Journal of Molecular Evolution (2018) 86:443–456
https://doi.org/10.1007/s00239-018-9857-5

ORIGINAL ARTICLE

Evolutionary Perspectives of Genotype–Phenotype Factors in *Leishmania* Metabolism

Abhishek Subramanian[1,2] · **Ram Rup Sarkar**[1,2]

Received: 13 February 2018 / Accepted: 13 July 2018 / Published online: 19 July 2018

Abstract

The sandfly midgut and the human macrophage phagolysosome provide antagonistic metabolic niches for the endoparasite *Leishmania* to survive and populate. Although these environments fluctuate across developmental stages, the relative changes in both these environments across parasite generations might remain gradual. Such environmental restrictions might endow parasite metabolism with a choice of specific genotypic and phenotypic factors that can constrain enzyme evolution for successful adaptation to the host. With respect to the available cellular information for *Leishmania* species, for the first time, we measure the relative contribution of eight inter-correlated predictors related to codon usage, GC content, gene expression, gene length, multi-functionality, and flux-coupling potential of an enzyme on the evolutionary rates of singleton metabolic genes and further compare their effects across three *Leishmania* species. Our analysis reveals that codon adaptation, multi-functionality, and flux-coupling potential of an enzyme are independent contributors of enzyme evolutionary rates, which can together explain a large variation in enzyme evolutionary rates across species. We also hypothesize that a species-specific occurrence of duplicated genes in novel subcellular locations can create new flux routes through certain singleton flux-coupled enzymes, thereby constraining their evolution. A cross-species comparison revealed both common and species-specific genes whose evolutionary divergence was constrained by multiple independent factors. Out of these, previously known pharmacological targets and virulence factors in *Leishmania* were identified, suggesting their evolutionary reasons for being important survival factors to the parasite. All these results provide a fundamental understanding of the factors underlying adaptive strategies of the parasite, which can be further targeted.

Keywords *Leishmania* metabolism · Evolutionary rate variation · Codon usage · Multi-functionality · Physiological flux-coupling · Principal component regression (PCR)

Introduction

Metabolism is one of the primary biological processes that underlie the survival of an organism within a given environment, due to its fundamental role in synthesis of biomass and energy generation. Even though the individual metabolic enzymes per se are highly conserved across species, adaptation to diverse environments brings about novel innovations in metabolic pathway function (Szappanos et al. 2016). Numerous features like horizontal gene transfer, gene expression, gene dispensability, gene duplications, and metabolic network structure are responsible for changes in metabolic function (Yamada and Bork 2009; Papp et al. 2011). The dominance of one feature over another largely depends on the nature, variations in the environment, and the effective contribution of a factor towards successful adaptation to that particular environment. In general, the change in metabolic function due to changes in a feature can either be selected in a population for its usefulness in adaptation or else it can be purged, if deleterious. This change in function is reflected within the coding sequence of a gene and is conventionally measured by assessing the number of non-synonymous substitutions per non-synonymous site relative to the

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00239-018-9857-5) contains supplementary material, which is available to authorized users.

✉ Ram Rup Sarkar
rr.sarkar@ncl.res.in

[1] Chemical Engineering and Process Development, CSIR-National Chemical Laboratory, Pune, Maharashtra, India

[2] Academy of Scientific & Innovative Research (AcSIR), CSIR-NCL Campus, Pune, India

number of synonymous substitutions per synonymous site, commonly referred as the evolutionary rates (Yang 1998). However, the knowledge of potential determinants of evolutionary rates of a metabolic enzyme within an organism still remains to be an open, unsolved problem.

Members of the *Leishmania* genus cause the widespread neglected tropical disease leishmaniasis in humans. Biologically, the *Leishmania* parasite exhibits a digenetic lifecycle, where the promastigote stages thrive within the midgut of the sandfly vector, and the amastigotes persist in the macrophage phagolysosome of the human host; the environments being largely antagonistic with respect to pH, temperature, and availability of carbon sources (Zilberstein and Shapira 1994; McConville and Naderer 2011). To ensure maximal survival, the parasites need to selectively adapt to these dual environmental constraints. This controlled biological setup provides us with a unique platform for investigating the contributory role of different genotypic and phenotypic factors in metabolic enzyme evolution.

Numerous genotype and phenotype factors are known to contribute to evolutionary rate variation in eukaryotes. The factors that are known to have a probable effect on protein evolution largely falls into two categories, namely, translation selection and functional constraint. Translation selection refers to the evolutionary selection of features that can increase efficiency of translation, whereas functional constraint of an enzyme refers to the degree at which random mutations are removed from the population by natural selection so as to avoid their deleterious effect on protein function (Zhang and Yang 2015). With respect to features explaining translation selection, gene expression, mRNA transcript length (or length of a coding sequence), and codon usage were demonstrated as important factors that explain the evolution of protein-coding genes in yeast and *Arabidopsis* (Kawaguchi and Bailey-Serres 2005; Drummond et al. 2006; Zhang and Yang 2015). With respect to features explaining functional constraint, pleiotropy of a gene due to multiple functional domains, involvement of enzymes in multiple biological processes, and multiple gene duplications can contribute to enzyme evolution thereby providing a dynamism to the metabolic network structure (Salathé et al. 2005; Warringer and Blomberg 2006; Chu et al. 2014; Chesmore et al. 2016). Another less studied functional constraint that affects the evolution of a metabolic enzyme is the role of an enzyme in the context of other enzymes within a metabolic network. As metabolic function is a result of stepwise transformation and utilization of different environmental metabolites through multiple pathways, it is not the effect of a single enzyme. Hence, more central proteins within a metabolic network are also resistant to functional change (Vitkup et al. 2006). Previous studies in yeast and human erythrocytes have also demonstrated that enzymes bearing higher metabolic flux tend to evolve slowly (Vitkup et al. 2006; Colombo et al. 2014). It was also demonstrated that co-regulation in metabolic genes is largely explained by flux-coupling within a metabolic network (Notebaart et al. 2008) suggesting it to be an important factor constraining metabolic function, and hence enzyme evolution.

Similar to other organisms, a few studies in *Leishmania* species also provide indirect hints towards the roles of translation selection and functional constraints on metabolic enzyme evolution. Stage-specific transcriptomics and proteomics studies identify variations in transcriptome and proteome abundances of metabolic genes across stages and species in *Leishmania* (Lahav et al. 2011; Nirujogi et al. 2014). Also, mutation pressure and translation selection are shown to preserve codons within genes which possess a high GC bias at the synonymous position and avoid the formation of mRNA secondary structures at the 5′ end of the mRNA; thereby indicating probable modes of translation regulation within genes (Subramanian and Sarkar 2015). Chromosomal aneuploidy is another well-known mechanism that causes variations in gene copy numbers across *Leishmania* species (Mannaert et al. 2012). Recent computational predictions of metabolic flux for different input metabolites and targeted ^{13}C-based metabolomics studies have identified that the *Leishmania* metabolome adapts to changing host environments through common metabolic routes, which are largely constrained by the inherent metabolic organization (Saunders et al. 2014; Subramanian and Sarkar 2017). The inherent metabolic organization also constrains enzyme evolution in *L. major* metabolism (Subramanian and Sarkar 2016).

The aforementioned studies in *Leishmania* have largely explored the genotype and phenotype complements of metabolism independently. The combined effects of these features on the disparate forces of conservation and divergence in enzyme evolution are yet to be tested. To establish their effects on evolutionary rates among metabolic enzymes, a comprehensive comparative strategy that can examine the relative effects of the different genotype and phenotype features simultaneously is required. In this study, we estimate the rate of non-synonymous substitutions per non-synonymous site (d_N), rate of synonymous substitutions per synonymous site (d_S), and their ratio ($\omega = d_N/d_S$) and for the first time, identify the potential determinants of d_N, d_S, and ω among orthologous singleton metabolic genes in three *Leishmania* species (*Leishmania major, Leishmania donovani*, and *Leishmania infantum*) using a principal component regression (PCR)-based analysis (Drummond et al. 2006; Jovelin and Phillips 2009; Alvarez-Ponce and Fares 2012; Alvarez-Ponce et al. 2017). Although it is possible to use these features for assignment of genes to the three *Leishmania* species using Bayesian classifiers and other techniques (Wang et al. 2007), the above regression-based analysis appropriately suits our objective of discerning the relationships of the genotype and phenotype features to

evolutionary rates of metabolic genes and their comparisons across the three *Leishmania* species. We introduce the flux-coupling potential of an enzyme within a metabolic network (Subramanian and Sarkar 2016), as a potential feature for regression along with other available features for *Leishmania* metabolism. Despite the unavailability of broad range of confounding cellular factors that influence both codon usage and protein evolutionary rates (for example, UTR length, recombination rate, gene essentiality, protein–protein interactions features) for *Leishmania* species, the results provided in this article highlight the significant contribution of codon usage, multi-functionality, gene duplications, and flux-coupling constraints as novel mechanisms underlying evolutionary divergence and conservation in *Leishmania* metabolic genes. Comparisons of gene clusters across the three species demonstrate that the same gene can be constrained by different features and hence, a unique set of species-specific genes governed by multiple features can occur across species. The targetable mechanisms and genes identified in this study can be further perused for designing novel strategies against parasite persistence.

