J Mol Evol (1999) 49:461–473

JOURNAL OF MOLECULAR EVOLUTION

The Evolutionary History of Carbamoyltransferases: A Complex Set of Paralogous Genes Was Already Present in the Last Universal Common Ancestor

Bernard Labedan,[1] Anne Boyen,[2,3] Margot Baetens,[2,3] Daniel Charlier,[3,4] Pingguo Chen,[2,3] Raymond Cunin,[2,3] Virginie Durbeco,[5] Nicolas Glansdorff,[2–4] Guy Herve,[6] Christianne Legrain,[4] Ziyuan Liang,[2,3] Christina Purcarea,[2,6] Martine Roovers,[2,3] Rony Sanchez,[2,3] Thia-Lin Toong,[2,3] Marc Van de Casteele,[3] Françoise van Vliet,[4] Ying Xu,[2,3] Yuan-Fu Zhang[2,*]

[1] Institut de Génétique et de Microbiologie, CNRS UMR 8621, Université Paris-sud, Bâtiment 409, F-91405 Orsay Cedex, France
[2] Microbiologie, Vrije Universiteit Brussel, 1 ave E. Gryson, B-1070 Brussels, Belgium
[3] Vlaams Interuniversitair Instituut voor Biotechnologie, 1 ave E. Gryson, B-1070 Brussels, Belgium
[4] Research Institute, CERIA-COOVI, 1 ave E. Gryson, B-1070 Brussels, Belgium
[5] Laboratoire de Microbiologie, Universite Libre de Bruxelles, 1 ave E. Gryson, B-1070 Brussels, Belgium
[6] Laboratoire de Biochimie des Signaux Régulateurs Cellulaires et Moléculaires, URA CNRS 1682-Université Pierre et Marie Curie, 96, Bd. Raspail, 75006 Paris, France

Abstract. Forty-four sequences of ornithine carbamoyltransferases (OTCases) and 33 sequences of aspartate carbamoyltransferases (ATCases) representing the three domains of life were multiply aligned and a phylogenetic tree was inferred from this multiple alignment. The global topology of the composite rooted tree (each enzyme family being used as an outgroup to root the other one) suggests that present-day genes are derived from paralogous ancestral genes which were already of the same size and argues against a mechanism of fusion of independent modules. A closer observation of the detailed topology shows that this tree could not be used to assess the actual order of organismal descent. Indeed, this tree displays a complex topology for many prokaryotic sequences, with polyphyly for Bacteria in both enzyme trees and for the Archaea in the OTCase tree. Moreover, representatives of the two prokaryotic Domains are found to be interspersed in various combinations in both enzyme trees. This complexity may be explained by assuming the occurrence of two subfamilies in the OTCase tree (OTC α and OTC β) and two other ones in the ATCase tree (ATC I and ATC II). These subfamilies could have arisen from duplication and selective losses of some differentiated copies during the successive speciations. We suggest that Archaea and Eukaryotes share a common ancestor in which the ancestral copies giving the present-day ATC II/OTC β combinations were present, whereas Bacteria comprise two classes: one containing the ATC II/OTC α combination and the other harboring the ATC I/OTC β combination. Moreover, multiple horizontal gene transfers could have occurred rather recently amongst prokaryotes. Whichever the actual history of carbamoyltransferases, our data suggest that the last common ancestor to all extant life possessed differentiated copies of genes coding for both carbamoyltransferases, indicating it as a rather sophisticated organism.

Key words: Carbamoyltransferases — ATCase — OTCase — Protein evolution — Gene duplication — Paralogous proteins — Last universal common ancestor — Molecular phylogeny

* *Present address:* Institute of Epidemiology and Microbiology, Beijing, The Peoples Republic of China
Correspondence to: B. Labedan; *e-mail:* labedan@igmors.u-psud.fr

Introduction

During the last 15 years, several important data have been gathered on the putative evolutionary history of two essential enzymes, ornithine carbamoyltransferase (OTCase; EC 2.1.3.3) and aspartate carbamoyltransferase (ATCase; EC 2.1.3.2). Sequence analysis and other properties strongly suggest that these carbamoyltransferases are the products of paralogous (Fitch 1970) genes, as follows.

Ubiquity and Functional Similarities

OTCases and ATCases, which are found in many bacteria and eukaryotes (see Cunin et al. 1986; Davis et al. 1986), as well as in archaea (Meile and Leisinger 1984; van de Casteele et al. 1990, 1997; Ruepp et al. 1995; Legrain et al. 1995; Purcarea et al. 1994, 1997; Roovers et al. 1997), catalyze analogous reactions. ATCase catalyzes the reaction between carbamoyl phosphate and aspartate that initiates the pathway for pyrimidine biosynthesis. OTCase serves two functions in arginine metabolism. (1) In arginine biosynthesis, it catalyzes the transfer of the carbamoyl moiety of carbamoyl phosphate to the 5-amino group of ornithine, thereby forming citrulline. In at least one case, that of *Escherichia coli* K12, there are two anabolic OTCases. The ArgF and ArgI proteins form true isoenzymes (Legrain et al. 1972) and are highly similar [86% identical amino acids (Van Vliet et al. 1988)]. The *ArgF* gene may have been acquired by lateral transfer, since it is flanked by two IS1 elements and both its GC content and its codon usage are not typical of *E. coli* (for a detailed commentary, see Van Vliet et al. 1988). (2) The arginine deiminase pathway of arginine catabolism, comprising arginine deiminase, OTCase, and carbamate kinase, is widely distributed among prokaryotic organisms (Cunin et al. 1986) but has not been found in eukaryotes, except for protozoan parasites, *Trichomonas vaginalis* and *Giardia intestinalis* (Linstead and Cranshaw 1983; Schofield et al. 1992), a feature that might support the organisms' position as a transition between prokaryotes and eukaryotes (Knodler et al. 1995). (3) In several cases (for instance, in *Pseudomonas aeruginosa*), two genes have been found to encode for a catabolic OTCase (the *arcB* gene) and an anabolic one (the *argF* gene), respectively. Remarkably, immunological studies have shown that the catabolic OTCase from *P. aeruginosa* shares common antigenic determinants with the anabolic ones from other bacterial species (especially enterobacteria) but not with the anabolic enzyme of the same species (Falmagne et al. 1985). This evidence and later sequence data suggested that anabolic enterobacterial OTCases, including the ArgF protein, descend from a catabolic OTCase, possibly related to that of *Aeromonas* (see Van Vliet et al. 1988). It appears that only a small number of mutations are needed to transform an allosteric catabolic OTCase into a nonallosteric, anabolic counterpart, or vice versa (Baur et al. 1990; Kuo et al. 1989).

Sequence Similarities

Both carbamoyltransferases, which are multimeric proteins, share a common functional domain, the carbamoyl phosphate binding domain (Wild and Wales 1990). This carbamoyl phosphate binding domain (N-terminal moiety) displays significant proportions of identical or similar residues for both carbamoyltransferases. In the C-terminal moiety, responsible for the binding of the amino acid substrate, the primary sequences are more divergent but significant conservation could be found at the level of secondary structures (Houghton et al. 1984; van Vliet et al. 1984). The actual structural similarity of the two carbamoyltransferases was recently confirmed by the three-dimensional analysis of the catabolic OTCase of *Pseudomonas aeruginosa* (the *arcB* gene) and comparison of its globular structure to that of the ATCase from *E. coli* (Villeret et al. 1995).

The evidence summarized above suggests that carbamoyltransferases arose by duplication of an ancestral gene. However, until recently, only bacterial and eukaryotic ATCase sequences and only one archaeal OTCase (Ruepp et al. 1995) had been reported. A critical testing of the paralogy issue now seems possible, with the characterization of two ATCases and one OTCase from thermophilic archaea by members of our group (see Table 1) and the sequencing of the *Methanococcus jannaschii* genome (Bult et al. 1996). Besides, several new bacterial sequences have been included, mostly from extremophilic carbamoyltransferases studied in our group (see Table 1).

We decided to use the now considerably expanded collection of sequence data, consisting of 33 OTCases and 31 ATCases, to reconstruct their respective gene trees, using one gene tree to root the other one. The comparison of paralogous proteins already present in the last common ancestor to the three domains of life (Woese et al. 1990) can be used to infer the actual order of organismal descent in cellular evolution (see, e.g., Gogarten et al. 1989; Iwabe et al. 1989; Benachenhou-Lahfa et al. 1993; Brown et al. 1994; Brown and Doolittle 1995). In the present paper, we further show that such paralogous proteins could be useful to reconstruct not only the early evolutionary history of genes present in the last university common ancestor but also very ancient events which may have occurred before the emergence of this last universal common ancestor.

Materials and Methods

DataBank Retrieval and Multiple Sequence Alignments

Newly determined sequences (Table 1) were used as queries to retrieve all the other sequences of OTCases and ATCases present in the databases using the Blastp and Fasta programs.

Table 1. Sequences of extremophilic carbamoyltransferases recently determined in our laboratories

Enzyme	Species	Domain[a]	Source(s)	Accession No.
Ornithine carbamoyltransferase (OTCase)	*Pyrococcus furiosus*	A	Roovers et al. (1997)	X99225
	Thermotoga maritima	B	M. Van de Casteele	Y10661
	Thermus thermophilus	B	Sanchez et al. (1997)	Y10266
	Vibrio species 2693	B	Z. Liang	Y11033
Asparate carbamoyltransferase (ATCase)	*Pyrococcus abyssi*	A	Purcarea et al. (1997)	U61765
	Sulfolobus solfataricus	A	V. Durbecq, D. Charlier	X99872
	Thermotoga maritima	B	C. Pingguo, M. Van de Casteele	Y10300
	Thermus aquaticus	B	Van de Casteele et al. 1997	Y09536
	Vibrio species 2693	B	M. Van de Casteele, Z. Yuan-Fu, Y. Xu, Z. Liang	Y09786

[a] A, domain of Archaea; B, domain of Bacteria.

