Metabolic flux is a determinant of the evolutionary rates of enzyme-encoding genes.

Martino Colombo, Hafid Laayouni, Brandon M Invergo, Jaume Bertranpetit and Ludovica Montanucci

Institute of Evolutionary Biology (CSIC- Pompeu Fabra University),
CEXS-UPF-PRBB, Dr. Aiguader 88, 08003 Barcelona, Catalonia, Spain.

Running Title: Metabolic flux and molecular evolutionary rates.

Manuscript type: Brief Communication

Word Count: 4492 words

Figure Count: 1 Figure

Table Count: 2 Tables

Supplementary Material: yes

Keywords: Molecular Evolution, Enzyme Evolution, Erythrocyte Metabolic Network, Network Dynamics, Selective pressures.

Corresponding authors: Montanucci L. ludovica.montanucci@upf.edu, Bertranpetit J. jaume.bertranpetit@upf.edu

Abstract

Relationships between evolutionary rates and gene properties on a genomic, functional, pathway or system level, are being explored to unravel the principles of the evolutionary process. In particular, functional network properties have been analyzed to recognize the constraints they may impose on the evolutionary fate of genes. Here we took as a case study the core metabolic network in human erythrocytes and we analyzed the relationship between the evolutionary rates of its genes and the metabolic flux distribution throughout it. We found that metabolic flux correlates with the ratio of non-synonymous to synonymous substitution rates. Genes encoding enzymes that carry high fluxes have been more constrained in their evolution, while purifying selection is more relaxed in genes encoding enzymes carrying low metabolic fluxes. These results demonstrate the importance of considering the dynamical functioning of gene networks when assessing the action of selection on system-level properties.

An ongoing area of research in the field of molecular evolution is in the understanding of the evolution of genes in relation to the structure and the function of the molecular system in which they participate. To gain insight into the evolution of molecular systems, much research has been devoted to relating evolutionary properties of the genes, such as their substitution rates or detectable events of selection, with measures of their position and role within the system in which they participate, generally through its representation as a network. This approach of assessing molecular systems as networks has revealed that a significant, though small, part of the variation in the evolutionary rates of the genes can be explained by their position within the network (Fraser et al. 2002; Papp et al. 2004; Hahn and Kern 2005; Kim et al. 2007; Greenberg et al. 2008; Jovelin and Phillips 2009; Alvarez-Ponce and Fares 2012; Alvarez-Ponce 2012). However, biological interpretation of position within networks and its significance from an evolutionary point of view has largely remained obscure.

Nevertheless, in the case of metabolic networks, insights into the biological and evolutionary relevance of position within networks have been achieved through the incorporation of information about their dynamics, such as flux control or metabolic flux (Kacser and Burns 1973; Kacser and Burns 1981). These measures characterize the dynamical functioning of a metabolic pathway, and therefore offer an insight into how individual enzymes contribute to the system behavior. Therefore, measures of network dynamics can play an important role in linking the function of each enzyme to system-level properties that are closer to the phenotype and that are likely to be directly targeted by selection.

Among many flux-related measures, flux control quantifies how much a change in the activity of one enzyme affects the global flux carried by each of the other enzymes of a pathway. Flux control has been proposed to account for the differences in evolutionary rates exhibited by enzymes in different positions within metabolic networks. To this end, it has been found that enzyme position in metabolic networks is related to the non-uniform distribution of metabolic flux control, with flux control being greater for enzymes lying at upstream positions of linear metabolic pathways and at branch points of branched pathways (Flower 2007, Wright and Rausher 2010, Rausher 2013). Interestingly, the genes encoding the enzymes at these positions have been found to be common targets of selection, showing both stronger purifying selection (Rausher et al. 1999, Lu and Rausher 2003, Rausher et al. 2008, Yang et al 2009, Livingstone and Anderson 2009) and a higher rate of positive selection events (Flower et al 2007, Olson-Manning et al. 2013). Thus, flux control may be a key system-level determinant of enzyme evolution.

Besides flux control, enzyme-specific metabolic flux itself may provide additional insights into the evolution of metabolic systems. Metabolic flux measures the rate at which metabolites are produced by a reaction, particularly as it is influenced by a given enzyme. The collection of the steady-state reaction rates of the enzymes of a metabolic network under a given environmental condition is the flux distribution and it is a systemic property of the whole metabolic pathway. The relationship between measures of the evolutionary pressures on the genes encoding the enzymes and the metabolic flux carried by those enzymes has been investigated in the yeast metabolic network (Vitkup et al 2006) and enzymes carrying higher metabolic fluxes have been found to evolve at slower rates, suggesting higher selective constraint for these genes. However, given the scarcity of experimentally determined fluxes, the metabolic flux distribution of that study (Vitkup et al. 2006) was computationally predicted through flux balance analysis. In fact, an analysis of a correlation between metabolic flux and the evolutionary rates of enzymes has not yet been determined using experimentally supported flux distributions, such as those directly measured or computed through kinetic models. In order to advance our understanding of the relationship between metabolic flux and evolutionary rates, such an analysis is needed in order to empirically assess whether such a relationship exists.

