ACADEMIC PRESS

Molecular Phylogenetics and Evolution 25 (2002) 429–444

MOLECULAR PHYLOGENETICS AND EVOLUTION

www.academicpress.com

Evolution of arginine deiminase (ADI) pathway genes

Manuel Zúñiga,[a] Gaspar Pérez,[a] and Fernando González-Candelas[b,*]

[a] *Instituto de Agroquímica y Tecnología de Alimentos-CSIC, Valencia, Spain*

[b] *Instituto Cavanilles de Biodiversidad y Biología Evolutiva, Universitat de Valencia, Aparado de Correos 2085, Valencia E-46071, Spain*

Received 18 October 2001; received in revised form 30 May 2002

Abstract

We have analyzed the evolution of the three genes encoding structural enzymes of the arginine deiminase (ADI) pathway, arginine deiminase (ADI), ornithine transcarbamoylase (OTC), and carbamate kinase (CK) in a wide range of organisms, including Archaea, Bacteria, and Eukarya. This catabolic route was probably present in the last common ancestor to all the domains of life. The results obtained indicate that these genes have undergone a complex evolutionary history, including horizontal transfer events, duplications, and losses. Therefore, these genes are not adequate to infer organismal relationships at deep branching levels, but they provide an insight into how catabolic genes evolved and were assembled into metabolic pathways. Our results suggest that the three genes evolved independently and were later assembled into a single cluster with functional interdependence, thus, providing support for the gene recruitment hypothesis. Furthermore, the molecular phylogenetic analysis of OTC suggests a new classification of these genes into three subfamilies.
© 2002 Elsevier Science (USA). All rights reserved.

1. Introduction

The biosynthesis and metabolism of arginine have attracted the interest of researchers because of the complexity and variety of the metabolic pathways used. Arginine and its precursors are involved in the biosynthesis of several metabolites such as polyamines and some antibiotics. Arginine metabolism is also linked to the pyrimidine biosynthetic pathway through carbamoyl phosphate. Multiple pathways for arginine degradation have been described in microorganisms and, occasionally, several of them are simultaneously present in the same organism (Abdelal, 1979; Cunin et al., 1986). Among these pathways, the arginine deiminase (ADI) is the most widespread anaerobic route for arginine degradation.

The arginine deiminase pathway comprises three reactions, catalyzed by arginine deiminase (ADI, EC 3.5.3.6), ornithine transcarbamoylase (OTC, EC 2.1.3.3), and carbamate kinase (CK, EC 2.7.2.2) (see Fig. 1), and performs the conversion of arginine to ornithine, ammonia, and CO_2, generating one mol of ATP per mol of arginine consumed. The ADI pathway constitutes a major source of energy for several microorganisms. ADI catalyzes the first step of the pathway, the deimination of arginine, yielding citrulline and ammonium. ADI genes have been sequenced from Bacteria, Archaea, and anaerobic eukaryotes such as *Giardia intestinalis* (Knodler et al., 1995). So far, no ADI genes or ADI activity has been reported for higher eukaryotes. At least one OTC-encoding gene is present in most organisms that have been studied so far. The inspection of the complete genome sequences available shows that OTC is absent only in some obligate parasitic bacteria. The OTC can also catalyze the reverse reaction, the synthesis of citrulline from ornithine and carbamoyl phosphate, one of the steps of the arginine biosynthetic pathway and the first step of the urea cycle. In this sense, some organisms with the ADI pathway, such as *P. aeruginosa*, carry genes encoding dedicated OTCs for the ADI pathway and the arginine biosynthetic pathway. CK catalyzes the hydrolysis of the carbamoyl phosphate to CO_2 and NH_4^+, while the phosphate group is used to phosphorylate ADP. CK-like genes have been sequenced in Bacteria, Eukarya, and Archaea, but they are not always involved in the ADI pathway. In this sense, *Haemophilus influenzae* possesses both OTC- and CK-encoding genes but no ADI, whereas *Escherichia*

* Corresponding author. Fax: +34-963-983-670.
E-mail address: fernando.gonzalez@uv.es (F. González-Candelas).

1055-7903/02/$ - see front matter © 2002 Elsevier Science (USA). All rights reserved.
PII: S1055-7903(02)00277-4

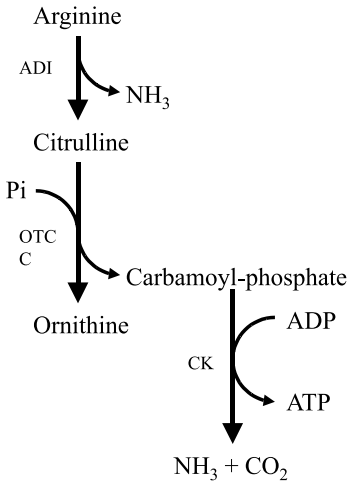

Fig. 1. ADI deiminase pathway.

coli harbors three different not characterized CK-like genes. On the other hand, *Pyrococcus furiosus* contains a CK gene, but biochemical evidence indicates that this enzyme actually catalyzes the opposite reaction, the synthesis of carbamoyl phosphate from ammonium, carbonate, and one molecule of ATP (Durbecq et al., 1997). Apart from these three genes encoding the catalytic activities, a number of additional genes can be found associated to ADI gene clusters. These genes include regulatory genes belonging to different families of transcriptional regulators, genes encoding a number of enzymatic activities, some not yet characterized, and genes encoding putative transport proteins.

The ADI pathway is widely distributed, being present in Bacteria (for a review, see Cunin et al., 1986), Archaea (Ruepp and Soppa, 1996), and Eukarya (Ludwig, 1993; Schofield et al., 1992; Yarlett et al., 1994). The ADI pathway is apparently confined to fermentative or facultative fermentative organisms. Consequently, within Eukarya the ADI pathway has been reported only for some amitochondrial organisms. This wide distribution suggests that the ADI pathway might already have been assembled in the last common ancestor of the three proposed domains of living organisms (Woese, 1987). Therefore, the ADI pathway constitutes an interesting model for the study of the evolution of metabolic pathways.

The first model to explain the evolution of metabolic pathways was proposed by Horowitz (1945). This model, usually referred to as the retrograde hypothesis, states that metabolic pathways were sequentially built up in a reverse order as found in extant pathways. When the substrate for an enzyme is depleted, a new enzyme evolves to supply this substrate from an available precursor. The discovery of operons led Horowitz (1965) to modify his model by postulating that clustering of genes encoding the enzymes of a given pathway could be explained by tandem duplication and divergence. A corollary of this hypothesis is that enzymes constituting a metabolic pathway should be structurally related. Nevertheless, although a few examples can be found where adjacent enzymes are homologous, there are many more examples where they are not. An alternative model is that of enzyme recruitment, whereby metabolic pathways were assembled by the recruitment of slow, inefficient enzymes of broad substrate specificity (Jensen, 1976). According to this model, new enzymes arise by gene duplication and divergence to allow for the adjustment of their substrate specificity and therefore enzymes catalyzing similar reactions in different pathways would be found to be homologous. Moreover, this model implies that enzyme evolution has been driven by retention of catalytic mechanisms and there is good evidence to suggest that this has occurred within many protein families (Copley and Bork, 2000).

The enzyme recruitment hypothesis provides a good framework to understand enzyme evolution; nevertheless, it does not provide an explanation for the frequent clustering of genes involved in a metabolic pathway observed in Bacteria and Archaea. Since the formulation of the concept of operon, several models have been proposed to explain the origin of operons (Glansdorff, 1999; Lawrence, 1999). These models can be classified into two classes: functional models postulate that gene clustering improves the performance of metabolic pathways and genetic models propose that gene clustering improves the propagation of the constituent genes. Among the first, the molarity model suggests that gene clustering results in a beneficially high local concentration of proteins. An extension of this model has been proposed by Glansdorff (1999): operons appeared in Archaea and Bacteria as a result of the adaptation to thermophily of their respective ancestors. In this view, operons would be favorably selected because they facilitate the interaction between the proteins participating in the pathway. These interactions would provide protection from thermodegradation to both thermolabile substrates and proteins. Among the second class, the selfish operon model (Lawrence and Roth, 1996) proposes that gene clustering confers a selective advantage to constituent genes because it allows their horizontal transfer. In this view, transferred genes would provide a selectable function to the host; thereby operons would be formed after transfer because genes not contributing to the selectable function can be deleted.

To elucidate the evolution of the ADI pathway, we present a phylogenetic analysis of all completely sequenced ADI, OTC, and CK genes. Other genes present in ADI gene clusters have not been included in this study, since the number of available sequences was too small to perform a meaningful analysis. The results obtained indicate that these genes have undergone a complex evolutionary history, probably involving several gene duplications and further losses, non-ortholo-

gous displacements, and lateral transfers. Therefore, these genes are not adequate to infer organismal relationships but they provide an insight into how catabolic genes evolved and were assembled to constitute catabolic pathways. In this respect, the results obtained suggest that the three genes evolved independently and were later assembled together. This is in accordance with the scenario depicted in the enzyme recruitment hypothesis (Jensen, 1976).

2. Material and methods

2.1. Sequence analysis

Sequences for OTC, ADI, and CK were obtained from GenBank and EMBL databases (release 108 and 59, respectively). Accession numbers and positions used in the analyses for all sequences are shown in Table 1.

Deduced amino acid sequences were aligned using the progressive alignment algorithm (Feng and Doolittle, 1987), as implemented in CLUSTAL W (Thompson et al., 1994), and rearranged further by visual inspection. Alignment regions with uncertain homology were discarded from the analysis. Multiple alignments for nucleotide sequences were obtained from the corresponding protein alignments. Complete alignments are available upon request.

The information content for evolutionary reconstruction from the protein sequences alignments was evaluated by likelihood mapping using Tree-Puzzle 5.0 (Strimmer and von Haeseler, 1996, 1997). Phylogenetic reconstructions were obtained from amino acid sequences by maximum likelihood using PAML (Yang, 2000), using an update to Dayhoff's PAM250 matrix distance among amino acids (Jones et al., 1992) and a discrete gamma model with shape parameter estimated from the data by iterative tree reconstruction to take into account heterogeneity in the rate of evolution among sites. The iteration started by obtaining an initial phylogenetic tree assuming a constant rate for all sites and then using this topology to estimate the shape parameter of a gamma distribution. Once this parameter is established, a new topology was obtained from the scratch. If this and the previous topology were coincident, the procedure was finished; otherwise, the iteration was continued until convergence was obtained.