Materials and Methods

Potential Determinants of Metabolic Enzyme Evolutionary Rates

In this study, a total of eight features representing the genotype and phenotype characteristics of the *Leishmania* parasite were computed. *Leishmania* species with known, curated metabolic reconstructions, namely, the *L. major* strain Friedlin reconstruction comprising of the 560 metabolic genes, the *L. donovani* BPK282A1 reconstruction comprising 604 metabolic genes and the *L. infantum* JPCM5 reconstruction with 556 genes were used for multivariate analysis (Chavali et al. 2008; Sharma et al. 2017; Subramanian and Sarkar 2017).

Genomic Features

The coding nucleotide sequences (CDS) of the metabolic genes curated within each metabolic reconstruction, obtained from the TriTrypDB database, v.8.1, release 32 (Aslett et al. 2010) were used for calculation of codon adaptation index (CAI), GC content, and the gene length. CAI values for each gene were computed using the EMBOSS package (Rice et al. 2000), with respect to a reference set of ribosomal protein-coding genes in each species (Subramanian and Sarkar 2015). The length and GC content for each CDS were computed using an in-house PERL script (Sect. 8A of Supplementary Text S1).

Gene Expression

Pre-calculated Fragments per million kilobases (FPKM) values were obtained for *L. major* promastigotes from an independent RNA sequencing study (Rastrojo et al. 2013). To maintain consistency, the total number of reads mapped onto each gene, reported in the Gene Expression Omnibus database for *L. donovani* (GEO ID: GSE48475) and *L. infantum* (GSE48394) were used for calculation of Reads per million kilobases (RPKM) values of each gene (Martin et al. 2014; Zhang et al. 2014). FPKM and RPKM are considered to be synonymous within the article. Further details provided in Sect. 7A of Supplementary Text S1.

Functional Constraint

Number of Processes and Functions

As the curated annotation GO processes and function IDs still remain unavailable for all genes in the three *Leishmania* species, computed Gene Ontology (GO) processes and functions associated with each gene was extracted from the TriTrypDB database (Aslett et al. 2010). The number of predicted processes (NumProcs) and functions (NumFuncs) was calculated from this information using an in-house PERL code (Sect. 8B of Supplementary Text S1). Further details provided in Sect. 7E of Supplementary Text S1.

Flux-Coupling Potential of an Enzyme

In this study, we introduce the flux-coupling potential of an enzyme as a proxy for quantifying the flux-based functional constraint imposed on a metabolic enzyme. The flux-coupling potential is calculated by the centrality of an enzyme (degree or number of flux-couplings, NCoup) and the tendency of an enzyme to cluster together with other enzymes with similar number of flux-couplings (local clustering coefficient, CCoFCA), within a flux-coupled subgraph of the metabolic network. Further details provided in Sect. 7B of Supplementary Text S1.

Sequence-Based Evolutionary Rates

For the estimation of the evolutionary rates, multiple sequence alignment of each gene in all the three species was performed with its orthologous sequences across five genomes, namely, *L. major, L. infantum, L. donovani, L. mexicana*, and *L. braziliensis* species. This captures the degree of sequence divergence across closely related species within the *Leishmania* lineage. These five *Leishmania* species were chosen, as their genomes are completely sequenced and assembled. The orthology information was available within the TriTrypDB database, v.8.1, release 32

(Aslett et al. 2010). The alignment was processed to remove sequence positions with gaps using a standalone version of the PAL2NAL program (Subramanian and Sarkar 2016). d_N, d_S, and ω (d_N/d_S) were estimated using the one-ratio M0 branch model implemented in the 'codeml' subroutine of the PAML package version 4.8a (Yang 1998, 2007).

Pre-processing the Datasets for Multivariate Analysis

For each species, the dataset of metabolic genes was pre-processed to remove—(a) genes with obsolete sequences, less than 200 codons, $d_S > 0.3$, (b) duplicates and (c) genes for which either of the targeted genomic, expression, or metabolic network-based features was unavailable. Finally, only 233 singletons common to the three species of *Leishmania* was considered for multivariate analyses. Details behind extraction of singleton genes are provided in Sect. 7C of Supplementary Text S1.

Multivariate Analysis and Clustering

Principal Component Regression

Principal Component Regression (PCR) analysis on metabolic networks of the three *Leishmania* species was used to identify the potential contribution of the genomic, gene expression and function-based features to the total variance in evolutionary rates among metabolic genes. The 'pls' package version 2.6 implemented in R was used to perform PCR with d_N and d_S as the response and the aforementioned eight parameters as the predictor variables. A subset of predictor variables with loadings of 0.45 or more was considered for interpretation of a principal component with respect to that subset (Tabachnick and Fidell 2007). Further details provided in Sect. 7D of Supplementary Text S1.

Selection of Minimum Principal Components for Regression

A randomization test approach was used to check whether the squared prediction errors of regression models with fewer components are significantly ($P < 0.01$) larger than the reference model predicting absolute minimum prediction accuracy or not, by generating a distribution of prediction errors in each model for comparison using 1000 random permutations (van der Voet 1994). Out of these significant models, the model with least number of principal components was chosen as the best model to predict d_N and d_S in all three species. The randomization test approach is implemented within the 'pls' package.

K-Means Clustering

K-means clustering of genes was performed in an n-dimensional space, where n represents the selected number of principal components. Clustering was performed so as to identify the groups of genes, governed by a particular set of principal components and thereby a subset of predictors. The number of clusters represented in each dataset was determined by computing the Akaike's Information Criterion (Manning et al. 2008) for every K clusters (AIC); where $K = 1–100$. The number of clusters corresponding to the model with least AIC was considered to be representative for each dataset.

Results

Features Associated with Evolutionary Rates are Also Inter-correlated in *Leishmania* Species

Performing a pairwise correlation analysis for the orthologous metabolic genes in *Leishmania major* Friedlin, *Leishmania donovani* BPK282A1 and *Leishmania infantum* JPCM5, it was identified that there is no significant correlation obtained between d_N and d_S, whereas ω is obviously correlated with both d_N and d_S (Fig. 1, Sect. 1 of Supplementary Text S1). A significant correlation is observed between the codon adaptation index (CAI) and evolutionary rates in all species, suggesting an obvious association of translation selection and enzyme evolution (Fig. 1, Sect. 1 of Supplementary Text S1). In comparison, features representing functional constraints demonstrate relatively weak species-specific associations with d_N, d_S and ω. In *L. major* (Fig. 1a), d_N and ω are negatively correlated with number of processes in which a gene is involved (NumProcs), indicative of a weak functional constraint (d_N:$r = -0.161$; $P = 0.014$, ω:$r = -0.172$, $P = 0.008$). Similarly, the number of flux-couplings per reaction associated with a gene (NCoup) is significantly associated with ω ($r = -0.152$; $P = 0.02$). In *L. donovani* (Fig. 1b), d_N seems to weakly correlate with NumProcs ($r = -0.16592$; $P = 0.011$). In *L. infantum* (Fig. 1c), ω demonstrates a weak positive association with a gene's tendency to occur in a flux-coupled module (CCoFCA) ($r = 0.165$; $P = 0.011$). It seems apparent from the pairwise correlation-based analysis that, with an exception of CAI and GC content, each of the aforementioned features was weakly correlated with evolutionary rates across the three *Leishmania* species.