Here we considered the human erythrocyte core metabolic network. This system has been deeply studied for nearly 30 years and kinetic, stochastic and constraint-based models have been determined for it (Joshi and Palsson 1989a,b, Price et al 2003). These studies provide an exceptional wealth of biochemical information for this system, including measures of metabolic flux (Wiback and Palsson 2002) and flux control (Schuster and Holzhütter 1995; Mulquiney et al. 1999). In particular, the determination of metabolic concentrations under physiological conditions, along with the *in vitro* determination of the kinetic rate law (that is, the curvilinear trend of the reaction rate as a function of the concentration of products and reagents) of all the enzymes and the experimental knowledge of chemical equilibrium parameters (Joshi and Palsson 1989b, Joshi and Palsson 1990b), have made possible the assembly of a comprehensive kinetic model of the pathway (Joshi and Palsson 1990b).

In the absence of detailed knowledge of individual reactions that would allow building a kinetic model, flux can be computationally predicted through flux balance analysis (Orth et al. 2010). Flux balance analysis is used to determine the flux distribution by applying the steady-state hypothesis and the mass balance principle to a pathway as a whole. However, these principles under-determine the system in that they can only delineate the set of all possible solutions. To estimate the physiological

steady-state flux distribution, optimization procedures must be applied. The optimal flux distribution within a metabolic network is thus defined as that which maximizes or minimizes a particular objective function (such as cell growth rate). Therefore, the validity of the determination of the flux distribution critically depends on the appropriateness of the chosen objective function, which is not obvious in cells from multicellular organisms.

In the case of the human erythrocyte metabolic network, the kinetic model was developed through the experimental determination of the rate parameters, which has allowed the derivation of a robust and predictive set of flux values with no need for any optimization against an arbitrary objective function (Joshi and Palsson 1990b; Ni and Savageau 1996; Wiback et al 2003). Flux values determined in this way agree with those directly measured in independent experiments (Joshi and Palsson 1990b). We took advantage of this exceptional dataset in order to investigate the relationship between the parameters that characterize this metabolic pathway, in terms of both its network structure and dynamic behavior, and the evolutionary dynamics of its genes.

METHODS

Genes The associations between the human genes and the proteins involved in the enzymatic reactions were determined through the iAB-RBC-283 database (Bordbar et al. 2011). This association is shown in Supplementary Table S1. In this data set there are 6 reactions that are associated to 2 genes each that code for two isoenzymes (R3: PFKL, PFKM; R4: ALDOA, ALDOC; R12: PKLR, PKLM; R14: LDHA, LDHB; R33: IMPA1, NT5C2; R36: PRPS1,PRPS1L1) and one reaction which is associated to 3 genes coding for 3 isoenzymes (R1: HK1, HK2, HK3). There are also 3 genes that are associated to 2 reactions each (BPGM: R8, R9; NT5C2: R16, R33; TKT: R30, R31). A total of 35 genes were found.

Orthologs We retrieved the one-to-one orthologs of each of the 35 human genes in three great apes species: Chimpanzee (*Pan troglodytes*), Gorilla (*Gorilla gorilla*) and Orangutan (*Pongo abelii*) from the Ensembl database (release 68). Four genes lacked an annotated one-to-one ortholog in at least one species: AMPD3 and PGM1 orthologs were missing in the Orangutan genome, ATP1A2 in the Chimpanzee genome, HK1 and TKT in the Gorilla genome. To amend cases of annotation errors, we predicted the orthologous sequences through the same procedure adopted in Montanucci et al (2011) and Invergo et al. (2013). This procedure roughly consists in predicting the structure of the gene

through GeneWise, after having performed a BLAT search against the whole genome sequence and controlling for synteny.

Multiple Sequence Alignment The multiple-sequence alignments of the orthologous genes were adapted from the corresponding GeneTree alignments (Vilella et al 2009), downloaded from Ensembl (release 68) through the Perl API, by filtering out other species and removing sequence gaps. In case of predicted orthologs, the alignment was performed through T-COFFEE (Notredame et al. 2000) version 9.03 with default parameters. The alignments were built on the protein sequences and then back-translated to the DNA sequence. All the alignments were then cleaned using Gblocks (Talavera and Castresana 2007) version 0.91 with default parameters and also visually inspected. The procedure adopted here is very stringent; however, given the incidence of sequencing errors in non-human genomes, it safeguards the analysis from spurious evolutionary signals originating from sequencing errors.

