Available online at www.sciencedirect.com
ScienceDirect
Biochimie 89 (2007) 555–573
BIOCHIMIE
www.elsevier.com/locate/biochi

Ordering events of biochemical evolution

C. Cunchillos [a,b], G. Lecointre [b,*]

[a] Institut Charles Darwin International, Place du Four, 81140 Puycelsi, France
[b] CNRS UMR 7138 ''Systématique, Adaptation, Evolution'', Département ''Systématique et Evolution'', Muséum National d'Histoire Naturelle, CP26, 57 rue Cuvier, 75231 Paris cedex 05, France

Received 5 September 2006; accepted 11 December 2006
Available online 4 January 2007

Abstract

Metabolic pathways exhibit structures resulting from an evolutionary process. Pathways have been inherited through time with modification, from the earliest periods of life. It is possible to compare the structure of pathways as done in comparative anatomy, i.e. for inferring ancestral pathways or parts of it (ancestral enzymatic functions), using standard phylogenetic reconstruction. Thus a phylogenetic tree of pathways provides a relative ordering of the rise of enzymatic functions. It even becomes possible to order the birth of each complete pathway in time. This particular ''DNA-free'' conceptual approach to evolutionary biochemistry is reviewed, gathering all the justifications given for it. Then, the method of assigning a given pathway to a time span of biochemical development is revisited. The previous method used an implicit ''clock'' of metabolic development that is difficult to justify. We develop a new clock-free approach, using functional biochemical arguments. Results of the two methods are not significantly different; our method is just more precise. This suggests that the clock assumed in the first method does not provoke any important artefact in describing the development of biochemical evolution. It is just unnecessary to postulate it. As a result, most of the amino acid metabolic pathways develop forwards, confirming former models of amino acid catabolism evolution, but not those for amino acid anabolism. The order of appearance of sectors of universal cellular metabolism is: (1) amino acid catabolism, (2) amino acid anabolism and closure of the urea cycle, (3) glycolysis and glycogenesis, (4) closure of the pentose-phosphate cycle, (5) closure of the Krebs cycle and fatty acids metabolism, (6) closure of the Calvin cycle.

Keywords: Metabolism; Metabolic pathways; Biochemical evolution

1. Introduction

1.1. Metabolic pathways and evolution

Cellular metabolism is a complex process made up of about a thousand chemical reactions catalyzed by globular proteins, enzymes. As any biological phenomenon, metabolism is the product of an evolutionary process. In 1945, Horowitz [1] postulated that if life began in a rich soup of organic molecules, the earliest biosynthetic pathways evolved in a backward direction (Fig. 1). If primitive cells were using a particular external nutrient, soon this organic molecule would be lacking in the environment. A selective advantage could be obtained by organisms able to synthesize this nutrient from an available precursor. Each biosynthetic step was therefore selected according to successive depletions of precursors in the environment. The first enzyme to appear in the biosynthetic pathway was the most downstream in the pathway. Confluence of pathways was selected because it saved energy. This energy is used for other needs that would be more difficult to satisfy for competitor cells without confluence. This pathway optimization is considered as a general rule of comparative biochemistry [2]. For these early anabolisms, common enzymes or common reactions shared by two (or more) synthetic pathways are downstream, and therefore provide evidence for a common ancestry of these pathways. Pathways sharing these enzymes are closer to each other—evolutionarily speaking—than to other pathways not using these enzymes.

* Corresponding author. Tel.: +33 1 40 79 37 51; fax: +33 1 40 79 38 44.
E-mail address: lecointr@mnhn.fr (G. Lecointre).

0300-9084/$ - see front matter © 2007 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.biochi.2006.12.007

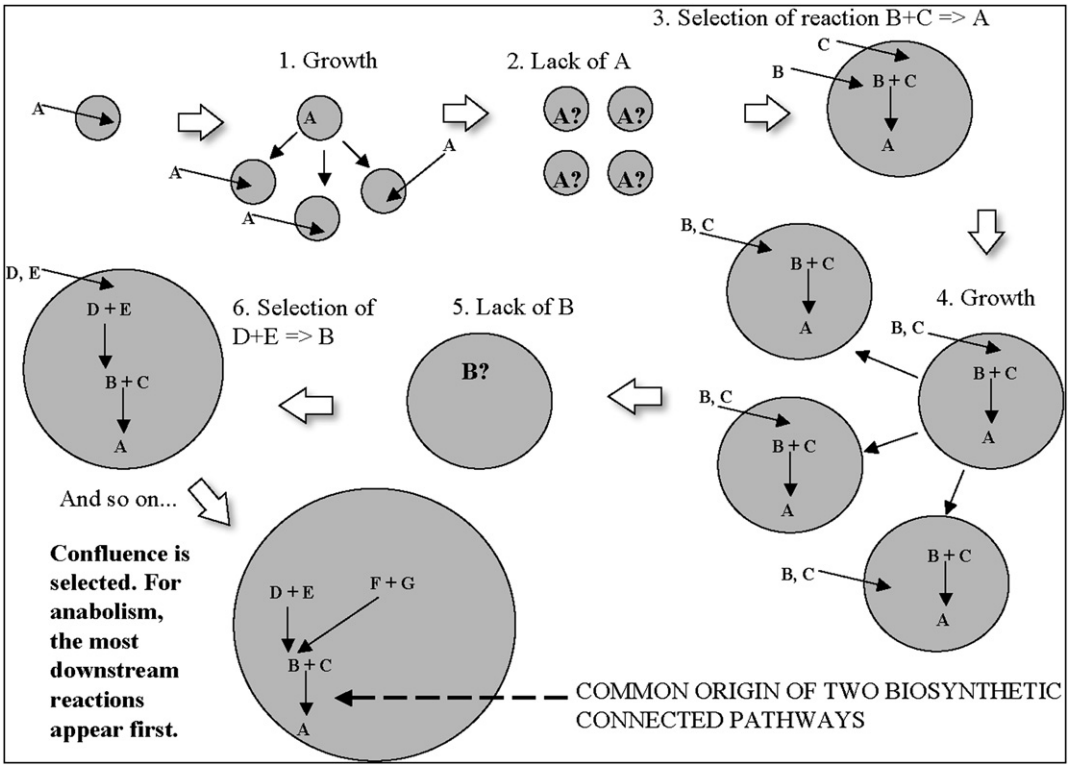

Fig. 1. Model of backward development of an anabolic pathway conceived by Horowitz [1]. The circle means the ''proto-cell''. A to F, chemical compounds. In a first step, proto-cells grow by exploiting the compound A. Soon the environment becomes depleted in A. This leads to the natural selection of the proto-cell able to use B and C to form A. Then those proto-cells grow until depletion in either B or C. Then a proto-cell able to obtain either B or C from other compounds (D and E) is selected. Selection of multiple abilities to obtain B (here from F and G) is favoured because it allows to face a depletion in either B, or D, or E. Confluence of pathways is therefore selected. As a general result the most downstream reaction is the first to appear in evolution (''Backward development''). The metabolic segment from B + C to A is a segment gained by common ancestry of the two pathways D + E to A and F + G to A. When two pathways share some enzymes or enzymatic functions, then it is conceivable to consider them as signs of common ancestry, justifying the coding of primary homologies into a matrix [20] and use of standard parsimony to reconstruct phylogenetic trees.

In 1990, Cordón [3] proposed a symmetrical scenario of catabolic pathways. Early forms of life extracted energy from the degradation of substrates available in the environment into a product. A selective advantage was obtained for those able to produce a supplementary reaction of deeper degradation of this product, therefore obtaining more energy from the original substrate (Fig. 2). Confluence is selected by obtaining the transformation of another substrate into an intermediate product already present in the proto-cell. The first reactions to appear in the evolution of catabolism are those upstream. The common distal elongation of two branched catabolic pathways is therefore a phenomenon which provides evidence for the common ancestry of these pathways. Two catabolic pathways sharing one or several downstream portions of catabolism are supposed to be more closely related to each other than to other pathways. But there is a risk with this theory, which consists in the late branching of an ''opportunistic'' catabolic pathway when the early catabolism onto which it branches is already complete. Common downstream portions in this case are not evidence for common ancestry, but rather convergences obtained by recruitment, a phenomenon recognized as having played a role in biochemical evolution [4–6]. Homoplasy, i.e. similarity without common ancestry (see below), appears when there are character conflicts due to similarities obtained by evolutionary convergences or reversions. The risk of homoplasy in data of comparative biochemistry depends on the relative time between the distal pathway elongation event and the branching event of a new pathway. An early branching event followed by downstream elongation will provide a good phylogenetic indicator. A late branching event (late in evolutionary time and/or late in the pathway) will probably bring homoplasy. As in any study of comparative biology, there are risks of homoplasy to carry on.

According to Cordón [3], these rules are not rigid and can change from one type of metabolite to another: the order of development of a given pathway depends on the position and availability of the initial substrate and/or final product. The above scenarios in the genesis of reactions in anabolic and catabolic pathways are both valid possibilities because the final product and the initial substrate respectively are imposed from the outside to the cell, at least initially. Alternative scenarios can be obtained for transformations starting from products already integrated into the cellular metabolism. New biosynthetic pathways can develop in a forward direction by the addition of new enzymes and reactions to pre-existing pathways. For example, the urea cycle uses the biosynthesis of arginine [7]. Thus, Cordón [3] proposed a forward development for amino acid catabolism, fatty acid anabolism and

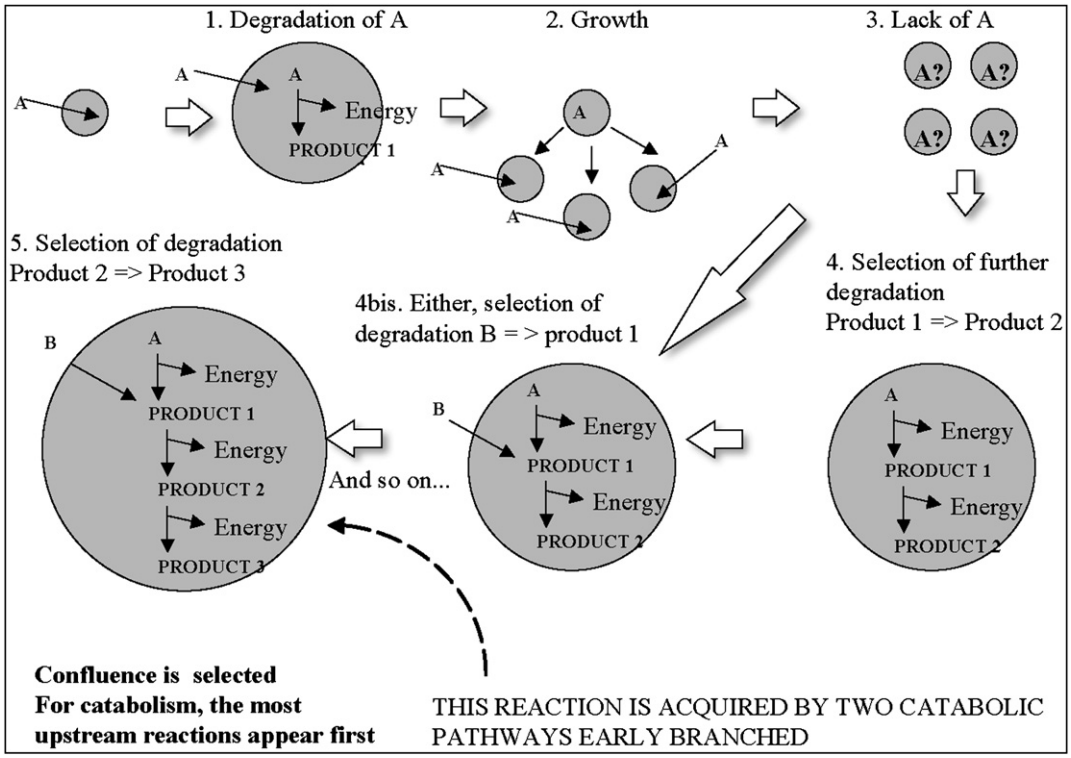

Fig. 2. Model of forward development of a catabolic pathway as conceived by Cordón [3]. When the compound A is lacking in the environment, the proto-cells selected are those able to further degrade the product 1 and gain more energy from it. As a result, the most upstream reactions are the first in evolution (‘‘forward’’ development). Confluence (B to product 1) is immediately selected because it allows facing the lack of A by having more product 1 without A, thus more energy. More downstream developments of the pathway (to product 3 and so on until product n) are signs of common ancestry of the pathways A to product n and B to product *n*.

glycogenesis, and a backward development for amino acid anabolism, fatty acid catabolism and glycolysis (confirmed in 1993 by Fothergill-Gilmore and Michels [8]).