The significance of the reconstructions was tested by several methods. First, branches were tested for being longer than zero using the corresponding standard errors derived from the information matrix (Yang, 2000). Branches were considered longer than zero when their lengths were larger than twice the standard error. Second, quartet puzzling was used to obtain the proportion of times for which each internal branch was recovered as significant in each possible quartet of sequences (Strimmer and von Haeseler, 1996). This gives an estimate of the total support for each internal branch. Third, a Kishino and Hasegawa (1989) test, as implemented in MOLPHY (Adachi and Hasegawa, 1996), was performed to compare the reconstructed tree from OTCs with that derived from the SSU rDNA of the same species. Finally, the same trees were tested for congruence using Page and Charleston (1997) method incorporated in program GeneTree (Page, 1998).

Relative rate tests of nucleotide substitutions were done using three different approaches. First, we used Wu and Li's (1985) method, complemented with Muse and Weir's (1992) modification as implemented in program K2WULI (Jermiin, 1996), in which multiple hits are considered using Kimura's two-parameter model only for transversions (Kimura, 1980). Second, we used the two-cluster test of Takezaki et al. (1995), as implemented in program PHYLTEST 2.0 (Kumar, 1996), and third, we employed program RRTREE 1.1, which allows the comparison of substitution rates among groups of sequences considering all kinds of substitutions or only fractions of them, such as transversions or nonsynonymous substitutions (Robinson et al., 1998).

3. Results

3.1. Phylogenetic analysis of the structural genes of the ADI pathway

Multiple alignments were obtained for 54 OTC, 18 ADI, and 20 CK protein sequences and 313, 261, and 280 amino acid positions were chosen, respectively, for further analysis. One of the ADI sequences from *Mycoplasma pneumoniae* (MPNEUMONI) showed a frameshift mutation that prevented a correct translation, as indicated in the annotation for that entry. This caused two potential, overlapping arginine deiminase genes to be described originally, each of them showing homology either to the 5′- or 3′-end of homologous sequences (Himmelreich et al., 1996). By comparison with ADI genes from other *Mycoplasma* species, we were able to identify the potential cause of the error, a T missing in position 4449 (see Table 1). The deduced amino acid sequence obtained after including this base in that gene was used in this analysis. Conversely, the *arcB* gene of this organism also harbors a frameshift mutation, causing the N-terminus of the protein to appear truncated. We identified the possible mistake as an extra T in position 3630 (see Table 1). The corrected *arcB* gene has been used for further analyses.

Likelihood mapping analysis (Fig. 2) showed a strong phylogenetic signal in the three data sets and, in consequence, we proceeded with the study. Evolutionary relationships among the different sequences were obtained by maximum likelihood using the JTT matrix of

Table 1
Sequences used in this study

Sequence	Accession Nos.	Species	Features	Positions
ADI				
BAFZELLI	AF008219	*Borrelia afzelii*		1595–2824
BBURGDORF	AE001183	*Borrelia burgdorferi*		120–1352
BLICHENIF	Y17554	*Bacillus licheniformis*		248–1489
CPERFRING	X97684	*Clostridium perfringens*		644–1885
GINTESTIN	U49236	*Giardia intestinalis*		221–1963
HSALINARI	X80931	*Halobacterium salinarium*		764–2224
LSAKEI	AJ001330	*Lactobacillus sakei*		571–1797
MARGININI	X54141	*Mycoplasma arginini*		228–1460
MARTHRITI	AF182646	*Mycoplasma arthritidis*		385–1614
MHOMINIS	D13314	*Mycoplasma hominis*		1–1227
MPNEUMONI	AE000052	*Mycoplasma pneumoniae*		5011–3798**
MPNEUMONI27	AE000027	*Mycoplasma pneumoniae*		12,526–13,842
MTUBERCUL	X93471	*Mycoplasma* sp.		1–1230
PAERUGINO	X14694	*Pseudomonas aeruginosa*		28–1284
PPUTIDA	U07185	*Pseudomonas putida*		85–1347
RETLI	AF025543	*Rhizobium etli*		825–2054
SCOELICOL	AL132991	*Streptomyces coelicolor*		32,710–33,972
SPYOGENES	X55659	*Streptococcus pyogenes*		37–1275
OTC				
AAEOLICUS	AE000752	*Aquifex aeolicus*	Anabolic	6818–7738
AFULGIDUS	AE001017	*Archaeoglobus fulgidus*	Anabolic	991–68
ANIGER	M19158	*Aspergillus niger*	Anabolic	581–1693
AORYZAE	AB020737	*Aspergillus oryzae*	Anabolic	982–2100
APERNIX	AP000063	*Aeropyrum pernix*	Anabolic	22,737–21,529
ATERREUS	Z67741	*Aspergillus terreus*	Anabolic	441–1526
ATHALIANA	AJ002524	*Arabidopsis thaliana*	Anabolic	14–1943 (5int)
BAFZELII	AF008219	*Borrelia afzelii*	Catabolic	2889–3872
BBURGDORF	AE001183	*Borrelia burgdorferi*	Catabolic	1417–2403
BLICHENIF	Y17554	*Bacillus licheniformis*	Catabolic	1518–2525
BSTEAROTH	U43091	*Bacillus stearothermophilus*	Anabolic	4991–5926
BSUBTILIS	X53360	*Bacillus subtilis*	Anabolic	477–1436
CGLUTAMIC	AF049897	*Corynebacterium glutamicum*	Anabolic	4597–5556
CHIRSUTUS	D26061	*Coriolus hirsutus*	Anabolic	735–2468 (4int)
CLINEATA	AF033562	*Canavalia lineata*	Anabolic	7–1086
CPERFRING	X97768	*Clostridium perfringens*	Catabolic	365–1360
DRADIODUR	AE001871	*Deinococcus radiodurans*	Anabolic	4616–3555
ECOLARGI	X00759	*Escherichia coli*	Anabolic	1–1005
ECOLIARGF	U14003	*Escherichia coli*	Anabolic	169,135–168,131
GINTESTIN	AF069576	*Giardia intestinalis*		278–1261
HINFLUENZ	U32741	*Haemophilus influenzae*		4471–3467
HPARAGALL	AF007428	*Haemophilus paragallinarum*		319–1323
HSALINARI	X81712	*Halobacterium salinarium*	Catabolic	91–978
HSAPIENS	K02100	*Homo sapiens*		136–1200
LPLANTARU	X99978	*Lactobacillus plantarum*		5105–6127
LSAKEI	AJ001330	*Lactobacillus sake*	Catabolic	1834–2844
MBOVIS	X64203	*Mycobacterium bovis*	Anabolic	184–1107
MJANNASCH	U67532	*Methanococcus jannaschii*		4738–3821
MMUSCULUS	M17030	*Mus musculus*		5–1069
MPNEUMONI	AE000052	*Mycoplasma pneumoniae*		3770–2729*
MTHERMOAU	AE000906	*Methanobacterium thermoautotrophicum*	Anabolic	1450–2355
MTUBERCUL	Z85982	*Mycobacterium tuberculosis*	Anabolic	26,575–27,498
NGONORRHO	M34930	*Neisseria gonorrhoeae*	Anabolic	141–1136
NMENINGIT	X64861-8	*Neisseria meningitidis*	Catabolic	<1–787>
NOSTOCSP	AF030524	*Nostoc punctiforme*		523–1443
PAERUGANA	M19939	*Pseudomonas aeruginosa*	Anabolic	357–1274
PAERUGCAT	X05637	*Pseudomonas aeruginosa*	Catabolic	64–1074
PABYSII	AJ248287	*Pyrococcus abyssi*		83,788–82,835
PFURIOSUS	X99225	*Pyrococcus furiosus*	Anabolic	1–945
PHORIKOSH	AP000003	*Pyrococcus horikoshii*		102,955–103,908
PSATIVUM	PS13684	*Pisum sativum*		38–1165
PSYRPHARE	M94049	*Pseudomonas syringae*	Phaseolotoxin-resistant	401–1384

Table 1 (*continued*)

Sequence	Accession Nos.	Species	Features	Positions
PSYRPHASE	X76945	*Pseudomonas syringae*	Phaseolotoxin-sensitive	524–1444
PTANNOPHI	X15412	*Pachysolen tannophilus*	Anabolic, mt location	972–1047–2012
RCATESBEI	M95193	*Rana catesbeiana*		112–1164
RETLI	AF025543	*Rhizobium etli*	Catabolic	2092–3096
RNORVEGIC	M11266	*Rattus norvegicus*		101–1162
SCEREVISI	M28301	*Saccharomyces cerevisiae*		522–1538
SCHIPOMBE	X63577	*Schizosaccharomyces pombe*	Anabolic	414–1397
SSCROFA	Y13045	*Sus scrofa*	Anabolic, mt location	<1–987
SYNECHOCY	D64006	*Synechocystis* sp.	Anabolic	88,663–87,737
TMARITIMA	Y10661	*Thermotoga maritima*		78–1019
TTHERMOPH	Y18353	*Thermus thermophilus*		2174-3079
VIBRIOSP	Y11033	*Vibrio* sp.	Anabolic	1–906
CK				
APERNIX	AP000063	*Aeropyrum pernix*		10,165–9215
BLICHENIFO	Y17554	*Bacillus licheniformis*		4007–4957
CPERFRINGE	X97768	*Clostridium perfringens*		2996–3940
ECOLIARCC	AE000158	*Escherichia coli*	*arcC*	3947–4840
ECOLIARCL	AE000370	*Escherichia coli*	*yqeA*	11,360–12,292
ECOLIARCM	AE000139	*Escherichia coli*	*yahI*	5044–5994
EFERACALIS	AJ223332	*Enterococcus faecalis*		1–933
GINTESTINA	AF017784.1	*Giardia intestinalis*		316–1269
HEXAMITA	AF107491	*Hexamita* sp.		73–1014
HINFLUENZA	U32741	*Haemophilus influenzae*		3453–2521
HSALINARIU	X80931	*Halobacterium salinarium*		2463–3389
LSAKEI	AJ001330	*Lactobacillus sakei*		2946–3887
MPNEUMONAR	AE000052	*Mycoplasma pneumoniae*		2726–1797
PABYSII	AJ248285.1	*Pyrococcus abyssi*		249,437–250,381
PAERUGINOS	X14693	*Pseudomonas aeruginosa*		110–1042
PFURIOSUS	AB016521	*Pyrococcus furiosus*		3244–4188
PHORIKOSHI	AP000005.1	*Pyrococcus horikoshii*		158,623–159,567
RETLI	AF025543	*Rhizobium etli*		3098–4030
SYNECHOCYS	D90917	*Synechocystis* sp.		126,282–125,356
TVAGINALIS	AF050082.1	*Trichomonas vaginalis*		183–1127

[*] T at position 3630 has been deleted from the alignment to avoid a frameshift in *arcB* amino acid sequence.
[**] T has been introduced at position 4449 to avoid a frameshift in *arcA* amino acid sequence.