Preliminary multiple alignments of the amino acid sequences were generated using the Clustal W program (Thompson et al. 1994) and the program AllAll, which is part of an interactive search system, the so-called DARWIN (Data Analysis and Retrieval With Indexed Nucleotide/Peptide Sequences) system available, at the Computational Biochemistry Research Group server at the ETH, Zurich, Switzerland (Gonnet et al. 1992). The crude multiple alignment obtained by the two automatic methods were manually optimized by reference to known tertiary structural data (Ke et al. 1984; Stevens et al. 1991; Villeret et al. 1995) using two biological sequence editors (SeqApp 1.9a and SeqPup 0.6f) developed by Don Gilbert (Biocomputing Office, Biology Department, Indiana University, Bloomington, IN 47405, USA).

Construction of Phylogenetic Trees

Two approaches were used to derive phylogenetic trees from the multiple alignment of the 64 sequences of OTCases and ATCases.

In the AllAll program, the sequences to be multiply aligned are organized as sets of evolutionarily connected components which are characterized by an evolutionary distance measured in PAM units. The PAM distance—the number of accepted point mutations per 100 residues separating two sequences—is based on (1) a mutation data matrix normalized to a distance of 250 PAM and recomputed for each new set of sequences and (2) a gap penality which is itself dependent on the PAM distance intrinsic to the set of sequences studied. The trees obtained using this program are based on the estimated PAM distances between each pair of sequences and the deduced evolutionary distance computed using a least-squares approach is weighted by computing the variance of the respective distance. Therefore, these distance trees are approximations to maximum-likelihood trees [see Gonnet et al. (1992) for additional details on the method and the booklet available at the Internet address http://cbrg.inf.ethz.ch/ServerBooklet, especially the subsection 2-3-5-1].

In the second approach, the multiple alignment obtained after manual improvement was used to implement more classical phylogenetic inference programs, using either maximum-parsimony or distance methods. The programs PROTPARS, FITCH, and NEIGHBOR from the PHYLIP package [version 3.55c (Felsenstein 1993)] were used. Moreover, the programs SEQBOOT, PROTPARS, and CONSENSE and the programs SEQBOOT, FITCH or NEIGHBOR, and CONSENSE were successively used to derive confidence limits, estimated by 100 bootstrap replicates, for each node in the maximum parsimony and distance tree, respectively. We further used a PAUP program [version 3.1.1 (Swofford 1993)] using the bootstrap method (heuristic search option, seed = 1) to analyze 500 replicates of smaller sets.

Results

Collecting Sequences of OTCase and ATCase Carbamoyltransferases

Table 1 reports a list of recently (November 1997) determined OTCase and ATCase genes sequences from different extremophilic prokaryotes (described in detail elsewhere). The predicted amino acid sequences displayed significant levels of identity with the complete sequences of the corresponding enzymes presently available in the databanks and also with the N-half of the other available carbamoyltransferases. This confirmed that the similarities between the OTCases and the ATCases are confined predominantly to the N-terminal half and, also, to the extreme C terminus of the molecule [Helix 12 (Honzatko et al. 1982)] corresponding to the carbamoyl phosphate binding domain, but that a large portion of the C-terminal half required for binding either ornithine or aspartate has diverged rather strongly.

Using this homology search we harvested 64 sequences of carbamoyltransferases corresponding to 33 OTCases and 31 ATCases (including our newly determined sequences). From the nearly identical seven sequences of OTCase from *Neisseria* spp., we used that of *Neisseria gonorrhoeae.* Surprisingly, there was no putative homologue of an ATCase-encoding gene in the completely determined sequence of the *Haemophilus influenzae* (strain Rd KW20) genome (Fleischman et al. 1995). We have checked this point using different approaches directly on the entire *H. influenzae* sequence (available on the Web server of The Institute of Genetic Research), but we could detect only faint similarities besides the putative homologue for an OTCase-encoding gene (called *arcB-B*). More recently, we obtained the same result with the complete sequence of *Mycoplasma pneumoniae.* A common feature of *H. influenzae* and *M. pneumoniae* is their parasitic lifestyle. It will be interest-

ing to see if other parasites are devoid of ATCase when more whole-genome sequences appear. (In the case of *M. genitalium*, both carbamoyltransferases were apparently absent.)

Establishing a Multiple Alignment

The 64 available sequences of OTCases and ATCases were further aligned, using two automatic procedures. Both the AllAll program and the Clustal W algorithm gave similar outputs, where the alignments were satisfactory for the N-halves but ambiguous for the C-halves. We tried to improve this multiple alignment by hand after visual examination relying mainly on the secondary structures. Indeed, the crystal structures of two ATCases, *E. coli* (Ke et al. 1984) and *B. subtilis* (Stevens et al. 1991) and the catabolic OTCase from *P. aeruginosa (*Villeret et al. 1995) have been determined. Their comparison showed that the respective positions of α-helices and β-strands are highly conserved in both carbamoyltransferases (see, e.g., Villeret et al. 1995). Therefore, starting from the experimentally determined secondary structures, we have been able to adjust progressively the alignment of all the diverged sequences, using as a guide the respective positions of the α-helices and β-sheets found in ATCases and OTCases. Moreover, a close scrutiny of all the alignments displayed in the Fasta and Blast outputs revealed that several sequences of OTCases and ATCases have maintained significant local identity in several amino acid blocks within their substrate-binding domain. This was the case, the example, for the ATCase from *Pyrococcus abyssi*, which displays in its C-half enough similarities to allow it to align along almost the entire sequence of the OTCase from *B. subtilis.* Using this approach, we could progressively realign all the sequences, including those which have diverged too far to be recognized by the similarity search programs. A partial view of this complete multiple alignment is shown in Fig. 1, where the species displaying both carbamoyltransferases plus our new sequences are exhibited. The complete alignment of the 64 sequences may be obtained by e-mail (labedan@igmors.u-psud.fr).

Constructing a Phylogenetic Tree

Two independent approaches were used to reconstruct a composite phylogenetic tree for all carbamoyltransferases. (1) The whole set of 64 sequences was used to implement the AllAll program (Gonnet et al. 1992) to obtain directly a rooted tree. (2) The reconstructed multiple alignment, taken as a whole, with the exception of a few regions corresponding to large gaps in a majority of sequences (e.g., positions 131–141, 170–176, and 311–328 in Fig. 1), was used to implement different programs from the PHYLIP package and also the maximum-parsimony program PAUP (Swofford 1993).

Figure 2 shows the obtained composite tree, where each paralogous enzyme is used as an outgroup to root the tree of the other one. This tree displays the following features:

General Topology. The same topology was obtained irrespective of the tree-making algorithm. This topology did not change when only the carbamoyl phosphate binding domains (positions 1 to 170 plus positions 355 to the end) were used instead of the complete sequences. The substrate binding domains (positions 171 to 354) also gave a similar topology except for the branching of several prokaryotic sequences in the OTCase tree (see below). The addition or deletion of various sequences did not substantially affect the global topology. Therefore, this tree appears to be fairly robust.

Characterization of Subfamilies. Although robust, the topology appears rather complex for each gene family. Putative subfamilies within each of the two protein sets were identified by searching for unique signature sequences. This led to the identification of two ATCase and two OTCase subfamilies.

(a) The ATC I subfamily corresponds to the proteins from the genera *Bacillus, Lactobacillus,* and *Pseudomonas* and from the species *Mycobacterium tuberculosis, Synechocystis* sp., and *Thermus aquaticus* ZO5 (only Bacteria up to now). The ATCase II subfamily, comprising all the other sequences, spans all three domains of life. If we exclude the *Drosophila* sequence, which in many places seems dubious, these two ATCase subfamilies display signatures at positions 106, 198, 199, 210, 266, 268, 301, 317, 342, 345, and 350 (106, 197, 198, 209, 260, 262, 294, 311, 330, 333, and 338, respectively, in Fig. 1) of the multiple alignment. Moreover, slightly degenerated signatures appear at positions 202, 315, and 344 (201, 309, and 332, respectively, in Fig. 1).

(b) As shown in Fig. 2, the OTC α subfamily corresponds to the proteins from various bacteria from *E. coli* to *Thermotoga maritima* and OTC β contains the other prokaryotic enzymes including the archaeal ones, that encoded by the chloroplast from *Pisum sativum,* and the eukaryotic sequences. These two subfamilies display signatures at positions 116, 156, 290, and 345 (116, 156, 283, and 338, respectively, in Fig. 1). Moreover, slightly degenerated signatures (where *T. maritima* is grouped with the other family) appear at positions 30, 49, and 346 (30, 49, and 334, respectively, in Fig. 1).

Location of the Species in Each Enzyme Tree. Each protein tree can, in principle, be rooted by its paralogous one. However, several features found in each tree make it impossible at this stage to interpret them as species trees and to use them to infer the actual order of organismal descent in cellular evolution.