As the history and interrelationships of living organisms is based on comparative anatomy, the history of metabolism must be reconstructed by the comparative analysis of its structural complexity [2,3,9,10]. Biochemists have recognized this necessity for a long time, but have never used suitable comparative methods that formally control the consistency of evolutionary biochemical theories [3,4,9,11–13]. It is clear from above that the evolutionary development of a pathway is linked to the darwinian concept of ‘‘descent with modification’’. This means that current similarities in the pathways, detected through shared enzymes and enzymatic reactions, can be interpreted as the product of a common ancestry, i.e. the result of evolutionary transformations of the pathways through time. It is therefore justified to use phylogenetic reconstruction [14,15] for ordering the events of metabolic evolution, as did Cunchillos and Lecointre [16–19].

1.2. Phylogenetic reconstruction: summarizing the standard comparative method

In systematics, the science of classification of living things, there are now standardized comparative methods to infer events of the past through a phylogeny, with a measurement of the consistency of this inference. The items or concepts among which we investigate interrelationships are called ‘‘terminals’’ or ‘‘taxons’’. Among the different taxon, we detect that some structures are the same, but with different versions. To formalize this observation, we create a column in a ‘‘matrix’’, in which a given version is associated with a dot (‘‘.’’) and the other one with 1 in the list of taxa to be compared (it can be also ‘‘coded’’ 2, 3, etc. depending on the number of versions found). The column is a character, and its version ‘‘.’’ or ‘‘1’’ is the character state. We make the primary hypothesis that the versions are homologous: similarities must come from common ancestry. That hypothesis is called ‘‘primary homology’’ [20]. Once the matrix is filled with all the character states for all taxa, a phylogenetic tree is retained according to a criterion explained below. Onto that tree, character state changes become hypothetical transformation events; and the tree yields a relative order of these events. This tree will show whether the primary hypothesis of homology is confirmed or not for each character. If a given character state appears only once in the tree, the primary hypothesis of homology is confirmed and is called a ‘‘secondary homology’’ [20], or a synapomorphy [21,22]: the state is present due to the common ancestry of those taxons that have it. If the state of a character is associated with more than one event, the primary hypothesis of homology is falsified. It is homoplasy, i.e. similarity without common ancestry. But how do we choose one tree over the others? For a given number of taxons, there are a limited, but high, number of possible trees that represent

different theories of the interrelationships. In science, theories are to be compared in terms of internal consistency. The theory that is the most consistent is the one that requires the fewest *ad hoc* hypotheses. The most parsimonious tree is the best theory (against alternative trees) because it contains the smallest number of *ad hoc* transformation hypotheses in its branches. This is the reason why one always associates the most parsimonious tree with the number of transformations (''number of steps''), and with measurements of internal consistency, like the consistency index (CI) and the retention index (RI). Such a parsimony analysis [14,15,21–23] has been shown to be useful to infer the relative timing of emergence of metabolic pathways in order to shed light on the earliest pathways and their different enzymatic functions.

Cunchillos and Lecointre [16–19] have used this method by defining each pathway as a taxon, from its initial substrate to its entry into the Krebs cycle. For example, the taxon ''dASP1'' is the catabolic pathway from aspartate to oxaloacetate. Enzymes and enzymatic functions along this pathway are the character states of this taxon. Enzymes and enzymatic functions through all the considered pathways are the characters (columns in the matrix). Such coding has nothing to do with the comparison of ''semantids'' of Zuckerkandl and Pauling [24], i.e. DNA sequences or protein sequences. Using them for ordering events of biochemical evolution would be of no help because they would lead to severe problems of linear sequence homology and would require the management of a high number of mutational changes that took place prior to the occurrence of these early metabolic pathways. Comparing differences in the metabolic pathways of a particular extant organism is of no help either, because all inferred the biochemical events will have occurred before any speciation of the extant organisms, and the biochemical differences among species will actually be linked to differences in metabolic regulation rather than to differences in pathway structure [2]. Finally, the phylogenetic analysis of biochemical pathways infers nothing about information storage, replication and the RNA world.

1.3. The universal core of metabolic pathways

Whatever the metabolic specializations found in diverse living organisms (heterotrophy, photosynthetic autotrophy, chemosynthetic autotrophy, various forms of respiration and fermentation), there is a universal core of about 50 metabolic pathways involving the anabolism and catabolism of amino acids, fatty acids, saccharides (glycolysis, glycogenesis, and the pentose phosphate pathway) and the Krebs cycle. Because the Krebs cycle is the point of confluence of all other metabolic pathways, it has sometimes been viewed as one of the earliest. But this view is challenged by several lines of evidence. Molecules entering the Krebs cycle (oxo acids and acyl CoAs) are intermediate metabolites that were unlikely to be readily available in primitive abiotic environments, and are most probably the products of a peripheral cellular metabolism. Schoffeniels [2,25], Gest [26,27], and Meléndez-Hevia et al. [12] considered the origin of the Krebs cycle to be secondary and composite. The latter authors [12] considered portions of the Krebs cycle as the evolutionary product of amino acid biosynthesis, instead of amino acid catabolism, because they excluded amino acid catabolism as a candidate origin right from the beginning. The main argument was that no pathway for amino acid and nitrogen-base degradation may have previously existed because ''the selective value of a mechanism to eliminate organic material hardly built was not obvious at all''. This argument is circular: it is based on the assumption that amino acid anabolism pre-dated amino acid catabolism, because glycolysis was the first anaerobic source of energy (as in Gest [26]). However, the presence of glucose in abiotic conditions is less well documented than the presence of amino acids. Even if we do not exclude *a priori* the role of glucose as the first energy source, this is not a sufficient reason to exclude the available amino acids. These considerations led us [16–19] to incorporate two portions of the Krebs cycle as taxons in the matrix for phylogenetic analysis, along with anabolism and catabolism of the three fundamental kinds of biomolecules: amino acids [16–18], fatty acids and saccharides (i.e., glycolysis and gluconeogenesis, Calvin and pentose-phosphate cycles) [19]. The development of the urea cycle is traditionally thought to be linked to the metabolism of arginine and might have evolved at the same time as arginine metabolism. This was confirmed by phylogenetic analysis [19]. The metabolism of aromatic amino acids is classically thought to have arisen after the appearance of aliphatic amino acid metabolism, a hypothesis which was also confirmed by these authors. The use of phylogenetic methods has been shown to be able to order these biochemical pathways through time [19], but also has the power to test a number of predictions made by Horowitz [1] and Cordón [3] relative to the evolutionary timing of pathway development from a theoretical point of view. For example, as free aliphatic amino acids were considered as one of the very first sources of molecules in abiotic environments, upstream reactions of amino acid catabolic pathways must have occurred before downstream reactions, while downstream reactions of amino acid anabolic pathways must have occurred before upstream reactions [16] (Figs. 1 and 2). Cunchillos and Lecointre [19] found that this prediction was verified for some amino acid pathways and not for others.

1.4. Refinements in the method of ordering biochemical evolution

One of the reasons for using standard phylogenetic analyses for studying the evolution of biochemical pathways was to use the resulting phylogenetic trees to order the appearance over time of various enzymatic activities in the earliest forms of life [16–19]. To reach this aim, the data matrix contained two kinds of characters: type I homologies when pathways shared the same enzyme with high degree of specificity for its substrate, and type II homologies when pathways shared the same family of enzymatic functions or the same family of reactions (see materials and methods). Along the branches of the phylogenetic tree, synapomorphies corresponding to

the rise of type II homologies were used as landmarks to define time spans. Each time span, delimited by two successive landmarks, was associated with a colour, allowing the description of the successive periods of development of the different reactions of universal metabolism. In the method used, a kind of clock was implicitly assumed. Indeed, when a tree is not fully asymmetrical, it is not possible to order events occurring on branches in the distal areas independent clades, except by assuming a clock-wise development of these events. This assumption can be questioned, particularly when we have no model or idea supporting a clock for enzymatic innovations. The purpose of the present study is to end our review of results in the field of phylogenetic analyses of biochemical pathways with a new approach to time spans. Here, we develop a new timing in the rise of enzymatic activities based on both the phylogenetic tree and on arguments of biochemical function, i.e. chemical knowledge external to the analysis. Then we compare the obtained periods of metabolic evolution with the periods found from the method that implicitly assumes a clock [19]. We explore whether the periods are the same or not. If yes, clockwise evolutionary development of enzymatic activities would be indirectly justified by biochemical arguments. If the periods are radically different, no conclusions can be drawn concerning a clockwise development.

2. Materials and methods

2.1. Taxons

To date the phylogenetic analyses of metabolic pathways has focused on the metabolism of the three main kinds of universal biochemical compounds, amino acids, fatty acids and monosaccharides. Analysis of the metabolism of these compounds required the coding of the following pathway assemblages: Krebs cycle, Calvin cycle, pentose-phosphate cycle, urea cycle, fatty acid anabolism and catabolism, amino acid anabolism and catabolism, glycolysis, and gluconeogenesis. These pathways are shared by all living things at least primitively with a few exceptions (Calvin cycle) possibly due to secondary losses in some species or species groups.

Enzymes acting on complex molecules have been excluded from this universal core of enzymatic activities. Complex molecules are assemblages of those elementary molecules whose metabolic evolution is being studied. Complex molecules include, for instance, coenzymes, triglycerides, phospholipids, nucleotides, and polymers (like glycogen, starch, proteins, DNA, RNA). They all are secondary products of the core metabolism studied by Cunchillos and Lecointre [19]. Purines and pyrimidines are considered as complex molecules themselves because they are never completely synthesized *in vivo*. Indeed, their precursors are already attached to other compounds (ribose or ribose-phosphates), so that their recognition as isolated compounds has no biological basis.

As in Cunchillos and Lecointre [16–19], taxons are defined from the tip of the pathway to its point of contact with the Krebs cycle. To name pathways, prefixes "d" and "s" are used to refer to degradation and synthesis, respectively. For example, dGLN is the set of enzymatic activities involved in converting glutamine to oxoglutarate, while sGLN is the synthetic pathway from oxoglutarate to glutamine. When degradation or synthesis of a compound can occur in different ways, the name of the pathway is numbered. For instance, cysteine can be degraded via mercaptopyruvate (dCYS2) or directly through pyruvate and acetyl-CoA (dCYS1). Taxons are listed in Table 2. Fatty acid catabolic and anabolic pathways stop at acetyl-CoA. Monosaccharide anabolic pathways (gluconeogenesis, pentose-phosphate cycle) stop at oxaloacetate, and monosaccharide catabolic pathways (glycolysis, Calvin cycle) stop at acetyl-CoA. Each amino acid anabolic and catabolic pathway starts from the amino acid itself and stops at different points of the Krebs cycle or at acetyl-CoA, depending on the amino acid (oxoglutarate, succinyl-CoA, oxaloacetate, acetyl-CoA). The Krebs cycle is divided into two parts, designated using the two main entrance points: KC1 from oxaloacetate to oxoglutarate, and KC2 from oxoglutarate to oxaloacetate. These two oxo acids are, among the metabolites of the Krebs cycle, the closest to amino acids, structurally speaking. The urea cycle is not delineated with reference to the Krebs cycle, but independently and includes the reactions of its own cycle.

2.2. Characters: homology criteria

If shared enzymes or similar enzymes are evidence for the common ancestry of metabolic pathways, similarities in the structure of active sites should be sufficient to formulate putative homologies. However, only a few active sites [28] are known in detail, in comparison with the number of known enzymatic species [10]. We are therefore led to consider similarities in catalytic reactions and enzymatic mechanisms as reflections of similarities in active sites. The higher the specificity, the more accurate this reflection is. In the same way, by considering the generally accepted idea that enzymes evolved from low specificities to high specificities [4,5,29–32], the putative common ancestry of pathways can be postulated not only on the basis of shared enzymes with high specificities; but also on the basis of very similar reactions. Functional similarities must correspond to an underlying similarity in the structure of active sites, a structural similarity that arises from common ancestry. Recognising that two metabolic pathways share the same reaction with a high specificity for a substrate uses a strict criterion of primary homology [20], while recognising a common family of reactions relaxes this criterion, with an associated risk of homoplasy obtained by convergence or recruitment. There is no reason to consider that this risk is any higher in the present case than in aligned DNA sequences or in classical morpho-anatomical matrices, both frequent in systematics [33]. The criteria for formulating a primary homology are fourfold: shared specific enzymatic activity (I), shared enzymatic function without shared specificity for a substrate (IIa), shared coenzymes (IIb), shared functional family (IIc), and a new kind of homology [19] that takes into account the pattern of recurrence of a set of reactions (IId).