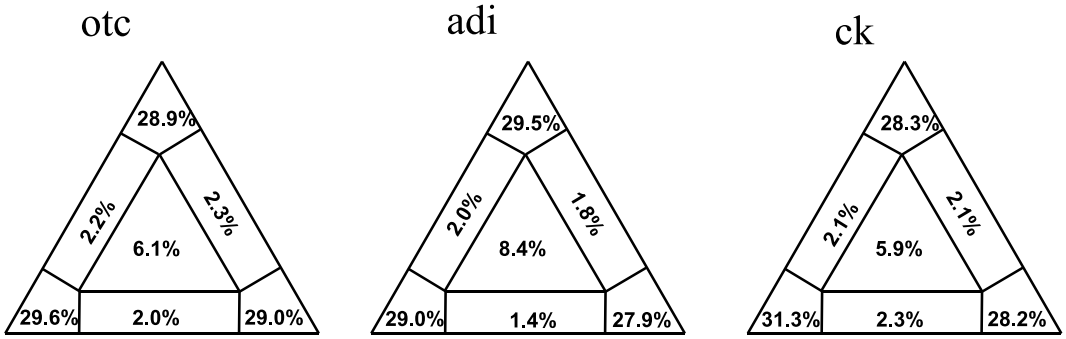

Fig. 2. Likelihood mapping analysis.

amino acid substitution and a gamma distribution to take the rate variation among sites into account. The three proteins presented different values of the shape parameter that characterizes the gamma distribution (OTC $= 0.8714 \pm 0.0669$, ADI $= 1.4591 \pm 0.1487$, and CK $= 0.9118 \pm 0.0857$). Figs. 3–5 represent the corresponding phylogenetic trees derived by this method along with the evaluation of the support for each node as described.

Arginine deiminases (Fig. 3) are characterized by the very high divergence of the two non-bacterial sequences (*G. intestinalis* and *H. salinarium*). Despite this, a number of conserved blocks can be identified in all sequences, indicating a common origin for these genes

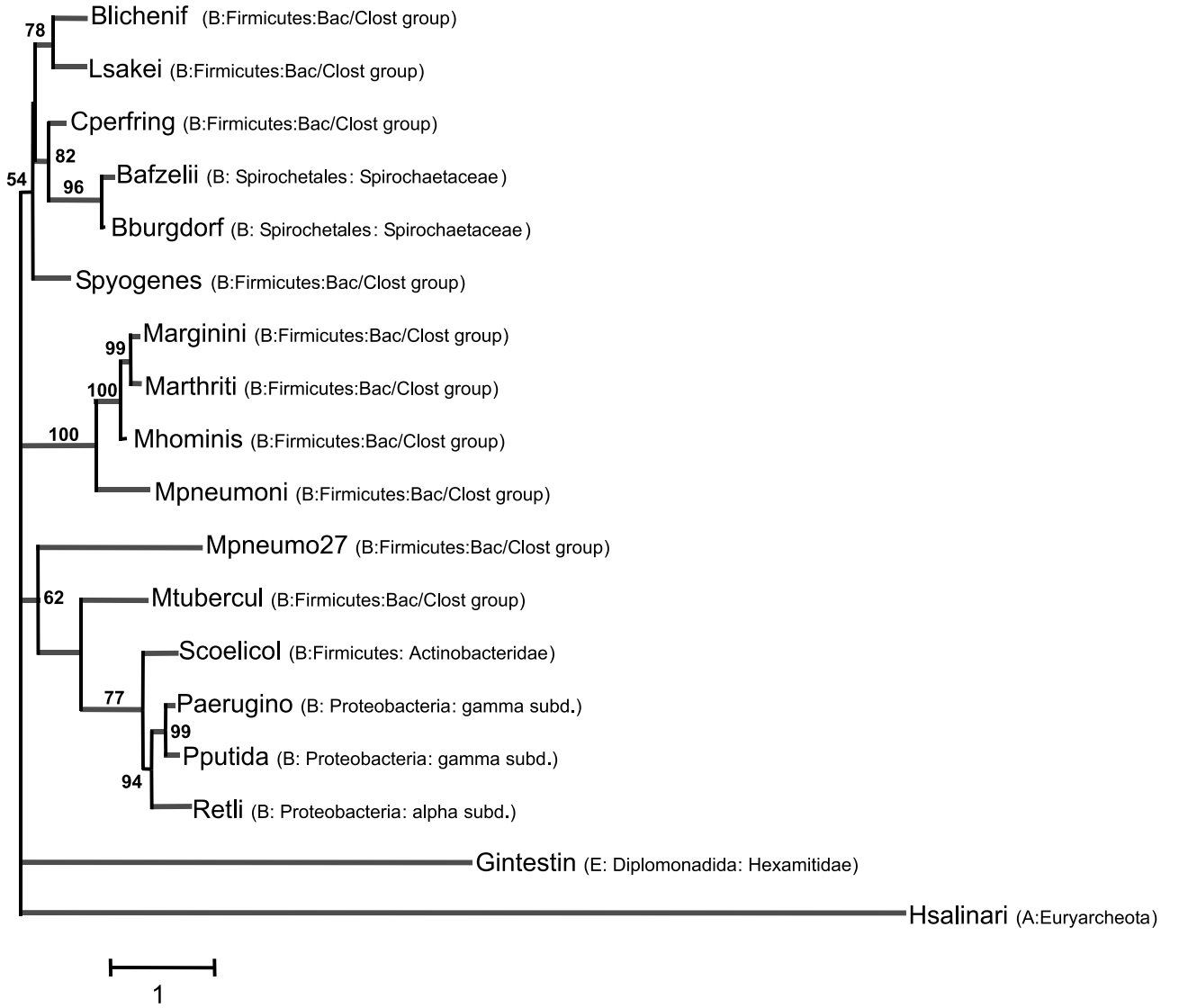

Fig. 3. Maximum likelihood phylogenetic tree for ADI sequences. Thick lines represent branches significantly longer than 0. Numbers along branches represent support values using the quartet-puzzling method when larger than 50%.

(Fig. 6). The phylogenetic tree shows that bacterial ADIs do not group according to the species' phylogeny, with several groups appearing as polyphyletic. Moreover, *M. pneumoniae* harbors two paralog genes appearing in two distinct groups.

Ornithine transcarbamoylases (Fig. 4) present two strongly supported groups and a non-statistically supported one. The OTC sequence from *G. intestinalis* does not cluster with any of these groups. The first supported group includes eukaryotic sequences from vertebrates and fungi, whereas the other encompasses some bacterial sequences. Neither archaeal nor bacterial sequences group monophyletically, and except for the previously indicated bacterial group that includes sequences from *Thermotogales*, *Firmicutes*, *Spirochaetales*, and *Proteobacteria*, most highly supported groupings correspond to terminal nodes. Labedan et al. (1999) also performed a phylogenetic analysis of OTCs. Based on their analysis, they defined two subfamilies: α-OTCs, including exclusively bacterial sequences, and β-OTCs, spanning eukaryal (including sequences of *Fungi* and *Metazoa*), archaeal, and bacterial OTCs. Our results indicate that OTCs from *Fungi* and *Metazoa* cluster significantly apart from the two previous subgroups and therefore could be considered as a different subfamily that we propose to denote γ-OTCs, following the nomenclature used by Labedan et al. (1999). Interestingly, we find that the OTC from *G. intestinalis* branches at the base of the OTC families, being the only sequence unassigned to any of these groups. The division of α- and β-OTC subfamilies can still be recognized, although the latter does not have enough statistical support. Moreover, the presence of a characteristic signature pattern in α-OTC (Fig. 7 and see alignment in Labedan et al., 1999), with