Apart from associations of the predictors with evolutionary rates, inter-correlations between predictors were also observed. As observed in a previous study (Subramanian and Sarkar 2015), CAI also correlates positively with GC content with varying strengths of associations in each species. GC content

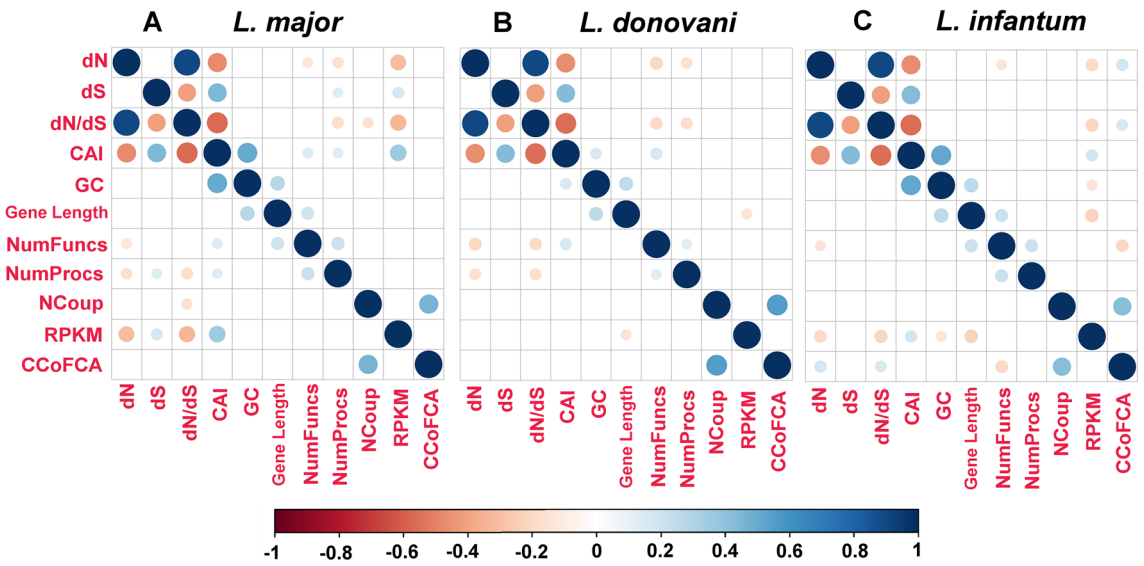

Fig. 1 Correlation dot plot demonstrating inter-correlations between the eight predictors and evolutionary rates for **a** *Leishmania major*, **b** *Leishmania donovani*, and **c** *Leishmania infantum*. This plot displays correlated pairs of features having significant correlation at $P<0.05$. Dots represent significant positive or negative correlations. Colors represent both the nature and degree of the association between any two features. The size of the dots represents the degree of the association between any two features. Pairwise correlation values are given in Sect. 1 of Supplementary Text S1. (Color figure online)

of a gene increases with larger gene lengths as indicated by their significant association across species (Fig. 1, Sect. 1 of Supplementary Text S1). In *L. major* and *L. donovani* (Fig. 1a, b), CAI of a gene is positively associated with NumFuncs (*L. major:* $r=0.145$, $P=0.026$; *L. donovani:* $r=0.173$, $P=0.008$) suggestive of multifunctional genes to contain more frequent codons. As popularly known, CAI correlates with mRNA abundance (measured in reads per million kilobases, RPKM) in *L. major* and *L. infantum* (Fig. 1a, c). In *L. donovani* and *L. infantum*, gene length and RPKM are negatively correlated suggesting expression of metabolic genes is probably limited by gene length in these species (Fig. 1b, c). Specifically in *L. infantum*, the number of functions associated with a gene (NumFuncs) demonstrates a weak negative association ($r=-0.20133$; $P=2\times10^{-3}$) with the tendency of a gene to cluster with genes demonstrating similar physiological fluxes (CCoFCA) hinting the role of multifunctional genes in routing fluxes within functional flux modules. The values of features for the selected genes in all three species are given in Supplementary File S1. This analysis further demonstrates that it is inappropriate to directly use these features to predict evolutionary rates of genes in *Leishmania* as they are not independent of each other.

Contribution of Features to the Variation Observed in Enzyme Evolutionary Rates

As indicated in Fig. 1, although many features are independently correlated with the evolutionary rates, some of them are also inter-correlated with each other. Hence, it is difficult to identify the potential contribution of each individual features to evolutionary rates. For this purpose, PCR was performed to identify independent principal components, which represent a linear combination of features, the coefficients representing the weight of a particular feature in explaining the variation in d_N, d_S, or ω (Drummond et al. 2006). The distribution statistics of evolutionary rates for the selected datasets is given in Sect. 2 of Supplementary Text S1. The identified principal components for the response d_N and d_S rates in the three *Leishmania* species are given in Supplementary File S2. PCR analysis with d_N and d_S in all three species indicates that the amount of variation explained by the principal components in the response variables (d_N and d_S) need not always be in descending order of the principal components (Jolliffe 1982). Additionally, it can also be observed that in most of the cases, a 90% variation in d_N and d_S, cannot be explained by considering only the first few components suggesting that no single factor dominates enzyme evolutionary rates. The pairwise correlation-based analysis fails to identify this observation, as only the effects of the strongest pairwise associations are highlighted. Furthermore, as there are inter-correlations among predictors, a combination of other related predictors probably outweighs the contribution of the apparent strongly associated codon usage/GC content features. Another important observation suggests that though the flux topological features explain a low variance in d_N, their occurrence within

the 1st principal component suggest that these features explain a majority of variation observed for metabolic genes in all three species.

With respect to d_N, it can be observed that the first two components (principal components 2, 3 of *L. major*, 2, 3 of *L. donovani* and 3, 7 of *L. infantum*), which cumulatively represent around 28.01% variance in *L. major* (Fig. 2a), 21.18% variance in *L. donovani* (Fig. 2d) and 23.41% variance in *L. infantum* (Fig. 2g) are dominated by genomic and gene expression features like CAI, GC, RPKM and gene length. In all the cases (Fig. 2a, d, g), the components dominated by flux-coupling potential and functional constraints explain a relatively small amount of variance in evolutionary rates. In *L. infantum* (Fig. 2g), a comparable amount of variation (7.92%) in d_N is explained by the principal components (1 and 8), which is dominated by metabolic flux-coupling potential of an enzyme, where the total variance explained in d_N by all the principal components is 38.21%.

With respect to d_S, it can be observed that the first two components (principal components 1, 8 of *L. major*, 2, 3 of *L. donovani* and 3, 8 of *L. infantum*), which cumulatively represent around 13.96% variance in *L. major* (Fig. 2b), 14.24% variance in *L. donovani* (Fig. 2e), and 21.37% variance in *L. infantum* (Fig. 2h) are dominated by genomic and gene expression features like CAI, GC, RPKM, and gene length. A relatively large amount of variance (7.2%) is also

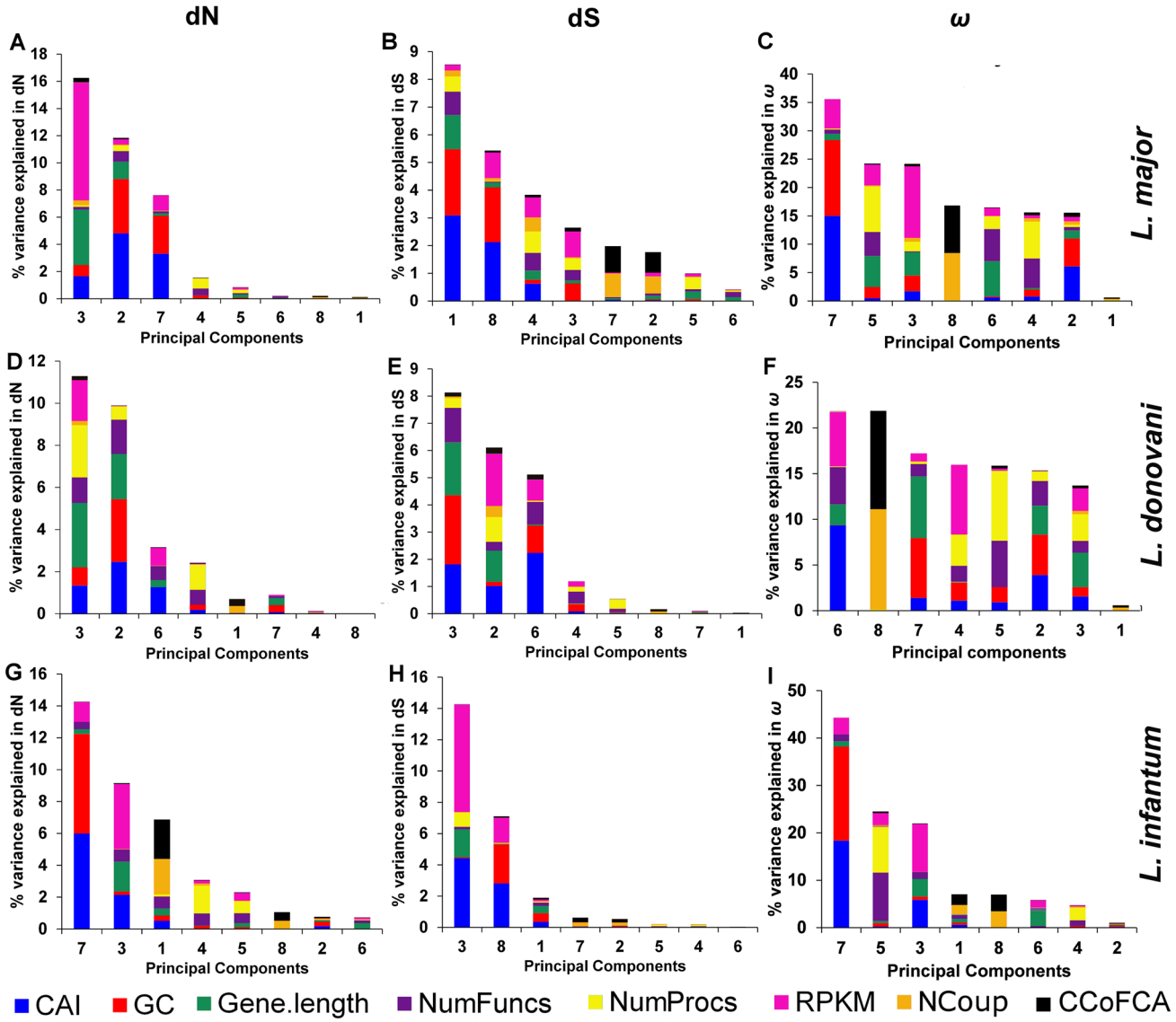

Fig. 2 Principal components regression on d_N (**a, d, g**), d_S (**b, e, h**), and ω (**c, f, i**) rates of 233 singleton orthologous metabolic genes in *L. major, L. donovani*, and *L. infantum* using eight different features. Each principal component represents a linear combination of the eight predictors, dominated by components that demonstrate a large variation in d_N and d_S. The colors correspond to the percentage variance explained by a particular feature, with respect to that principal component. (Color figure online)