(a) In the ATCase tree, bacteria appear clearly poly-

OTC α

1 *Escherichia coli argI*
2 *Escherichia coli argF*
3 *Salmonella typhimurium argI*
4 *Pseudomonas aeruginosa arcB*
5 *Thermotoga maritima*

OTC β

6 *Synechocystis species*
7 *Mycobacterium bovis*
8 *Pseudomonas aeruginosa argF*
9 *Vibrio species*
10 *Thermus thermophilus*
11 *Halobacterium salinarium*
12 *Bacillus subtilis argF*
13 *Pyrococcus furiosus*
14 *Methanococcus jannaschii*
15 *Schizosaccharomyces pombe ARG3*
16 *Saccharomyces cerevisiae ARG3*
17 *Homo sapiens OTC*

ATC I

18 *Pseudomonas aeruginosa pyrB*
19 *Synechocystis species pyrB*
20 *Bacillus subtilis pyrB*
21 *Thermus aquaticus pyrB*
22 *Mycobacterium tuberculosis pyrB*

ATC II

23 *Thermotoga maritima pyrB*
24 *Salmonella typhimurium pyrB*
25 *Escherichia coli pyrB*
26 *Vibrio species pyrB*
27 *Pyrococcus abyssi pyrB*
28 *Sulfolobus solfataricus pyrB*
29 *Methanococcus jannaschii pyrB*
30 *Saccharomyces cerevisiae PYR1*
31 *Schizosaccharomyces pombe PYR1*
32 *Homo sapiens PYR1*

C Consensus for all sequences

```
                10        20        30        40        50        60        70        80        90       100       110       120
 1  YHKHFLKLLDFTPAELNSLLQLAAKLKADKKSGKE-----EAK-LTGKNIALIFEKDSTRTRCSFEVAA--YDQGARVTYL--GPSGSQIGHKESIKDTARVLGRM--YDGIQYRGYGQE  108
 2  YKKHFLKLLDFTPAQFTSLLTLAAQLKADKKNGKE-----VQK-LTGKNIALIFEKDSTRTRCSFEVAA--FDQGARVTYL--GPSGSQIGHKESIKDTARVLGRM--YDGIQYRGHGQE  108
 3  YQKPFLKLLDFTASELTALLQLAAKLKADKKNGKE-----EQK-LVGKNIALIFEKDSTRTRCSFEVAA--YDQGARVTYL--GSSGSQIGHKESIKDTARVLGRM--FDGIQYRGYGQE  108
 4  HNRNLLSLMHHSTRELRYLLDLSRDLKRAKYTGTE-----QQH-LKRKNIALIFEKTSTRTRCAFEVAA--YDQGANVTYI--DPNSSQIGHKESMKDTARVLGRM--YDAIEYRGFKQE  108
 5  KGRSLLTLLDFSPEEIRYLLDISKQVKMENRSKLR-----TER-FKGMTLAMIFEKRSTRTRLAFETAF--AEEGGHPIFL--SPNDIHLGAKESLEDTARVLGRM--VDAIMFRGYKQE  108

 6  AGRDLLAIADLTIEEMKSLLQLAADLKSGVL--KP---------HCRKILGLLFYKASTRTRVSFTAAM--YQLGGQVLDL--NPSVTQVGRGEPIQDTARVLDRY--IDILAVRTFKQT  103
 7  -IRHFLRDDDLSPAEQAEVLELAAELKKDPVSRRP-----LQGP---RGVAVIFDKNSTRTRFSFELGI--AQLGGHAVVV--DSGSTQLGRDETLQDTAKVLSRY--VDAIVWRTFGQE  105
 8  SVRHFLSFMDYSPEELIGLIRRGSELKDLRNRGVL-----YEP-LKSRVLGMVFEKASTRTRLSFEAGM--IQLGGQAIFL--SPRDTQLGRGEPIGDSARVMSRM--LDGVMIRTFAHA  108
 9  --ENLLSVKDLSKQQILDLLALAKAVKANPAE--------YSQALAGKSIVTIYEKPSLRTRVTFDIGI--HKLGGHAVYL--DAQNGAIGERETVKDFAANISRW--ADGIVARVVSHK  104
10  LPKDLLDFSGYGPKELQALLDLAERLKRERYRGED---------LKGKVLALLFEKPSLRTRTTLEVAM--VHLGGHAVYL--DQKQVGIGEREPVRDVAKNLERF--VEGIAARVFRHE  105
11  --EHLVDINDVESEEIEQLLDLAASMKENPGE--------FSGVMDNKSLVMLFAKPSTRTRLSFETGM--TQLGGHGIFF--EMGSSQLSRGEPISDVSQVMSRY--EDAIMARLFEHD  104
12  YGKDLLTLKDLSEEDINALLAEAGELKQ-NKI---------QPIFHGKTLAMIFEKSSTRTRVSFEAGM--AQLGGSALFL--SQKDLQLGRGETVADTAKVLSGY--VDAIMIRTFEHE  104
13  AGRDLLCLQDYTAEEIWTILETAKMFKIWQKIGKP------HRLLEGKTLAMIFQKPSTRTRVSFEVAM--AHLGGHALYL--NAQDLQLRRGETIADTARVLSRY--VDAIMARVYDHK  108
14  ---HLLDLDVLSREDVLKIIEYGIYFKK-NRRKHE-------KILEGKSVAILFEKPSTRTRMSFDIAV--YELGGHPLIM--NQNEIHLGKKESIKDTAKVMGRY--VDTIVARVYKHR  103
15  -PRHLLSIRDLSRGEIVKLIDRSSEIKQAYKQNFQNRRSVQMSGLSSQNVAMIFSKRSTRTRVSVESAV--SCLGGNAMFL--GKDDIQLGVNESLYDTSKVISSM--VSGIVARVNKYS  113
16  SLRHLISIKDLSDEEFRILVQRAQHFKNVFKANKTNDFQSNHLKLLGRTIALIFTKRSTRTRISTEGAA--TFFGAQPMFL--GKEDIQLGVNESFYDTTKVVSSM--VSCIFARVNKHE  114
17  KGRDLLTLKNFTGEEIKYMLWLSADLKFRIKQKGE-----YLPLLQGKSLGMIFEKRSTRTRLSTETGL--ALLGGHPCFL--TTQDIHLGVNESLTDTARVLSSM--ADAVLARVYKQS  109

 C  ..:+:+:. +.:.:+.  :+..+..::   :...        ..+.++..+.:*.::S*RT*.**+.+:   .+*:.:..:  ....:..:::E::.D*:::::::  :*:+:.R:....

18  QLRHFISLDGLPRELLTEILDTADSFLEVGARAVK-----KVPLLRGKTVCNVFFENSTRTRTTFELAA--QRLSADVISL--NVSTSSTSKGETLTDTLRNLEAM-AADMFVVRHSDSG  110
19  TRNHILDLTDWRGEELDIVLQTATTFQQVLSGQTK-----KVPALQGQVVTNLFFEPSTRTRSSFELAA--KRLSADVMNF--SPGTSSLTKGETILDTAKTYLAM-GSDIFVIRHQQAG  110
20  --KHLTTMSELSTEEIKDLLQTAQEL----KSGKT-----DNQ-LTGKFAANLFFEPSTRTRFSFEVAE--KKLGMNVLNL--DGTSTSVQKGETLYDTIRTLESI-GVDVCVIRHSEDE  103
21  --RHLLDFQGWTRPEVESLLDTARVMREVLERPVK-----KVPALQGFTVATVFFEPSTRTRISFELAA--RRMSADVVSF--AAQTSSLQKGESYKDTLLTLEAM-GVDAYVIRADSAG  108
22  TPRHLLTAADLSRDDATAILDDADRFAQALVGRDI----KKLPTLRGRTVVTMFYENSTRTRVSFEVAGKW--MSADVINV--SAAGSSVGKGESLRDTALTLRAA-GADALIIRHPASG  111

23  -----------FLYEKTRELKQ--RWYSGEDVSDF------RIKKRNVGIYIVFVEPSTRTKESFINAAKFHSGPNVKVNVFD-SEHSSFNKQESYTDTFSMLTGYSDYSIFVVRTRLEG  100
24  YQKHIISINDLSRDDLNLVLATAAKL----KANPQ-----PE-LLKHKVIASCFFEASTRTRLSFETSM--HRLGASVVGFSDSANTSLGKKGETLADTISVISTY--VDAIVMRHPQEG  106
25  YQKHIISINDLSRDDLNLVLATAAKL----KANPQ-----PE-LLKHKVIASCFFEASTRTRLSFETSM--HRLGASVVGFSDSANTSLGKKGETLADTISVISTY--VDAIVMRHPQEG  106
26  FRKHIVSINDISRNELELIVKTAAKL----KKQPQ-----PE-LLKNKVIASCFFEASTRTRLSFETAI--QRLGGTVIGFDNASNTSLAKKGETLADSISVISSY--VDAFVMRHPQEG  106
27  KGRDVISIRDFSKEDIETVLATAERLERELKEKGQ-----LE-YAKGKILATLFFEPSTRTRLSFESAM--HRLGGAVIGFAEA-STSSVKKGESLRDTIKTVEQY--CDVIVIRHPKEG  109
28  -LKHIISAYNFSRDELEDIFALTDKY---SKNLND-----TRKILSGKTISIAFFEPSTRTYLSFQKAI--INLGGDVIGFSGEESTS-VAKGENLADTIRMLNNY--SDGIVMRHKYDG  106
29  --KHLISMKDIGKEEILEILDEARKMEELLNTKRP------LKLLEGKILATVFYEPSTRTRLSFETAM--KRLGGEVITMTD-LKSSSVAKGESLIDTIRVISGY--ADIIVLRHPSEG  107
30  -GRHILSIKQFKRSDFHVLFAVAQELRAAVAREGV-----LD-LMKGHVITTIFFEPSTRTCSSFIAAM--ERLGGRIVNVNPL--VSSVKKGETLQDTIRTLACY--SDAIVMRHSEEM  107
31  YRKHIISVHQVTRSDLHVLFAIAHQMRIIVERQGV-----ID-LCYGKLLCTMFFEPSTRTSSSFDAAM--QRLGGKVVAVTAS--ASSVNKGESLADTIRTLGCY--GDAIVLRHPSIE  108
32  VGQHILSVQQFTKDQMSHLFNVAHTLRMMVQKERS-----LD-ILKGKVMASMFYEVSTRTSSSFAAAM--ARLGGAVLSFSEA--TSSVQKGESLADSVQTMSCY--ADVVVLRHPQPG  108
```

Fig. 1. A partial view of the multiple alignment of the ornithine and aspartate carbamoyltransferases. OTCases and ATCases are separated by their consensus (line C, with decreasing order of degeneration as follows: letter corresponding to strict conservation, *, +, :). Only species displaying both carbamoyltransferases plus our new sequences are shown. Signatures specific to each subfamily (OTC α, OTC β, ATC I, ATC II) are shown in *boldface*. Slightly degenerated signatures are *underlined*.