Evolutionary rates For each gene, the non-synonymous (*dN)* and synonymous (*dS)* substitution rates and their ratio (*dN/dS)* were estimated by maximum likelihood, using the *codeml* program from the PAML package (version 4.5) (Yang 2007) with model M0. Two tests of positive selection were performed by likelihood-ratio tests of model M1a versus M2a, which is a more conservative test, and M7 versus M8, which is a more sensitive test (Wong et al 2004). The two tests were always in agreement. Each PAML model was run starting from 3 different initial *dN/dS* values: 0.1, 1 and 2. Each of these models, for each *dN/dS* starting value, was run 5 times to ensure results were reproducible. These computations were automated with Python through the use of Biopython's Bio.Phylo.PAML module (Talevich et al 2012). Effective number of codons (ENC) was computed with the software CodonW.

Statistical Computations Statistical computations were carried out with IBM SPSS Statistics v19 (http://www-01.ibm.com/software/analytics/spss).

All the variables were log-transformed to approximate them to normality. This transformation did not change the magnitude nor the significance of any of the tests performed.

To correct for multiple testing, we applied a Bonferroni-based correction method to account for the correlation of multiple endpoints (Shi et al. 2012). The approach is a straightforward examination of correlated data. This correction is based on the calculation of an adjusted number of statistical tests to correct for (g*) given the correlation between the tests performed (g) g* =(g+1)- (1+(g-1)* ICC) where

g is the number of statistical tests to be performed. g* is an adjusted value for g that incorporates the correlation of the g tests. ICC is the sample-based estimate of the intraclass correlation. Hence, the "corrected α" level is α/g* and the corrected adjusted p-value equals to unadjusted p-value × g*. When ICC = 0, the ICC correction factor (g*) reduces to the standard Bonferroni adjustment, namely, g* =(g+1)- (1+(g-1)* ICC)=(g+1)-(1+(g-1)*0)= g. When ICC = 1, g* = 1, no correction is needed. In our case, we have 3 measures of flux, g= 3; ICC= 0.88; g* = 1.24

Network analysis For each node we computed the connectivity, or the number of the edges of the node, and two measures of centrality: closeness and betweenness. All the centrality measurements were computed with the NetworkX Python package on the undirected graph (version 1.7).

RESULTS AND DISCUSSION

Enzymes of the Erythrocyte Core Metabolic Network

The core metabolic network of erythrocytes was compiled by Joshi and Palsson (1989a) and, in successive works on the biochemical characterization of this system (Ni and Savageau 1996, Wiback and Palsson 2002, Wiback et al. 2003), it was extended through the addition of peripheral transport reactions. The network compiled for the present study results from the manual integration of all the cited works and comprises 36 metabolites and 48 reactions. Of these, 30 reactions are enzymatic, 8 are co-factor-balance reactions and 10 are transport reactions. A list of the reactions is shown in Supplementary Tables S1 and S2. From a functional point of view, the system can be divided into three main sub-pathways: glycolysis, responsible for the energy source for the cell; the pentose phosphate pathway, responsible for the production of NADPH and pentose sugars; and the partial nucleotide metabolism, mainly responsible for the synthesis and degradation of adenine (see Supplementary Table S1).

We retrieved the 35 genes encoding the proteins responsible for the 30 enzymatic reactions in humans and we retrieved from Ensembl the multiple-sequence alignments of the human sequences with orthologous sequences from three great apes: chimpanzee, gorilla and orangutan. The multiple alignments were then used to estimate the evolutionary rates using CODEML (Yang 2007). Genetic evolutionary rates were estimated by maximum likelihood as a unique *dN/dS* ratio for the whole phylogenetic tree and across the entire gene length, where *dN* and *dS* are the non-synonymous and the synonymous divergence levels respectively. The *dN/dS* values, which provide an estimate of the

strength of the purifying selection, range from nearly 0 to 0.69, with a mean *dN/dS* value of 0.13 and 50% of the genes in the center of the distribution ranging from 0.02 and 0.16. These low values indicate that this system is evolving under high evolutionary constraint. In addition, we performed two likelihood-ratio tests per gene to detect sequence footprints of positive selection events by assessing two pairs of nested site-models implemented in the PAML package (see methods). No events of adaptive selection were found for any of the genes in the pathway during the time of divergence of great apes (see Table S3 for the Λ statistic and *p*-values of the test for each gene).

We then checked whether the genes belonging to different sub-pathways showed any differences in evolutionary rates. We found that the *dN/dS* values among the three sub-pathways are significantly different (ANOVA test, $F_{2,32} = 3.97$ $p = 0.029$). Only one pairwise comparisons among the genes of the three sub-pathways remained statistically significant after applying Bonferroni correction for multiple testing, and it was between genes involved in the glycolysis and nucleotide metabolism sub-pathways ($t_{32} = 2.81$, $p = 0.025$). Genes involved in the glycolysis sub-pathway show significantly lower *dN/dS* values than those involved in the nucleotide metabolism sub-pathway.