explained in d_S rate of enzymes in *L. major* by two principal components governed by the flux-coupling potential, where the total variance explained in d_S by all the principal components is 25.6% (Fig. 2b).

As observed for the d_N and d_S rates, the largest percentage of the total variation in ω is explained by features related to translation selection (CAI, GC content), as indicated by the 6th or 7th principal components in all the three species (variations − 35.5% in *L. major*, 21.88% in *L. donovani*, and 44.33% in *L. infantum;* Fig. 2c, f, i). But, as observed for *L. major* and *L. donovani*, multi-functionality and flux topology features explain larger variations in ω as compared to their contributions to individual d_N and d_S rates (the heights of orange/black bars representing flux topology and purple/yellow representing multi-functionality are greater in ω as compared to d_N and d_S in all three species). An almost equal variation in ω is explained by flux-coupled features (NCoup and CCoFCA) in *L. donovani* (8th principal component − 21.86%, Fig. 2f). The second largest percentage of variance is explained by the variable related to multi-functionality in *L. major* (24.19%, Fig. 2c) and *L. infantum* (24.51%, Fig. 2i). Similar to the d_N and d_S rates, no single component is alone enough to explain more than 90% of the variation in ω.

Selection of Components for Predicting Enzyme Evolutionary Rates

A set of principal components were shortlisted for predicting evolutionary rates using a randomization test approach (see "Materials and Methods"). The principal components selected for regression are given in Sect. 3 of Supplementary Text S1. Features with loadings greater than 0.45 were considered for interpreting a principal component (Table 1). Most of the principal components explaining any variation in d_N or d_S can be interpreted on the basis of three distinct classes of features—(*a*) codon usage (CAI) and GC content, (*b*) multi-functionality (NumProcs, NumFuncs), and (*c*) flux phenotypic

Table 1 Contribution of the eight predictors to the selected principal components (loading cut-off > 0.45) and hence, the $\log_{10}(d_N)$ and $\log_{10}(d_S)$ rates in *L. major, L. donovani* and *L. infantum*

Component	$\log_{10}(d_N)$			$\log_{10}(d_S)$		
	L. major	*L. donovani*	*L. infantum*	*L. major*	*L. donovani*	*L. infantum*
1	NCoup (+), CCoFCA (+)	NCoup (+), CCoFCA (+)	NCoup (−), CCoFCA (−)	CAI (+), GC (+)	NCoup (+), CCoFCA (+)	GC (+), GeneLength (+)
	0.0059	0.0146	− 0.046***	0.02***	0.0013	0.0094*
2	CAI (−), GC (−)	CAI (−), GC (−), GeneLength (−)	CAI (+), GC (+)	NCoup (−), CCoFCA (−)	RPKM (−)	NCoup (−), CCoFCA (−)
	0.0641***	0.065***	− 0.017	− 0.0097*	− 0.0185***	− 0.0053
3	GeneLength (−), RPKM (+)	GeneLength (−), NumProcs (+)	CAI (−), GeneLength (+), RPKM (−)	–	CAI (+), GC (+), GeneLength (+)	CAI (+), RPKM (+)
	− 0.088***	− 0.073***	0.065***		0.0223***	0.029***
4	NumFuncs (+), NumProcs (+)	–	NumFuncs (−), NumProcs (−)	–	–	–
	− 0.0281*		0.04**			
5	GeneLength (+), NumProcs (−)	–	NumFuncs (+), NumProcs (−), RPKM (+)	–	–	–
	− 0.0244		− 0.042**			
6	–	–	GeneLength (−), NumFuncs (+)	–	–	–
			− 0.025			
7	–	–	CAI (−), GC (+)	–	–	–
			0.156***			
8	–	–	–	–	–	–

The positive and negative signs in brackets indicate the nature of their contributions to the principal component as demonstrated by the principal component loadings. The numbers below each combination of features indicates the regression coefficients associated with that principal component. The regression coefficients corresponding to each principal component were obtained after regressing the chosen principal components to the response evolutionary rates. *P* values of regression coefficients: ***$P < 0.001$, **$P < 0.01$, *$P < 0.05$. Genes with positive or negative scores with respect to a principal component correspond to the positive or negative contribution of features of those genes as indicated by the loadings on that component. The dash (–) in the table indicates that the corresponding principal component was not selected for regression, as identified by the randomization test approach

features (NCoup, CCoFCA). Most importantly, in all species (except *L. infantum*), effect of CAI and GC content of a gene on evolutionary rates can be interpreted by the same principal component suggesting their combinatorial effect in constraining d_N and d_S. To explain d_N rate of a gene, two principal components (2 and 7) involving CAI and GC content as principle features can be observed in *L. infantum*, where GC content negatively contributes to d_N in the 2nd principal component and positively contributes to d_N in the 7th principal component. Additionally, the 7th component has a relatively large role in explaining d_N as compared to the 2nd component. In all species, CAI negatively relates to d_N and positively relates to d_S. In all species, number of processes associated with a gene (NumProcs) negatively contributes to d_N. Further, no principal component can be interpreted solely on the basis of gene length, to explain both d_N and d_S.

Gene expression (RPKM) positively contributes to d_S rate in *L. donovani* and *L. infantum* and negatively contributes to d_N rate in *L. major* and *L. infantum*. In case of *L. major*, it can be seen that distinct principal components (2 and 3) can be interpreted using CAI and RPKM, respectively, suggesting weak associations with each other and their independent associations with d_N (Table 1). Most of these relationships corroborate with the pairwise correlation-based analysis performed above (Fig. 1).

To explain d_S in *L. donovani*, it can be seen that distinct principal components (3 and 2) can be interpreted using CAI and RPKM, respectively, suggesting their independent relationships with d_S and no association with each other (Table 1). On the contrary, in *L. infantum*, principal component 3 can be interpreted by both CAI and RPKM suggesting their inter-relatedness. Interestingly, an important observation points out that synonymous substitution rates are not constrained by the multifunctional potential of a gene (NumFuncs, NumProcs). Flux topological features significantly contribute to d_N rates of genes in *L. infantum* and d_S rates of genes in *L. major*. Patterns common to both d_N and d_S are observed with respect to the ω rate across the three *Leishmania* species (Sect. 4 of Supplementary Text S1). Features related to translation selection (CAI, GC, RPKM, gene length) demonstrate a significant association with ω in all the three species. In *L. major*, translation selection is the only factor affecting ω. Multi-functionality (NumFuncs, NumProcs) is significantly associated negatively with ω in *L. donovani* and *L. infantum*. Further, in *L. infantum*, the flux topological features (NCoup, CCoFCA) are also significantly associated with the ω rate.

Relationship Between Physiological Flux Coupling and Enzyme Evolutionary Rates

The pairwise correlation analysis indicated a weak correlation between flux-coupling features and evolutionary rates in *L. major* and *L. infantum* (Fig. 1). But, in the above analysis, it was found that across *Leishmania* species, physiological flux coupling potential seems to be a poor predictor of evolutionary rates (Table 1). This relationship between evolutionary rates and flux-coupling potential can be affected because certain enzymes demonstrate no flux coupling with other reactions within the network. Apart from explaining variations, PCR analysis also allows us to classify genes into two clusters, with respect to the contribution of the predictor features of the genes (interpreted through a principal component) to a response. It was observed that the potential of an enzyme to be physiologically coupled to other enzymes within metabolism or not can be classified only using scores of enzymes loaded on the first principal component (PC1) associated with the three evolutionary rates in all the species (Insets, Fig. 3a–i).