OTC α

1 *Escherichia coli argI*
2 *Escherichia coli argF*
3 *Salmonella typhimurium argI*
4 *Pseudomonas aeruginosa arcB*
5 *Thermotoga maritima*

OTC β

6 *Synechocystis species*
7 *Mycobacterium bovis*
8 *Pseudomonas aeruginosa argF*
9 *Vibrio species*
10 *Thermus thermophilus*
11 *Halobacterium salinarium*
12 *Bacillus subtilis argF*
13 *Pyrococcus furiosus*
14 *Methanococcus jannaschii*
15 *Schizosaccharomyces pombe ARG3*
16 *Saccharomyces cerevisiae ARG3*
17 *Homo sapiens OTC*

ATC I

18 *Pseudomonas aeruginosa pyrB*
19 *Synechocystis species pyrB*
20 *Bacillus subtilis pyrB*
21 *Thermus aquaticus pyrB*
22 *Mycobacterium tuberculosis pyrB*

ATC II

23 *Thermotoga maritima pyrB*
24 *Salmonella typhimurium pyrB*
25 *Escherichia coli pyrB*
26 *Vibrio species pyrB*
27 *Pyrococcus abyssi pyrB*
28 *Sulfolobus solfataricus pyrB*
29 *Methanococcus jannaschii pyrB*
30 *Saccharomyces cerevisiae PYR1*
31 *Schizosaccharomyces pombe PYR1*
32 *Homo sapiens PYR1*

C Consensus for all sequences

```
                130       140       150       160       170       180       190       200       210       220       230       240
 1 IVETLAEYAS-----------VPVWNGL-TNEFHPTQLLADLLTMQEHL-------PGKAFNEMTLVYAGDARNN-MGNSMLEAAALTG-LDLRLVAPQACWP--EAALVTECRALAQQN  205
 2 VVETLAQYPG-----------VPVWNGL-TNEFHPTQLLADLMTMQEHL-------PGKAFNEMTLVYAGDARNN-MGNSMLEAAALTG-LDLRLLAPKACWP--EESLVAECSALAEKH  205
 3 IVETLAEYSG-----------VPVWNGL-TDEYHPTQLLADLLTMQEHL-------PGKAFNEMTLVYAGDARNN-MGNSMLEAAALTG-LDLRLVAPKACWP--QAALVAECSAMAKKN  205
 4 IVEELAKFAG-----------VPVFNGL-TDEYHPTQMLADVLTMREHS--------DKPLHDISYAYLGDARNN-MGNSLLLIGAKLG-MDVRIAAPKALWP--HDEFVAQCKKFAEES  204
 5 TVEKLAEYSG-----------VPVYNGL-TDEFHPTQALADLMTIEEN---------FGRLKGVKVVFMGDTRNN-VATSLMIACAKMG-MNFVACGPEELKP-RSDVFKRCQEIVKETD  204

 6 DLQTFADHAK-----------MPIINAL-SDLEHPCQILADLQTIKECF---------GKLEGLTVTYLGDG-NN-VAHSLILGGVMMG-MTVRVATPKNYEP------------IAAPG  187
 7 RLDAMASVAT-----------VPVINAL-SDEFHPCQVLADLQTIAER---------KGALRGLRLSYFGDGANN-MAHSLLLGGVTAG-IHVTVAAPEGFLP-DPSVRAAAERRAQDTG  201
 8 TLTEFAAHSK-----------VPVINGL-SDDLHPCQLLADMQTFHEH---------RGSIQGKTVAWIGDG-NN-MCNSYIEEEMKFD-FQLRVACPEGYEP-----KAEFVALA---G  196
 9 TLVGLVEHGS-----------VPVVNSL-CDLYHPCQALADFLTISEHY---------EDVSNLKLAYVGEG-NN-ATHSLMLTGAILG-AEVTAVCPRGSSP-DAQIVKQAMALAEISG  199
10 TVEALARHAK-----------VPVVNAL-SDRAHPLQALADLLTLKEVF---------GGLAGLEVAWVGDG-NN-VLNSLLEVAPLAG-LKVRVATPKGYEP-DPGLLKRANA------  194
11 EMMELAENAD-----------VPVVNGL-TDFLHPCQALTDMFTMQEK----------DRLD-TL-AFVGDG-NN-VAHSLMQASAKMG-VDCRIATPEGMEP-DEEIQDRVS------D  190
12 KVEELAKEAD-----------IPVINGL-TDKYHPCQALADLLTIKEI---------KGKLKGVKVAYIGDG-NN-VAHSLMIGCAKMG-CDISIASPKGYEVLDEAAEAAKT-YALQSG  199
13 DVEDLAKYAT-----------VPVINGL-SDFSHPCQALADYMTIWEK---------KGTIKGVKVVYVGDG-NN-VAHSLMIAGTKLG-ADVVVATPEGYEP-DEKVIKWAEQNAAESG  203
14 HLEEMAKYSS-----------VPVINAL-SDLAHPCQILADLMTIKEY---------KGKFKGLKIAYLGDG-NN-VCNSLILGSALVG-MDTYVGTPKGYEPNAKVVLKAKEIINNYGE  199
15 DVATLAKHAS-----------CPVINGL-CDTFHPLQALADLLTIKETF---------KSFDGLKVAWVGDA-NN-VLHDLMIANAKVG-IHTSVAKPKDVNV-RDDILSIVNEAANENG  208
16 DILAFCKDSS-----------VPIINSL-CDKFHPLQAICDLLTIIENFNISLDEVNKGINSKLKMAWIGDA-NN-VINDMCIACLKFG-ISVSISTPPGIEM-DSDIVDEAKKVAERNG  218
17 DLDTLAKEAS-----------IPIINGL-SDLYHPIQILADYLTLQEHY---------SSLKGLTLSWIGDG-NN-ILHSIMMSAAKFG-MHLQAATPKGYEP-DASVTKLAEQYAKENG  204

 C ....:+....           +*++N::::: .HP:Q:*:D+.*+.*..          + :.+:.:::.:G*..::::.:*:..::::..+ .......P..:.:..  . ....  .   .

18 AAHFIAEHVSP----------NVAVINGGDGRHAHPTQGMLDMLTIRRHK---------GNFEQLSVAIVGDILHSRVARSNMLALKTLGCPDIRVIAPATLLP-----IGLEEQY----G  203
19 VPHFIASQM------DRLQTGVKVLNAGDGQHEHPSQGLLDLFTICSQF--APDNPAIQCLQGKKIAVVGDILHSRVARSNLWSLTTAGADVHLAGPPTLLPK----EFQQLTLAP--GS  216
20 YYEELVSQV-----------NIPILNAGDGCGQHPTQSLLDLMTIYEEF---------NTFKGLTVSIHGDIKHSRVARSNAEVLTRLGARVLFS-GPSEWQ--------DEEF-----N  189
21 VPHQATRWVK-----------GAVINGGDGRRAHPTQALLDAYTLLEAL---------GTLEGKKVAIVGDILHSRVARSNAELLPLLGAQVFCA-GPPSLLPQ------------SLPG  195
22 AAHLLAQWTG------AHNDGPAVINAGDGTHEHPTQALLDALTIRQRL---------GGIEGRRIVIVGDILHSRVARSNVMLLDTLGAEVVLV-APPTLLPV------------GVTG  203

23 VCRLLERRISEFASRNGIE-VPSFINAGDGKHEHPTQELLDEYTFLEQNG--------FDNSFIHVALVGDLLHGRTVHSKVNGLKIFKNVKVDLVAPEELMMPEHYVEKMKKN--GFEV  209
24 AARLATEFSG----------QVPVLNAGDGSNQHPTQTLLDLFTIQETQ---------GRLDNLHIAMVGDLKYGRTVHFAKRTLAKFSGNRFYFIAPDALAMPQYILDMLDEK---GMA  204
25 AARLATEFSG----------NVPVLNAGDGSNQHPTQTLLDLFTIQETQ---------GRLDNLHVAMVGDLKYGRTVHSLTQALAKFDGNRFYFIAPDALAMPQYILDMLDEK---GIA  204
26 AARLASEFS-----------NVPVINGGDGSNQHPTQTLLDLFSIYETQ---------GCLDNLNIALVGDLKYGRTVHSLAQALAKFSGCKFYFIAPDALAMPEYICDELDEH---NVS  203
27 AARLAAEVA-----------EVPVINAGDGSNQHPTQTLLDLYTIKKEF---------GRIDGLKIGLLGDLKYGRTVHSLAEALT-FYDVELYLISPELLRMPRHIVEELREK---GMK  205
28 ASRFASEIS-----------DIPVINAGDGKHEHPTQAVIDIYTINKHF---------NTIDGLVFALLGDLKYARTVNSLLRILTRFRPKLVYLISPQLLRARKEILDELN----YPVK  202
29 AARLASEYS-----------QVPIINAGDGSNQHPTQTLLDLYTIMREI---------GRIDGIKIAFVGDLKYGRTVHSLVYALSLFENVEMYFVSPKELRLPKDIIEDLKAK---NIK  204
30 SVHIAAKYS-----------PVPIINGGNGSREHPTQAFLDLFTIREEI---------GTVNGITVTFMGDLKHGRTVHSLCRLLMHYQ-VRINLVSPPELRLPEGLREELRKA---GLL  203
31 SARIAANFS-----------PVPIINGGNGSKEHPTQAFLDLYTIREEL---------GSVNGLTITFIGDLKYGRTVHSLARLLAFWH-VELHLVSPEQLALPDDVKDDIRAN---GLN  204
32 AVELAAKHC-----------RRPVINAGDGVGEHPTQALLDIFTIREEL---------GTVNGMTITMVGDLKHGRTVHSLACLLTQYR-VSLRYVAPPSLRMPPTVRAFVASR---GTK  204
```

Fig. 1. Continued.