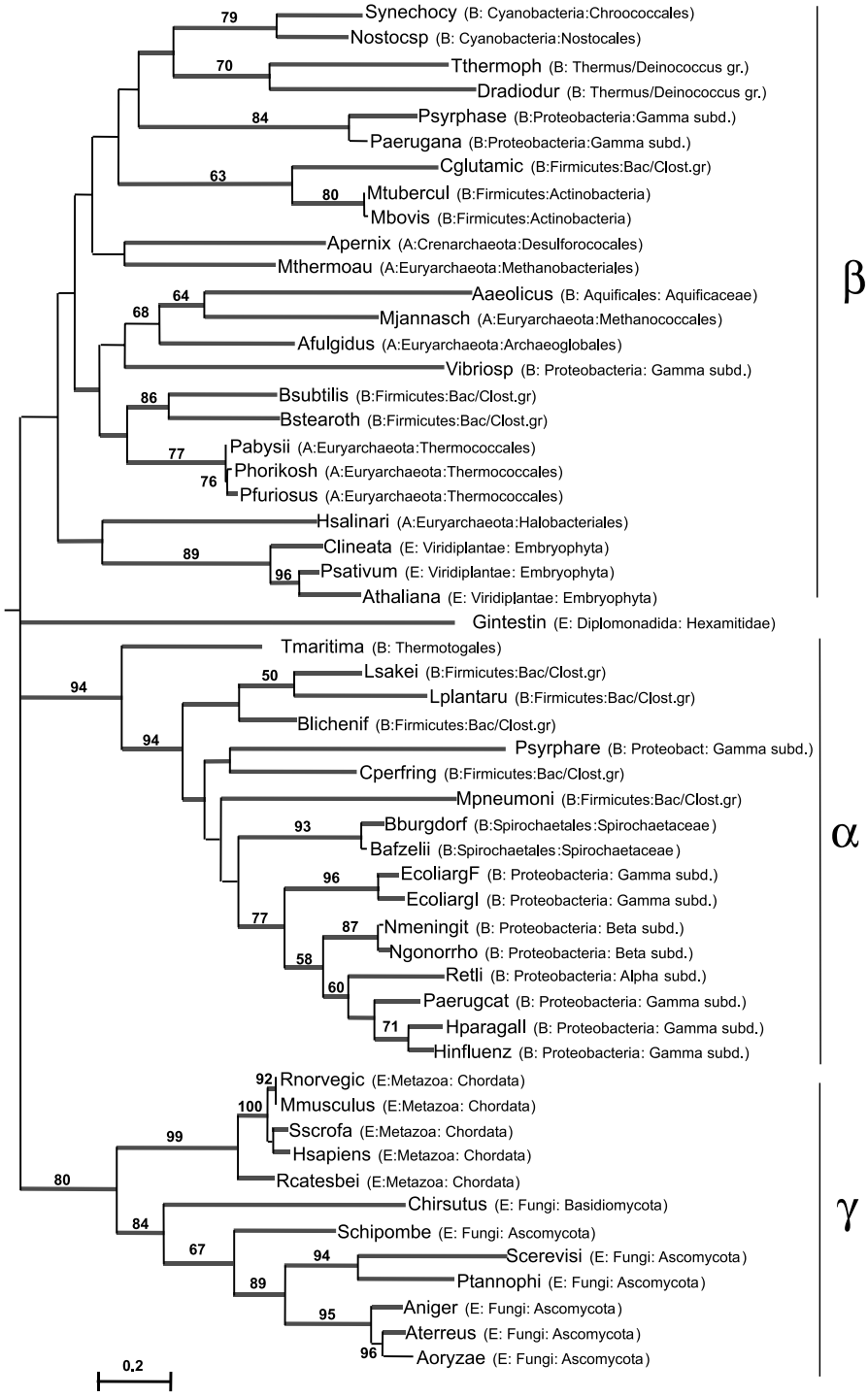

Fig. 4. Maximum likelihood phylogenetic tree for OTC sequences. Thick lines represent branches significantly longer than 0. Numbers along branches represent support values using the quartet-puzzling method when larger than 50%. The root of the tree was inferred by using. ATC sequences in Labedan et al. (1999).

the only exception of *Thermotoga maritima*, supports further that it should be considered a distinct subfamily. In summary, while the α-OTC subfamily could remain as originally proposed by Labedan et al. (1999), β-OTCs should be split into two subfamilies, β-OTCs *sensu stricto*, spanning bacterial, archaeal, and plant sequences, and γ-OTCs, comprising fungal and metazoan sequences.

Additionally, we have used the reconciled trees method (Page and Charleston, 1997) and the Kishino–Hasegawa test to compare the species tree (obtained from SSU rDNA sequences, not shown) with the OTC

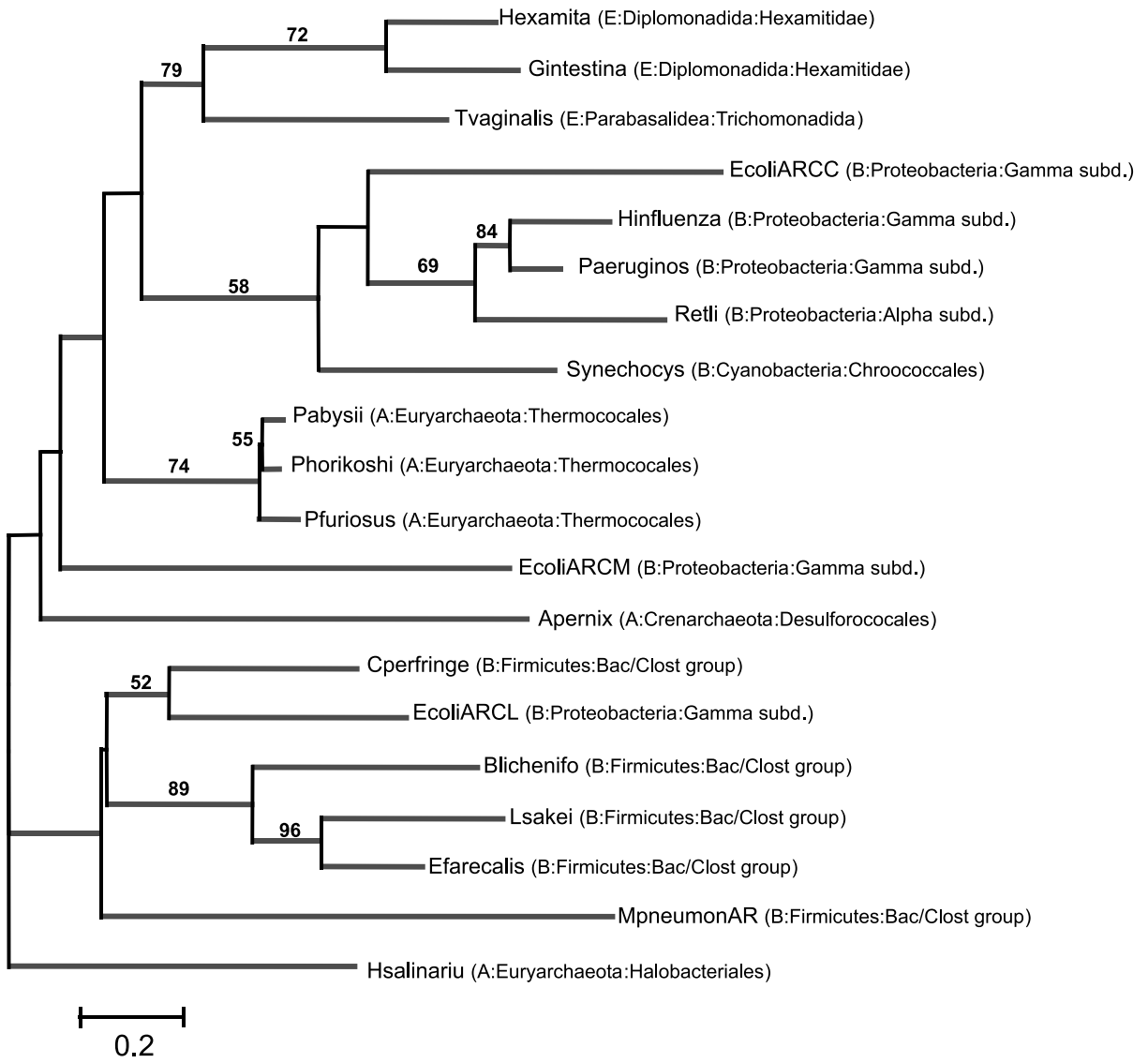

Fig. 5. Maximum likelihood phylogenetic tree for CK sequences. Thick lines represent branches significantly longer than 0. Numbers along branches represent support values using the quartet-puzzling method when larger than 50%.

shown in Fig. 4. We have verified that the OTC tree is statistically significant (by comparison with 1000 random trees obtained with the same sequences) and better for OTC sequences than the species tree (RELL support in 100% cases, LRT = 1209.4 with a standard error of 95.2). The reconciled trees approach indicates that the OTC tree requires at least 15 independent duplications and 97 paralogous gene losses to be explained without invocation to horizontal gene transfer events (not allowed in the program).

Carbamate kinases (Fig. 5) present relatively good support for an eukaryote clade encompassing three genes and a lack of monophyletic groups for either bacterial or archaeal sequences. The three genes for CK present in *E. coli* belong to very different groups and these data do not support a close relationship or a recent duplication. One of these genes ECOLIARCC is related, although without high statistical support, to other *Proteobacteria* sequences; the second ECOLIARCM does not show close similarity to any other CK gene and the last one ECOLIARCL is included in a non-supported cluster of *Firmicutes*.

To gain further information on the joint evolution of the ADI pathway genes, we proceeded to a more detailed analysis of those species for which sequences of the three genes are available. There are eight such species included in this study, six Bacteria, one Archaea, and one eukaryote. Two species present two genes for one of the enzymes. *P. aeruginosa* has two OTCs and *M. pneumoniae* has two ADIs. Based on their position in the previously described phylogenetic trees, we decided to exclude PAERUGANA, the anabolic OTC from *P. aeruginosa*, and MPNEUMO27, one of the arginine deiminases from *M. pneumoniae*. The inferred pairwise

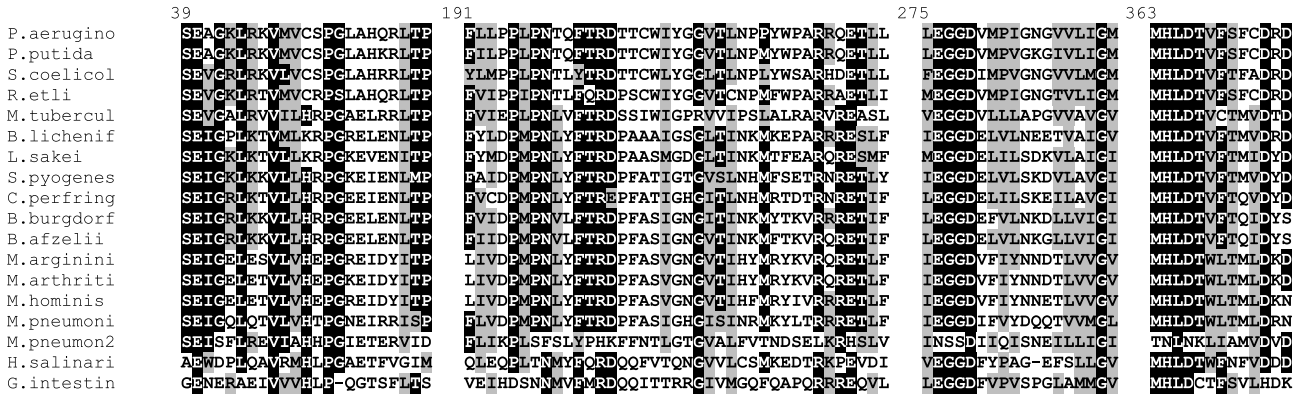

Fig. 6. Partial alignment of ADI sequences illustrating conserved blocks. The numbering above indicates the positions in the complete sequence alignment. Identical residues are indicated by black shading and conservative substitutions are indicated by gray shading.