Network Topological Analysis

In previous studies (Ni and Savageau 1996, Wiback and Palsson 2002, Wiback et al. 2003), the metabolic network was represented as a substrate graph. In substrate graphs, nodes represent metabolites (substrates and products), while edges represent reactions (Wagner and Fell 2001, Ravasz et al. 2002). Because our focus is on the enzymes rather than the substrates, we converted this network into its corresponding reaction graph (Montañez et al. 2010), in which nodes represent reactions and an edge is established between two nodes if the two corresponding reactions are consecutive (the product of one reaction is taken as substrate for the other reaction). Adopting a common criterion (Vitkup et al. 2006), the so-called "currency" metabolites, which are the most widely used cellular metabolites (in this network: ATP, ADP, AMP, NADH, NAD^+, NADPH, $NADP^+$), were not considered in determining the edges of the reaction graph. Since currency metabolites participate in a large number of different reactions, including them in the reaction graph would have connected reactions involved in unrelated functions, with the effect of both dominating the structure of the network and hiding functional patterns in network representation. The resulting network consists of 39 nodes and 74 edges (Figure 1). For a figure of the complementary substrate graph see Figure 1S in supplementary material and also Wiback

et al. 2003.

Each reaction was then associated with the gene (or genes) that encodes the enzyme (or enzymes) that catalyzes that reaction. To characterize the position of each gene within the network, we computed, for each node, three commonly used measures of centrality (Scardoni et al. 2009). These measures relate the position of a node with respect to the global topological organization of the network. We computed connectivity (also known as "degree"), which is the number of the node's edges; closeness (the inverse of the node's average shortest-path distance to all other nodes), in which, like in the common usage of the word "central", the most central nodes are on average closer to all of the other nodes; and betweenness (the fraction of all the shortest paths on the network that pass through a given node), which gives information on whether a protein acts as a sort of "information bridge". We then sought whether these graph-theoretic measures correlated with evolutionary rates. Connectivity, betweenness and closeness centralities were assessed against the *dN/dS* ratio and the *dN* and *dS* rates separately. Positive relationships were observed between the centrality measurements and the rates of evolution (Table 1). While the direction of these relationships is opposite to what has been found in previous studies of metabolic networks (Vitkup et al. 2006; Montanucci et al. 2011), the correlations between evolutionarily rates and the considered topological measures are non-significant. Hence, for this particular system and unlike for many others found in literature, we cannot conclude that network structure imposes constraints on the evolution of the pathway genes.

Metabolic Flux Correlates with Evolutionary Rates

We next investigated the influence of the system's dynamics on molecular evolutionary constraint. We initially considered flux control coefficients that have been determined for some reactions of this network (Schuster and Holzhütter 1995; Mulquiney et al. 1999), but it was not possible to use them given the high percentage of missing data (data not shown). We instead focused on the metabolic flux distribution. Three flux distributions were determined for this pathway (Joshi and Palsson 1990b; Ni and Savageau 1996; Wiback et al. 2003), which were derived from three successive improvements and refinements of the kinetic model (Joshi and Palsson 1990b; Ni and Savageau 1996; Jamshidi et al. 2001). We retrieved the metabolic flux distribution from the most recent work (Wiback et al. 2003), determined on the basis of the most updated kinetic model (Jamshidi et al. 2001). This flux distribution is derived for the nominal state of the erythrocyte, in which metabolite concentrations are those

experimentally measured in the physiologically unstressed state. As the kinetic model for this system also allows simulating other environmental states, such as environmental loads through *in silico* alteration of metabolite concentrations, other flux distributions corresponding to these simulated states of the cell (Wiback et al. 2003) were also considered (see next section). The calculated flux values are reported in Supplementary Table S1.

We began by investigating the relationship between the flux values in the nominal state carried by the pathway enzymes and their evolutionary rates. The comparisons between metabolic flux and both *dN/dS* ratios and *dN* of the genes showed significant negative correlations, r=-0.415 p=0.013 for *dN/dS*, r=-0.400 p=0.017 for *dN* (see also Table 1; Figure 1c;1d). *dS* showed non-significant correlations with flux (Table 1; Figure 1e). This means that genes encoding enzymes carrying high fluxes are uniformly more constrained in their evolution. Also, this effect is uniquely due to a slower rate of non-synonymous substitution. In Figure 1a and 1b the network is presented with nodes colored in a gradient of blue according to *dN/dS* values (Figure 1a) and of flux (Figure 1b). It can be noted that reactions carrying high fluxes (dark blue nodes in Figure 1b) correspond to reactions associated to genes with low *dN/dS* (light blue/white nodes in Figure 1a). Thus, a higher flux capacity constrains the evolution of a gene while evolution under a more relaxed constraint is found more often in genes encoding enzymes carrying low fluxes.