2.2.1. Type I homology

Several pathways share the same enzyme with high specificity for its substrate. The enzyme itself is the hypothesis of primary homology. For reversible enzymes, this is valid for a given reaction and the reverse reaction. The absence is coded "." and the presence "1". For instance, the catabolism of aspartate and asparagine both use aspartate aminotransferase which transforms aspartate into oxaloacetate. The anabolism of these two amino acids use the same enzyme for the reverse reaction which transforms oxaloacetate into aspartate. The character is therefore called "aspartate aminotransferase" (character 34). It is coded "1" for taxa dASP2, dASN2, sASP2, sASN2, and "." for dGLU, sGLU1, dALA, sALA, for example. This type of homology is used in characters 33 to 202 (Tables 1 and 2).

2.2.2. Type II homology

2.2.2.1. IIa: shared enzymatic functions. Several pathways utilize the same enzymatic functions, i.e. exhibit the same kind of chemical transformation, without considering the specificity of each enzyme for its substrate. The underlying hypothesis is that similarity in enzymatic function must correspond to similarity in structure of active sites, with the hypothesis that enzymes must have evolved from generalist active sites to specialized ones [29–32]. When the substrate is present but the function is not performed, the character state is coded ".". When the substrate is present and the function is performed, the character state is coded "1". When the required substrate is not available, the character state is coded "?". For example, the catabolism of alanine and aspartate both perform transaminations which use exactly the same enzymatic mechanisms involving pyridoxalphosphate, and two similar enzymes that differ in their specificities for their respective substrates: alanine aminotransferase and aspartate aminotransferase. The reverse anabolic reactions use the same enzymes with high specificities for their respective oxoacids. The character state is coded "1" in dASP2, dALA, sASP2, sALA and other pathways where aminotransferases occur, "." in dSER, dGLY, sASP1, sASN1, where transaminations occur with a different mechanism that uses NAD, and "?" for portions of the Krebs cycle where transaminations are impossible. This type of homology is used in characters 2–10, 13, 15, 17, 19, 20, 22–32.

2.2.2.2. IIb: shared cofactors. Shared cofactors reflect similarity in enzymatic mechanisms, which in this case do have the same functional meaning. If a common cofactor is used without similarity in the enzymatic mechanisms, it is considered that the use of this cofactor has been gained independently, and each enzymatic mechanism is thus coded as type IIa homologies. For example, this criterion applies to pyridoxal phosphate (PLP), which functions in deamination/amination. The character state is coded "." when the deamination or the amination is not performed although possible, or when it is performed but using another cofactor, "1" when direct deamination or amination occurs or when a transamination uses PLP, and "?" for portions of the Krebs cycle where neither deaminations/aminations nor transaminations are possible.

This criterion of homology is restricted in its application. First, when the use of a cofactor is too specific to a given enzyme, coding this cofactor as a character would lead to coding the character associated with the enzyme (type I homology) twice. For instance, the use of biotin is restricted to propionyl carboxylase. In such a case, the character "biotin" is not taken into account, otherwise it would give a double weight to the character "propionyl carboxylase". Second, when a cofactor is specific to an enzymatic function, there is also a risk of over weighting a type IIa character homology. For example, thiamin is specific to alphadecarboxylations. Third, coding a ubiquitous cofactor brings risks of homoplasy. It is necessary in this case to consider the kind of reaction, that is, to come back to a type IIa homology. For instance, NAD is used in a wide range of enzymatic functions: NAD-deaminations, NAD-aldehyde acid dehydrogenases, NAD-betaoxydations, NAD-alphadecarboxylations. Recoding each of these functions is useless: they are already coded as IIa homologies. Consequently, it appears that this criterion of homology is only useful for three characters (characters 1, 16, 21).

2.2.2.3. IIc: shared functional family. In principle, this homology criterion is the same as IIa, just relaxed. It is an extension of the above idea that enzymes must have evolved from generalists to specialists. The character state is coded "." when the reaction is not performed though possible, "1" when the reaction is performed and "?" when the reaction cannot be performed, considering the chemical groups present in the metabolitic pathway ("?" for deaminations/aminations in KC1 and KC2). This criterion actually concerns three main families of reactions: decarboxylations (character 11), deaminations/aminations (character 12), and phosphorylations (character 14).

2.2.2.4. IId: recurrence. A new kind of type II homology was proposed by Cunchillos and Lecointre [19]. Pathways can be similar in the recurrence of a set of reactions made by the same enzymes from different substrates. This homology is used in the metabolism of fatty acids (character 18) and called "IId".

2.3. Characters: names

All characters are named in Table 1, along with the number from the enzymatic nomenclature and the type of homology involved. In Figs. 3 and 4, characters involving type II homologies are numbered from 1 to 32. It must be stressed that, although type I homologies are named strictly following the international enzymatic nomenclature, type II homologies are not. For example, hydratase (character 29) is used here in a wider meaning than in international nomenclature. A reaction involving the pyridoxal-phosphate is coded as involving a hydratase because a molecule of water is implied, while it is not the case for international nomenclature. Conversely, our delineation of type II homologies is more precise for

Table 1
Names of characters, with the corresponding number of international nomenclature [37] for the sake of precision, and homology types defined in the text

[1] Pyridoxalphosphate enzymes; IIb.
[2] Deamination (NAD); IIa.
[3] Transamination (PLP); IIa.
[4] Amide deamination; IIa.
[5] Aldehyde dehydrogenation (NAD); IIa.
[6] α-Decarboxylation; IIa.
[7] Deamination (PLP); IIa.
[8] Carboxylation (biotine); IIa.
[9] No Alcohol, no aldehyde NAD dehydrogenases; IIa.
[10] β-Decarboxylation; IIa.
[11] Decarboxylation; IIc.
[12] Deamination; IIc.
[13] Acid-ammonia ligases; IIa.
[14] Phosphorylation; IIc.
[15] Acyl-CoA synthases; IIa.
[16] Thiamine pyrophosphate enzymes; IIb
[17] Transketolases; IIa.
[18] β-Oxidation/reduction sequence repeat; IId.
[19] β-Oxidation sequence; IIa.
[20] No niacin, no flavin oxidases; IIa.
[21] Tetrahydrofolate enzymes; IIb.
[22] Transaldolases; IIa.
[23] Acetyltransferases; IIa.
[24] Succinyltransferases; IIa.
[25] Malonyltransferases; IIa.
[26] Ammonia lyases; IIa.
[27] Alcohol dehydrogenation (NAD); IIa.
[28] FAD dehydrogenation; IIa.
[29] Hydratases; IIa.
[30] Isomerases; IIa.
[31] Hydrolases; IIa.
[32] Phosphoribosyltransferase; IIa.
[33] *Amino acid dehydrogenase*: 1.4.1.5; I.
[34] *Aspartate aminotransferase*: 2.6.1.1; I.
[35] *Asparaginase*: 3.5.1.1; I.
[36] *Glutamate dehydrogenase*: 1.4.1.3; I.
[37] *Pyruvate dehydrogenase*: 1.2.4.1; I.
[38] *Serine deaminase*: 4.3.1.19; I.
[39] *Serine hydroxymethyltransferase*: 2.1.2.1; I.
[40] *Cysteine desulphhydrase*: 4.4.1.1; I.
[41] *Propionyl-CoA carboxylase*: 6.4.1.3; I.
[42] *Branched-chain-amino acid aminotransferase*: 2.6.1.42; I.
[43] *Branched-chain ketoacid dehydrogenase:* 1.2.4.4; I.
[44] *Acetyl-CoA acetyltransferase*: 2.3.1.9; I.
[45] *Aminoadipate-semialdehyde dehydrogenase*: 1.2.1.31; I.
[46] *Aminoadipate aminotransferase*: 2.6.1.39; I.
[47] *Saccharopine dehydrogenase 1*: 1.5.1.9; I.
[48] *Alanine aminotransferase*: 2.6.1.2; I.
[49] *Glutaminase*: 3.5.1.2; I.
[50] *Glutamine-oxo-acid aminotransferase*: 2.6.1.15; I.
[51] *ω-Amidase*: 3.5.1.3; I.
[52] *Arginase*: 3.5.3.1; I.
[53] *Ornithine-oxto-acid aminotransferase*: 2.6.1.13; I.
[54] *Pyrroline-carboxylate reductase*: 1.5.1.2; I.
[55] *Cysteine aminotransferase*: 2.6.1.3; I.
[56] *Mercaptopyruvate sulphurtransferase*: 2.8.1.2; I.
[57] *Cysteine dioxygenase*: 1.13.11.20; I.
[58] *Aspartate decarboxylase*: 4.1.1.12; I.
[59] *Homocysteine methyltransferase*; 2.1.1.10; I.
[60] *Cystathionine synthase*: 4.2.1.22; I.
[61] *Hydroxymethylbutyryl-CoA dehydrogenase*: 1.1.1.178; I.
[62] *Methylcrotonoyl-CoA carboxylase*: 6.4.1.4; I.
[63] *Methylglutaconyl-CoA hydratase*: 4.2.1.18; I.
[64] *Hydroxymethylglutaryl-CoA lyase*: 4.1.3.4; I.

Table 1 (*continued*)

[65] *Ketoacid CoA-transferase*: 2.8.3.5; I.
[66] *Methylmalonate-semialdehyde dehydrogenase*: 1.2.1.27; I.
[67] *Saccharopine dehydrogenase 2*: 1.5.1.7; I.
[68] *Citrate synthase*: 2.3.3.1; I.
[69] *Citrate dehydratase*; 4.2.1.3; I.
[70] *Isocitrate dehydrogenase*: 1.1.1.41; I.
[71] *α-Oxoglutarate dehydrogenase*: 1.2.4.2; I.
[72] *Succinyl-CoA hydrolase*: 3.1.2.3; I.
[73] *Succinate dehydrogenase*; 1.3.99.1 I.
[74] *Fumarate hydratase*: 4.2.1.2; I.
[75] *Malate dehydrogenase*; 1.1.1.37; I.
[76] *Asparagine synthase (ADP-forming)*: 6.3.1.4; I.
[77] *Asparagine synthase (glutamine-hydrolysing)*: 6.3.5.4; I.
[78] *Glutamine synthetase*: 6.3.1.2; I.
[79] *Amino acid acetyltransferase*: 2.3.1.1; I.
[80] *Acetylglutamate kinase*: 2.7.2.8; I.
[81] *N-acetyl-glutamyl-P reductase*: 1.2.1.38; I.
[82] *Acetylornithine aminotransferase*: 2.6.1.11; I.
[83] *Acetylornithine deacetylase*: 3.5.1.16; I.
[84] *Ornithine carbamoyltransferase*; 2.1.3.3; I.
[85] *Argininosuccinate synthetase*: 6.3.4.5; I.
[86] *Argininosuccinate lyase*: 4.3.2.1; I.
[87] *Glutamate kinase*; 2.7.2.11; I.
[88] *Glutamate semialdehyde dehydrogenase*; 1.5.1.12; I.
[89] *Oxaloacetate decarboxylase*: 4.1.1.3; I.
[90] *Phosphoenolpyruvate carboxykinase (GTP)*: 4.1.1.32; I.
[91] *Enolase*: 4.2.1.11; I.
[92] *Phosphoglycerate phosphomutase*; 5.4.2.1; I.
[93] *Glycerate kinase*: 2.7.1.31; I.
[94] *Hydroxypyruvate reductase*: 1.1.1.81; I.
[95] *Serine-pyruvate aminotransferase*: 2.6.1.51; I.
[96] *Phosphoglycerate dehydrogenase*: 1.1.1.95; I.
[97] *Phosphoserine aminotransferase*: 2.6.1.52; I.
[98] *Phosphoserine phosphatase*: 3.1.3.3; I.
[99] *Acyl-CoA dehydrogenase*: 1.3.99.3; I.
[100] *Serine acetyltransferase*: 2.3.1.30; I.
[101] *Cysteine synthase*: 2.5.1.47; I.
[102] *Aspartate kinase*: 2.7.2.4; I.
[103] *Aspartate-semialdehyde dehydrogenase*: 1.2.1.11; I.
[104] *Homoserine dehydrogenase*: 1.1.1.3; I.
[105] *Homoserine succinyltransferase;* 2.3.1.46; I.
[106] *Cystathionine synthase:* 2.5.1.48; I.
[107] *Cystathionine lyase*: 4.4.1.8; I.
[108] *Homoserine kinase*: 2.7.1.39; I.
[109] *Threonine synthase*: 4.2.3.1; I.
[110] *Acetolactate synthase*: 2.2.1.6; I.
[111] *Keto-acid reductoisomerase*:1.1.1.86. I.
[112] *Dihydroxyacid dehydratase*: 4.2.1.9. I.
[113] *Isopropylmalate synthase*: 2.3.3.13; I.
[114] *Isopropylmalate isomerase*; 4.2.1.33; I.
[115] *Isopropylmalate dehydrogenase*: 1.1.1.85; I.
[116] *Dihydrodipicolinate synthase*: 4.2.1.52; I.
[117] *Dihydrodipicolinate reductase*: 1.3.1.26; I.
[118] *Piperideindicarboxylate succinyltransferase:* 2.3.1.117; I.
[119] *Succinyl-diaminopimelate aminotransferase*: 2.6.1.17; I.
[120] *Succinyl-diaminopimelate desuccinylase*: 3.5.1.18; I.
[121] *Diaminopimelate epimerase*: 5.1.1.7; I.
[122] *Diaminopimelate decarboxylase*: 4.1.1.20; I.
[123] *Homocitrate synthase*: 2.3.3.14; I.
[124] *Homoaconitate hydratase*: 4.2.1.36; I.
[125] *Homoisocitrate dehydrogenase:* 1.1.1.155; I.
[126] *Phosphoglycerate kinase:* 2.7.2.3; I.
[127] *Triosephosphate dehydrogenase:* 1.2.1.12; I.
[128] *Triose-phosphate isomerase:* 5.3.1.1; I.
[129] *Fructose-bisphosphate aldolase:* 4.1.2.13; I.