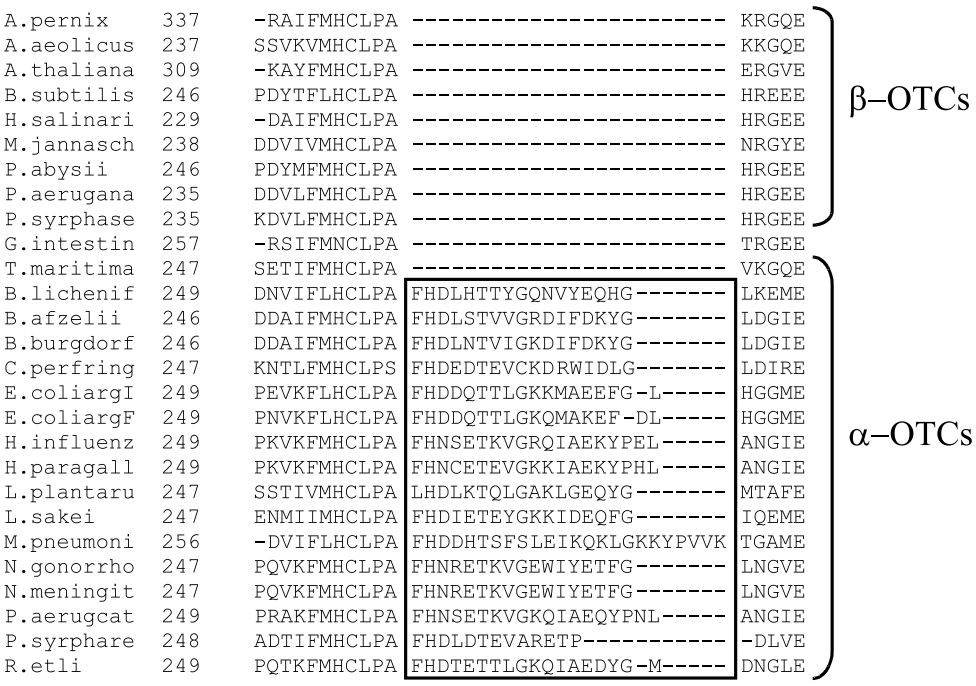

Fig. 7. Signature sequence in ornithine carbamoyl-transferase proteins. The large indel common to most members of β-OTCs subfamily is boxed. Dashes indicate gaps introduced in the alignment.

distances among the remaining proteins are shown in Table 2.

The three proteins have evolved at different rates in these species, with ADI showing the highest rate (0.6925 substitutions/site on average) and OTC the lowest (0.5176). This result holds, even when one or both non-bacterial sequences are removed from the analysis (OTC = 0.4314, ADI = 0.5636, and CK = 0.5424, when *G. intestinalis* and *H. salinarium* are removed). However, these evolutionary rates are not completely independent, as shown in Fig. 8. There is a clear correlation among the evolutionary rates in the three genes and this is not even larger due to the slower rate for CK than for ADI and OTC in *Giardia* and *H. salinarium*. Correlation coefficients for all sequences vary from 0.913 for OTC–ADI to 0.517 for OTC–CK, with CK–ADI showing 0.544. When only bacterial sequences are considered, the corresponding correlation coefficients are 0.831, 0.879, and 0.919 for OTC–ADI, OTC–CK, and CK–ADI, respectively.

The different methods used for testing the constancy of evolutionary rates yielded largely compatible results and only those obtained with the procedure of Robinson et al. (1998) will be commented. A summary of the results obtained with program RRTREE is shown in Table 3. For the arginine deiminases, the significant differences appear in the comparisons between *H. salinarium* and both bacterial subgroups *Firmicutes* and *Proteobacteria*. For OTCs, the only significant differ-

Table 2
Pairwise distances (amino acid substitutions per site) for the three genes in the ADI pathway for the eight species for which all the sequences are available

	B. lichenifo	*C. perfringe*	*G. intestinal*	*L. sakei*	*M. pneumoni*	*P. aeruginos*	*R. etli*
ADI							
C. perfringe	0.4153						
G. intestina	0.7955	0.7827					
L. sakei	0.3514	0.3898	0.8211				
M. pneumoni	0.5942	0.5974	0.8243	0.6198			
P. aeruginos	0.6198	0.6454	0.7955	0.6454	0.6901		
R. etli	0.6038	0.6390	0.8019	0.6390	0.6837	0.3195	
H. salinari	0.8626	0.8722	0.9073	0.8626	0.8754	0.8626	0.8722
OTC							
C. perfringe	0.3908						
G. intestinal	0.6245	0.6398					
L. sakei	0.2912	0.4598	0.6513				
M. pneumoni	0.4291	0.4521	0.6513	0.4904			
P. aeruginos	0.4368	0.4559	0.6284	0.4789	0.4866		
R. etli	0.4368	0.4444	0.6284	0.4713	0.4904	0.2567	
H. salinari	0.5632	0.6015	0.59	0.6322	0.6322	0.5785	0.6015
CK							
C. perfringe	0.4893						
G. intestina	0.5929	0.5857					
L. sakei	0.4143	0.4893	0.5536				
M. pneumoni	0.575	0.5286	0.5964	0.5679			
P. aeruginos	0.5643	0.5821	0.5464	0.55	0.6357		
R. etli	0.6071	0.6143	0.6214	0.6143	0.6321	0.3143	
H. salinari	0.5236	0.4836	0.5709	0.5236	0.5855	0.5564	0.5891

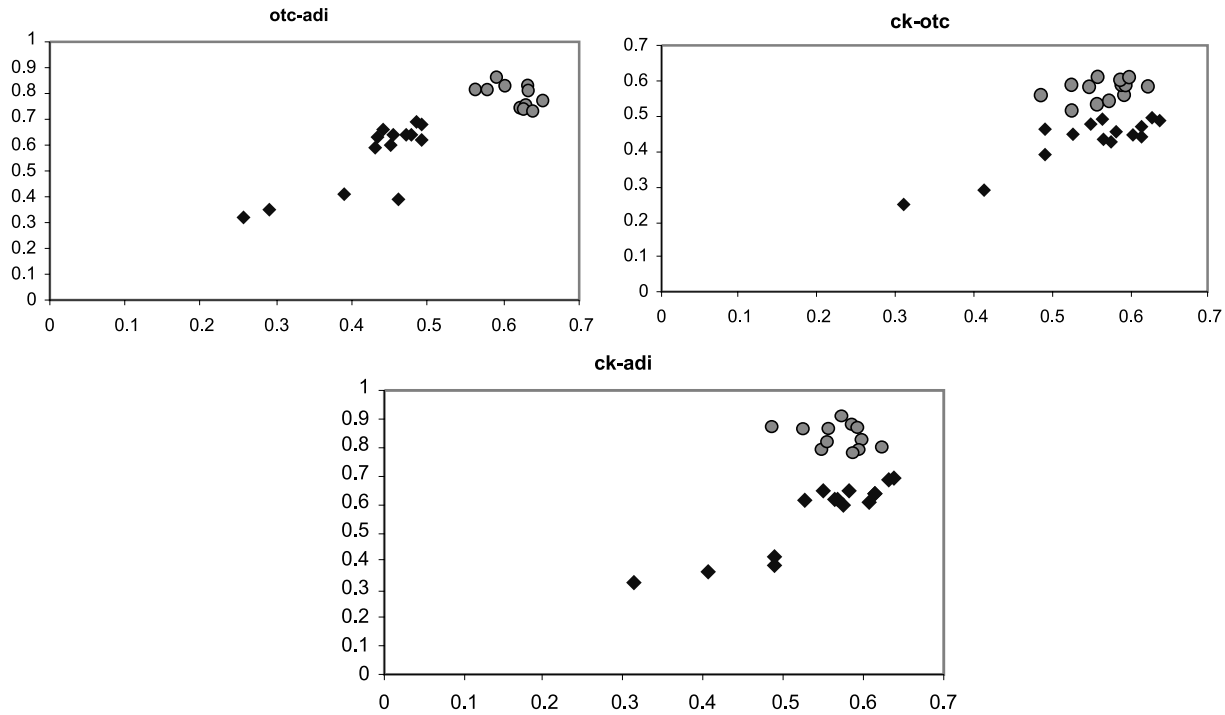

Fig. 8. Pairwise correlations among ADI–OTC–CK for the eight species with sequenced genes for all of them. Circles represent points in which either *Giardia* or *Halobacterium* is one member of the pair of species being compared.

ences are found in the comparison of non-synonymous sites between *H. salinarium* and the bacterial *Firmicutes*, using either *G. intestinalis* or the *Proteobacteria* as outgroup. Finally, none of the comparisons among non-synonymous sites for carbamate kinases showed significantly different evolutionary rates.

