvious, although for prokaryotes comparable likelihoods of loss and HGT seem to be a viable possibility [14,27]. This hardly could be the case for multicellular eukaryotes because of the barrier presented by the separation of the germline from the soma. Furthermore, recent analysis of a large set of genes from the cnidarian *Acropora millepora*, a basal metazoan, revealed many genes (although not those encoding chemical-messenger-metabolism enzymes) that were shared with chordates but missing in worm and fly, emphasizing the case for multiple gene losses [28]. Nevertheless, likely cases of late HGT accumulate (e.g. the discovery of cellulose synthase of apparent bacterial origin in the urochordate *Ciona intestinalis* [29,30]).

Figure 3 presents two contrasting scenarios for the origin of the messenger-metabolism enzymes. In Figure 3a, HGT between eukaryotic lineages is forbidden but the likelihood of HGT from bacteria to any eukaryotic lineage is considered to be comparable with that of lineage-specific gene loss. Thus, the parsimony principle required postulating the occurrence of HGT from bacteria to the last common ancestor of the eukaryotic lineages that have the given enzyme. For a comparison, Figure 3b shows the scenario, in which any HGT from bacteria to eukaryotes, except for HGT from the proto-mitochondrion, is forbidden. Under this scenario, nearly all of the enzymes involved in messenger metabolism would have been lost in plants, and many additional losses would have to be postulated in fungi and animals. The obvious corollary is that the last common ancestor of the eukaryotic crown group would contain the full complement of genes encoding these enzymes in chordates.

Concluding remarks

Ruling out the possibility of late HGT, even at the price of multiple gene losses, leads to the rather implausible (but not, technically, impossible) notion of an ancestral organism whose functional diversity was substantially greater than that of the extant organisms (each of the extant organisms having lost some of the ancestral genes). Thus, although the acquisition of bacterial genes by advanced, multicellular eukaryotes might be rare, such events seem to be neither impossible nor inconsequential. The present analysis suggests that many enzymes involved in first-messenger metabolism might have been acquired through independent HGT events from different bacterial groups, subsequent to the divergence of the plants and the animal–fungi clade. The possibility of an even later HGT can not be refuted for several of these enzymes.

Remarkably, just a few comparatively late HGTs from bacteria, which occurred independently over the course of >100 million years (MY) of evolution, might have been responsible for the emergence of some of the complex vertebrate functions. Similar scenarios involving lateral transfer events from bacteria have been proposed previously for the evolution of several regulatory systems that function in development, differentiation and programmed cell death in animal and plants [9,31,32]. Sequencing and comparative analysis of genomes from a representative sample of eukaryotic lineages will test these ideas further.

Note added in proof

While this article was being prepared for publication, the availability of new genomic sequence enabled us to detect among the proteins of the planctomycete Pirellula (gi numbers: 32475373, 32475747), the first bacterial homologs of DBH, which is described as an animal-specific enzyme in this article. These proteins are secreted like their animal counterparts, combine an N-terminal thioredoxin fold domain with a C-terminal hydroxylase domain and are likely to function as oxygenases. Although it is difficult to infer the direction of HGT in the absence of additional bacterial representatives, the detection of these DBH homologs in bacteria underscore the prevalence of HGT in the evolution of messenger-metabolism enzymes.