(*continued on next page*)

Table 1 (*continued*)

[130] *Fructose-bisphosphatase:* 3.1.3.11; I.
[131] *Glucose-6-phosphate isomerase:* 5.3.1.9; I.
[132] *6-Phosphofructokinase:* 2.7.1.11; I.
[133] *Pyruvate kinase:* 2.7.1.40; I.
[134] *Glycolaldehydetransferase:* 2.2.1.1; I.
[135] *Sedoheptulose diphosphatase:* 3.1.3.37; I.
[136] *Ribulose-phosphate 3-epimerase:* 5.1.3.1; I.
[137] *Ribulose-5-phosphate isomerase:* 5.3.1.6; I.
[138] *Phosphoribulokinase:* 2.7.1.19; I.
[139] *Ribulose diphosphate carboxylase:* 4.1.1.39; I.
[140] *Glucose-6-phosphate 1-dehydrogenase:* 1.1.1.49; I.
[141] *Phosphogluconolactonase:* 3.1.1.31; I.
[142] *Phosphogluconic acid dehydrogenase:* 1.1.1.44; I.
[143] *Dihydroxyacetonetransferase:* 2.2.1.2; I.
[144] *Dihydrolipoamide dehydrogenase:* 1.8.1.4; I.
[145] *Acetate thiokinase:* 6.2.1.13; I.
[146] *Acyl-CoA oxidase:* 1.3.3.6; I.
[147] *Enoyl-CoA hydratase:* 4.2.1.17; I.
[148] *β-Hydroxyacyl dehydrogenase:* 1.1.1.35; I.
[149] *Acetyl-CoA C-acyltransferase:* 2.3.1.16; I.
[150] *Trans-2-enoyl-CoA reductase (NAD*): 1.3.1.44; I.
[151] *Acyl-ACP-hydrolase:* 3.1.2.14; I.
[152] *Acetyl-CoA carboxylase:* 6.4.1.2; I.
[153] *Acyl-carrier-protein] S-malonyltransferase:* 2.3.1.39; I.
[154] *Acyl-carrier-protein] S-acetyltransferase:* 2.3.1.38; I.
[155] *β-Ketoacyl synthetase:* 2.3.1.41; I.
[156] *β-Ketoacyl reductase:* 1.1.1.100; I.
[157] *Enoyl acyl carrier protein hydratase:* 4.2.1.58; I.
[158] *Acyl-ACP dehydrogenase:* 1.3.1.10; I.
[159] *β-Hydroxydecanoate dehydrase:* 4.2.1.60; I.
[160] *Glutaryl-CoA dehydrogenase:*1.3.99.7; I.
[161] *Phenylalanine 4-monooxygenase*:1.14.16.1; I.
[162] *Hydroxyphenylpyruvate dioxygenase*:1.13.11.27; I.
[163] *Dihydroxyphenylacetate 2,3-dioxygenase*:1.13.11.15; I.
[164] *Maleylacetoacetate isomerase*: 5.2.1.2; I.
[165] *Fumarylacetoacetase*: 3.7.1.2; I.
[166] *Tryptophan 2,3-dioxygenase*: 1.13.11.11; I.
[167] *Kynurenine formamidase*: 3.5.1.9; I.
[168] *Kynurenine hydroxylase*: 1.14.13.9; I.
[169] *Kynureninase*: 3.7.1.3; I.
[170] *Hydroxyanthranilate oxygenase*: 1.13.11.6; I.
[171] *Picolinic acid carboxylase*: 4.1.1.45; I.
[172] *Aminomuconate-semialdehyde dehydrogenase*: 1.2.1.32; I.
[173] *Histidine ammonia-lyase*: 4.3.1.3; I.
[174] *Urocanate hydratase*: 4.2.1.49; I.
[175] *Imidazolone propionic acid hydrolase*: 3.5.2.7; I.
[176] *Glutamate formyltransferase*: 2.1.2.5; I.
[177] *Aromatic amino acid aminotransferase*: 2.6.1.57; I.
[178] *Deoxy-7-phosphoheptulonate synthase*: 2.5.1.54; I.
[179] *Dehydroquinate synthase*: 4.2.3.4; I.
[180] *Dehydroquinate dehydratase*: 4.2.1.10; I.
[181] *Shikimate 5-dehydrogenase*: 1.1.1.25; I.
[182] *Shikimate kinase*: 2.7.1.71; I.
[183] *Riboflavin synthase*: 2.5.1.9; I.
[184] *Anthranilate synthase*: 4.1.3.27; I.
[185] *Anthranilate phosphoribosyltransferase*: 2.4.2.18; I.
[186] *Phosphoribosylanthranilate isomerase*: 5.3.1.24; I.
[187] *Indoleglycerol phosphate synthetase*: 4.1.1.48; I.
[188] *Tryptophan synthase*: 4.2.1.20; I.
[189] *Prephenate dehydrogenase*: 1.3.1.12; I.
[190] *Chorismate mutase*: 5.4.99.5; I.
[191] *Prephenate dehydratase*: 4.2.1.51; I.
[192] *Methylmalonyl-CoA epimerase*: 5.1.99.1; I.
[193] *Methylmalonyl-CoA mutase*: 5.4.99.2; I.
[194] *Carbamate kinase*: 2.7.2.2; I.
[195] *ATP phosphoribosyltransferase*: 2.4.2.17; I.

Table 1 (*continued*)

[196] *Phosphoribosyl-ATP diphosphatase*: 3.6.1.31; I.
[197] *Phosphoribosyl-AMP cyclohydrolase*: 3.5.4.19; I.
[198] *Phosphoribosylformiminoaminophosphoribosylimidazolecarboxyamide isomerase*: 5.3.1.16; I.
[199] *Imidazoleglycerol-phosphate dehydratase*: 4.2.1.19;I.
[200] *Histidinol-phosphate transaminase*: 2.6.1.9; I.
[201] *Histidinol-phosphatase*: 3.1.3.15; I.
[202] *Histidinol dehydrogenase*: 1.1.1.23; I

From character 1–32, homologies of type II; from characters 33–202, homologies of type I.

dehydrogenases (or reductases) involving NAD. Alcohol-NAD-dehydrogenases, aldehyde-NAD-dehydrogenases, deaminase-NAD-dehydrogenases and FAD-dehydrogenases have been separated into different characters.

2.4. The matrix

How should one manage the presence of reverse reactions in the same matrix? Because the biosynthesis of a given amino acid can follow a different path than its degradation, some enzymes are used in catabolism, but they have no meaning for the anabolism of the same amino acid. In this case the catabolic enzyme will be coded with a question mark in the anabolic pathway. In the same way, question marks will be present for some anabolic enzymes in catabolic pathways where they have no meaning. The total matrix contains 75 taxons and 202 characters (Table 2).

2.5. Phylogenetic reconstruction

2.5.1. Tree search

Characters were treated as unordered and unweighted in the search of the most parsimonious tree. Heuristic searches were conducted with NONA [34] as implemented in WINCLADA [35], using TBR branch swapping. For a better exploration of trees, the Parsimony Ratchet (Hopper Islands [36]) was used. The proportion of data to be reweighted was set between 25% and 50% and the number of iterations progressively increased from 25,000 to 150,000 (option ambpoly =). This increase in iterations was used to check that the number of supplementary MP trees gained each time was decreasing or null. Each time the number of trees was recorded after collapsing all unsupported nodes in all trees (“hard collapse”).

2.5.2. Rooting

The tree was rooted using an all-zero hypothetical ancestor (HYPANC). This is justified by the fact that, in the coding of character states, a dot was given to the absence of enzymes, or to the absence of performance of particular functions (even in presence of a putative suitable substrate), or to the absence of use of a cofactor. Such a rooting option will automatically put the simplest pathways closer to the root. However, this does not make any assumption of the nature of the corresponding enzymatic reactions. The aim was to produce explicit

Table 2
Matrix containing 75 taxons (rows) and 202 characters (columns)