Table 3
Relative rate tests for ADI pathway genes in selected taxa

Taxon 1	Taxon 2	Outgroup	ΔK_a	SD (ΔK_a)	Probability
ADI					
Firmicutes	Proteobacteria	*Giardia*	−0.073831	0.081362	0.364
Firmicutes	*Halobacterium*	*Giardia*	−0.83014	0.286253	0.003737
Proteobacteria	*Halobacterium*	*Giardia*	−0.756309	0.288083	0.008664
Firmicutes	Proteobacteria	*Halobacterium*	0.067871	0.112837	0.547516
Firmicutes	*Giardia*	*Halobacterium*	−0.481061	0.282292	0.088367
Proteobacteria	*Giardia*	*Halobacterium*	−0.548932	0.289237	0.057722
OTC					
Firmicutes	Proteobacteria	*Giardia*	0.040878	0.035269	0.246448
Firmicutes	*Halobacterium*	*Giardia*	0.157456	0.046235	0.000664
Proteobacteria	*Halobacterium*	*Giardia*	0.116578	0.04277	0.006424
Firmicutes	Proteobacteria	*Halobacterium*	0.056312	0.031743	0.076071
Firmicutes	*Giardia*	*Halobacterium*	0.088741	0.046798	0.057935
Proteobacteria	*Giardia*	*Halobacterium*	0.032429	0.027587	0.73866
CK					
Firmicutes	Proteobacteria	*Giardia*	−0.036337	0.040614	0.370961
Firmicutes	*Halobacterium*	*Giardia*	−0.015216	0.039097	0.697144
Proteobacteria	*Halobacterium*	*Giardia*	0.021121	0.048327	0.662081
Firmicutes	Proteobacteria	*Halobacterium*	−0.074207	0.039139	0.05797
Firmicutes	*Giardia*	*Halobacterium*	−0.074593	0.040778	0.067369
Proteobacteria	*Giardia*	*Halobacterium*	−0.000386	0.04825	0.993616

The results shown correspond to comparisons of non-synonymous substitutions in non-synonymous sites (K_a) between the relevant sets of sequences (Firmicutes = *B. licheniformis*, *C. perfringens*, *L. sakei*, *M. pneumoniae*; Proteobacteria = *P. aeruginosa*, *R. etli*).

3.2. The organization of ADI genes

The increasing number of genome sequences publicly available has revealed an unexpected diversity of both gene arrangement and composition of the ADI gene clusters, even for closely related microorganisms. Most bacteria studied so far have ADI-related genes organized in one or more clusters (Fig. 9). So far, only four archaean species containing the complete set of genes for the ADI pathway have been found: *H. salinarium* and *Halobacterium* sp. NRC-1 harbor the three structural genes and a putative regulator organized in one cluster, whereas *Thermoplasma acidophilum* and *T. volcanium* harbor the three structural genes dispersed in their chromosomes.

In the Actinomycetales *M. tuberculosis* and *S. coelicolor* an *arcA* gene has been found, but it is not associated with other ADI-related genes. Inspection of the

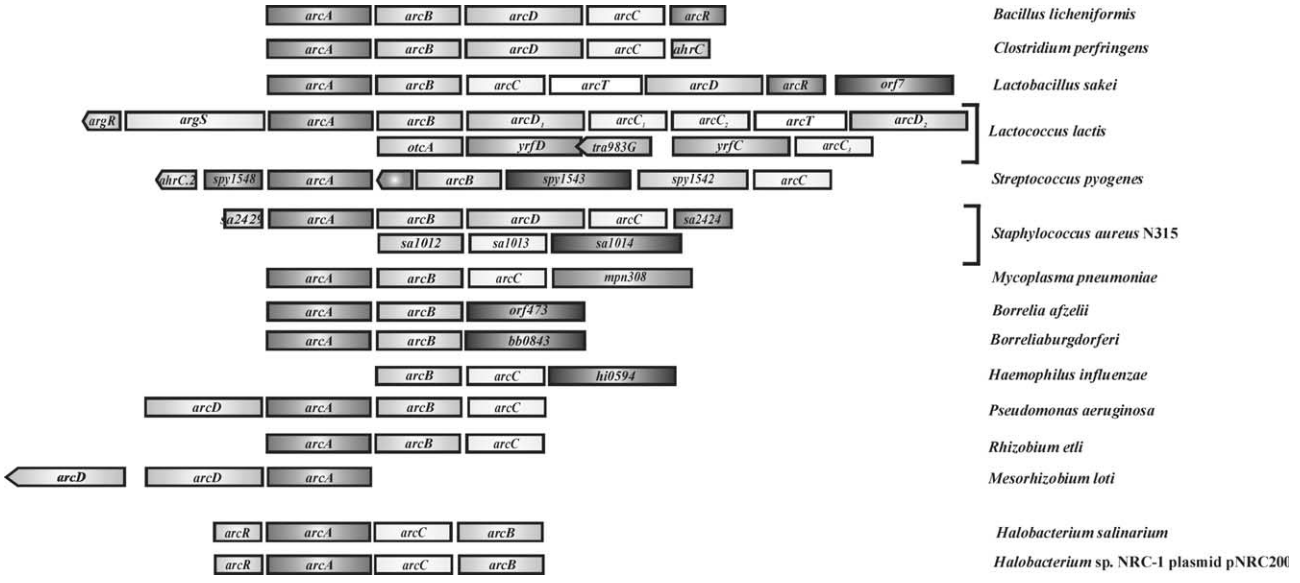

Fig. 9. ADI cluster organization in representative microbial genomes.

whole *M. tuberculosis* genome reveals that no *arcC* homolog is present, whereas the only OTC-encoding gene (*argF*) is located in the arginine biosynthetic pathway operon *argCJBDFRGH*.

Despite the wide organizational diversity observed in other bacteria, the three main structural genes preserve the order *arcA*, *arcB*, and *arcC* (Fig. 9). Several genes have been subsequently added and/or removed to this core along evolution. Clusters lacking at least one of the main structural genes are not rare: *H. influenzae*, *L. lactis*, and *S. aureus* contain a cluster lacking *arcA* (note that the two latter contain an additional complete cluster); *arcC* is missing in *B. afzelii* and *B. burgdorferii* while *Mesorhizobium loti* lacks both *arcB* and *arcC*, although it harbors a putative biosynthetic OTC-encoding gene elsewhere. Most ADI gene clusters contain at least one gene encoding a putative amino acid transport protein. Moreover, some clusters span additional genes that are mostly not yet fully identified (see Fig. 9).

4. Discussion

The phylogenetic analysis of the three structural genes of the ADI pathway clearly shows that the evolution of these genes does not agree with the order of organismal descent, inferred from the analyses of either rRNAs (Woese, 1987) or other protein markers such as RecA (Eisen, 1995). These incongruities have been previously observed for a large number of proteins (see, for example, Brown and Doolittle, 1997) and have led to a lively debate among researchers favoring a view in which horizontal gene transfer has played a major, continued role in evolution (Doolittle, 1999; Jain et al., 1999) and others who consider that unrecognized paralogy explains many of these anomalous phylogenies (Forterre and Philippe, 1999; Glansdorff, 2000). Moreover, some authors have pointed out that the limitations of the current methods of phylogenetic analysis may jeopardize the conclusions derived from such analyses (Philippe and Forterre, 1999). Our analyses indicate that paralogy accounts for many of the anomalies observed in the phylogenetic reconstructions of the three genes. This had been previously observed for OTCs by Labedan et al. (1999). Moreover, we suggest that non-orthologous displacements have also occurred during evolution.

The phylogenetic analysis of the ADI-encoding genes shows a marked divergence of archaeal and eukaryal genes; nevertheless, the conservation of several blocks along the entire sequence of these genes indicates a common origin (Fig. 6). Moreover, these differences suggest that *arcA* was not transferred from a bacterial donor to either *H. salinarium* or *G. intestinalis*, hence, favoring the presence of an ancestral ADI in the putative last common ancestor (LCA). Within Bacteria, the topology of the tree shows that low G + C gram-positive bacteria and *Borrelia* on one hand, and high G + C gram-positive bacteria and gram-negative bacteria on the other, constitute paraphyletic clusters. Our analysis does not allow to determine whether this clustering is due to horizontal gene transfers or hidden paralogy. Nevertheless, paralogy is a compelling hypothesis, since *M. pneumoniae* actually harbors two paralog *arcA* genes (Himmelreich et al., 1996), one of which (MPNEUMONI) is included in a ADI cluster (Fig. 9) whereas the other (MPNEUMO27) is located next to an arginine (AGA) tRNA encoding gene and is separated by three unrelated genes transcribed in the opposite orientation, an arginyl-tRNA synthetase encoding gene (*argS*). However, it is uncertain whether this gene actually codes for an ADI, since most conserved domains present in the remaining ADIs are missing in this sequence (Fig. 6). A horizontal gene transfer from an ancestor of low G + C gram-positive bacteria to an ancestor of *Borrelia* might explain the clustering of *arcA* genes of *Borrelia* with their counterparts of low G + C gram-positive bacteria. Nevertheless, we have found that the *arcB* genes of *Borrelia* are more closely related to their counterparts of gram-negative bacteria (α-OTC subfamily) and therefore we should assume an additional transfer event to explain this. Moreover, there is evidence indicating that *Borrelia* displays a remarkably low level of horizontal transfer (Dykhuizen and Baranton, 2001). Therefore, we suggest that hidden paralogy might account better for this anomalous clustering. The clustering of the *Actinomycetales* with gram-negative bacteria is more controversial, since *Actinomycetales* sequences branch at the base of the group. Nevertheless, OTC-encoding genes from *Actinomycetales* show a similar branching pattern (Fig. 4), although these OTCs are probably involved in the arginine biosynthetic pathway (see below). To our knowledge, there is no report of ADI activity in *Actinomycetales* but, interestingly, de la Fuente et al. (1996) reported an arginase-OTC coupled activity for the OTCs from *Streptomyces clavuligerus* and *Nocardia lactamdurans*. We suggest that this activity might be due to contamination of the samples with arginine deiminase, since the authors noted that their enzyme preparations were systematically contaminated with a minor protein of approximately 44.5 kDa, a size that agrees with that expected for ADI. These authors determined the arginase activity by measuring the production of citrulline from arginine, assuming that the ornithine produced from the degradation of arginine was carbamoylated by OTC. This conclusion was supported by the fact that carbamoyl phosphate was required to detect the degradation of arginine. We think that there is an alternative explanation for this observation: the OTCs from *S. clavuligerus* and *N. lactamdurans* may work both in the biosynthesis and in catabolism of arginine. When carbamoyl phosphate was absent, citrulline produced from

the deimination of arginine was degraded to ornithine and carbamoyl phosphate, but when carbamoyl phosphate was added to the reaction, conversion of citrulline was no longer possible and citrulline accumulated. This would be the first example of a bifunctional OTC and might explain the presence of an *arcA* gene in *Actinomycetales* without an associated catabolic OTC.