OTC α

1 *Escherichia coli argI*
2 *Escherichia coli argF*
3 *Salmonella typhimurium argI*
4 *Pseudomonas aeruginosa arcB*
5 *Thermotoga maritima*

OTC β

6 *Synechocystis species*
7 *Mycobacterium bovis*
8 *Pseudomonas aeruginosa argF*
9 *Vibrio species*
10 *Thermus thermophilus*
11 *Halobacterium salinarium*
12 *Bacillus subtilis argF*
13 *Pyrococcus furiosus*
14 *Methanococcus jannaschii*
15 *Schizosaccharomyces pombe ARG3*
16 *Saccharomyces cerevisiae ARG3*
17 *Homo sapiens OTC*

ATC I

18 *Pseudomonas aeruginosa pyrB*
19 *Synechocystis species pyrB*
20 *Bacillus subtilis pyrB*
21 *Thermus aquaticus pyrB*
22 *Mycobacterium tuberculosis pyrB*

ATC II

23 *Thermotoga maritima pyrB*
24 *Salmonella typhimurium pyrB*
25 *Escherichia coli pyrB*
26 *Vibrio species pyrB*
27 *Pyrococcus abyssi pyrB*
28 *Sulfolobus solfataricus pyrB*
29 *Methanococcus jannaschii pyrB*
30 *Saccharomyces cerevisiae PYR1*
31 *Schizosaccharomyces pombe PYR1*
32 *Homo sapiens PYR1*

C Consensus for all sequences

```
                250       260       270       280       290       300       310       320       330       340       350       360
 1  GGNITLTEDVAK-GVEGADFIYTDVWVSMGEAKEKWAERIALLREYQVNSKMMQLT-GNPEVKFLHCLPAFHDDQTTLGKKMAEEF-GLHGGMEVTDEVFESAASIVFDQAENRMHTIKAVMVATL 328
 2  GGKITLTEDVAA-GVKGADFIYTDVWVSMGEAKEKWAERIALLRGYQVNAQMMALT-DNPNVKFLHCLPAFHDDQTTLGKQMAKEF-DLHGGMEVTDEVFESAASIVFDQAENRMHTIKAVMMATL 328
 3  GGAITLTED---------------------------------------------------------------------------------------------------------------- 214
 4  GAKLTLTEDPKE-AVKGVDFVHTDVWVSMGEPVEAWGERIKELLPYQVNMEIMKAT-GNPRAKFMHCLPAFHNSETKVGKQIAEQYPNLANGIEVTEDVFESPYNIAFEQAENRMHTIKAILVSTL 328
 5  GSVSFTSNLEE--ALAGADVVYTDVWASMGEEDKE-KERMALLKPYQVNERVMEMT-GKSETIFMHCLPA----------------VKGQEVTYEVIEG-KQSRVWDEAENRKHTIKAVMIATL 307

 6  GKVELTDDPKA--AAQGSHILYTDVWASMGQEDLA-DSRIPIFQPYQINQELLALA--DPEAIVLHCLPA----------------HRGEEITDAVMEG-PQSRLWDQAENRMHAQKALMVALL 289
 7  ASVTVTADAHA--AAAAADVLVTDTWTSMGQENDG-LDRVKPFRPFQLNSRLLALA--DSDAIVLHCLPA----------------HRGDEITDAVMDG-PASAVWDEAENRLHAQKAVLVWLL 303
 8  DRLRVVRDPRE--AVAGAHLVSTDVWASMGQEDEA-AARIAMFRPYQVNAALLD--GAADDVLFMHCLPA----------------HRGEEISEELLDD-PRSVAWDQAENRLHAQKALLELLI 298
 9  GKINVTDNLDD---IVDYDVIYGDTWVSMGDDTPL-AQVKEKYMPYQINKALLMRT---GIKHVLHCQPA----------------HRELEITSEVMDG-EHSLIFDQAENRMHAQNAVLLTLL 299
10  ---FFTHDPKE--AALGAHALYTDVWTSMGQEAER-AKRLRDFQGFQVNGELLK--LLRPEGVFLHCLPA----------------HYGEETTEEAVHG-PRSRVFDQAENRLHTAKAVLLTLL 293
11  ANVTVTNDPYE--AVDGATAVYGDVFVSMGEEEQR-EEKLAEFDGFQIDQDLMD--AARDDAIFMHCLPA----------------HRGEEVTAEVADG-PQSVIFDQAENRMHVQKAIVHTLV 292
12  SSVTLTDDPIE--AVKDADVIYSDVFTSMGQEAEE-QERLAVFAPYQVNAALVS--HAKPDYTFLHCLPA----------------HREEEVTAEIIDG-PNSAVFQQAENRLHVQKALLKAIL 301
13  SFELL-HDPVK--AVKDADVIYTDVWASMGQEAEA-EERRKIFRPFQVNKDLVK--HAKPDYMFMHCLPA----------------HRGEEVTDDVID-SPNSVVWDQAENRLHAQKAVL-ALV 303
14  GSLTLTNDPIE--AAEDADVLYTDVWISMGDDKDK-EEVLKIFPPFQINSKLLE--YAKDDVIVMHCLPA----------------NRGYEITDDVIDG-EHSVVYDEAENRLHVQKGVFKFIF 301
15  STFEIVNDPKV--AVKNADIVVTDTWISMGQEAEK-EQRLKQFTGFQVTGEIMK--LAKPSCKFMHCLPR----------------HP-EEVSDEVFYG-ENSLVFQEAENRKWTTVAVLEALL 309
16  ATFELTHDSLK--ASTNANILVTDTFVSMGEEFAK-QAKLKQFKGFQINQELVS--VADPNYKFMHCLPR----------------HQ-EEVSDDVFYG-EHSIVFEEAENRLYAAMSAIDIFV 319
17  TKLLLTNDPLE--AAHGGNVLITDTWISMGQEEEK-KKRLQAFQGYQVTMKTAK--LAAPNWIFLHCLPR----------------KP-EEVDDEVFY-SPRSLVFPEAENRKWTIMAVMVSLL 305

 C  . .. . :. .  :..:::....:::::+:.: .   :... .....::  .. .  :::...::*:+*:                    .:::::.... :. :.+ ::.++:+:.:.::*

18  VRVFT-NADE-GLKDVDVVIMLRLQRERMQGGLLP----SEGEFFKLYGLTEKRLKLAKPDAIVMHPGPI----------------NRGVEIESAVADG-AQSVILNQVTYGIAIRMAVLSMAM 304
19  GKLHCHWQLQPALEGADIVMTLRLQKERMTAHLLP----SLREYHHYFGITHDRLKVCQPGVKVLHPGPV----------------NRGVEISSELMDDPDISLIQDQVTSGVAIRMALLYLLG 320
20  TFGTYVSMDE-AVESSDVVMLLRIQNERHQSAVSQ------EGYLNKYGLTVERAERMKRHAIIMHPAPV----------------NRGVEIDDSLVES-EKSRIFKQMKNGVFIRMAVIQCAL 289
21  AHLTPRLEE--ALEEADAVMVLRLQKERMEAGLVH-----LEDYIARYQVTEKRLALAKPQAPLLHPGPM----------------NRDVELEGTLADS-ARSLVNRQVQNGVAVRMAVLYHLL 295
22  WPATVSHDFDAELPAADAVLMLRVQAERMNGGFFP----SVREYSVRYGLTERRQAMLPGHAVVLHPGPM----------------VRGMEITSSVADS-SQSAVLQQVSNGVQVRMAVLFHVL 306

23  RIFSSIREYLDQKDVAKIWYFTRLQLERMGEDIL---EKVH-VLREAVTFKKEYLDALPEGVKFYHPLPR----------------HKVYPTIPNFLDTL--PLNGWETQARNGYWVRIVLLSMFG 313
24  WSLHGSIEE--VMADVDILYMTRVQKERLDPS-----EYAN-VKAQFVLRPDLN--GARENMKVLHPLPR----------------IDEITTDVDKT--PHAWYFQQAGNGIFAAQALLALVL 299
25  WSLHSSIEE--VMAEVDILYMTRVQKERLDPS-----EYAN-VKAQFVLRASDLHN-AKANMKVLHPLPR----------------VDEIATDVDKT--PHAWYFQQAGNGIFARQALLALVL 300
26  YACYNSIEE--VVPEIDVLYMTRVQKERFDET-----EYQH-MKAGFILSASSLKH-AKDNLKVLHPLPR----------------VDEIAVDVDKT--PYAYYFQQAENGVYAREALLALVL 299
27  VVETTTLED--VIGKLDVLYVTRIQKERFPDE----QEYLK-VKGSYQVNLKVL-EKAKDELRIMHPLPR----------------VDEIHPEVDNT--KHAIYFRQVFNGVPVRMALLALVL 302
28  EVENPFE----VINEVDVLYVTRIQKERFVDE----MEYEK-IKGSYIVSL-DLANKMKKDSIILHPLPR----------------VNEIDRKVDKT--TKAKYFEQASYGVPVRMSILTKIY 297
29  FYEKESLDD--LDDDIDVLYVTRIQKERFPDP----NEYEK-VKGSYKI-KREYVE-GKK-FIIMHPLPR----------------VDEIDYDVDDL--PQAKYFKQSFYGIPVRMAILKKLI 299
30  GVE-SIELTPHIISKTDVLYCTRVQEERFNSP----EEYAR-LKDTYIV-DNKILAHAKENMAIMHPLPR----------------NEIKEEVDYD--HRAAYFRQMKYGLFVRMALLAMVM 300
31  FIE-HRELTKEVVAQSDVLYCTRVQKERFASV----DEYEK-LKDSFIV-DNSVLASAKSHCIVMHPLPR----------------NREISEEVDFDQ-RRAAYFRQMRYGLYIRMALLACVM 303
32  QEEFESIEE--ALPDTDVLYMTRIQKERFGST----QEYEA-CFGQFIL-TPHIMTRAKKKMVVMHPMPR----------------VNEISVEVDSD--PRAAYFRQAENGMYIRMALLATVL 301
```

Fig. 1. Continued.

