

The Conservation and Evolutionary Modularity of Metabolism

Jose Peregrin-Alvarez and John Parkinson

Program in Molecular Structure and Function, Hospital for Sick Children, Toronto, Canada
Departments of Biochemistry and Medical and Molecular Genetics, University of Toronto,
Canada

Email: jparkin@sickkids.ca

With the generation of vast amounts of sequence data from a diverse collection of organisms, there has been considerable interest in applying these datasets to examine biological networks from an evolutionary perspective [1,2]. One of the more amenable networks for such studies is the ensemble of enzymatic reactions that together make up cellular metabolism. The limited number and phylogenetic range of organisms with fully sequenced genomes have largely precluded detailed systematic analyses of metabolism within eukaryotes. Recently, however we have assembled so called partial genomes from sets of expressed sequence tags, to form complementary resources for comparative analyses [3]. Here we use partial genome datasets from 193 different eukaryotes to present a first in depth view of the conservation of metabolism across the three domains of life. Building on previous studies, we also present a comprehensive multi-level analysis of the co-conservation of enzymes that reveals a spectrum of evolutionary modularity.

Of the 1,474 enzymes analysed in this study, 933 were conserved in all six major taxonomic groups (Archaea, Bacteria, Protists, Fungi, Metazoa and Plants), demonstrating that enzymes are significantly more highly conserved than random proteins. Organizing these enzymes within the context of the Kyoto Encyclopedia of Gens and Genomes (KEGG) functional pathways, we identified a spectrum of conservation (see Figure 1). At one end, pathways were found to be highly conserved across a majority of species. These include pathways involved in carbohydrate, energy, amino acid and nucleotide metabolism. At the other end of the spectrum, several pathways were revealed to be specific to individual taxa, these included pathways involved in glycan metabolism and the generation of secondary metabolites. Interestingly, enzymes involved in these latter pathways were poorly represented in the complete genome datasets, but well represented within the partial genomes. This shows the bacterial bias of the former and highlights the need for informed choices of genomes when performing evolutionary analyses.

Having identified a high degree of conservation among enzymes but a greater variability within pathways, we next investigated the extent to which enzymes from a functional pathway are co-conserved across different genomes (i.e. exhibit modularity). In general, pathways (as defined by KEGG) were not highly modular except within smaller taxonomic groups. Examining the conservation of enzymes within the context of the global metabolic network (Figure 2), we found that highly conserved enzymes tend to be involved in multiple reactions across different pathways. Hence while these enzymes may be well conserved, the observed lack of pathway modularity suggests that the biological context within which they operate is not (i.e. multi-functional enzymes may be performing different roles in different organisms). We were also able to identify small sets of functionally related enzymes which demonstrated evolutionary modularity. These results show that while sets of enzymes may exhibit evolutionary modularity, such modularity is limited, operating below the level of defined KEGG pathways. Hence we suggest that studies examining evolutionary modularity are more informative when they operate at the level of the network rather than in the context of previously defined functional entities.