OTCs play a role in both the biosynthesis of arginine and its degradation via the ADI pathway. Nevertheless, this functional difference is not reflected at the genetic level: genes encoding biosynthetic OTCs can be found in the α-, β-, and γ-subfamilies. On the other hand, all catabolic OTCs are included in the α-subfamily, with the only exception of *H. salinarium*. The phylogenetic analysis of the OTCs shows a striking, complex picture, where bacterial, archaeal, and eukaryal sequences are intermingled. The extremely complex topology of the OTC tree suggests that duplication and loss of paralogs and/or horizontal transfers have played a major role in the evolutionary history of OTCs. In fact, several genes encoding OTC can be found in a number of extant organisms like *E. coli*, *P. aeruginosa*, and *P. syringae* and the recently sequenced *L. lactis* (Bolotin et al., 2001) and *S. aureus* (Kuroda et al., 2001). To explain the anomalous phylogeny of OTCs, Labedan et al. (1999) proposed that the LCA harbored two genes encoding OTC, from which the current α- and β-OTCs were derived, respectively. This model fails to explain the topology of the β-OTC subfamily unless a large number of horizontal gene transfer events and/or duplication and losses of paralogs are postulated. Moreover, it does not consider the existence of the γ-OTC subfamily. Additionally, we have have verified that the OTC tree is statistically significant and better (for OTC sequences) than the species tree (from SSU rDNA sequences) and that it requires at least 15 independent duplications and 97 paralogous gene losses to be explained without invocation to horizontal gene transfer events. Hence, we conclude that the obtained topology is not the result of sampling error or insufficient phylogenetic signal in the data and that it cannot be explained without assuming that a certain number of gene duplication and losses and/or horizontal gene transfers have actually occurred.

Current data do not allow sorting out definitely whether an α-OTC was already present in the LCA. There are a number of arguments favoring the fact that α-OTCs appeared from an early duplication within Bacteria and were not present in the LCA: first, α-OTCs have been found so far only in Bacteria; second, all catabolic OTCs belong to the α subfamily, except for the only archaeal catabolic OTC characterized up to now, that of *H. salinarium*; third, most α-OTCs contain a characteristic signature sequence in their C-terminal part that is not found in either β- or γ-OTC (Fig. 7 and see also the alignment in Labedan et al., 1999). Only the gene from *T. maritima*, a species belonging to an early branching lineage within Bacteria, lacks this signature motif.

We postulate that α-OTCs arose from a duplication of an ancestral OTC before the divergence of the extant bacterial lineages and they specialized in a catabolic role. This would explain the conservation of these paralogs in organisms that either contain a functional ADI pathway or have lost it in a relatively recent past and their absence in those lineages that lost the ADI pathway early in evolution. Under this scenario, bacterial α-OTCs acquired their characteristic insertion after *Thermotogales* had already constituted a differentiated lineage. On the other hand, the catabolic OTC present in *H. salinarium* would have independently evolved a catalytic role from an anabolic β-OTC. Later, the catabolic gene occasionally displaced the anabolic one in some bacterial lineages. These non-orthologous displacements must have occurred independently in several lineages, since α-OTCs playing an anabolic role can be found in some lineages whereas in other closely related ones a β-OTC plays this role. For instance, in gram-negative bacteria, *E. coli* and *Neisseria gonorrhoeae* harbor anabolic α-OTCs whereas *Moritella* (Xu et al., 2000) or *Pseudomonas* use a β-OTC for this function. The same scenario is found in low G + C gram-positive bacteria, where *Bacillus subtilis* and *Bacillus stearothermophilus* harbor an anabolic β-OTC and *Lactobacillus plantarum* uses an anabolic α-OTC. This displacement of function could have occurred easily, since it has been observed that a point mutation can convert a catabolic enzyme into an anabolic one (Baur et al., 1990).

The location of plant OTCs also deserves a closer scrutiny, since they do not group neither with cyanobacterial sequences, which would support a chloroplastic origin for plant OTCs with the accepted origin for plastids in Cyanobacteria, nor with eukaryotic sequences, hence, favoring an independent and common origin for eukaryotic OTCs. Their grouping with the Archaea *H. salinarium* at the base of β-OTCs, although not statistically well supported, also suggests their close relatedness and common origin, giving further support to the presence of β-OTCs in the LCA.

Summarizing, we hypothesize that the LCA harbored several genes belonging to the β-OTC subfamily and at least one belonging to the γ subfamily. This is in accordance with the model proposed by Woese (1998) for the LCA as a community of genomes. Evolution of the different lineages was accompanied by the loss of most paralogs, although a certain level of redundancy must have been maintained along evolution, at least in Bacteria. This hypothesis does not rule out horizontal gene transfer events after the differentiation of the main domains of organisms, but these transfers would be mainly limited to closely related organisms.

The analysis of CKs also shows that paralogy severely hinders the phylogenetic analysis of these genes.

Nevertheless, *Firmicutes* and *M. pneumoniae arcC* genes conform a monophyletic cluster in accordance with the organismal phylogeny, with the exception of the *E. coli* ARCM sequence. On the other hand, archaeal, eukaryal, and gram-negative bacterial CKs form a polyphyletic cluster. The available data do not allow us to determine whether this anomalous clustering is due to horizontal gene transfer, hidden paralogy, or both. Nevertheless, paralogs of CK-encoding genes are currently found in *E. coli*, *L. lactis*, and *S. aureus*. The phylogenetic analysis indicates that the paralogs found in *E. coli* are due neither to recent duplications nor to horizontal transfers from closely related organisms. Therefore, ancient duplications and subsequent losses should be seriously considered as possible explanation for these results.

The phylogenetic analysis of the three structural genes of the ADI pathway reveals a complex evolutionary story. Nevertheless, the organization of the ADI gene clusters suggests an early assemblage of these genes in operons. Moreover, our data suggest that the assemblage occurred independently in Bacteria and Archaea. All the ADI gene clusters determined so far from Bacteria preserve the ABC gene order, although deletions and insertions have occurred along evolution. Within the archaeal species studied so far, only those belonging to the genus *Halobacterium* harbor an *arcACB* ADI gene cluster. The set of regulatory genes associated to ADI also points to an independent assemblage. In this sense, all the regulatory genes associated either with or involved in the regulation of the expression of the ADI gene cluster in Bacteria belong to the ArgR or CRP/FNR protein families, whereas in Archaea the only regulator associated with ADI belongs to the IclR family of transcriptional regulators. Finally, the phylogenetic analysis of the structural genes present in ADI clusters also indicates an independent origin: first, the *arcA* gene present in *H. salinarium* has diverged much more from its bacterial counterparts than any of the other structural genes; second, all bacterial OTCs present in ADI clusters belong to the subfamily α-OTC, while the one present in *H. salinarium* belongs to the β subfamily.

Within Bacteria, the organization of ADI clusters shares distinct characteristics within the higher taxa: gram-negative bacteria containing the ADI pathway usually harbor clusters encompassing the structural genes, but no regulatory gene associated to these clusters has been found so far. *Spirochaetales* also contain a cluster with no regulatory gene associated. High-G + C gram-positive bacteria studied so far do not contain ADI clusters, although ADI-encoding genes are present. Low G + C gram-positive bacteria contain the most complex clusters, including dedicated regulatory genes and additional genes encoding for proteins of unknown function or not directly involved in the ADI pathway. These coincidences can be best explained if an ADI cluster was already present before the diversification of each of these taxa.

The three structural genes constituting the ADI pathway do not share any significant similarity among one another. Therefore, it is unlikely they may have arisen from the duplication of an ancestral gene. More likely, they evolved independently from ancestral genes encoding proteins of broad specificity with deiminase, carbamoylase, and kinase activities, respectively, and their functional relatedness favored their clustering.

We postulate that the genes required for the ADI pathway were already present in the LCA, and probably, the pathway was functional in this organism. We base this hypothesis on the fact that ADI-encoding genes can be found in organisms belonging to the three domains of life and, although its presence is limited, the presence of ADI-encoding genes in Eukarya and Archaea cannot be explained by horizontal transfer from Bacteria. Hence, the presence of this gene can be considered an ancestral feature. Interestingly, only amitochrondrial eukaryotes have conserved the arginine deiminase pathway. Although there is compelling evidence indicating that lack of mitochondria in many of these organisms is secondary (Bui et al., 1996; Hashimoto et al., 1998; Roger et al., 1998), the presence of genes such as those encoding for ADI and CK, not found so far in other eukaryotes, clearly points to the exceptionality of these organisms within Eukarya. To reconcile these disparate results, Chihade et al. (2000) have proposed that loss of mitochondria in these lineages occurred before the genetic exchange between the organellar and nuclear genomes was complete. The presence of the ADI pathway in these lineages fits well within this scenario, since the acquisition of the highly efficient aerobic functions provided by mitochondria would render the ADI pathway dispensable.

Acknowledgments

This research was funded by CICYT project ALI95-0038 (to GP) and DGES project PB98-1436 (FGC). We thank B. Labedan for his ATC-OTC alignment data and two anonymous reviewers for their comments on a previous version of the manuscript.

References

Abdelal, A.T., 1979. Arginine catabolism by microorganisms. Annu. Rev. Microbiol. 33, 139–168.

Adachi, J., Hasegawa, M., 1996. MOLPHY, programs for molecular phylogenetics based on maximum likelihood, Version 2.3. The Institute of Statistical Mathematics, Tokyo.

Baur, H., Tricot, C., Stalon, V., Haas, D., 1990. Converting catabolic ornithine carbamoyltransferase to an anabolic enzyme. J. Biol. Chem. 265, 14728–14731.