HYPANC	.....
sASP1	.1.?.?...??1.....??.....1.1.1.....
sASP2	1.1?.?1..??1.....??.....1.1..1.....
sASN1	.1.1.?...??111...??.....1.1.1.....1.....
sASN2	1.11.?1..??111...??.....1.1..1.....1.....
sASN3	.1.1.?...??111...??.....1.1.1.....1.....
sASN4	1.11.?1..??111...??.....1.1..1.....1.....
sGLU1	.1.?.?...??1.....??.....1.1....1.....
sGLU2	1.1?.?1..??1.....??.....1.1..1.....
sGLN1	.1.1.?...??111...??.....1.1....1.....1.....
sGLN2	1.11.?1..??111...??.....1.1..1.....1.....
sARG1	111?1?1..??1.11..??...1.....1.1....1.....11111111.....
sARG2	1.1?1?1..??1.11..??...1.....1.1..1.....11111111.....
sARG3	111?1?1..??1.1...??.....1.1....1.....1.....11111.....
sARG4	1.1?1?1..??1.1...??.....1.1..1.....1.....11111.....
sPRO1	.1.?1?..1??1.1...??.....1.1....1.....11.....
sPRO2	1.1?1?1.1??1.1...??.....1.1..1.....1.....11.....
sALA1	1.1?.?1..111.....??.....1.1.1.....1.....1.....
sSER1	1.1?.?1..111.1...??.....1.111.....111111.....
sSER2	1.1?.?1..111.1...??.....1.111.....111...111..
sGLY1	1.1?.?1..111.1...??.1.....1.111.....1.....111111.....
sGLY2	1.1?.?1..111.1...??.1.....1.111.....1.....111...111..
sCYS1	1.1?.?1..111.11..??...1...1.111.....111111....1
sCYS2	1.1?.?1..111.11..??...1...1.111.....111...111.1
sMET1	11.?1?...??1.11..??.1..1....1.1.1.....1.....
sMET2	1.1?1?1..??1.11..??.1..1....1.1..1.....1.....
sTHR1	11.?1?...??1.1...??.....1.1.1.....
sTHR2	1.1?1?1..??1.1...??.....1.1..1.....
sILE1	111?1?1..??1.1.1.??.....1.1.1....1...1.....
sILE2	1.1?1?1..??1.1.1.??.....1.1..1...1...1.....
sLEU	1.1?.?1..111..11.??...1...1.111.....1.....1.....
sVAL	1.1?.?1..111...1.??.....1.1.1.....1.....1.....
sLYS1	111?1?1..111.11..??....1..1.111.1.....
sLYS2	1.1?1?1..111.11..??....1..1.111..1.....
sLYS3	111?1?1.1111.11..??...1...1.1.1.....111.....1.....
dASP1	.1.??..?...1..?..??.....1.1.1.....
dASP2	1.1??.1?...1..?..??.....1.1..1.....
dASN1	.1.1?..?...1..?..??.....1.1.1.1.....
dASN2	1.11?.1?...1..?..??.....1.1..11.....
dGLU	.1.??..?...1..?..??.....1.1....1.....
dGLN1	.1.1?..?...1..?..??.....1.1....1.....1.....
dGLN2	1.11?.1?...1..?..??.....1.1.....11.....
dARG	.1.?1..?...1..?..??.....1.1....1.....1.....
dPRO	.1.?1..?1..1..?..??.....11.1....1.....1.....1.....
dALA	1.1??11?..11..?1.??.....11.1.....1.....1.....
dSER	1..??11?..11..?1.??.....1.....11.....
dGLY	1..??11?..11..?1.??.1.....11.....111.....
dCYS1	1..??11?..11..?1.??.....11.....1..1.....
dCYS2	1.1??11?..11..?1.??.....11.1.....1.....11.....
dCYS3	1.1??11?..11..?1.??1.....11.1.....1.....1.....11.....
dMET	1..??111..11.1?1.??.1.....111.....1..11.....11.....
dTHR1	1..??111..11.1?1.??.....1.1.....11..1.....
dTHR2	1..??11?..11..?1.??.1.....11.....111.....
dILE	1.1??111..11.1?1.?1.....11111.....111.....1....1.....1.
dLEU	1.1??111..11.111.??...11..111.1.....111.....11111.....1.
dVAL	1.1??111.111.1?1.?1.....11111.....111.....1.....1.
dLYS	111?111?1111...1.?1...1...111.1.....1111.....1...1.....
1Krebs Cycle	?????.?..11???1..??...1...1.1.....111.....
2Krebs Cycle	??????1?..11??..1.?1.....111.1.....11111.....
Glyconeogenesis	????11??..1??1.1.??..1.....111.....1.....11.....
Gluconeogenesis	????1.?..11??1...??..1....1.111.....111.....
Calvin cycle	????11??.11??1.11??..1....11111.....1.....11.....
Pentose ph cycle	????1.?..11??1.11??..1....1.111.....111.....
dFatty acids	????????...??11..11...1...111.....
s2Fatty acids	???????..????11..11...1...1.1.1.....1.....
s1Fatty acids	????????11????11..11...1.1.1.1.1.....
dTYR	1.1??11?..11.....??1.....111.....1.....
dPHE	1.1??11?..11.....??1.....111.....1.....
dTRP	111?111?1.11...1.??1..1...111.1.....1.....1...1.....1.....
dHIS	1.1??.1?...1..?..??.1....1..1.1..1.....
sTYR	1?1?1?1.111??1.11??..1....1.111.....111.....
sPHE	1?1?1?1.111??1.11??..1....1.111.....111.....
sTRP	1?1?1?1.111??1.11??..1....1.1111.....111.....
sHIS	1?1?1?1.111??1.11??..1....1.1111.....111.....
Urea Cycle	111?1?1..??1.1...??.....1.1....1.....1.....11111111.....

(continued on next page)

Table 2 (*continued*)

HYPANC	.....
sASP1	.....
sASP2	.....
sASN1	.....
sASN2	.....
sASN3	.....
sASN4	.....
sGLU1	.....
sGLU2	.....
sGLN1	.....
sGLN2	.....
sARG1	.....1.....
sARG2	.....1.....
sARG3	.....1.....
sARG4	.....1.....
sPRO1	.....
sPRO2	.....
sALA1	.....
sSER1	.....
sSER2	.....
sGLY1	.....
sGLY2	.....
sCYS1	1.....
sCYS2	1.....
sMET1	.111111.....
sMET2	.111111.....
sTHR1	.111...11.....
sTHR2	.111...11.....
sILE1	.111...11111.....
sILE2	.111...11111.....
sLEU	.....111111.....
sVAL	.....111.....
sLYS1	.11.....1111111.....
sLYS2	.11.....1111111.....
sLYS3	.....111.....
dASP1	.....
dASP2	.....
dASN1	.....
dASN2	.....
dGLU	.....
dGLN1	.....
dGLN2	.....
dARG	.....
dPRO	.....
dALA	.....1.....
dSER	.....1.....
dGLY	.....1.....
dCYS1	.....1.....
dCYS2	.....1.....
dCYS3	.....1.....
dMET	.....1.....1.....11.....
dTHR1	.....1.....11.....
dTHR2	.....1.....
dILE	.....1..1.....11.....
dLEU	.....1.....
dVAL	.....1..1.....11.....
dLYS	.....1..11.....1.....
1Krebs Cycle	.....
2Krebs Cycle	.....1.....
Glycolysis	.....1111.111.....1.....
Gluconeogenesis	.....111111.....
Calvin cycle	.....1111111111111....1.....
Pentose ph cycle	.....111111..1.11..1111.....
dFatty acids	.....11111.....
s2Fatty acids	.....11111.....
s1Fatty acids	.....111111111.....
dTYR	.....11111.....1.....
dPHE	.....1111.....1.....
dTRP	.....1..11.....1....1111111.....
dHIS	.....1111.....
sTYR	.....111111..1.11..1111.....1111111.....1.....
sPHE	.....111111..1.11..1111.....1111111.....11.....
sTRP	.....111111..1.11..1111.....11111111111.....
sHIS	.....111111..1.11..1111.....1.....11111111
Urea Cycle	.....1.....

Each taxon (row) is a pathway of degradation (“d”) or a pathway of synthesis (“s”). Dots “.” and “1” refer to the character state found in the corresponding taxon. Question marks are assigned to character states when the enzyme or the enzymatic function is not applicable to the taxon: the required substrate is not available in the pathway. Characters (columns) are numbered following their order from left to right: the character number one is the first column; the character number 26 is the 26th column. All these characters are named in Table 1.

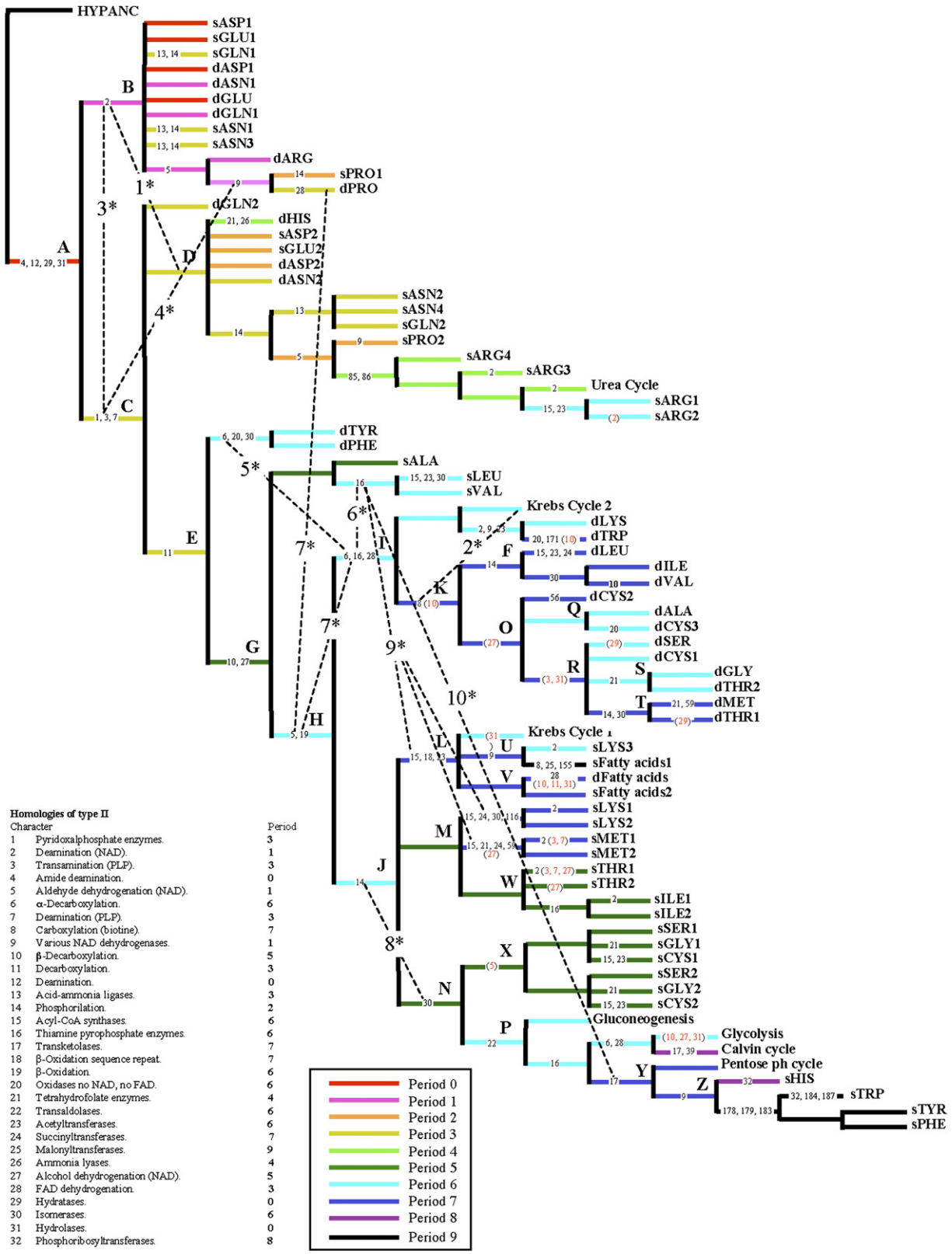

Fig. 3. Strict consensus of 20 equi-parsimonious trees [19]. Each tree is of 347 steps (CI = 0.58, RI = 0.79). Colours refer to time spans defined using an implicit clock [19]. Transversal lines joining branches are time linkages using the new functional criteria explained in the text, used in the second method for defining time spans.

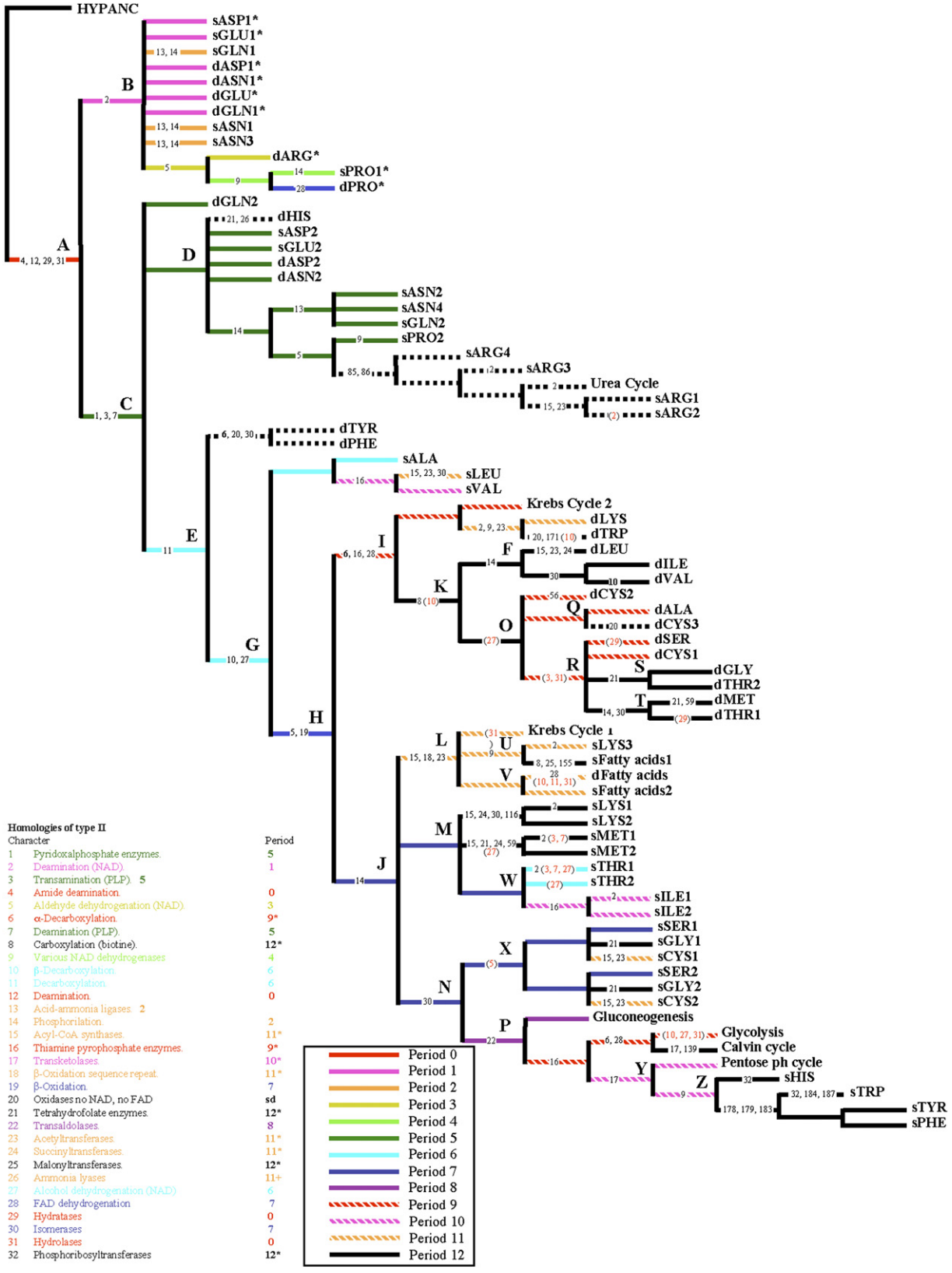

Fig. 4. The same tree as in Fig. 3, with time spans resulting from the second method, i.e. using time linkages among branches based on the new functional criteria explained in the text.

hypotheses so that the character coding and trees produced by the present study would be useful even if another way of rooting is used. They will just have to be considered again in the light of the new root. In other words, the topology (the way branches are connected to each other) could still be the same, just differently rooted.