To verify correlates of evolutionary rates, other known genomic determinants must be taken into account. In particular, it is known that both expression level and gene length are major determinants of evolutionary rates, accounting for a large part of their variance (Drummond et al. 2006). A partial correlation analysis was carried out considering gene length and effective number of codons (ENC) to account for codon usage bias, which is known to be positively correlated with expression. Additionally, to disentangle the role of topological parameters, we sought whether there was a relationship between flux and topological parameters. A negative correlation between flux and closeness centrality (r=-0.35, p=0.039) indicates that metabolic flux has a non-random distribution over this network, with central genes showing lower flux values under physiological conditions. Hence we also included closeness centrality in the partial correlation analysis. When controlling for all of these variables at the same time, the correlation between flux distribution and *dN/dS maintains statistical* significance r=-0.364, p=0.040. However, the correlation between flux distribution and *dN* lost significance (r=-0.331, p=0.64). Thus, the partial correlation analysis shows that the correlation between flux and *dN/dS*

maintains significance after controlling for closeness, suggesting an independent relationship between flux and evolutionary rates and discarding the possibility of being a byproduct of the relationship between evolutionary rates and centrality measures.

Flux under simulated stressed environmental conditions

As the more recent kinetic model of this pathway is available as a software module that allows simulations (Jamshidi et al. 2001), it was also used (in Wiback et al. 2003) to derive flux distributions for two simulated stressed environmental conditions. NADH load and NADPH load were *in silico* simulated to mimic two biologically relevant states of the cell: the cell's response to the need of increasing hemoglobin concentration and to oxidative stress, respectively. To investigate whether either of these two conditions have been relevant to the evolution of the enzymes, we have also investigated the correlations of evolutionary rates with fluxes under these conditions.

The two flux values computed from the kinetic model for these conditions highly correlate with that of the nominal state (r=0.98 and $p<0.0001$ for NADH load, and r=0.84 and $p<0.0001$ for NADPH load). The small differences between these flux distributions and the nominal distribution are an increase in pyruvate uptake, a reduction in the lactate dehydrogenase flux and lactate export under the NADH load condition, and an increase in flux through the oxidative pentose phosphate reactions under the NADPH load condition (Wiback et al. 2003). This suggests that the flux distribution of the two latter conditions is the result of a biochemical fine-tuning of the system, whose flux distribution does not undergo strong modifications under stressed conditions. (All retrieved flux values are reported in Supplementary Table S1.)

We then investigated the relationship between evolutionary rates and flux in these two simulated conditions. The correlations between *dN/dS* and flux are r=-0.413 with $p=0.014$ in NADH load and r=-0.410 with $p=0.015$. Thus, we also find a significant negative correlation between *dN/dS* and flux for these two simulated cell states. Given that we tested three different flux distributions (in the whole study), we corrected the *p*-values for multiple testing through a Bonferroni-based correction method (Shi et al. 2012) that takes into account correlations between the three flux distributions. The corrected correlation *p*-values are 0.016, 0.017 and 0.019 for the nominal, NADH load and NADPH load condition respectively (see Table 2). Hence, even with this correction, the correlation between any of the fluxes and *dN/dS* holds significance. When *dN* is considered separately, the correlations with flux

are r=-0.390 with *p*=0.021 for NADH load and r=-0.338 with *p*=0.047 for NAPH load. After multiple test correction the *p*-values for the three flux conditions became 0.021, 0.026 and 0.058 for the nominal, NADH load and NADPH load condition respectively. So significance of the correlation between *dN* and flux is maintained in the nominal and NADH load condition. Conversely, *dS* shows no significant correlation with any flux measure (see Table 2). As expected, the strongest statistical significance is found for the nominal condition, whose values correspond to the experimentally measured metabolite concentrations.

When we perform partial correlations controlling for gene length, ENC and closeness centrality, the correlation between the *dN/dS* ratios and flux become r=-0.34 with p=0.057 for NADH load condition and r=-0.36 with p=0.045 for NADPH load condition. This result shows that when two particular stressed conditions of the cell are considered, the relationship between flux and evolutionary rates become marginally significant and this may suggest that these two conditions have not been determinants in the evolution of this system.

Concluding remarks

Many factors have been investigated as determinants of molecular evolutionary rates, each one giving hints about a particular aspect of the evolutionary process. In particular, the identification of system-level determinants of evolutionary rates helps to unravel how selection on individual enzymes relates to the evolution of the system as a whole. Here we show that metabolic flux, which quantifies the rates at which metabolites flow through the system, is a determinant of genetic evolutionary rates taking as an example the core metabolic network of erythrocytes.