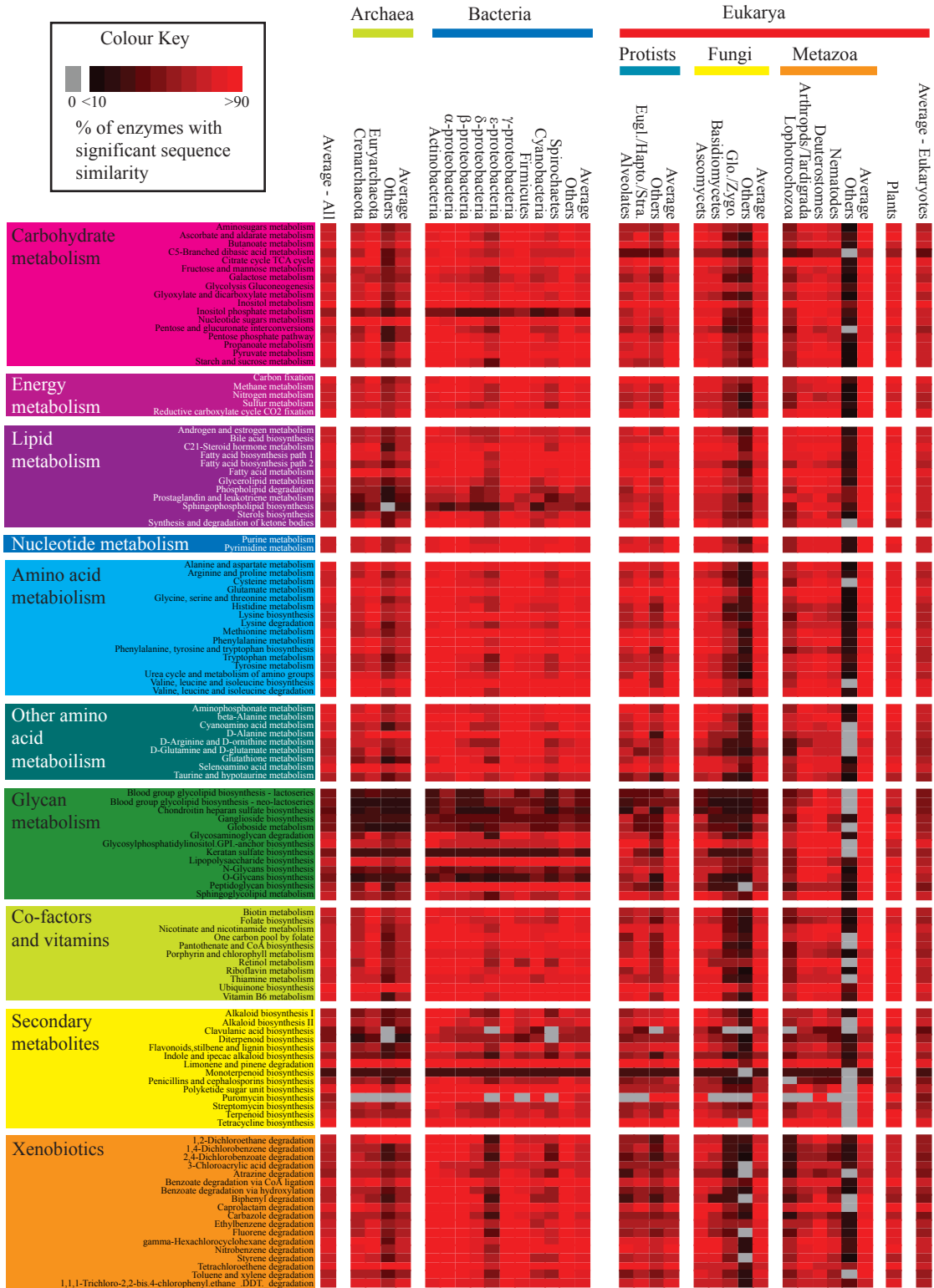


Figure 1. Conservation of metabolic pathways across the three domains of life. For each pathway we have identified the number of enzymes which are present across a broad range of phylogenetic groups. Many phylogenetically related differences are observed. For example, compared to the well conserved glycolysis, glycan metabolism is well conserved in deuterostomes (which include vertebrates) and arthropods, but relatively poorly conserved in bacteria.