Bolotin, A., Wincker, P., Mauger, S., Jaillon, O., Malarme, K., Weissenbach, J., Ehrlich, S.D., Sorokin, A., 2001. The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ssp. *lactis* IL1403. Genome Res. 11, 731–753.

Brown, J.R., Doolittle, W.F., 1997. Archaea and the prokaryote-to-eukaryote transition. Microbiol. Mol. Biol. Rev. 61, 456–502.

Bui, E.T., Bradley, P.J., Johnson, P.J., 1996. A common evolutionary origin for mitochondria and hydrogenosomes. Proc. Natl. Acad. Sci. USA 93, 9651–9656.

Chihade, J.W., Brown, J.R., Schimmel, P.R., Ribas de Pouplana, L., 2000. Origin of mitochondria in relation to evolutionary history of eukaryotic alanyl-tRNA synthetase. Proc. Natl. Acad. Sci. USA 97, 12153–12157.

Copley, R.R., Bork, P., 2000. Homology among (βα)8 barrels: implications for the evolution of metabolic pathways. J. Mol. Biol. 303, 627–640.

Cunin, R., Glansdorff, N., Piérard, A., Stalon, V., 1986. Biosynthesis and metabolism of arginine in bacteria. Microbiol. Rev. 50, 314–352.

de la Fuente, J.L., Martín, J.F., Liras, P., 1996. New type of hexameric ornithine carbamoyltransferase with arginase activity in the cephamycin producers *Streptomyces clavuligerus* and *Nocardia lactamdurans*. Biochem. J. 320, 173–179.

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

Durbecq, V., Legrain, C., Roovers, M., Pierard, A., Glansdorff, N., 1997. The carbamate kinase-like carbamoyl phosphate synthetase of the hyperthermophilic archaeon *Pyrococcus furiosus*, a missing link in the evolution of carbamoyl phosphate biosynthesis. Proc. Natl. Acad. Sci. USA 94, 12803–12808.

Dykhuizen, D.E., Baranton, G., 2001. The implications of a low rate of horizontal transfer in *Borrelia*. Trends Microbiol. 9, 344–350.

Eisen, J.A., 1995. The RecA protein as a model molecule for molecular systematic studies of Bacteria: comparison of trees of RecAs and 16S rRNAs from the same species. J. Mol. Evol. 41, 1105–1123.

Feng, D.-F., Doolittle, R.F., 1987. Progressive sequence alignment as a prerequisite to correct phylogenetic trees. J. Mol. Evol. 25, 351–360.

Forterre, P., Philippe, H., 1999. Where is the root of the universal tree of life? Bioessays 21, 871–879.

Glansdorff, N., 1999. On the origin of operons and their possible role in evolution toward thermophily. J. Mol. Evol. 49, 432–438.

Glansdorff, N., 2000. About the last common ancestor, the universal life-tree and lateral gene transfer: a reappraisal. Mol. Microbiol. 38, 177–185.

Hashimoto, T., Sánchez, L.B., Shirakura, T., Müller, M., Hasegawa, M., 1998. Secondary absence of mitochondria in *Giardia lamblia* and *Trichomonas vaginalis* revealed by valyl-tRNA synthetase phylogeny. Proc. Natl. Acad. Sci. USA 95, 6860–6865.

Himmelreich, R., Hilbert, H., Plagens, H., Pirkl, E., Li, B.C., Herrmann, R., 1996. Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. Nucl. Acids Res. 24, 4420–4449.

Horowitz, N.H., 1945. On the evolution of biochemical synthesis. Proc. Natl. Acad. Sci. USA 31, 153–157.

Horowitz, N.H., 1965. The evolution of biochemical syntheses—retrospect and prospect. In: Bryson, V., Vogel, H.J. (Eds.), Evolving Genes and Proteins. Academic Press, New York, pp. 15–23.

Jain, R., Rivera, M.C., Lake, J.A., 1999. Horizontal gene transfer among genomes: the complexity hypothesis. Proc. Natl. Acad. Sci. USA 96, 3801–3806.

Jensen, R.A., 1976. Enzyme recruitment in evolution of new function. Annu. Rev. Microbiol. 30, 409–425.

Jermiin, L.S., 1996. K2WuLi version 1.0. Australian National University, Camberra.

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

Kimura, M., 1980. Average time until fixation of a mutant allele in a finite population under continued mutation pressure: studies by analytical, numerical, and pseudo-sampling methods. Proc. Natl. Acad. Sci. USA 77, 522–526.

Kishino, H., Hasegawa, M., 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. J. Mol. Evol. 29, 170–179.

Knodler, L.A., Schofield, P.J., Edwards, M.R., 1995. L-Arginine transport and metabolism in *G. intestinalis* support its position as a transition between prokaryotic and eukaryotic kingdoms. Microbiology 149, 2063–2070.

Kumar, S., 1996. PHYLTEST: a program for testing phylogenetic hypothesis, version 2.0. Institute of Molecular Evolutionary Genetics and Department of Biology, The Pennsylvania State University.

Kuroda, M., Ohta, T., Uchiyama, I., Baba, T., Yuzawa, H., Kobayashi, I., Cui, L., Oguchi, A., Aoki, K., Nagai, Y., Lian, J., Ito, T., Kanamori, M., Matsumaru, H., Maruyama, A., Murakami, H., Hosoyama, A., Mizutani-Ui, Y., Kobayashi, N., Sawano, T., Inoue, R., Kaito, C., Sekimizu, K., Hirakawa, H., Kuhara, S., Goto, S., Yabuzaki, J., Kanehisa, M., Yamashita, A., Oshima, K., Furuya, K., Yoshino, C., Shiba, T., Hattori, M., Ogasawara, N., Hayashi, H., Hiramatsu, K., 2001. Whole genome sequencing of meticillin-resistant *Stapylococcus aureus*. The Lancet 357, 1225–1240.

Labedan, B., Boyen, A., Baetens, M., Charlier, D., Chen, P., Cunin, R., Durbeco, V., Glansdorff, N., Herve, G., Legrain, C., Liang, Z., Purcarea, C., Roovers, M., Sánchez, R., Toong, T.-L., Van de Casteele, M., van Vliet, F., Xu, Y., Zhang, Y.-F., 1999. The evolutionary history of carbamoyltransferases: a complex set of paralogous genes was already present in the last universal common ancestor. J. Mol. Evol. 49, 461–473.

Lawrence, J., 1999. Selfish operons: the evolutionary impact of gene clustering in prokaryotes and eukaryotes. Curr. Opin. Genet. Dev. 9, 642–648.

Lawrence, J., Roth, J.R., 1996. Selfish operons: horizontal transfer may drive the evolution of gene clusters. Genetics 143, 1843–1860.

Ludwig, R.A., 1993. *Arabidopsis* chloroplasts dissimilate L-arginine and L-citrulline for use as nitrogen source. Plant Physiol. 101, 429, 434.

Muse, S.V., Weir, B.S., 1992. Testing for equality of evolutionary rates. Genetics 132, 269–276.

Page, R.D.M., 1998. GeneTree: comparing gene and species phylogenies using reconciled trees. Bioinformatics 14, 819–820.

Page, R.D.M., Charleston, M.A., 1997. From gene to organismal phylogeny: reconciled trees and the gene tree/species tree problem. Mol. Phylogen. Evol. 7, 231–240.

Philippe, H., Forterre, P., 1999. The rooting of the universal tree is not reliable. J. Mol. Evol. 49, 509–523.

Robinson, M., Gouy, M., Gautier, C., Mouchiroud, D., 1998. Sensitivity of the relative-rate test to taxonomic sampling. Mol. Biol. Evol. 15, 1091–1098.

Roger, A.J., Svard, S.G., Tovar, J., Clark, C.G., Smith, M.W., Gillin, F.D., Sogin, M.L., 1998. A mitochondrial-like chaperonin 60 gene in Giardia lamblia: evidence that diplomonads once harbored an endosymbiont related to the progenitor of mitochondria. Proc. Natl. Acad. Sci. USA 95, 229–234.

Ruepp, A., Soppa, J., 1996. Fermentative arginine degradation in *Halobacterium salinarium* (formerly *Halobacterium halobium*): genes, gene products, and transcripts of the arcRACB gene cluster. J. Bacteriol. 178, 4942–4947.

Schofield, P.J., Edwards, M.R., Matthews, J., Wilson, J.R., 1992. The pathway of arginine catabolism in *Giardia intestinalis*. Mol. Biochem. Parasitol. 51, 29–36.

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

Strimmer, K., von Haeseler, A., 1997. Likelihood mapping: a simple method to visualize phylogenetic content of a sequence alignment. Proc. Natl. Acad. Sci. USA 94, 6815–6819.

Takezaki, N., Rzhetsky, A., Nei, M., 1995. Phylogenetic test of the molecular clock and linearized trees. Mol. Biol. Evol. 12, 823–833.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucl. Acids Res. 22, 4673–4680.

Woese, C.R., 1987. Bacterial evolution. Microbiol. Rev. 51, 221–271.

Woese, C.R., 1998. The universal ancestor. Proc. Natl. Acad. Sci. USA 95, 6854–6859.

Wu, C.-I., Li, W.-H., 1985. Evidence for higher rates of nucleotide substitution in rodents than in man. Proc. Natl. Acad. Sci. USA 82, 1741–1745.

Yang, Z., 2000. Phylogenetic Analysis by Maximum Likelihood (PAML), Version 3.0. University College London, London, England.

Yarlett, N., Lindmark, D.G., Goldberg, B., Moharrami, M.A., Bacchi, C.J., 1994. Subcellular localization of the enzymes of the arginine dihydrolase pathway in *Trichomonas vaginalis* and *Tritrichomonas foetus*. J. Eukaryotic Microbiol. 41, 554–559.

Xu, Y., Liang, Z., Legrain, C., Ruger, H.J., Glansdorff, N., 2000. Evolution of arginine biosynthesis in the bacterial domain: novel gene–enzyme relationships from psychrophilic *Moritella* strains (*Vibrionaceae*) and evolutionary significance of *N*-α-acetyl ornithinase. J. Bacteriol. 182, 1609–1615.