With respect to this coupled set of enzymes (cluster 1 in insets, Fig. 3a–i), a negative relationship is observed between d_N or ω and the number of couplings associated with an enzyme with varying strengths (Fig. 3). In all three species, no association was observed between d_S and number of couplings (Fig. 3b, e, h). With respect to the number of couplings, the association between d_N or ω and NCoup decreases as *L. major* < *L. donovani* < *L. infantum*. In *L. major* (Fig. 3a, c), the association, although weak, is statistically significant at $P < 0.01$ (d_N:$r = -0.252$, $P = 0.007$; ω:$r = -0.291$, $P = 0.002$). In *L. donovani* (Fig. 3d, f), the association is weaker than *L. major* (d_N:$r = -0.159$, $P = 0.094$, ω:$r = -0.198$, $P = 0.036$). In *L. infantum* (Fig. 3g, i), the association is the weakest and seems to be a purely chance phenomenon (d_N:$r = -0.019$, $P = 0.83$; ω:$r = -0.094$; $P = 0.29$). The associations become weaker from *L. major* to *L. infantum* due to the gain or loss of flux-couplings by enzymes across species. This gain or loss is affected by the coupling between duplicated and singleton genes in unique subcellular locations across species (Supplementary File S3). Furthermore, the number of flux-couplings observed for duplicated genes is much higher as compared to singletons (Supplementary File S3).

Hence, we asked the question whether gene duplications affect the relationship between d_N or ω and number of couplings associated with an enzyme or not? Comparing the distributions of number of couplings associated with duplicated enzymes in the three species revealed that most of the duplicated enzymes are coupled to a less number of other enzymes within the metabolic network of *Leishmania* species (Fig. 3j). But, the variance in the number of couplings of duplicated enzymes is notably higher in *L. major* with some duplicated enzymes displaying a large number of couplings. On the contrary, the variance drastically reduces in *L. donovani* and *L. infantum* as compared to *L. major*. Similar to duplicated enzymes, comparing the distributions of number of couplings associated with

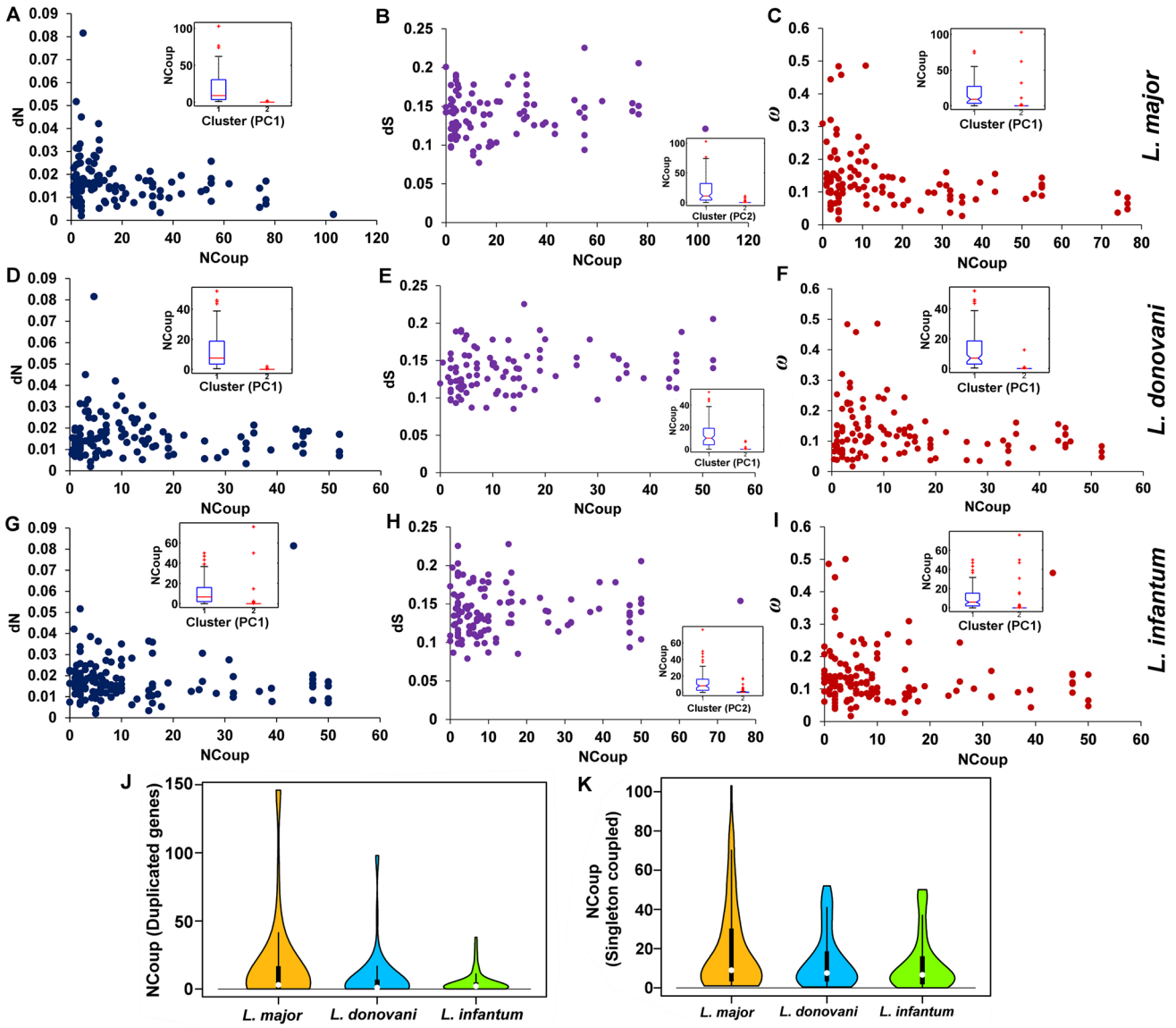

Fig. 3 Association between rates of protein evolution and number of couplings (NCoup) is affected by gene duplications. Relationship between d_N rates and NCoup of flux-coupled set of enzymes is given for **a** *L. major*; **d** *L. donovani;* and **g** *L. infantum*. Relationship between d_S rates and NCoup of flux-coupled set of enzymes is given for **b** *L. major*; **e** *L. donovani;* and **h** *L. infantum*. Relationship between ω and NCoup is given for **c** *L. major*; **f** *L. donovani*; and **i** *L. infantum*. **j** Violin plot demonstrating the differences in the variance of number of couplings associated with duplicated genes between *L. major* (median = 3), *L. donovani* (median = 1.03), and *L. infantum* (median = 2); **k** Violin plot demonstrating the differences in variance of singleton genes between *L. major* (median = 9), *L. donovani* (median = 7.55), and *L. infantum* (median = 6.64). Insets represent the two clusters of metabolic enzymes that are flux-coupled (1) and uncoupled (2)

coupled set of singleton enzymes in the three species also revealed that most of the singleton enzymes are coupled to a less number of other enzymes within the metabolic network of *Leishmania* species (Fig. 3k), with decreasing variance from *L. major* to *L. infantum*. This decreasing variance relates to the decreasing association of flux coupling potential with evolutionary rates of metabolic genes from *L. major* to *L. donovani* to *L. infantum* (Fig. 3a–i).

Comparing the variance in the number of flux-couplings across species in both the duplicated and singleton cases using Levene's test of homogeneity of variances (Martin and Bridgmon 2012) indicated that the variance in number of couplings significantly differs between species at $P < 0.001$ (duplicated: $F = 10.968$, $P = 3.25 \times 10^{-5}$, singletons: $F = 8.54$, $P = 2.6 \times 10^{-4}$). The similarity in distributions of number of couplings between duplicated enzymes and

singletons indicates that more gene duplications might indirectly create new flux coupling associations with singletons, under stoichiometry, reversibility, and environmental constraints, thereby promoting the association of the evolutionary rate with number of couplings associated with singleton genes. Furthermore, variance in number of couplings from *L. major* to *L. infantum* decreases at a slower rate in singletons as compared to duplicated enzymes indicating that the association between evolutionary rates and number of couplings in singletons is not promoted equally by all gene duplication events across species.