Figure 3. Two opposing scenarios for the origin of first-messenger-metabolism pathways. **(a)** The first scenario postulates horizontal gene transfer (HGT) from bacteria to the last common ancestor of the eukaryotic lineages containing the given enzyme. The boxes with crosses indicate gene losses. Arrows show postulated HGT from a bacterial lineage to a eukaryotic lineage. The broken oval shapes in orange indicate uncertainty as to the source of the bacterial gene. The red stars indicate lineage-specific innovations and the gray stars indicate lineages-specific duplications. The number of events in the most parsimonious scenario depends on the topology of the tree that is used for the reconstruction. For example, we assumed that the coelomate topology of the animal tree applied (which we consider to be most plausible), whereby arthropods and chordates form a clade to the exclusion of nematodes [34,35]; accordingly, the absence of an ancestral gene in insects and nematodes requires two losses. By contrast, if the ecdysozoan topology is adopted, whereby nematodes and arthropods form a clade to the exclusion of chordates [36], only one loss would be necessary. The position of slime molds in the tree is also uncertain; the slime mold–metazoa clade is assumed here on the basis of phylogenetic analysis of multiple protein sequences [37] (L. Aravind, unpublished observations) and additional analysis of evolution of protein domain architectures (L. Aravind, unpublished observations); however, other analyses yielded alternative topologies where slime molds branched off before the animal–fungi divergence or even outside the crown group [38,39]. **(b)** A second scenario of lineage-specific gene loss, with HGT allowed only from the proto-mitochondrial endosymbiont to an ancestral eukaryote. The tree topology is the same as in Figure 3a. Abbreviations: AAH, amino-acid hydroxylases; Ach, acetylcholine; AADC, amino-acid decarboxylase; AANAT, arylalkylamine N-acetyltransferase; COMT, catechol methyltransferase; DBH, dopamine β-hydroxylase; GAD, glutamate decarboxylase; HIOMT, hydroxyindole-O-methyltransferase; HDC, histidine decarboxylase; HNMT, histamine N-methyltransferase; MAO, monoamine oxidase; NOS, nitric oxide synthase; PMNT, phenylethanolamine N-methyltransferase; TyBH, tyrosine β-hydroxylase.

Acknowledgements
We thank Floyd Bloom for critical reading of an early version of the manuscript and for thoughtful comments.