It has to be noted that while we focused on the activity of enzymes in the core metabolic network of erythrocytes, they are in fact active in many cell types in the organism. Each cell type performs its own unique function and, therefore, imposes its own evolutionary constraints on the encoding genes. It is thus impressive that, despite this potentially great source of variance in evolutionary rates, a relationship with the dynamic properties exhibited in a single cell type could still be identified. The pathway analyzed is rather small and correlation is clearly significant. It is expected that statistical power would be much higher if dynamical data for a bigger pathway was available for analysis.

In addition, we tested this correlation both in the nominal unstressed condition and in two simulated stressed conditions of the cell; the strongest significance was found for the nominal

condition, which is the state in which the cell is thought to prominently work. As suggested in Vitkup et al. (2006), the strongest significance may help in identifying the condition that has dominated the evolution of the network. According to this observation we can conclude here that the nominal condition is the most relevant from an evolutionary point of view.

Unlike many other metabolic pathways for which significant correlations were found between evolutionary rates and topological parameters, such a correlation was not found in this study. When such correlations are found, they usually only explain a small part of the variance in evolutionary rates, so the lack of signal of a relationship between evolutionary rate and topology may be due to its complex topological structure. Also, it has been proposed that the uneven distribution of selective pressures over networks may be due to other systemic properties such as flux control. Here we show that the metabolic flux carried by each enzyme, which reflects the global organization of the flux over the entire network, imposes constraints in the evolution of the genes, with genes carrying high fluxes being highly constrained in their evolution. The introduction of relevant features that capture the dynamics of the network may allow the understanding of distinct patterns in the distribution of selective pressures over networks that are found for different pathways.

Finally, an obvious limit of this approach is the lack of availability of quantitative species-specific measurements of system-level properties. The flux values considered here have been determined for human erythrocytes and are assumed to be largely conserved throughout the great apes. Given the short evolutionary distance of the species considered in this study, the extension of these measures is a reasonable assumption. However, future system-evolutionary studies would greatly benefit from the determination of species-specific data for systemic properties.

ACKNOWLEDGEMENTS

We thank David Alvarez-Ponce for fruitful discussions on an early version of this manuscript and the two anonymous reviewers for their insightful comments and suggestions that allowed us to greatly improve this work. We are thankful to Christopher Wheat who helped us to improve the last version of this manuscript. This research was funded by grants BFU2010-19443 (subprogram BMC) awarded by Ministerio de Ciencia y Tecnología (Spain) and by the Direcció General de Recerca, Generalitat de Catalunya (Grup de Recerca Consolidat 2009 SGR 1101). BI is supported by a FI-DGR from AGAUR, Generalitat de Catalunya (2011F1 B1 00275). LM acknowledges funding from the Juan de la Cierva

Program of the Spanish Ministry of Science and Innovation (MICINN).

REFERENCES

Alvarez-Ponce, D. 2012. The relationship between the hierarchical position of proteins in the human signal transduction network and their rate of evolution. 2012. BMC Evol Biol. 12:192.

Alvarez-Ponce, D., M. A. Fares. 2012. Evolutionary rate and duplicability in the Arabidopsis thaliana protein-protein interaction network. Genome Biol Evol. 4(12):1263-74.

Bordbar, A., N. Jamshidi, and B .Ø. Palsson. 2011. iAB-RBC-283: A proteomically derived knowledge-base of erythrocyte metabolism that can be used to simulate its physiological and patho-physiological states. BMC Syst Biol. 5:110.

Drummond, D. A., A. Raval, C. O. Wilke. 2006. A single determinant dominates the rate of yeast protein evolution. Mol Biol Evol. 23(2):327-37.

Flowers, J. M., E. Sezgin, S. Kumagai, D. D. Duvernell, L. M. Matzkin, P. S. Schmidt, and W. F. Eanes. 2007. Adaptive evolution of metabolic pathways in Drosophila. Mol Biol Evol. 24(6):1347–54.

Fraser, H. B., A. E. Hirsh, L. M. Steinmetz, C. Scharfe, and M. W. Feldman. 2002. Evolutionary rate in the protein interaction network. Science 296(5568):750–2.

Greenberg, A. J., S. R. Stockwell, and A. G. Clark. 2008. Evolutionary constraint and adaptation in the metabolic network of Drosophila. Mol Biol Evol. 25(12):2537–46.

Hahn, M. W., and A. D. Kern. 2005. Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. Mol Biol Evol. 22(4):803–6.