Identification of Metabolic Genes Constrained by Translation Selection, Multi-functionality, and Flux Topology

From Table 1, it is possible to identify principal components that can be interpreted by the independent features namely, CAI, Number of processes (NumProcs) and number of flux-couplings (NCoup) associated with a gene and the nature of their contributions to the evolutionary rates. Each of these features explains the role of translation selection, multi-functionality, and flux topology respectively on evolutionary rates of metabolic genes. Observing the centroids of the clusters (Sect. 5 of Supplementary Text S1, Supplementary File S4), the gene clusters that are associated with contributions of such principal components can be identified. Likewise, the d_N rate of genes in cluster numbers 4 and 14 in *L. major*, 9, 12, 13, 17 in *L. donovani* and 8, 18 in *L. infantum* are dominated by non-zero values of NCoup (positive scores on respective principal component in *L. major, L. donovani* and negative scores on respective principal component in *L. infantum*) and low values of NProcs (positive scores on principal component in *L. major, L. donovani* and negative scores on principal component in *L. infantum*) and CAI (negative scores on respective principal component in *L. major, L. donovani* and positive scores on respective principal component in *L. infantum*). Multi-functionality, which is represented by NumProcs or NumFuncs does not appear to be a dominant predictor in explaining the d_S rate and hence, does not occur as a major contributor in any of the selected components (Table 1). Hence, those gene clusters whose evolutionary rates can be interpreted by CAI and flux topology alone were identified. Likewise, the d_S rate of genes in cluster numbers 2, 3, 4, 7, 8, 11 in *L. major*, 1, 2, 3 10, 15, 18 in *L. donovani* and 2, 3, 10, 18 in *L. infantum* are associated with high values of CAI (positive scores on respective principal component in all three species) and NCoup (positive scores on respective principal component in *L. major* and *L. infantum* and negative scores on respective principal component in *L. donovani*). Comparison of chosen genes between the species indicates five genes in all species, whose evolutionary rates are dominated by all the three factors—translation selection, multi-functionality and flux topology, whereas 13 genes whose evolutionary rates are governed by translation selection and flux topology (Fig. 4).

There is a larger overlap of genes between the *L. major* and *L. donovani* species with respect to d_N as compared to d_S. Further, the overlap between *L. donovani* and *L. infantum* is restricted with respect to d_N as compared to d_S. In all species, there are also a unique set of genes whose evolutionary rates are specifically explained by the identified independent features (Fig. 4, Sect. 6 of Supplementary Text S1).

Discussion

Owing to its parasitic nature and the long-standing evolutionary association with hosts, *Leishmania* species experience a largely constrained metabolic environment. For

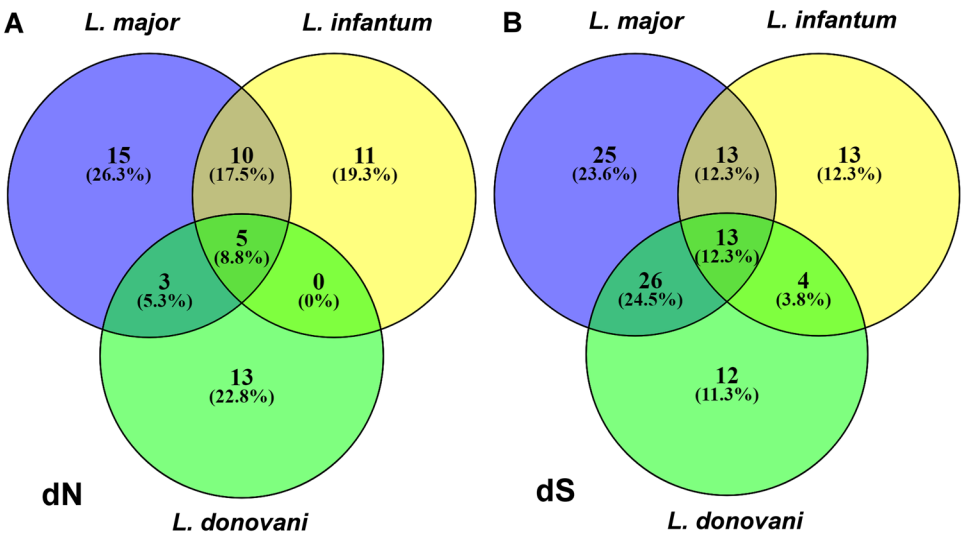

Fig. 4 Comparison of genes demonstrating high values of independent dominant factors namely, codon adaptation, number of biological processes, and number of flux-coupling associations between species with respect to **a** d_N and **b** d_S

efficient adaptation within the host, both translation selection and functional constraint might constrain evolution of enzymes within *Leishmania* metabolism. To our knowledge, there is no study available till date in *Leishmania* parasites that compares these heterogeneous potential determinants in predicting non-synonymous (d_N) and synonymous (d_S) substitution rates in metabolic enzymes simultaneously, on a single platform. Also, the inter-relationship between these factors and their differences across species is seldom explored. As used in other eukaryotes (Drummond et al. 2006; Yang and Gaut 2011; Alvarez-Ponce et al. 2017), the present study integrates the available, potential features of metabolic enzymes into a principal component-based regression model to identify the unknown confounding factors that explain observed variation in the evolutionary rates and compares them across three *Leishmania* species.

As observed in other eukaryotes (Drummond et al. 2006), codon usage negatively correlates with d_N, ω, and positively correlates with d_S in all species, signifying translation selection to be an important constraint in *Leishmania* metabolic enzyme evolution. This can also be observed from the highest percentage of variation explained by the principal component dominated by CAI. Furthermore, GC content also occurs as a dominating factor of the same principal component as CAI, indicating their relatedness, supporting previous observations (Subramanian and Sarkar 2015). But, as observed in all the three *Leishmania* species, neither a single principal component is enough to explain a significant proportion of variation among evolutionary rates nor does a single set of similar features explain sufficient variation across principal components, indicating that multiple features potentially contribute to enzyme evolution in *Leishmania* species. Hence, more than one principal component was observed to be selected for regression (van der Voet 1994). Although with an exception in *L. infantum*, results indicate that gene expression (RPKM) does not always occur in the same principal component as CAI, suggesting their independent roles in governing evolutionary rates of enzymes. This is contrary to the observations in yeast and *E. coli*, where gene expression complements CAI as a dominant factor governing evolutionary rates (Drummond et al. 2006). This also contrasts observations in *Trypanosoma brucei*, an evolutionary-related Trypanosomatid, where codon usage is demonstrated to affect global mRNA levels (Jeacock et al. 2018). This might be due to the weak association observed between mRNA and protein abundances in *Leishmania* species (Lahav et al. 2011); CAI being an important predictor of protein abundance (Subramanian and Sarkar 2015). Similarly, the occurrence of CAI, multi-functionality and flux-coupling features as dominant features on distinct principal components suggests that these features affect evolutionary rates independently. Further, the multi-functionality of a gene (NumProcs, NumFuncs) contributes only to the non-synonymous substitution rate (d_N) and is negatively associated with d_N. Hence, as observed in yeast (Salathé et al. 2005), genes (enzymes) with multiple processes or functions evolve slowly as compared to genes associated with low number of functions in the *Leishmania* species as well.

As the parasite stages live in fixed host environments, the pathways used to metabolize resources across stages remain strikingly similar (Subramanian and Sarkar 2017). Thus, enzymes (reactions) that are more coupled to other enzymes within the metabolic network might be constrained evolutionarily as opposed to enzymes that are less or not coupled to other enzymes. Hence, for the first time, we introduce the notion of the flux-coupling potential of an enzyme within its metabolic network and investigate whether it is an important determinant of evolutionary rate in *Leishmania* species or not. Although the associations of the flux-coupling features with evolutionary rates are weak, unlike multi-functionality, the occurrence of flux topological features in the first principal component and the selection of their associated principal component for regression against evolutionary rates explains their important contribution to variation in both d_N and d_S rates. Supporting this factor, a significant amount of variation in the d_S rate of enzymes in *L. major* and d_N rate of enzymes in *L. infantum* is also sufficiently explained by these features. Considering only the flux-coupled set of enzymes in all three species, a weak negative association can be observed between d_N, ω and number of couplings associated with an enzyme (NCoup). Flux-coupling reaction subsets capture the total number of paths of metabolite distribution under defined uptake constraints, as they can explain co-regulation between metabolic genes (Notebaart et al. 2008). A negative association was observed between ω and metabolic flux through an enzyme in yeast, human RBCs and *L. major* (Vitkup et al. 2006; Colombo et al. 2014; Subramanian and Sarkar 2016). This suggests that an enzyme is slow-evolving if it is coupled to large number of other enzymes by flux (hubs) within the flux-coupled network when compared to enzymes with low number of couplings. Further, few numbers of enzymes with high number of flux-couplings are observed as compared to enzymes with low number of flux-couplings. This indicates that a hierarchical organization of fluxes within *Leishmania* metabolism is largely constrained during evolution.