2.6. Ordering events

2.6.1. Previous criteria for defining time spans and ordering events

From the root to the tip of branches, phylogenetic trees provide a relative order of transformations (here enzymatic innovations) through time. We call an ''upstream node'' a deep, more inclusive internal branch or clade, and a ''downstream node'' a more terminal, less inclusive internal branch or clade. Time spans (or ''periods'' or ''phases'' [16–19]) in metabolism are defined as the time along the tree separating two type II character changes and take into account the following criteria [16–19]:

2.6.1.1. The order of nodes. In the absence of other criteria, two sister-nodes are of the same relative period. There is one exception: within in the last period it cannot be said that reactions are of the same period: they are just reactions later than all others.

2.6.1.2. The nature of enzymatic changes. If a branch is followed by downstream branches that do not bear changes in the type II homologies, the downstream branches are of the same period. When a new type II homology occurs on a branch, it defines the next period. When the type II homology occurs in the next branch, but is already present in earlier periods (homoplasy), it does not define a new period. For example, the node H is light blue (Fig. 3, period 6), and the downstream node J exhibits a change in type II homology (homology II no. 14) but remains in light blue. There is no new period here because the phosphorylation is already available in a previous period (period 2, within clade B: sPRO1).

2.6.1.3. No homoplasy. When a character exhibits homoplasy, one classically has a choice in localising the changes in the tree (convergences, reversions or more complex optimizations). Only transformations with unambiguous localisations (i.e., when the placement on the tree is not open) will be taken into account in the use of the second criterion.

Classically, type II homologies involve several enzymes with different specificities. However, for 14 cases, single enzymes with the same high specificity innovate new enzymatic mechanisms. They are recorded as type I homologies, but as they also correspond to enzymatic changes used to define type II homologies, they are taken into account for defining periods. In Fig. 3, they are shown with asterisks.

2.6.2. Polytomies

It must be stressed that both the tree topology and the new non-homoplastic changes in type II homologies are involved in the definition of periods. How can this be used when parts of the tree are unresolved? A polytomy is just an absence of resolution. A polytomy normally contains branches of different periods, simply because there are different numbers of new type II homologies occurring on them. For example, within clade B (Fig. 3), a polytomy involves branches from the same period as the origin of clade B (period 1, pink) and branches from later periods, periods 2 (orange) and 3 (yellow). Some of these branches are from period 3 (yellow) because they exhibit two additional type II homologies (13 and 14). Type II homologies are therefore cumulative in defining periods.

2.6.3. The problem of question marks

Question marks bring complications when defining time spans. As a consequence of using the parsimony criterion, some type II homologies are focused in a node deeper than the one where the derived state is actually observed because of question marks in the rest of the clade terminals. For example, in Fig. 3, character 8 (carboxylations using biotin) is optimized to occur on node K. Among the 12 pathways included within K, only 5 (dLEU, dILE, dVAL, dMET and dTHR1) really exhibit this function; all others are coded ''question mark'' for that function. Thus the early position of the appearance of character 8 is due to the parsimony criterion used to manage question marks in character optimization, while the true observation of this function would require two appearances, in nodes T and F only. This causes no problem, except for ordering time spans. K is of period 7 because of the gain of this type II homology. However, 7 downstream pathways (dALA, dCYS2, dCYS3, dSER, dCYS1, dGLY and dTHR2) do not really have to wait for that period to be achieved because they just do not need to use this function, for which they are coded ''?''. Therefore, assigning their complete development to period 7 is an optimization artefact that affects time span assignment. Because these pathways are actually possible as early as the previous period (period 6), they have been assigned to that period. This is the reason why there can be an apparent contradiction in observing a node or a branch to which the assigned period is earlier than an upstream (more inclusive) branch (dCYS1 is possible in period 6 while node K is in period 7). Such situations are met four times in the analysis, and are always due to question marks.

2.6.4. New criteria

As the previous criteria have been set [16–19], assuming that a given period follows another one the same in two independent parts of the tree implicitly assumes a clockwise development of type II homologies. This assumption is rather dubious, in particular when we have no model or idea supporting a clock of enzymatic innovations. This is the reason why new functional criteria can be proposed to anchor the order of periods based on some biochemical knowledge.

1. When the enzymatic transformation involved in the type II homology depends on another one, the latter is anterior in time to the former. For example, character 8 involves biotin-carboxylation, which cannot be performed without

previous phosphorylation (character 14). This must be taken into account in the mutual ordering of time spans whose limits involve characters 8 and 14.

2. When, paradoxically, a degradation pathway ends with a short step of synthesis. In extant metabolism, this corresponds to a situation where a catabolic pathway ends with a product that cannot be used directly without a synthetic step which permits the catabolic process to continue through another previously available pathway. In this case, the synthetic step is interpreted as a connecting event between two pre-existing catabolic sequences, therefore posterior to them. For instance, the catabolic pathways of group IV amino acids (Met, Thr, Ile, Val) end with the transformation of propionyl-CoA into succinyl-CoA, which is a synthetic step because a three-carbon compound is transformed into a four-carbon compound. This step finally connects the catabolic pathways of amino acids of group IV to the Krebs cycle (the already available CK2).
3. When a coenzyme (pyridoxalphosphate) appears to be a structural and functional specialization of another one (NAD). In this case, the appearance of the former in the tree is posterior to the rise of the latter.
4. When the same biochemical change is performed through several possible enzymatic functions, and one function appears to be simpler than the other. In this case, simpler means so easy to obtain that the reaction can occur almost spontaneously; and then the simpler is anterior in the evolutionary time to the other one. For example, the beta-decarboxylation (character 10) is a spontaneous reaction that can occur without any enzyme. In that sense it is simpler than alpha-decarboxylation (character 6).
5. When a transformation yields a by-product that is used by another one, then the former is anterior to the latter. For example, the hydroxyethyl-thiamine-pyrophosphate is a by-product of the alpha-decarboxylation using thiamine-pyrophosphate. When thiamine-pyrophosphate enzymes (character 16) perform without previous alpha-decarboxylation (character 6) in their own pathway, they have to use hydroxyethyl-thiamine-pyrophosphate produced by alpha-decarboxylations from other pathways. Then the ability to use hydroxyethyl-thiamine-pyrophosphate is posterior to alpha-decarboxylations (character 6).
6. Some taxons are identified as different, while they only differ in a single feature, like a single reaction. Indeed, there can be sets of reactions (e.g. dASP1, dASN1, sASP1, sASN1, sASN3, sGLU1, sGLN1, sPRO1) distinguished from other sets (dASP2, dASN2, sASP2, sASN2, sASN4, sGLU2, sGLN2, sPRO2) while they are very similar, differing in a single reaction (NAD-deamination for the first set, PLP-transamination for the second set). As a consequence, the tree topologies within each set are very similar. When two such clades exhibit the same series of transformations along their branches, the structure of the temporal spans can be transferred from one to the other. Actually there is a single case to deal with: anabolic pathways of clade B and clade D (Fig. 3, Table 3). As an example, let us just consider the anabolisms of asparagine sASN1 and sASN3 in clade B. The structure of these anabolic pathways is the same as sASN2 and sASN4 in clade D, except a difference in the two latter which use PLP-transamination (character 3, node C). All anabolic pathways of clade B are actually in the same situation of similarity with regard to those of clade D. Therefore the structure of the time spans are transferred from D to B, though pathways of D are fully completed after those of B because of changes in the node C. By doing this, we transfer temporal information from the most resolved clade to the less resolved clade.

These criteria lead one to draw lines joining branches of the tree (Fig. 3), either to show that they must belong to the same time span or to show that one is posterior to the other. Each line will be explained below.

3. Results

3.1. The tree

From the data matrix (Table 2), the heuristic search [19] using the Parsimony Ratchet stabilized the number of equi-parsimonious trees to 20 (347 steps, CI = 0.58, RI = 0.79) [19]. The strict consensus is shown Fig. 3. Colours on that tree are those implicitly assuming a clockwise development of metabolism. Anyway the first enzymes to occur (node A) are hydrolases, hydratases, and those involving the deamination of amino acids. Therefore, degradation of amino acids must have preceded all other kind of reactions.

3.2. New time spans

The temporal links across character changes in the tree (Fig. 3) as defined as above are numbered and can be described as follows.

- *Link 1*. See the sixth criterion above. Although the pathways of clade D are fully achieved after those of clade B, they have the same time span structure. Indeed, all events are common to pathways of both clades, except those of node C (all concerning PLP), which later lead to the complete pathways of node D. Moreover, if a majority-rule consensus tree is used instead of the strict consensus tree, it is noticeable that clades B and D exhibit exactly the same internal tree topology.
- *Link 2*, between the change in character 8 on node K and CK2. See the second criterion above. The catabolic pathways of group IV amino acids (Met, Thr, Ile, Val) end with the transformation of propionyl-CoA into succinyl-CoA (character 8), which is a synthetic step because a three-carbon compound is transformed into a four-carbon compound. This synthetic step is interpreted as a late connecting event between two pre-existing catabolic sequences: the pathways of group IV amino acids and the Krebs cycle (the already available CK2).