References

1 Rho, J.M. and Storey, T.W. (2001) Molecular ontogeny of major neurotransmitter receptor systems in the mammalian central nervous system: norepinephrine, dopamine, serotonin, acetylcholine, and glycine. *J. Child Neurol.* 16, 271–281
2 Von Bohlen *et al.* (2002) *Neurotransmitters and Neuromodulators: Handbook of Receptors and Biological Effects*, John Wiley
3 Coon, S.L. *et al.* (1995) Pineal serotonin N-acetyltransferase: expression cloning and molecular analysis. *Science* 270, 1681–1683
4 Klein, D.C. (1999) Serotonin N-acetyltransferase. A personal historical perspective. *Adv. Exp. Med. Biol.* 460, 5–16
5 Ganguly, S. *et al.* (2001) Characterization of the *Saccharomyces cerevisiae* homolog of the melatonin rhythm enzyme arylalkylamine N-acetyltransferase (EC 2.3.1.87). *J. Biol. Chem.* 276, 47239–47247
6 Brown, J.R. and Doolittle, W.F. (1997) Archaea and the prokaryote-to-eukaryote transition. *Microbiol. Mol. Biol. Rev.* 61, 456–502
7 Andersson, S.G. *et al.* (2003) On the origin of mitochondria: a genomics perspective. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 358, 165–179
8 Chandok, M.R. *et al.* (2003) The pathogen-inducible nitric oxide synthase (iNOS) in plants is a variant of the P protein of the glycine decarboxylase complex. *Cell* 113, 469–482
9 Aravind, L. *et al.* (2003) Evolutionary connections between bacterial and eukaryotic signaling systems: a genomic perspective. *Curr. Opin. Microbiol.* 6, 490–497
10 Roelofs, J. and Van Haastert, P.J. (2001) Genes lost during evolution. *Nature* 411, 1013–1014
11 Blanchard, J.L. and Lynch, M. (2000) Organellar genes: why do they end up in the nucleus? *Trends Genet.* 16, 315–320
12 Doolittle, W.F. (1999) Lateral genomics. *Trends Cell Biol.* 9, M5–M8
13 Koonin, E.V. *et al.* (2001) Horizontal gene transfer in prokaryotes: quantification and classification. *Annu. Rev. Microbiol.* 55, 709–742
14 Gogarten, J.P. *et al.* (2002) Prokaryotic evolution in light of gene transfer. *Mol. Biol. Evol.* 19, 2226–2238
15 Doolittle, W.F. *et al.* (2003) How big is the iceberg of which organellar genes in nuclear genomes are but the tip? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 358, 39–58
16 Kurland, C.G. *et al.* (2003) Horizontal gene transfer: a critical view. *Proc. Natl. Acad. Sci. U. S. A.* 100, 9658–9662
17 Doolittle, W.F. (1998) You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. *Trends Genet.* 14, 307–311
18 Andersson, J.O. *et al.* (2003) Phylogenetic analyses of diplomonad genes reveal frequent lateral gene transfers affecting eukaryotes. *Curr. Biol.* 13, 94–104
19 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. U. S. A.* 99, 12246–12251
20 Lander, E.S. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860–921
21 Andersson, J.O. *et al.* (2001) Genomics. Are there bugs in our genome? *Science* 292, 1848–1850
22 Stanhope, M.J. *et al.* (2001) Phylogenetic analyses do not support horizontal gene transfers from bacteria to vertebrates. *Nature* 411, 940–944
23 Salzberg, S.L. *et al.* (2001) Microbial genes in the human genome: lateral transfer or gene loss? *Science* 292, 1903–1906
24 Genereux, D.P. and Logsdon, J.M. Jr (2003) Much ado about bacteria-to-vertebrate lateral gene transfer. *Trends Genet.* 19, 191–195
25 Eisen, J.A. and Fraser, C.M. (2003) Phylogenomics: intersection of evolution and genomics. *Science* 300, 1706–1707
26 Koonin, E.V. (2003) Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nat. Rev. Microbiol.* 1, 127–136
27 Mirkin, B.G. *et al.* (2003) Algorithms for computing parsimonious evolutionary scenarios for genome evolution, the last universal common ancestor and dominance of horizontal gene transfer in the evolution of prokaryotes. *BMC Evol. Biol.* 3, 2 (http://www.biomedcentral.com/1471-2148/3/2)
28 Kortschak, R.D. *et al.* (2003) EST analysis of the cnidarian *Acropora millepora* reveals extensive gene loss and rapid sequence divergence in the model invertebrates. *Curr. Biol.* 13, 2190–2195
29 Dehal, P. *et al.* (2002) The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. *Science* 298, 2157–2167
30 Matthysse, A.G. *et al.* (2004) A functional cellulose synthase from ascidian epidermis. *Proc. Natl. Acad. Sci. U. S. A.* 101, 986–991
31 Aravind, L. *et al.* (2001) Apoptotic molecular machinery: vastly increased complexity in vertebrates revealed by genome comparisons. *Science* 291, 1279–1284
32 Koonin, E.V. and Aravind, L. (2002) Origin and evolution of eukaryotic apoptosis: the bacterial connection. *Cell Death Differ.* 9, 394–404
33 Hintermann, E. *et al.* (1996) Cloning of an arylalkylamine *N*-acetyltransferase (aaNAT1) from *Drosophila melanogaster* expressed in the nervous system and the gut. *Proc. Natl. Acad. Sci. U. S. A.* 93, 12315–12320
34 Blair, J.E. *et al.* (2002) The evolutionary position of nematodes. *BMC Evol. Biol.* 2, 7 (http://www.biomedcentral.com/1471-2148/2/7)
35 Wolf, Y.I. *et al.* (2004) Coelomata and not ecdysozoa: evidence from genome-wide phylogenetic analysis. *Genome Res.* 14, 29–36
36 Aguinaldo, A.M. *et al.* (1997) Evidence for a clade of nematodes, arthropods and other moulting animals. *Nature* 387, 489–493
37 Loomis, W.F. and Smith, D.W. (1995) Consensus phylogeny of *Dictyostelium*. *Experientia* 51, 1110–1115
38 Baldauf, S.L. and Doolittle, W.F. (1997) Origin and evolution of the slime molds (Mycetozoa). *Proc. Natl. Acad. Sci. U. S. A.* 94, 12007–12012
39 Hedges, S.B. (2002) The origin and evolution of model organisms. *Nat. Rev. Genet.* 3, 838–849