Chromosomal aneuploidy in *Leishmania* gives rise to significant variations in copy numbers of genes across species that might increase genomic plasticity, gene dosage, and rescue of essential functions from deleterious mutations (Mannaert et al. 2012). In addition to the aforementioned roles, for the first time, we document an observation indicating a possible species-specific involvement of duplicated metabolic enzymes in increasing the evolutionary constraints on other metabolic enzymes within a network, through re-wiring of

physiological flux dependencies within the metabolism. This is typically indicated by a higher variance in the number of couplings associated with singleton and duplicated enzymes and relatively stronger associations between number of couplings associated with singletons and evolutionary rates. With decrease in the variance of number of couplings of duplicated enzymes from *L. major* → *L. donovani* → *L. infantum*, the strength of associations between number of couplings and evolutionary rates also reduces. A similar re-wiring of fluxes due to cross-compartmentalized metabolism was also hypothesized for glycolysis and isoprenoid biosynthesis in other Trypanosomatids (close evolutionary relatives of *Leishmania*) and other protists (Ginger et al. 2010). Interestingly, not all gene duplications are highly flux-coupled with other enzymes in the network, suggesting that the species-specific metabolic network structure dynamically constrains the choice of unique gene duplications occurring at multiple subcellular locations for flux re-wiring, thereby imposing evolutionary constraints on other singletons associated with them.

Previously, codon bias, pleiotropy, and centrality within a biomolecular network were implicated to impose relatively strong evolutionary constraints on enzymes that are important pharmacological targets for a disease (Searls 2003; Pál et al. 2006; Gladki et al. 2013; Lv et al. 2016). As mentioned above, codon adaptation, multi-functionality, and flux topological constraints independently affect evolutionary rates; each of these features being negatively associated with d_N. Comparison of genes with the d_N rate dominated by these factors leads to the identification of both common and species-specific enzymes, which are evolutionarily constrained by multiple genotype–phenotype factors, reckoning them to be important enzymes. Likewise, this analysis was able to identify enzymes like trypanothione reductase, aspartate carbamoyltransferase, orotidine-5-phosphate decarboxylase, and dihydrolipoamide dehydrogenase common to all three species. Among the enzymes common to the three *Leishmania* species, trypanothione reductase, the sole enzyme in the *Leishmania* parasite to combat oxidative stress (Tovar et al. 1998), aspartate carbamoyltransferase and orotidine-5-phosphate decarboxylase, involved in production of pyrimidines, like ump and cmp, (Mukherjee et al. 1988; Bello et al. 2007) are previously speculated pharmacological targets in *Leishmania* and other eukaryotes. On the other hand, unique enzymes majorly belonging to energy metabolism and conservation (C), Carbohydrate transport and metabolism (G), Amino acid transport and metabolism (E), and Nucleotide transport and metabolism (F) were also identified for each species (Sect. 6 of Supplementary Text S1). Among these unique enzymes, known virulence factors like trypanothione synthetase, phosphomannose isomerase and GDP-mannose pyrophosphorylase were specifically identified for *L. major;* dihydrofolate-reductase/thymidylate synthase, pyrroline-5-carboxylate reductase and phosphomannomutase were identified for *L. infantum* and tyrosine aminotransferase for *L. donovani* (Mukherjee et al. 1988; Titus et al. 1995; Tovar et al. 1998; Garami and Ilg 2001b, a; Scott et al. 2008; Moreno et al. 2014; Mantilla et al. 2015). Their role in virulence probably makes them more resistant to change. From this analysis, few more novel species-specific enzymes were also predicted (Tables G, H, Sect. 6 of Supplementary Text S1). These can be used as potential drug targets because they are governed by unique evolutionary constraints. Their biological role in virulence, survival or visceralization of the parasite needs to be experimentally investigated.

Although the results provided here are limited by the unavailability of genome-scale metabolic networks for multiple known species, strains, and isolates of *Leishmania* (Cantacessi et al. 2015), the use of such comprehensive multivariate analyses in teasing apart the known confounding factors of enzyme evolution provides a broad insight into the organization of *Leishmania* metabolism and the underlying factors governing its change. Additionally, this work also provides a multitude of hypotheses that can be tested experimentally in *Leishmania*. Furthermore, identification of the role of multiple factors in constraining evolutionary divergence within metabolic enzymes suggests that the survival and adaptation of the parasite within the host are a complex problem. This emphasizes the need for systems-level experiments to identify other features, like UTR length, recombination rate, gene essentiality, protein–protein interactions features, etc. unavailable at an organismal level for *Leishmania* species and to analyze their integrated effect. The integration of these diverse features can thus provide the complete knowledge of the strategies employed by the parasite for survival and virulence, which can help the community to combat this largely neglected tropical parasitic infection.

Conclusion

For the first time, we measure the relative contribution of eight inter-correlated genotype, phenotype predictors on the evolutionary rates of singleton metabolic genes and further compare them across three *Leishmania* species. Codon usage, multi-functionality, and flux-coupling potential of an enzyme independently constrain evolution of metabolic genes in *Leishmania*. This seems to be a unique feature of *Leishmania* metabolic evolution which was previously not reported. Our observations suggest that occurrence of duplicated genes in novel subcellular locations can create new species-specific flux routes through certain singleton flux-coupled enzymes, thereby constraining their evolution. This observation asserts the role of gene duplications in contributing to evolutionary innovations of *Leishmania* metabolism.

Our results reveal that although *Leishmania* metabolic genes are very similar with respect to their sequence information, the systems-level function of metabolic genes can affect metabolic enzyme evolution. The unique and common enzymes identified for all the three species from our analysis were previously reported to govern important biological roles for *Leishmania* metabolism and virulence. Moreover, some of these were pharmacological targets experimentally reported for related *Leishmania* species. Unique enzymes whose evolutionary rates are affected by a high contribution of dominating factors can explain species-specificity and the reasons for within-host adaptation. Most importantly, these might be perused as mechanisms to be targeted for in vivo control or as important causes of parasite visceralization.

Acknowledgements This work was supported by a Grant from the Department of Biotechnology, Government of India [BT/PR14958/BID/7/537/2015] provided to RRS. AS also acknowledges the Senior Research Fellowship from DBT-BINC. The authors are thankful to the anonymous reviewers for their critical comments and suggestion to improve the quality of the paper.

References

Alvarez-Ponce D, Fares MA (2012) Evolutionary rate and duplicability in the *Arabidopsis thaliana* protein-protein interaction network. Genome Biol Evol 4:1263–1274. https://doi.org/10.1093/gbe/evs101

Alvarez-Ponce D, Feyertag F, Chakraborty S (2017) Position matters: network centrality considerably impacts rates of protein evolution in the human protein–protein interaction network. Genome Biol Evol 9:1742–1756. https://doi.org/10.1093/gbe/evx117

Aslett M, Aurrecoechea C, Berriman M et al (2010) TriTrypDB: a functional genomic resource for the Trypanosomatidae. Nucleic Acids Res 38:D457–D462. https://doi.org/10.1093/nar/gkp851

Bello AM, Poduch E, Fujihashi M et al (2007) A potent, covalent inhibitor of orotidine 5'-monophosphate decarboxylase with antimalarial activity. J Med Chem 50:915–921. https://doi.org/10.1021/jm060827p

Cantacessi C, Dantas-Torres F, Nolan MJ, Otranto D (2015) The past, present, and future of *Leishmania* genomics and transcriptomics. Trends Parasitol 31:100–108. https://doi.org/10.1016/j.pt.2014.12.012

Chavali AK, Whittemore JD, Eddy JA et al (2008) Systems analysis of metabolism in the pathogenic trypanosomatid *Leishmania major*. Mol Syst Biol 4:177. https://doi.org/10.1038/msb.2008.15

Chesmore KN, Bartlett J, Cheng C, Williams SM (2016) Complex patterns of association between pleiotropy and transcription factor evolution. Genome Biol Evol 8:3159–3170. https://doi.org/10.1093/gbe/evw228

Chu S, Wang J, Cheng H et al (2014) Evolutionary study of the isoflavonoid pathway based on multiple copies analysis in soybean. BMC Genet 15:1–12. https://doi.org/10.1186/1471-2156-15-76

Colombo M, Laayouni H, Invergo BM et al (2014) Metabolic flux is a determinant of the evolutionary rates of enzyme-encoding genes. Evolution 68:605–613. https://doi.org/10.1111/evo.12262

Drummond DA, Raval A, Wilke CO (2006) A single determinant dominates the rate of yeast protein evolution. Mol Biol Evol 23:327–337. https://doi.org/10.1093/molbev/msj038

Garami A, Ilg T (2001a) The role of phosphomannose isomerase in *Leishmania mexicana* glycoconjugate synthesis and virulence. J Biol Chem 276:6566–6575. https://doi.org/10.1074/jbc.M009226200

Garami A, Ilg T (2001b) Disruption of mannose activation in *Leishmania mexicana*: GDP-mannose pyrophosphorylase is required for virulence, but not for viability. EMBO J 20:3657–3666. https://doi.org/10.1093/emboj/20.14.3657

Ginger ML, McFadden GI, Michels PAM (2010) Rewiring and regulation of cross-compartmentalized metabolism in protists. Philos Trans R Soc B 365:831–845. https://doi.org/10.1098/rstb.2009.0259

Gladki A, Kaczanowski S, Szczesny P, Zielenkiewicz P (2013) The evolutionary rate of antibacterial drug targets. BMC Bioinform. https://doi.org/10.1186/1471-2105-14-36