- *Link 3*, between the change in character 2 (NAD-deaminations) and changes in characters 1, 3 and 7 (all involving PLP). The coenzyme pyridoxalphosphate appears to be a structural and functional specialization of the coenzyme NAD. The rise of the former in the tree is posterior to the rise of the latter. Indeed, the reactive sites of both coenzymes are highly similar. Compared to NAD, PLP has just a supplementary aldehydic carbon attached to the C6 of the niacin ring, but the reactive mechanisms are the same in the two cases. Otherwise other differences do not concern the reactive site itself, but parts of the coenzymes involving the peculiarities of free cellular coenzymes (in the case of NAD) or peculiarities of coenzymes behaving as a prosthetic group linked to the enzyme (PLP). PLP appears as a further specialization of NAD because it avoids the need for an independent synthetic transformation for reinitializing the form of the coenzyme after the deamination, which is the case for NAD. In PLP, the first step of the transamination is a deamination similar to that of NAD, but then the initial form of the coenzyme is automatically reacquired by the next step of amination. PLP seems more economic than NAD and can be interpreted as selected after the rise of NAD.
- *Link 4*, according to the previous argument, changes in characters involving PLP (characters 1, 3, 7) are necessarily posterior to the rise of NAD as a cofactor (character 2). Characters 5 and 9 represent diversification steps in the use of NAD. Although it is possible to consider that the rise of the PLP cofactor is possible at these steps, it is very likely that the selection of a new specialized cofactor was possible only under the constraint of exhaustion of the previously available cofactors. The hypothesis is then that the use of pyridoxalphosphate (character changes 1, 3, 7) is posterior to the phase of diversification in the use of NAD (character changes 5 and 9).
- *Link 5*. See the fourth criterion above. The degradation of tyrosine and phenylalanine both share an alpha-decarboxylation using hydroxyphenylpyruvate dioxygenase (enzyme 1.13.11.27, character 162), which is a highly specific and sophisticated decarboxylation that uses dioxygen and involves a restructuring of the whole molecule, an enzymatic mechanism far more complex than standard alpha-decarboxylations (character 6 on node I) using thiamine-pyrophosphate (character 16). The more sophisticated alpha-decarboxylation using dioxygen is therefore posterior to the standard one. It is not possible to infer how much posterior, so the period remains unknown.
- *Link 6*. See the fifth criterion above. The syntheses of leucine and valine involve thiamine-pyrophosphate enzymes (character 16) to transfer hydroxyethyl, without previous alpha-decarboxylation (character 6) in their own pathway. The reaction requires an external hydroxyethyl, which is provided by the hydroxyethyl-thiamine-pyrophosphate, a by-product of the alpha-decarboxylation using thiamine-pyrophosphate performed elsewhere. The synthesis of leucine and valine are therefore posterior to node I, where character 6 occurs.
- *Link 7*. See the first criterion above. FAD dehydrogenation (character 28) is necessary to perform beta-oxidation (character 19). The beta-oxidation of node H is therefore simultaneous or posterior to the degradation of proline and to node I which involves FAD dehydrogenation.
- *Link 8*. In terms of type II homologies, the isomerase occurring on node N under character change 30 is actually phosphoglycerate phosphomutase (character 92, a type I homology, not shown on the tree in Fig. 1), a particular isomerase whose enzymatic mechanism is similar to a phosphorylation (character 14, already available on node J and within clade B). For this reason, there is no change in period.
- *Link 9*. See the fourth criterion above. Various acyl-CoA-synthases occurring within clade J (character changes 15) require Acyl-CoAs, which are products of various alpha-decarboxylations (character 6) posterior to the possibility to incorporate hydroxyethyl-thiamine-pyrophosphate and other hydroxyacyl-TPPs into an alpha-decarboxylation (character 16 available on the branch leading to the anabolisms of leucine and valine). Therefore, clade L and two branches within clade M are posterior to the period of the branch leading to syntheses of leucine and valine.
- *Link 10*. The transketolases occurring on node Y under character change 17 (a type II homology) perform transformations in two steps, the first is similar to an alpha-decarboxylation (character 6) and the second uses TPP under the form dihydroxyethyl-TPP, a homolog of hydroxyethyl-thiamine-pyrophosphate of character change 16 that occurs on the branch leading to the syntheses of leucine and valine. Therefore node Y is of the same period as the node common to the syntheses of leucine and valine. Note that the type II homology numbered as 16 on the node just upstream to Y refers to TPPs used in the alpha-decarboxylations of the sister-group of Y that do not need external hydroxyacyl-TPPs, and to TPPs in Y that precisely require hydroxyacyls from elsewhere.

Periods have been assigned according to these constraints (Fig. 4). For a branch assigned to a given period, the corresponding enzymatic changes are also included in the same period. Then it is possible to reconstruct the steps of metabolic development (Fig. 5) as performed previously [19].

3.3. The development of metabolism

In the deepest parts of the tree, the earliest enzymatic functions are mostly linked to amino acid catabolism: deaminations, transaminations, then decarboxylations. Cordón [3] defined four groups of amino acids, according to common reactions in their metabolism and the point of entry into the Krebs cycle: group I enters by oxaloacetate, group II by ketoglutarate, group III by pyruvate and acetyl-CoA, and group IV by succinyl-CoA. In our tree (Fig. 4), amino acids of Cordón's groups I and II develop first. At period 6 (light blue), it is possible to synthesize and degrade all these amino acids (Asp, Asn, Arg, Pro, Gln, Glu). The complete metabolism of

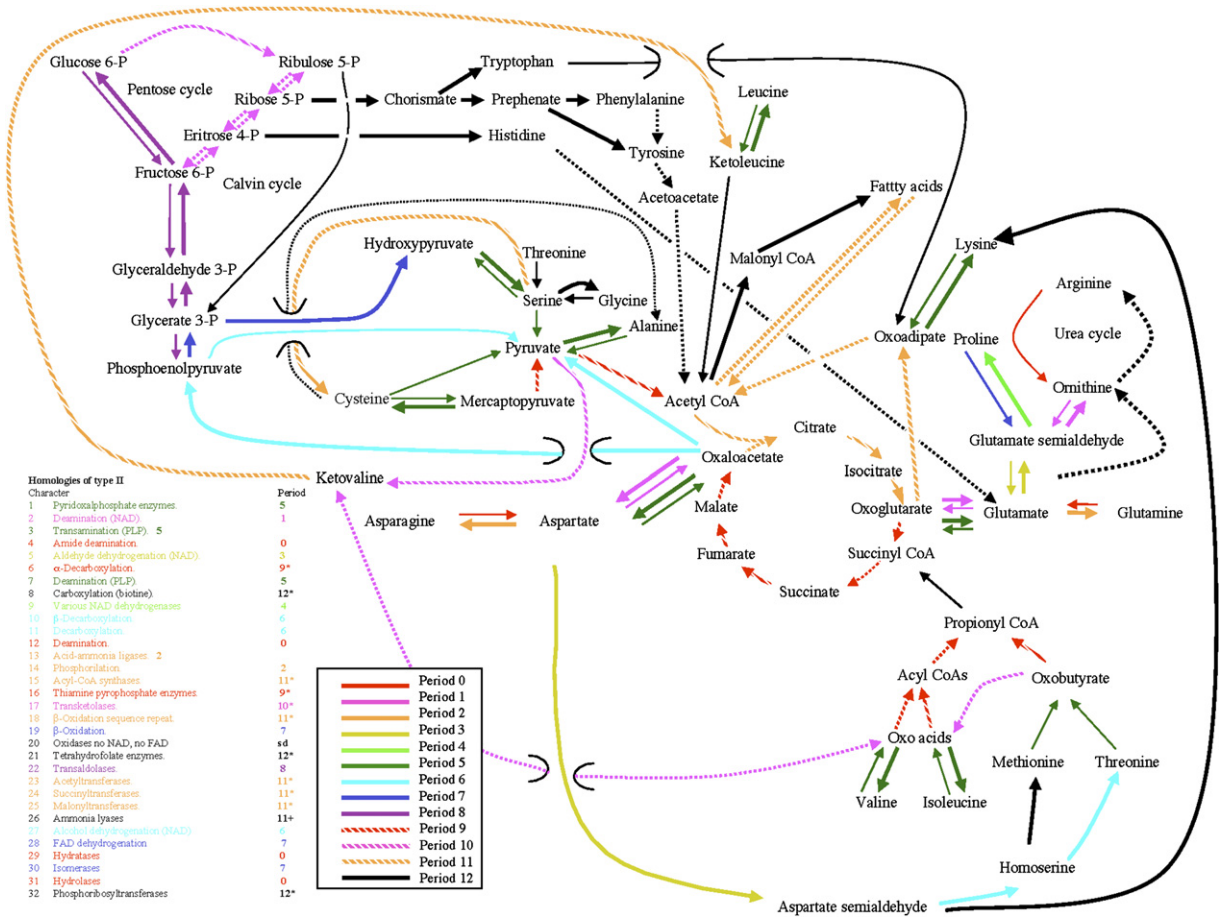

Fig. 5. General view of metabolic pathways and their connecting points to the Krebs cycle. Colours are successive time spans (or "periods") as inferred from the tree Fig. 4. Reactions in black (period 12) do not mean that all these reactions appeared within the same time span. These are just reactions later than the period 11. Black dashed arrows indicate that no period could be assigned to these reactions. However, their period is posterior to the period assigned to the upstream branch. For example, the closure of the urea cycle is posterior to the period 5. Thin arrows are for degradations, fat arrows for syntheses.

Cordón's groups III (Ser, Gly, Cys, Ala) and IV (Thr, Val, Ile, Met) starts at period 5 (deep green) and develops until period 11. The amino acids for which catabolism was developed prior to anabolism are Asn, Gln, Arg. The amino acids for which anabolism was developed before full catabolism are Pro, Thr, Ala, Ser, Val, Ile, Leu, His, Phe, Trp, Tyr. The amino acids for which both anabolism and catabolism seem to have been developed during the same time span are Asp, Glu, Lys, Cys, Met, Gly. The amino acid pathways whose development was forward are sAsn, dAsn, sGln, dGln, sPro, sArg, dAla, dSer, sCys, dCys, sMet, dMet, dThr, dIle, dVal, dLeu, dLys, sTyr, sPhe, sHis, sTrp. The amino acid pathways whose development was backward are dPro, sAla, sLys. This question is not resolved for the remaining amino acid pathways.

The two parts of the Krebs cycle did not develop within the same period. The segment CK1 developed in period 9, whereas CK2 developed later in period 11. The relative order of glycolysis and the closure of the Krebs cycle are clarified; glycolysis is achieved in period 9, before the closure of the Krebs cycle in period 11. Period 11 is also the termination period for almost all aliphatic amino acid syntheses (except for lysine, methionine, and glycine whose syntheses are achieved later) and fatty acid metabolism. To summarize, the order of general metabolic development based on Fig. 4 is as follows (Fig. 5): some amino acid catabolisms are fully achieved as early as period 0, some amino acid anabolisms at period 2, closure of the urea cycle (but with an uncertain time span assignment), glycolysis and glycogenesis are achieved at period 8, closure of the pentose–phosphate cycle occurs during period 10, closure of the Krebs cycle and fatty acid metabolism occurs in period 11, and closure of the Calvin cycle is achieved in period 12.

Some branches are in an earlier period than one of their upstream branches due to question mark optimization (see above). This is the case for the optimizations of characters 4, 8, 18 and 19. Questions marks in character 8 (carboxylations using biotin) imply that terminal branches dALA, dSER, dCYS1, and branches Q and R are found in period 9 while their upstream branches K and O are from period 12. Question marks explain why one can find new type II homologies in a branch without changing the period: the assignment of the period is one step later because of the new type II homology, and one step earlier because of a question mark associated with another character. For example, within clade B, the node on which character "5" occurs remains in period 3

because there is one step forward due to the new homology "5" and one step backward because of question marks associated with character "4" for dARG, sPRO1, dPRO. Another example is node J which remains in period 6 in spite of the new homologies "10" and "27".

4. Discussion

4.1. Comparison of results of the two ordering methods

A few type II homologies (actually 5 over 32, Table 1) disturb the order defined by Cunchillos and Lecointre [19]. However there is a good global congruence between the periods defined when a clock is implicitly assumed [19] and the periods based on the new criteria (Table3). Moreover, the latter criteria offer a more precise ordering of type II homologies. The good correlation obtained for 27 homologies between the two time frames shows that, operationally, the clock implicitly assumed in the previous study was efficient to describe biochemical evolution. However, it does not really demonstrate the clockwise behaviour of in the acquisition of new functions as we have no model to justify such a clock at short time scales; except by postulating that type II homologies are regularly obtained *on average* at large time scales. The existence of the five disturbing homologies shows that the assumption of a clockwise appearance of new functions (type II homologies) must not be made, even at large time scales.