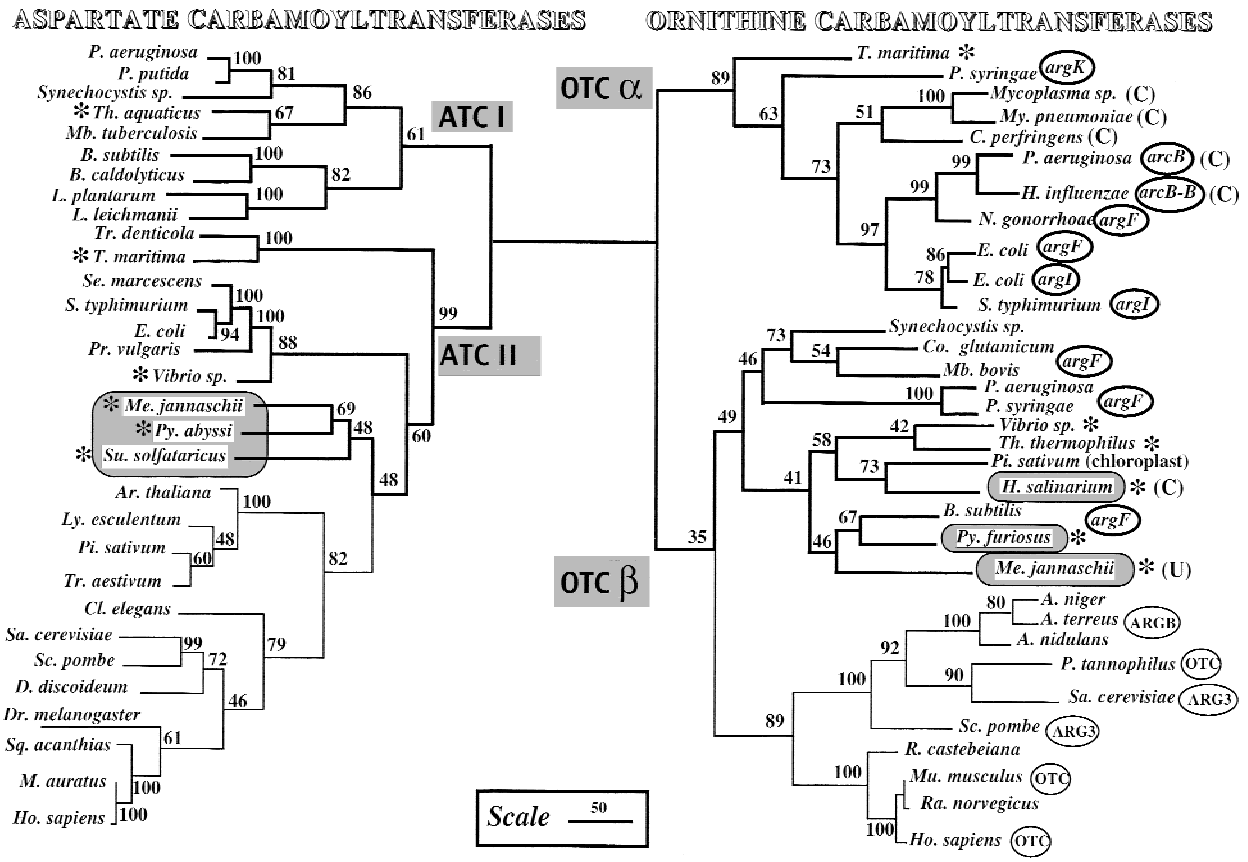

Fig. 2. A composite phylogenetic tree for ornithine and aspartate carbamoyltransferases. The multiple alignment was used to reconstruct a tree using the maximum-parsimony method. Each enzyme tree was rooted by its paralogous one. The branch lengths (computed using the PAUP program) are drawn to scale: *thick lines*, prokaryotes; *thin lines*, eukaryotes. Extremophiles are indicated by an *asterisk* next to the species name (which is *framed* in the case of Archaea). In the OTCase tree, the enzymes which are *not* anabolic are indicated by either a C (catabolic) or a U (unknown). Each node was tested using the bootstrap approach and its strength, given as its percentage presence in the output of 100 random trees, is shown as *boldface numbers* next to each internal node. The bootstrap values were determined using the maximum-parsimony (PROTPARS) method. Slightly different values were obtained when using the distance (e.g., NEIGHBOR) method (may be obtained by e-mail: labedan@igmors.u-psud.fr).

phyletic since they are found in both subfamilies. In ATC II subfamily, bacteria still appear polyphyletic, with an unexpected, paraphyletic grouping of the evolutionarily distantly related bacteria *Treponema denticola* and *T. maritima*. Interestingly, this grouping is also reflected at the level of the tertiary structure: in both organisms the gene encoding ATCase has been found to be a fusion product between a proximal part homologous to the gene for the basic catalytic monomer (*pyrB* in *E. coli*) and a distal part bearing a strong similarity to the gene encoding the regulatory subunit (*pyrI*), also found in enterobacteria and the archaea *Sulfolobus solfataricus*, *Py. abyssi*, and *M. jannaschii* (Van de Casteele et al. 1994, 1997). These archaea, *Py. abyssi*, *M. jannaschii*, and *S. solfataricus*, form a monophyletic group and share a common node with all the eukaryotes. This cluster archaea/eukaryotes is generally found branching with the enterobacteria. However, this topology (which is reminiscent of the 16S RNA tree topology) is not strongly supported by the bootstrap value attached to the node common to archaea and eukaryotes.

(b) The OTCase tree appears to be even more problematic. Indeed, the subfamily OTC β contains the three available archaeal sequences, but they are both polyphyletic and interspersed within various bacteria which are themselves polyphyletic. The hyperthermophilic *Py. furiosus* shares a common node with *B. subtilis*, and this pair is paraphyletic to *M. jannaschii*. The extreme halophile *Halobacterium salinarium* is unexpectedly close to the sequence from the *Pisum sativum* chloroplast and this pair shares a common ancestor with two extremophilic bacteria, the psychrophile *Vibrio* sp. and the thermophile *Th. thermophilus*. However, it must be emphasized that several nodes of this particular cluster have low bootstrap values and that the OCT β subfamily is the unique portion of the tree, whose topology changes when only the substrate-binding domain of both carbamoyltransferases is used to reconstruct a tree.

(c) A striking feature appears when comparing the relative positions of prokaryotes in the OTCase and ATCase trees. For example, a similar sampling of bacteria is found in both the ATC I and the OTC β subfamilies,

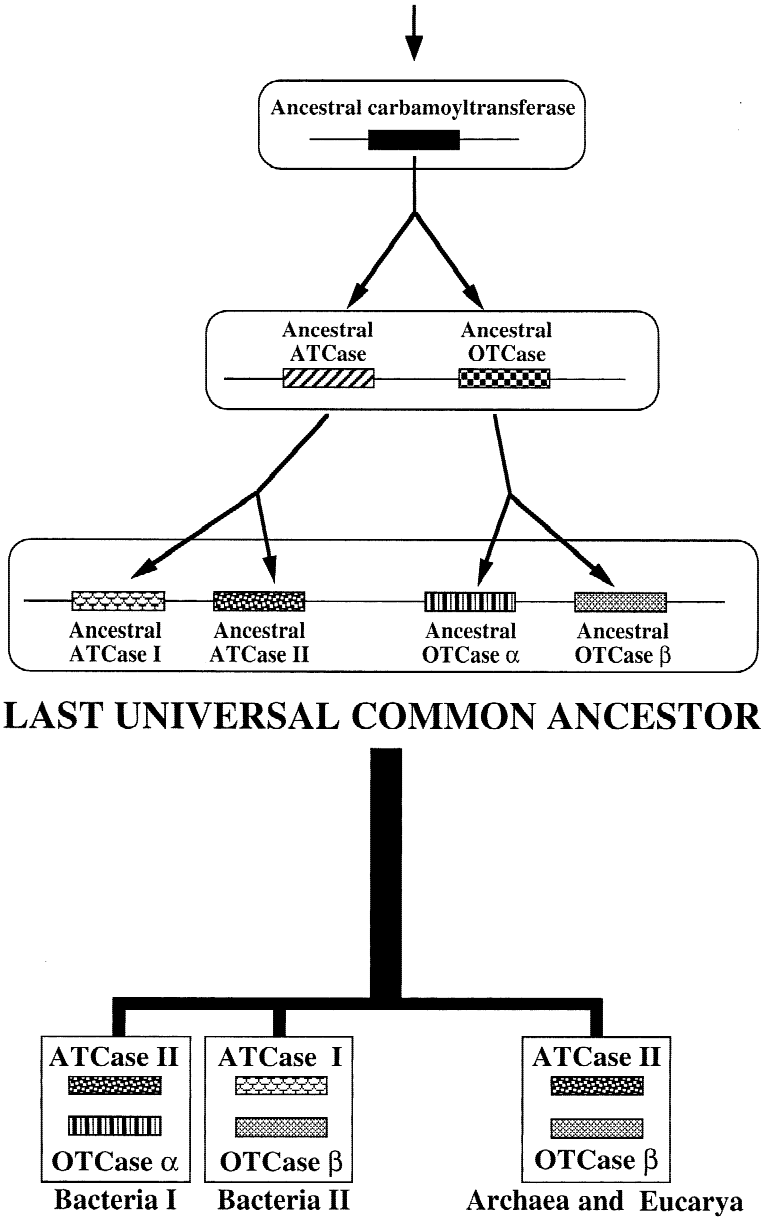

Fig. 3. A possible scenario for the evolutionary history of carbamoyltransferases. The putative successive events of gene duplication leading to the presence of two ancestral ATCases and two ancestral OTCases in the last universal common ancestor are schematized **(top)** as the simplest explanation for the sampling of both carbamoyltransferases in present-day species **(bottom).**

such as *B. subtilis*, *Pseudomonas*, *Mycobacterium*, and *Thermus* species. Likewise, comparison of the ATC II and OTC α subfamilies shows common species such as *T. maritima* and enterobacteria. It therefore appears puzzling that archaea have opposite positions in the two trees. In the case of the ATCase tree, they group with enterobacteria and appear very distant from *B. subtilis* and other bacteria belonging to the ATC I subfamily. On the contrary, in the OTCase tree, they cluster with *B. subtilis* and appear very distant from *E. coli* and other bacteria belonging to the OTC α subfamily. This strongly suggests that successive duplications have occurred and that the ancestors of these subfamilies kept only one copy of the duplicated genes (see Fig. 3 and Discussion).