Invergo, B. M., L. Montanucci, H. Laayouni, J. Bertranpetit. 2013. A system-level, molecular evolutionary analysis of mammalian phototransduction. BMC Evol Biol. 13:52.

Jamshidi, N., J. S. Edwards, T. Fahland, G. M. Church, B. Ø. Palsson. 2001. Dynamic simulation of the human red blood cell metabolic network. Bioinformatics. 17(3):286-7.

Joshi, A., and B. Ø. Palsson. 1989a. Metabolic Dynamics in the Human Red Cell. Part I – A Comprehensive Kintec Model. J Theor Biol. 141(4):515–528.

Joshi, A., and B. Ø. Palsson. 1989b. Metabolic dynamics in the human red cell. Part II--Interactions with the environment. J Theor Biol. 141(4):529–45.

Joshi, A., and B. Ø. Palsson. 1990a. Metabolic Dynamics in the Human Red Cell. Part III -

Metabolic Reaction Rates. J Theor Biol. 142(1):41–68.

Joshi, A., and B. Ø. Palsson. 1990b. Metabolic Dynamics in the Human Red Cell. Part IV - Data Prediction and Some Model Computations. J Theor Biol. 142(1):69–85.

Jovelin, R., and P. C. Phillips. 2009. Evolutionary rates and centrality in the yeast gene regulatory network. Genome Biol. 10(4):R35.

Kacser H., and J. A. Burns. 1973. The control of flux. Symp Soc Exp Biol. 27:65-104.

Kacser H., and J. A. Burns. 1981 . The molecular basis of dominance. Genetics. 97(3-4):639-66.

Kim, P. M., J. O. Korbel, and M. B. Gerstein. 2007. Positive selection at the protein network periphery: evaluation in terms of structural constraints and cellular context. P Natl A Sci USA. 104(51):20274–9.

Livingstone, K., and S. Anderson. 2009. Patterns of variation in the evolution of carotenoid biosynthetic pathway enzymes of higher plants. J Hered. 100(6):754-61.

Lu, Y., and M. D. Rausher. 2003. Evolutionary rate variation in anthocyanin pathway genes. Mol Biol Evol. 20:1844-53.

Montanucci, L., H. Laayouni, G. M. Dall'Olio, and J. Bertranpetit. 2011. Molecular evolution and network-level analysis of the N-glycosylation metabolic pathway across primates. Mol Biol Evol. 28(1):813–23.

Montañez, R., M. A. Medina, R. V. Solé, and C. Rodríguez-Caso. 2010. When metabolism meets topology: Reconciling metabolite and reaction networks. BioEssays 32:246-256.

Mulquiney P. J., W. A. Bubb, P. W. Kuchel. 1999. Model of 2,3-bisphosphoglycerate metabolism in the human erythrocyte based on detailed enzyme kinetic equations: in vivo kinetic characterization of 2,3-bisphosphoglycerate synthase/phosphatase using 13C and 31P NMR. Biochem. J. 342 Pt 3:567-80.

Ni, T., and M. A. Savageau. 1996. Application of Biochemical Systems Theory to Metabolism in Human Red Blood Cells. *J Biol Chem*. 271(14):7927–7941.

Notredame, C., D. G. Higgins, and J. Heringa. 2000. T-Coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol. 302(1):205–17.

Olson-Manning C.F., C.R. Lee, M.D. Rausher, T. Mitchell-Olds. 2013. Evolution of flux control

in the glucosinolate pathway in Arabidopsis thaliana. Mol Biol Evol. 30(1):14-23.

Orth, J. D., I. Thiele, B. Ø. Palsson. 2010. What is flux balance analysis? Nat Biotechnol. 28(3):245-8.

Papp, B., C. Pál, and L. D. Hurst. 2004 Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. Nature 429(6992):661-4.

Price, N. D., J. L. Reed, J. A. Papin, S. J. Wiback, and B. Ø. Palsson. 2003. Network-based analysis of metabolic regulation in the human red blood cell. J Theor Biol. 225(2):185–194.

Rausher, M. D. 2013. The evolution of genes in branched metabolic pathways. Evolution 67(1):34-48.

Rausher, M. D., Y. Lu, and K. Meyer. 2008. Variation in constraint versus positive selection as an explanation for evolutionary rate variation among anthocyanin genes. J Mol Evol. 67(2):137–44.

Rausher, M. D., R. E. Miller, and P. Tiffin. 1999. Patterns of evolutionary rate variation among genes of the anthocyanin biosynthetic pathway. Mol Biol Evol. 16(2):266–74.

Ravasz, E., A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A. L. Barabási. 2002. Hierarchical organization of modularity in metabolic networks. Science 297:1551-1555.

Scardoni, G., M. Petterlini, C. Laudanna. 2009. Analyzing biological network parameters with CentiScaPe. Bioinformatics. 2009 Nov 1;25(21):2857-9.