4.2. Development of amino acid metabolism

In previous models of biochemical evolution [3], aliphatic amino acid catabolisms were generally thought to develop before the corresponding anabolisms. From the present work, full catabolism is achieved before full anabolism for Asn, and Gln in period 1, and Arg in period 3. Full anabolism is achieved before full catabolism for Pro in period 2, Thr and Ala in period 6, Ser in period 7, Val and Ile in period 10, Leu in period 11, and His, Phe, Trp, and Tyr later on. No ordering between completion of full anabolism and full catabolism can be made for Asp, and Glu in period 1, Lys in period 5, Cys period 11 and Met and Gly later on. The amino acids for which catabolism is achieved before anabolisms are the same as those found by Cunchillos and Lecointre [19], but the ordering is now ambiguous for Asp and Glu. However, there is a contradiction concerning the aromatic amino acids Tyr, Trp, His and Phe. From the present work, their anabolism is achieved before their catabolism. Full anabolism appears before full catabolism for the same amino acids as in the previous work (Cordon's groups III and IV: Ala, Cys, Gly, Ser, Leu, Val, Ile, Lys, Thr, Met), except for Lys, Cys, Met, and Gly whose ordering becomes ambiguous. In previous models

Table 3
Comparison of periods based on a clock (left) with periods based on functional links (right)

Periods using an implicit clock	**Periods according to new functional Criteria, with the same colours as in left for comparison**
Period 0: 4 Amide deamination; 12 Deamination; 29 Hydratases; 31 Hydrolases.	Period 0: **4** Amide deamination; 12 Deamination; 29 Hydratases; 31 Hydrolases.
Period 1: 2 Deamination (NAD); 5 Aldehyde dehydrogenation (NAD); 9 Various NAD dehydrogenases.	Period 1: **2 Deamination (NAD).**
Period 2: 14 Phosphorylation.	Period 2: *13 Acid-ammonia ligases;14 Phosphorylation.*
Period 3: 1 Pyridoxalphosphate enzymes; 3 Transamination (PLP); 7 Deamination (PLP); 11 Decarboxylation; 13 Acid-ammonia ligases; 28 FAD dehydrogenation.	Period 3: 5 Aldehyde dehydrogenation (NAD).
Period 4: 21 Tetrahydrofolate enzymes; 26 Ammonia lyases.	Period 4: 9 Various NAD dehydrogenases.
Period 5: 10 β-Decarboxylation; 27 Alcohol dehydrogenation (NAD).	Period 5: 1 Pyridoxalphosphate enzymes; 3 Transamination (PLP); 7 Deamination (PLP).
Period 6: 6 α-Decarboxylation; 15 Acyl-CoA synthases; 16 Thiamine-pyrophosphate enzymes; 19 β-Oxidation; 20 No NAD, no FAD Oxidases; 22 Transaldolases; 23 Acetyltransferases; 30 Isomerases.	Period 6: 11 Decarboxylation; 10 β-Decarboxylation; 27 Alcohol dehydrogenation (NAD).
Period 7: 8 Carboxylation (biotine); 17 Transketolases; 18 β-Oxidation sequence repeat; 24 Succinyltransferases.	Period 7: 19 β-Oxidation; *28 FAD dehydrogenation;* 30 Isomerases.
Period 8: 32 Phosphoribosyltransferases.	Period 8: 22 Transaldolases.
Period 9: 25 Malonyltransferases.	Period 9: 6α-Decarboxylation; 16 Thiamine pyrophosphate enzymes.
	Period 10: *17 Transketolases;* 16 no 6 (Thiaminepyrophosphate enzymes without α-descarboxilation).
	Period 11: 15 Acyl-CoA synthase; 23 Acetyltransferases; 18 β-Oxidation sequence repeat; 24 Succinyltransferases.
	Period 12: 8 Carboxylation (biotine); *21 Tetrahydrofolate enzymes;* 32 Phosphoribosyltransferases; 25 Malonyltransferases.
	No assignment: 20 No NAD no FAD oxidases; 26 Amonia lyases.

Numbers and enzymatic functions cited are homologies of type II, arising during the period. Colours are used according to [19] and Fig. 3 to show correspondences in time spans between the two methods of time span assignment. Homologies of type II showing discrepancies in time span assignment are shown in italics.

[1,3], aliphatic amino acid catabolisms develop in a forward fashion, and anabolisms in a backward fashion. The situation here seems more complex. Most of aliphatic amino acid degradations (11 of them) develop forwards (confirming the model), and only one backwards (dPro). For the remaining 2 (dArg, dGly), no ordering could be obtained. The ordering in the development of aliphatic amino acid anabolisms is more difficult: for 6 of them the development is ambiguous, for 6 others the development is forward, and for 2 it is backwards (sAla, sLys). Note that if the ancient model seems to be confirmed for catabolisms [3], it is not for anabolisms [1,3]. Indeed, very few pathways have been identified to develop backwards (dPro, sAla, sLys), confirming the results of Cunchillos and Lecointre [19]. Actually in that paper the only pathways developing backwards were dPro, dHis, dTrp, sLys, to which sAla and sSer should be added because they were forgotten in the list (note that dAla and dMet, which were wrongly interpreted to develop backwards actually develop forward in [19]). The differences between the two studies are simply linked to the fact that the developmental order of dHis, dTrp and sSer has become ambiguous here. This lack of identification of "backward anabolic pathways" could come from the late opportunistic connecting of different pathways, which provokes difficulties in ordering pathway development.

4.3. Ordering the metabolism through time

The present approach in defining time spans slightly changes the ordering of the main parts of the metabolism, as was defined in Cunchillos and Lecointre [19]:

1. Amino acid catabolism
2. Amino acid anabolism
3. Closure of the urea cycle
4. Closure of the Krebs cycle, glycolysis, glycogenesis
5. Closure of the pentose-phosphate cycle, fatty acids
6. Closure of the Calvin cycle

In the present paper, we conclude:

1. Amino acid catabolism
2. Amino acid anabolism
?. Closure of the urea cycle
3. Glycolysis, glycogenesis
4. Closure of the pentose-phosphate cycle
5. Closure of the Krebs cycle, fatty acids
6. Closure of the Calvin cycle

The two studies agree on the facts that amino acid catabolisms and anabolisms were the first to have evolved [16–18], that the closure of the urea cycle arose soon after, that the closure of the Calvin cycle is the last part to have evolved, and that fatty acid metabolism was possible just one step before the Calvin cycle. The studies slightly differ in the ordering of the intermediate steps. Interestingly, while it was not possible to order the rise of glycolysis and glycogenesis with regard to the Krebs cycle in previous studies [19], now glycolysis and glycogenesis appear to be possible before the Krebs cycle is closed, confirming the scenario of Melendez-Hevia et al. [12]. Note that the only loss of precision concerns the closure of the urea cycle, which was thought in the previous study to be possible soon after the start of amino acid metabolism. Now the closure of the urea cycle is still possible as soon as amino acid metabolism is complete, but its exact ordering among the next steps remains ambiguous.

4.4. Methodological interest

By proposing a new kind of taxon, the metabolic pathway, and by using enzymes, enzymatic functions, cofactors and families of enzymatic functions as characters, it is possible to propose phylogenetic analyses as a general analytical framework for the study of metabolic pathway evolution. Hypotheses in biochemical (metabolic) evolution become explicit, based on patterns only, and parsimonious.

Acknowledgements

We are grateful to Karen McCoy who significantly improved the manuscript. We thank Philippe Lopez for helpful discussions, and Patrick Tort for help to C.C. The Muséum National d'Histoire Naturelle and the company Saint-Gobain are acknowledged for support.

References

[1] N.H. Horowitz, On the evolution of biochemical syntheses, Proc. Natl. Acad. Sci. USA 31 (1945) 153–157.
[2] E. Schoffeniels, Biochimie Comparée, Masson, Paris, 1984.
[3] F. Cordón, Tratado Evolucionista de Biología, Aguilar, Madrid, 1990.
[4] R.A. Jensen, Enzyme recruitment in evolution of new function, Annu. Rev. Microbiol. 30 (1976) 409–425.
[5] P. Petsko, G.L. Kenyon, J.A. Gerlt, D. Ringe, J.W. Kozarich, On the origin of enzymatic species, Trends Biochem. Sci. 18 (1993) 372–376.
[6] S.D. Copley, Evolution of a metabolic pathway for degradation of a toxic xenobiotic: the patchwork approach, Trends Biochem. Sci. 25 (2000) 261–265.
[7] M. Takiguchi, T. Matsubasa, Y. Amaya, M. Mori, Evolutionary aspects of urea cycle enzyme genes, BioEssays 10 (1989) 163–166.
[8] L.A. Fothergill-Gilmore, P.A.M. Michels, Evolution of glycolysis, Prog. Biophys. Mol. Biol. 59 (1993) 105–235.
[9] C. Cunchillos, Les grands axes de l'évolution du métabolisme cellulaire, in: P. Tort (Ed.), Pour Darwin, Presses Universitaires de France, Paris, 1997, pp. 425–447.
[10] G. Michal, Biochemical Pathways, John Wiley and Sons, New York, 1999.
[11] T. Cavalier-Smith, The origin of cells: a symbiosis between genes, catalysts and membranes, Symp. Quant. Biol. LII (1987) 805–824.
[12] E. Meléndez-Hevia, T.G. Waddell, M. Cascante, The puzzle of the Krebs citric acid cycle: assembling the pieces of chemically feasible reactions, and opportunism in the design of metabolic pathways during evolution, J. Mol. Evol. 43 (1996) 293–303.
[13] W. Martin, M. Müller, The hydrogen hypothesis for the first eukaryote, Nature 392 (1998) 37–41.
[14] P.L. Forey, C.J. Humphries, I.L. Kitching, R.W. Scotland, D.J. Siebert, D.M. Williams, Cladistics. A Practical Course in Systematics, Clarendon Press, Oxford, 1992.

[15] P. Darlu, P. Tassy, Reconstruction Phylogénétique, Concepts et Méthodes, Masson, Paris, 1993.
[16] C. Cunchillos, G. Lecointre, L'histoire du catabolisme des acides aminés aliphatiques inférée par l'analyse cladistique de deux nouveaux types de caractères: l'enzyme et la réaction enzymatique, in: V. Barriel, T. Bourgoin (Eds.), Biosystema 18, Publication de la Société Française de Systématique, Paris, 2000, pp. 87–106.
[17] C. Cunchillos, G. Lecointre, Early steps of metabolism evolution inferred by cladistic analysis of amino acid catabolic pathways, C.R. Biologies 325 (2002) 119–129.
[18] C. Cunchillos, G. Lecointre, Evolution of amino acid metabolism inferred through cladistic analysis, J. Biol. Chem. 278 (2003) 47960–47970.
[19] C. Cunchillos, G. Lecointre, Integrating the universal metabolism into a phylogenetic analysis, Mol. Biol. Evol. 22 (2005) 1–11.
[20] M.C.C. de Pinna, Concepts and tests of homology in the cladistic paradigm, Cladistics 7 (1991) 367–394.
[21] W. Hennig, Phylogenetic Systematic, University of Illinois Press, Urbana and Chicago, IL, 1966.
[22] W. Hennig, Grundzüge einer Theorie der Phylogenetischen Systematik, Deuscher Zentralverlag, Berlin, 1950.
[23] J.S. Farris, The logical basis of phylogenetic analysis, in: N. Platnick, V.A. Funk (Eds.), Advances in Cladistic, vol. 2, Columbia University Press, New York, 1983, pp. 7–36.
[24] E. Zukerkandl, L. Pauling, Molecules as documents of evolutionary history, J. Theoret. Biol. 8 (1965) 357–366.
[25] E. Schoffeniels, Les Cahiers de Biochimie, Vaillant-Carmanne, Maloine, Paris, 1981.
[26] H. Gest, Evolution of the citric acid cycle and respiratory energy conversion in prokaryotes, FEMS Microbiol. Lett. 12 (1981) 209–215.
[27] H. Gest, Evolutionary roots of the citric acid cycle in prokaryotes, Biochem. Soc. Symp 54 (1987) 3–16.
[28] J.R. Knowles, Enzyme catalysis: not different, just better, Nature 350 (1991) 121–124.
[29] J.T. Holden, Evolution of transport systems, J. Theoret. Biol. 21 (1968) 97–102.
[30] J.-M. Jallon, Les glutamate-déshydrogénases de *Escherichia coli* a l'homme, in: G. Hervé (Ed.), L'Evolution des Protéines, Masson, Paris, 1983, pp. 92–103.
[31] A. Pierard, Évolution des systèmes de synthèse et d'utilisation du carbamoylphosphate, in: G. Hervé (Ed.), L'Evolution des Protéines, Masson, Paris, 1983, pp. 53–66.
[32] P.J. O'Brien, D. Herschlag, Catalytic promiscuity and the evolution of new enzymatic activities, Chem. Biol. 6 (1999) 91–105.
[33] M.J. Sanderson, H. Hufford, Homoplasy, Academic Press, Inc., New York, 1966.
[34] P. Goloboff, Nona Computer Program and Software, published by the author, Tucumán, Argentina, 1998.
[35] Nixon, K.C., Winclada (BETA), Versión 0.9.9, published by the author, Ithaca, New York, 1999.
[36] K.C. Nixon, The Parsimony ratched, a new method for rapid parsimony analysis, Cladistics 15 (1999) 407–414.
[37] IUPAC/IUB, Enzyme Nomenclature. Recommendations (1992) of the International Union of Pure and Applied Chemistry and the International Union of Biochemistry, Academic Press, San Diego, CA, 1992.