(d) Another level of complexity shows up when we take into account the metabolic properties of prokaryotic OTCases. The catabolic OTCase from *P. aeruginosa* (the *arcB* gene) shares a most recent common ancestor with several anabolic ones inside the subfamily OTC α, but appears very distant from the anabolic one (the *argF* gene) of the same species, which belongs to subfamily OTC β. Note that the position of this catabolic OTCase is very coherent with previous immunological and structural studies (Falmagne et al. 1985; Villeret et al. 1995). This could be due to a recent event which changed the metabolic properties independently of the history of each protein.

Discussion

The description of many new sequences of OTCases and ATCases, belonging to various organisms from bacteria

to humans, strongly supports previous observations which suggested that these two carbamoyltransferases are ancestrally related entities (see the Introduction). This emerges very clearly from the multiple alignment of the presently available 33 OTCases and 31 ATCases. All these sequences show significant sequence similarities. Some of the blocks of similarities are specific to one family but others appear to be common to both families, helping to anchor the multiple alignment. These similarity data, now extended to the three domains of life (Bacteria, Archaea, and Eucarya), strongly suggest that ATCases and OTCases are encoded by paralogous genes descending from one ancestral gene which had already duplicated in one of the common ancestors to all present-day organisms. This gene may have coded for an ancestral carbamoyltransferase displaying substrate ambiguity (Jensen 1976) and each of the copies would have diverged subsequently to become the ancestors of two more specific carbamoyltransferases which have further evolved by speciation to give the present-day proteins (Fig. 3).

This ancestral paralogy was confirmed when the multiple alignment of the 64 sequences of both OTCases and ATCases was used to reconstruct a phylogenetic tree (see Fig. 2). Indeed, the two enzymes clearly form two interconnected trees corresponding to each protein. Although several of the nodes have low bootstrap values, this tree topology seems rather robust since it was consistently obtained using entirely different methods (maximum parsimony or distance or approximation to maximum likelihood) and since it did not change substantially when adding or subtracting various sequences. Moreover, the same tree was obtained when using partial sequences (conserved residues of the carbamoylphosphate binding domain) instead of complete ones (thus including the more divergent residues of the substrate binding domain). This strongly supports the notion that the present-day genes derive from ancestral ones which already had the same size rather than from the fusion of independent modules leading to specialized carbamoyltransferases.

Therefore, it would seem reasonable to use each tree, rooted by the outgroup made of the paralogous carbamoyltransferase and containing representatives of the three domains of life, as species trees to challenge the topology of the universal tree of life. However, a close inspection showed that these gene trees exhibit features (polyphyly of Bacteria and of Archaea) which render them unsuitable as phylogenetic probes to infer organismal trees.

The polyphyly of Bacteria (in both trees) and of Archaea (in the OTCase tree) as well as the differences in the respective positions of Archaea and Bacteria in ATCase and OTCase trees, may be interpreted in two ways.

(i) Secondary duplications of the ancestral paralogous genes occurred before the divergence of Bacteria and Archaea, leading to an ancestor cell containing two genes coding for aspartate carbamoyltransferase and two genes coding for ornithine carbamoyltransferase (Fig. 3). Let us call these genes *atc1, atc2, otc*α, and *otc*β, respectively. The history of present-day species for which we know both carbamoyltransferase sequences would be as follows. One of the last ancestors to Archaea and Eukaryotes would have kept only the copies *atc2* and *otc*β. One single event is sufficient to explain this choice, and this would correspond with current views according to which Archaea and Eukaryotes would have had a time interval of their evolutionary history in common. In the case of Bacteria, the situation appears more complex since we have two categories, one having the pair *atc2* and *otc*α (example: *E. coli*) and the other having the pair *atc1* and *otc*β (example: *B. subtilis*). Such a phenomenon of secondary paralogy followed by the specific loss of one copy at different moments of their evolutionary history could explain the unexpected associations of unrelated prokaryotic species (Fig. 3).

(ii) The differences in the respective positions of archaea and bacteria in the ATCase and OTCase trees could alternatively be explained by some mechanism of horizontal transfer. Such an event appears to have occurred recently in the case of *E. coli* K12, which harbors two genes coding for ornithine carbamoyltransferase [see Van Vliet et al. (1988) for a detailed discussion]. Such a mechanism of horizontal transfer could also explain two features of the OTCase tree: the unexpected association of the halophile *H. salinarium* with the *Pi. sativum* chloroplast (presumably of bacterial descent) and the apparent mixture of catabolic (indicated by C in Fig. 2) and anabolic enzymes. From the mutagenesis experiments referred to in the Introduction, the conversion of regulated catabolic OTCases into Michaelian anabolic ones, or vice versa, appears to require very few mutations. Horizontal transfer followed by divergence could therefore have occurred recently and this functional difference may have been acquired independently in several organisms, after the divergence of the new species. Besides, for those catabolic OTCases which are not allosteric, the question of metabolic specialization arises not at the level of structure–function relationships but at the level of transcriptional regulation.

Finally, it must be considered that mutational saturation (Meyer et al. 1986) at some sites could contribute to hindering the analysis of the early steps in the evolutionary history of these enzymes. Nevertheless, our data clearly indicate that the last common ancestor to all extant life had genes coding for the key enzymes ATCase and OTCase and even strongly suggest that this universal ancestor had two differentiated copies of each gene coding for these carbamoyltransferases (Fig. 3). Even if we do not know how refined these ancestral enzymes were, our data support the hypothesis (see, e.g., Forterre et al. 1993) that this universal ancestor must have been a rather sophisticated organism already, possibly endowed with a

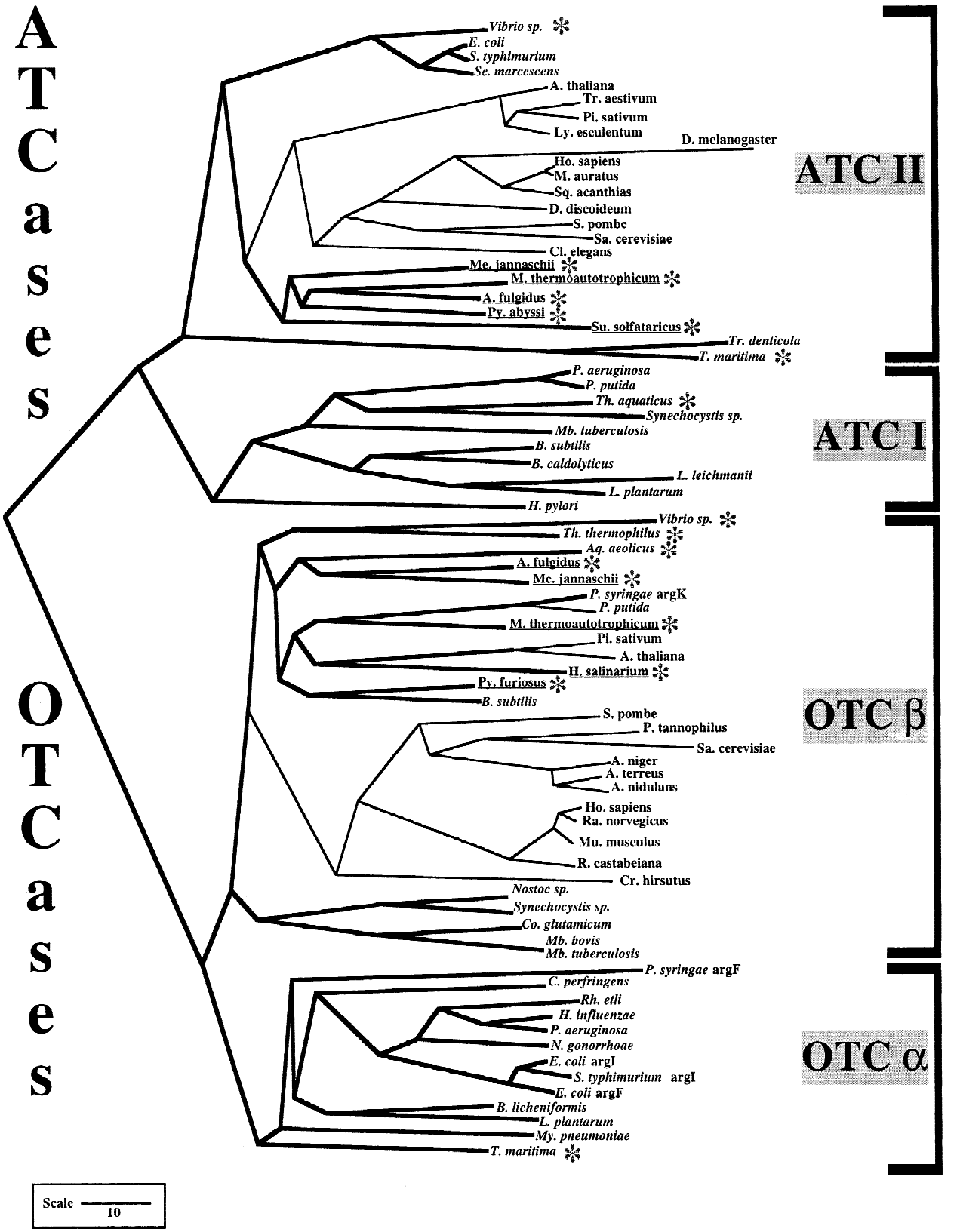

Fig. 4. An updated (15 March 1999) tree. The newly appeared sequences of ATCases and OTCases have been added to our multiple alignment. The obtained new alignment has been used to build a tree using the DARWIN method (see Materials and Methods). In this rooted distance tree each subfamily (ATC I, ATC II, OTC α and OTC β) has been indicated using thick frames. The branch lengths are drawn to scale: *thick lines*, prokaryotes; *thin lines*, eukaryotes. Archaeal sequences are underlined and bacterial sequences are in italics. As in Fig. 2, extremophiles are indicated by an *asterisk* next to the species name.

fair amount of genetic redundancy. At any rate, the data suggest that the primary duplication having produced the two differentiated genes from an ancestral, substrate-ambiguous carbamoyltransferase was already an ancient event in the evolution of the universal ancestor (Fig. 3).