Jeacock L, Faria J, Horn D (2018) Codon usage bias controls mRNA and protein abundance in trypanosomatids. Elife 7:e32496

Jolliffe IT (1982) A note on the use of principal components in regression. Appl Stat. https://doi.org/10.2307/2348005

Jovelin R, Phillips PC (2009) Evolutionary rates and centrality in the yeast gene regulatory network. Genome Biol 10:R35. https://doi.org/10.1186/gb-2009-10-4-r35

Kawaguchi R, Bailey-Serres J (2005) mRNA sequence features that contribute to translational regulation in *Arabidopsis*. Nucleic Acids Res 33:955–965. https://doi.org/10.1093/nar/gki240

Lahav T, Sivam D, Volpin H et al (2011) Multiple levels of gene regulation mediate differentiation of the intracellular pathogen *Leishmania*. FASEB J 25:515–525. https://doi.org/10.1096/fj.10-157529

Lv W, Xu Y, Guo Y et al (2016) The drug target genes show higher evolutionary conservation than non-target genes. Oncotarget 7:4961–4971. https://doi.org/10.18632/oncotarget.6755

Mannaert A, Downing T, Imamura H, Dujardin J-C (2012) Adaptive mechanisms in pathogens: universal aneuploidy in *Leishmania*. Trends Parasitol 28:370–376. https://doi.org/10.1016/j.pt.2012.06.003

Manning CD, Raghavan P, Schütze H (2008) Introduction to information retrieval. Cambridge University Press, Cambridge

Mantilla BS, Paes LS, Pral EMF et al (2015) Role of Δ1-pyrroline-5-carboxylate dehydrogenase supports mitochondrial metabolism and host-cell invasion of *Trypanosoma cruzi*. J Biol Chem 290:7767–7790. https://doi.org/10.1074/jbc.M114.574525

Martin WE, Bridgmon KD (2012) Quantitative and statistical research methods: from hypothesis to results. Wiley, Hoboken

Martin JL, Yates PA, Soysa R et al (2014) Metabolic reprogramming during purine stress in the protozoan pathogen *Leishmania donovani*. PLoS Pathog 10:e1003938. https://doi.org/10.1371/journal.ppat.1003938

McConville MJ, Naderer T (2011) Metabolic pathways required for the intracellular survival of *Leishmania*. Annu Rev Microbiol 6:543–561. https://doi.org/10.1146/annurev-micro-090110-102913

Moreno MA, Alonso A, Alcolea PJ et al (2014) Tyrosine aminotransferase from *Leishmania infantum*: a new drug target candidate. Int J Parasitol Drugs Drug Resist 4:347–354. https://doi.org/10.1016/j.ijpddr.2014.06.001

Mukherjee T, Ray M, Bhaduri A (1988) Aspartate transcarbamylase from *Leishmania donovani*. A discrete, nonregulatory enzyme as a potential chemotherapeutic site. J Biol Chem 263:708–713

Nirujogi RS, Pawar H, Renuse S et al (2014) Moving from unsequenced to sequenced genome: reanalysis of the proteome of *Leishmania donovani*. J Proteom 97:48–61. https://doi.org/10.1016/j.jprot.2013.04.021

Notebaart RA, Teusink B, Siezen RJ, Papp B (2008) Co-regulation of metabolic genes is better explained by flux coupling than by network distance. PLoS Comput Biol 4:e26. https://doi.org/10.1371/journal.pcbi.0040026

Pál C, Papp B, Lercher MJ (2006) An integrated view of protein evolution. Nat Rev Genet 7:337–348. https://doi.org/10.1038/nrg1838

Papp B, Notebaart RA, Pál C (2011) Systems-biology approaches for predicting genomic evolution. Nat Rev Genet 12:591–602. https://doi.org/10.1038/nrg3033

Rastrojo A, Carrasco-Ramiro F, Mart'\in D et al (2013) The transcriptome of *Leishmania major* in the axenic promastigote stage: transcript annotation and relative expression levels by RNA-sEq. BMC Genom 14:223. https://doi.org/10.1186/1471-2164-14-223

Rice P, Longden I, Bleasby A et al (2000) EMBOSS: the European molecular biology open software suite. Trends Genet 16:276–277. https://doi.org/10.1016/S0168-9525(00)02024-2

Salathé M, Ackermann M, Bonhoeffer S (2005) The effect of multifunctionality on the rate of evolution in yeast. Mol Biol Evol 23:721–722. https://doi.org/10.1093/molbev/msj086

Saunders EC, Ng WW, Kloehn J et al (2014) Induction of a stringent metabolic response in intracellular stages of *Leishmania mexicana* leads to increased dependence on mitochondrial metabolism. PLoS Pathog 10:e1003888

Scott DA, Hickerson SM, Vickers TJ, Beverley SM (2008) The role of the mitochondrial glycine cleavage complex in the metabolism and virulence of the protozoan parasite *Leishmania major*. J Biol Chem 283:155–165. https://doi.org/10.1074/jbc.M708014200

Searls DB (2003) Pharmacophylogenomics: genes, evolution and drug targets. Nat Rev Drug Discov 2:613. https://doi.org/10.1038/nrd1152

Sharma M, Shaikh N, Yadav S et al (2017) A systematic reconstruction and constraint-based analysis of *Leishmania donovani* metabolic network: identification of potential antileishmanial drug targets. Mol Biosyst 13:955–969. https://doi.org/10.1039/c6mb00823b

Subramanian A, Sarkar RR (2015) Comparison of codon usage bias across *Leishmania* and Trypanosomatids to understand mRNA secondary structure, relative protein abundance and pathway functions. Genomics 106:232–241. https://doi.org/10.1016/j.ygeno.2015.05.009

Subramanian A, Sarkar RR (2016) Network structure and enzymatic evolution in *Leishmania* metabolism: a computational study. In: BIOMAT 2015: Proceedings of the international symposium on mathematical and computational biology, p 1. https://doi.org/10.1142/9789813141919_0001

Subramanian A, Sarkar RR (2017) Revealing the mystery of metabolic adaptations using a genome scale model of *Leishmania infantum*. Sci Rep 7:10262. https://doi.org/10.1038/s41598-017-10743-x

Szappanos B, Fritzemeier J, Csörg\Ho B et al (2016) Adaptive evolution of complex innovations through stepwise metabolic niche expansion. Nat Commun 7:11607. https://doi.org/10.1038/ncomms11607

Tabachnick BG, Fidell LS (2007) Using multivariate statistics, 5th edn. Allyn and Bacon, New York

Titus RG, Gueiros-Filho FJ, de Freitas LA, Beverley SM (1995) Development of a safe live *Leishmania* vaccine line by gene replacement. Proc Natl Acad Sci 92:10267–10271. https://doi.org/10.1073/pnas.92.22.10267

Tovar J, Wilkinson S, Mottram JC, Fairlamb AH (1998) Evidence that trypanothione reductase is an essential enzyme in *Leishmania* by targeted replacement of the tryA gene locus. Mol Microbiol 29:653–660

van der Voet H (1994) Comparing the predictive accuracy of models using a simple randomization test. Chemom Intell Lab Syst 25:313–323. https://doi.org/10.1016/0169-7439(94)85050-X

Vitkup D, Kharchenko P, Wagner A (2006) Influence of metabolic network structure and function on enzyme evolution. Genome Biol 7:R39. https://doi.org/10.1186/gb-2006-7-5-r39

Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol 73:5261–5267

Warringer J, Blomberg A (2006) Evolutionary constraints on yeast protein size. BMC Evol Biol 6:61. https://doi.org/10.1186/1471-2148-6-61

Yamada T, Bork P (2009) Evolution of biomolecular networks—lessons from metabolic and protein interactions. Nat Rev Mol Cell Biol 10:791–803. https://doi.org/10.1038/nrm2787

Yang Z (1998) Synonymous and nonsynonymous rate variation in nuclear genes of mammals. J Mol Evol 46:409–418. https://doi.org/10.1007/PL00006320

Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586–1591. https://doi.org/10.1093/molbev/msm088

Yang L, Gaut BS (2011) Factors that contribute to variation in evolutionary rate among *Arabidopsis* genes. Mol Biol Evol 28:2359–2369. https://doi.org/10.1093/molbev/msr058

Zhang J, Yang J-R (2015) Determinants of the rate of protein sequence evolution. Nat Rev Genet 16:409. https://doi.org/10.1038/nrg3950

Zhang WW, Ramasamy G, McCall L-I et al (2014) Genetic analysis of *Leishmania donovani* tropism using a naturally attenuated cutaneous strain. PLoS Pathog 10:e1004244. https://doi.org/10.1371/journal.ppat.1004244

Zilberstein D, Shapira M (1994) The role of pH and temperature in the development of *Leishmania* parasites. Annu Rev Microbiol 48:449–470