Shi, Q., E. S. Pavey, R. E. Carter. 2012. Bonferroni-based correction factor for multiple, correlated endpoints. Pharm Stat. 11(4):300-9.

Schuster, R., H.G. Holzhütter. 1995. Use of mathematical models for predicting the metabolic effect of large-scale enzyme activity alterations. Application to enzyme deficiencies of red blood cells. Eur J Biochem. 229(2):403-18.

Talavera, G., and J. Castresana. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol. 56(4):564–77.

Talevich, E., B. M. Invergo, P. J. Cock, and B. A. Chapman. 2012. Bio.Phylo: A unified toolkit for processing, analyzing and visualizing phylogenetic trees in Biopython. BMC Bioinformatics 13:209.

Vilella A. J., J. Severin, A. Ureta-Vidal, L. Heng, R. Durbin, E. Birney. 2009. EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. Genome Res. 19(2):327-35.

Vitkup, D., P. Kharchenko, and A. Wagner. 2006. Influence of metabolic network structure and function on enzyme evolution. Genome Biol. 7(5):R39.

Wagner, A., and D. A. Fell. 2001. The small world inside large metabolic networks. Proc Biol Sci. 268:1803-1810.

Wiback, S. J., R. Mahadevan, and B. Ø. Palsson. 2003. Reconstructing metabolic flux vectors from extreme pathways: defining the α-spectrum. J Theor Biol. 224(3):313–324.

Wiback, S. J., and B. Ø. Palsson. 2002. Extreme pathway analysis of human red blood cell metabolism. Biophys J. 83(2):808–818.

Wong, W. S., Z. Yang, N. Goldman, R. Nielsen. 2004. Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. Genetics. 168(2):1041-51.

Wright, K. M., and M. D. Rausher. 2010. The evolution of control and distribution of adaptive mutations in a metabolic pathway. Genetics 184(2):483-502.

Yang, Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24(8):1586-91.

Yang, Y. H., F. M. Zhang, and S. Ge. 2009. Evolutionary rate patterns of the Gibberellin pathway genes. BMC Evol Biol. 9:206.

TABLES

Table 1 Correlations between evolutionary rates and centrality measures and flux. Significant values are written in bold text.

	dN/dS	***dN***	***dS***
degree	r=0.049, *p*=0.782	r=0.024, *p*=0.889	r=-0.067, *p*=0.701
closeness	r=0.124, *p*=0.477	r=0.159, *p*=0.363	r=-0.005, *p*=0.979
betweenness	r=-0.014, *p*=0.937	r=0.004, *p*=0.984	r=-0.145, *p*=0.407
$flux_{nominal}$	r=-0.415, *p*=**0.013**	r=-0.400, *p*=**0.017**	r=0.054, *p*=0.757

Table 2 Correlations between evolutionary rates and flux values in three different environmental conditions. The *p*-values between parentheses are corrected for multiple testing through a Bonferroni-based correction according to Shi et al. (2012). Significant values are written in bold text.

	dN/dS	***dN***	***dS***
$flux_{nominal}$	r=-0.415 *p*=**0.013** (**0.016**)	r=-0.400 *p*=**0.017** (**0.021**)	r=0.054 *p*=0.757 (0.939)
$flux_{NADH\ load}$	r=-0.413 *p*=**0.014** (**0.017**)	r=-0.390 *p*=**0.021** (**0.026**)	r=0.090 *p*=0.606 (0.751)
$flux_{NADPH\ load}$	r=-0.410 *p*=**0.015** (**0.019**)	r=-0.338 *p*=**0.047** (0.058)	r=0.155 *p*=0.375 (0.465)

FIGURE LEGENDS

Figure 1 Distribution of evolutionary rates and flux over the reaction graph. **a)** Reaction graph of the erythrocyte core metabolic network. Nodes represent reactions (the numbers on the nodes refer to the number of the reactions as listed in Supplementary Table S1 and S2), and edges represent the sharing of a common metabolite between two reactions. White nodes with underlined reaction names correspond to transport reactions (and therefore have no associated genes). Each node is colored according to the *dN/dS* ratios of the gene encoding the enzyme that catalyze the corresponding reaction. The color ranges from white (*dN/dS* equal to 0) to dark blue (*dN/dS* equal to 0.69) **b)** Reaction graph colored according to a gradient of flux ranging from white (low flux, toward 0) to dark blue (high flux, toward the maximum value of 2.6). It can be seen that the region of the graph characterized by lower *dN/dS* values (node color toward white in figure a) are characterized by high values of flux (node color toward dark blue in figure b). **c) d) e)** Scatter plot of evolutionary rates versus flux measures in the nominal condition: c) *dN/dS*; d) *dN*; e) *dS*.