No data are presently available on putrescine carbamoyltransferase. Yet considering, on the one hand, the chemical relatedness between ornithine and putrescine and, on the other hand, the basic and therefore probably primeval functions of polyamines, it would be interesting to see which genealogical pattern would emerge from the comparison between ornithine and putrescine carbamoyltransferases and to compare the results of such an analysis with the present one.

Since this paper, which was accepted for publication over 1 year ago, had its publication unexpectedly delayed for "unwanted" reasons, we have found it necessary to update our results at the present date (15 March 1999). Accordingly, we have added to our multiple alignment the new sequences of carbamoyltransferases which have appeared recently in the different databases. The new tree made from this updated alignment of 77 sequences (33 ATCases and 44 OTCases) was found to confirm all our previous conclusions. It appears that the tree topology shown in Fig. 4 is in agreement with that shown in Fig. 2 and thus supports our working hypothesis (see above) summarized in Fig. 3. More information can be obtained by e-mail: labedan@igmors.u-psud.fr.

Acknowledgments. This work was supported by the Vlaams Actieprogramma Biotechnologie, a University Concerted Action, and the Belgian Foundation for Joint and Fundamental Research. We thank the reviewers for helpful and constructive comments on an early version of the manuscript.

References

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

Benachenhou-Lahfa N, Forterre P, Labedan B (1993) Evolution of glutamate dehydrogenase genes: evidence for two paralogous protein families and unusual branching patterns of the archaebacteria in the university tree of life. J Mol Evol 36:335–346

Brown JR, Doolittle WF (1995) Root of the universal tree of life based on ancient aminoacyl-tRNA synthetase gene duplication. Proc Natl Acad Sci USA 92:2441–2445

Brown JR, Masuchi Y, Robb FT, Doolittle WF (1994) Evolutionary relationships of bacterial and archaeal glutamine synthetase genes. J Mol Evol 38:566–576

Bult CJ, White O, Olsen GJ, et al. (1996) Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii.* Science 273:1058–1073

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

Davis R (1986) Compartmental and regulatory mechanisms in the arginine pathway of *Neurospora crassa* and *Saccharomyces cerevisiae.* Microbiol Rev 50:280–313

Falmagne P, Portetelle D, Stalon V (1985) Immunological and structural relatedness of catabolic ornithine carbamoyltransferase and the anabolic enzymes of enterobacteria. J Bacteriol 161:714–719

Felsenstein J (1989) PHYLIP—A phylogeny inference package (version 3.2). Cladistics 5:164–166

Fitch WD (1970) Distinguishing homologous from analogous proteins. Syst Zool 19:99–113

Fleischman RD, Adams MD, White O, et al. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science 269:496–512

Forterre P, Benachenhou N, Confalonieri F, Duguet M, Elie C, Labedan B (1993) The nature of the last universal ancestor and the root of the tree of life, still open questions. BioSystems 28:15–32

Gogarten JP, Kibak H, Dittrich P, et al. (1989) Evolution of the vacuolar H^+-ATPase: Implications for the origin of eukaryotes. Proc Natl Acad Sci USA 86:6661–6665

Gonnet GH, Cohen MA, Benner SA (1992) Exhaustive matching of the entire protein sequence database. Science 256:1443–1445

Honzatko RB, Crawford JL, Monaco HL, et al. (1982) Crystal and molecular structures of native and CTP-liganded aspartate transcarbamylase from *Escherichia coli.* J Mol Biol 160:219–263

Houghton JE, Bencini DE, O'Donovan GA, Wild JR (1984) Protein differentiation: A comparison of aspartate transcarbamoylase and ornithine transcarbamoylase from *Escherichia coli* K12. Proc Natl Acad Sci USA 81:4864–4868

Iwabe N, Kuma K, Hasegawa M, Osawa S, Miyata T (1989) Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. Proc Natl Acad Sci USA 86:9355–9359

Jensen R (1976) Enzyme recruitment in evolution of new function. Annu Rev Microbiol 30:409–425

Ke H-M, Honzatko RB, Lipscomb WN (1984) Structure of unligated aspartate transcarbamylase of *Escherichia coli* at 2.6 Å resolution. Proc Natl Acad Sci USA 81:4037–4040

Knodler LA, Schofield PJ, Edwards MR (1995) L-Arginine transport and metabolism in *Giardia intestinalis* support its position as a transition between the prokaryotic and eukaryotic kingdoms. Microbiology 141:2063–2070

Kuo LC, Zambidis I, Caron C (1989) Triggering of allostery in an enzyme by a point mutation: Ornithine transcarbamoylase. Science 245:522–524

Legrain C, Halleux P, Stalon V, Glansdorff N (1972) The dual genetic control of ornithine carbamoyltransferase in *Escherichia coli:* A case of bacterial hybrid enzymes. Eur J Biochem 27:93–102

Legrain C, Demarez M, Glansdorff N, Piérard A (1995) Ammonia-dependent synthesis and metabolic channeling of carbamoyl phosphate in the hyperthermophilic archaeon *Pyrococcus furiosus.* Microbiology 141:1093–1099

Linstead D, Cranshaw MA (1983) The pathway of arginine catabolism in the parasitic flagellate *Trichomonas vaginalis.* Mol Biochem Parasitol 8:241–252

Meile L, Leisinger T (1984) Enzymees of arginine biosynthesis in methanogenic bacteria. Experientia 40:899–900

Meyer TE, Cusanovich MA, Kamen MD (1986) Evidence against use of bacterial amino acid sequence data for construction of all-inclusive phylogenetic trees. Proc Natl Acad Sci USA 83:217–220

Purcarea C, Erauso G, Prieur D, Hervé G (1994) The catalytic and regulatory properties of aspartate transcarbamoylase from *Pyrococcus abyssi*, a new deep-sea hyperthermophilic archaebacterium. Microbiology 140:1967–1975

Purcarea C, Hervé G, Ladjimi MM, Cunin R (1997) Aspartate transcarbamylase from the deep-sea hyperthermophilic archaeon *Pyrococcus abyssi:* Genetic organization, structure, and expression in *Escherichia coli.* J Bacteriol 179:4134–4157

Roovers M, Hethke C, Legrain C, Thomm M, Glansdorff N (1997) Isolation of the gene encoding *Pyrococcus furiosus* ornithine carbamoyltransferase and study of its expression profile *in vivo* and *in vitro.* Eur J Biochem 247:1038–1045

Ruepp A, Holger NM, Lottspeich F, Soppa J (1995) Catabolic ornithine

transcarbamylase of *Halobacterium halobium* (*salinarium*): Purification, characterization, sequence determination, and evolution. J Bacteriol 177:1129–1136

Sanchez R, Baetens M, Van De Casteele M, Roovers M, Legrain C, Glansdorff N (1997) Ornithine carbamoyltransferase from the extreme thermophile *Thermus thermophilus.* Analysis of the gene and characterisation of the protein. Eur J Biochem 248:466–474

Schofield PJ, Edwards MR, Matthews J, Wilson JR (1992) The pathway of arginine catabolism in *Giardia intestinalis.* Mol Biochem Parasitol 51:29–36

Stevens RC, Reinisch KM, Lipscomb WN (1991) Molecular structure of *Bacillus subtilis* aspartate transcarbamoylase at 3.0 Å resolution. Proc Natl Acad Sci USA 88:6087–6091

Swofford DL (1993) PAUP: Phylogenetic analysis using parsimony, Version 3.1.1. Smithsonian Institute, Washington, DC

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Van de Casteele M, Demarez M, Legrain C, Glansdorff N, Piérard A (1990) Pathways of arginine biosynthesis in extreme thermophilic archaeo- and eubacteria. J Gen Microbiol 136:1177–1183

Van de Casteele M, Chen PG, Roovers M, Legrain C, Glansdorff N (1997) Structure and expression of a pyrimidine gene cluster from the extreme thermophile *Thermus* strain ZO5. J Bacteriol 179: 3470–3481

Van Vliet F, Cunin R, Jacobs A, et al. (1984) Evolutionary divergence of genes for ornithine and aspartate carbamoyltransferases—Complete sequences and mode of regulation of the *Escherichia coli argF* gene; Comparison of *argF* with *argI* and *pyrB.* Nucleic Acids Res 12:6277–6289

Van Vliet F, Boyen A, Glansdorff N (1988) On interspecies gene transfer: The case of *argF* gene of *Escherichia coli.* Ann Inst Pasteur/Microbiol 139:493–496

Villeret V, Tricot C, Stalon V, Dideberg O (1995) Crystal structure of *Pseudomonas aeruginosa* catabolic ornithine transcarbamoylase at 3.0-Å resolution: A different oligomeric organization in the transcarbamoylase family. Proc Natl Acad Sci USA 92:10762–10766

Wild J, Wales M (1990) Molecular evolution and genetic engineering of protein domains involving aspartate carbamoylase. Annu Rev Microbiol 44:193–218

Woese CR, Kandler O, Wheelis ML (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci USA 87:4576–4579