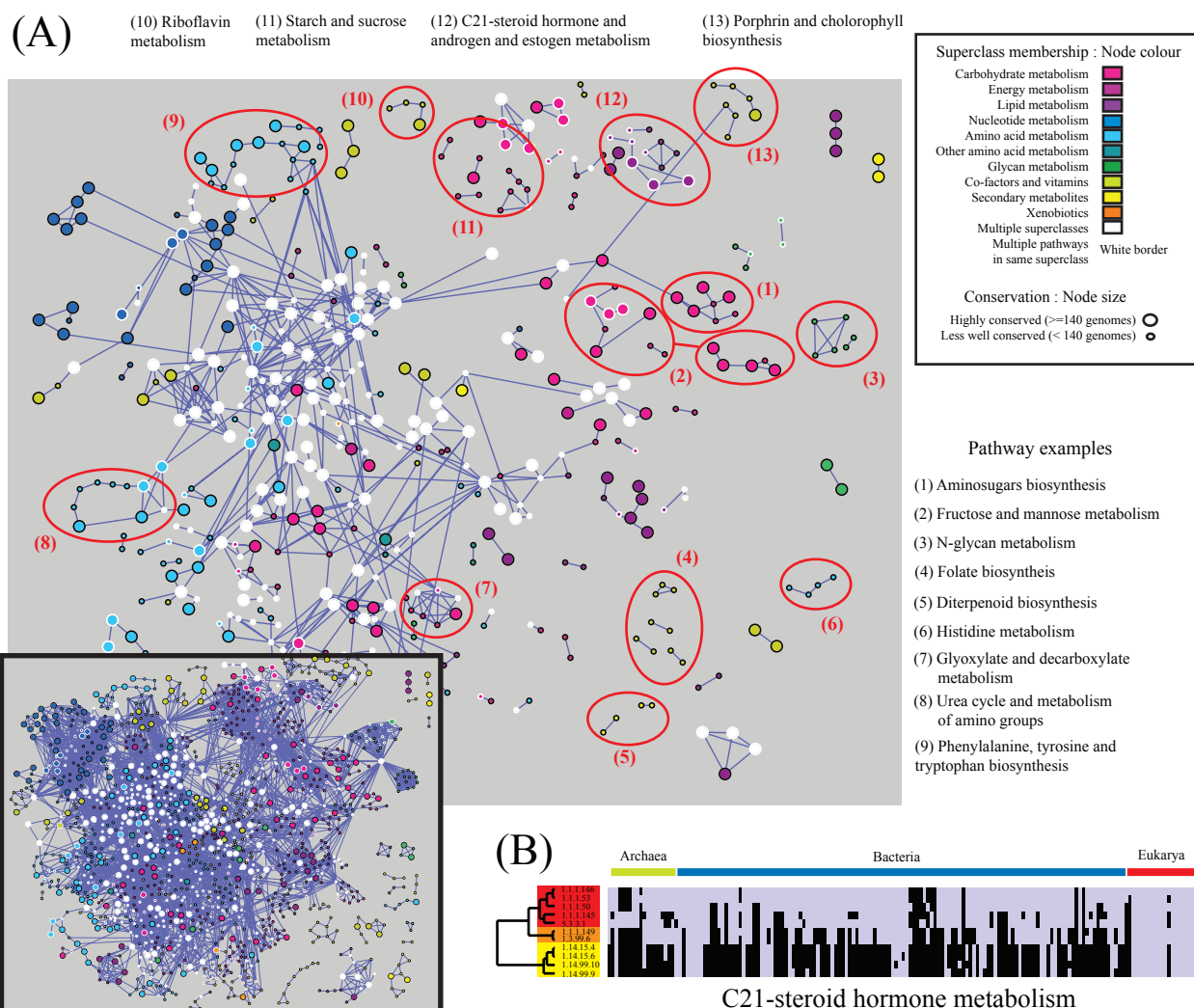


Figure 2. Evolutionary modularity of the metabolic network

(A) In the main figure, pairs of enzymes were identified which were significantly co-conserved (as measured by the Jaccard coefficient). In general most enzymes pairs were associated with the 'core' of the metabolic network, representing highly conserved enzymes (large nodes) that were involved in multiple pathways (white nodes). However, in addition to enzymes involved in the core - several groups of less well conserved enzymes were identified which also displayed a high degree of modularity. The pathway involvement of 13 sets of these enzymes are given. Inset is a global view of the entire metabolic network, constructed from the ensemble of sets of enzymatic reactions obtained from KEGG. Nodes are coloured by KEGG superclass membership.

(B) From the above diagram we noted that a large number of enzymes in the C21-steroid hormone metabolic pathway appear to be co-conserved. This clustergram shows the phylogenetic profiles, presence (grey) and absence (black) of an enzyme across 163 different completely sequenced genome). Three distinct clusters of enzymes with similar profiles can be observed, suggesting that these clusters form discrete functional units within the C21-steroid hormone metabolic pathway.

References:

1. Wuchty S. (2004) Evolution and topology in the yeast protein interaction network. *Genome Res.* 14(7):1310-4.
2. Sharan R, Suthram S, Kelley RM, Kuhn T, McCuine S, Uetz P, Sittler T, Karp RM, Ideker T. (2005) Conserved patterns of protein interaction in multiple species. *Proc Natl Acad Sci U S A.* 8;102(6):1974-9.
3. Peregrin-Alvarez JM, Yam A, Sivakumar G, Parkinson J. (2005) PartiGeneDB--collating partial genomes. *Nucleic Acids Res.* 33:D303-7.