

# Asymmetric relationships between proteins shape genome evolution

Richard A Notebaart<sup>✉</sup>, Philip R Kensche<sup>✉</sup>, Martijn A Huynen and Bas E Dutilh

Address: Center for Molecular and Biomolecular Informatics, Nijmegen Center for Molecular Life Sciences, Radboud University Nijmegen Medical Center, Geert Grooteplein 26-28, 6525 GA, Nijmegen, The Netherlands.

✉ These authors contributed equally to this work.

Correspondence: Martijn A Huynen. Email: huynen@cmbi.ru.nl

Published: 12 February 2009

Genome **Biology** 2009, **10**:R19 (doi:10.1186/gb-2009-10-2-r19)

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2009/10/2/R19>

Received: 20 November 2008

Revised: 28 January 2009

Accepted: 12 February 2009

© 2009 Notebaart et al.; licensee BioMed Central Ltd.

This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Abstract

---

**Background:** The relationships between proteins are often asymmetric: one protein (A) depends for its function on another protein (B), but the second protein does not depend on the first. In metabolic networks there are multiple pathways that converge into one central pathway. The enzymes in the converging pathways depend on the enzymes in the central pathway, but the enzymes in the latter do not depend on any specific enzyme in the converging pathways. Asymmetric relations are analogous to the “if->then” logical relation where A implies B, but B does not imply A (A->B).

**Results:** We show that the majority of relationships between enzymes in metabolic flux models of metabolism in *Escherichia coli* and *Saccharomyces cerevisiae* are asymmetric. We show furthermore that these asymmetric relationships are reflected in the expression of the genes encoding those enzymes, the effect of gene knockouts and the evolution of genomes. From the asymmetric relative dependency, one would expect that the gene that is relatively independent (B) can occur without the other dependent gene (A), but not the reverse. Indeed, when only one gene of an A->B pair is expressed, is essential, is present in a genome after an evolutionary gain or loss, it tends to be the independent gene (B). This bias is strongest for genes encoding proteins whose asymmetric relationship is evolutionarily conserved.

**Conclusions:** The asymmetric relations between proteins that arise from the system properties of metabolic networks affect gene expression, the relative effect of gene knockouts and genome evolution in a predictable manner.

---

## Background

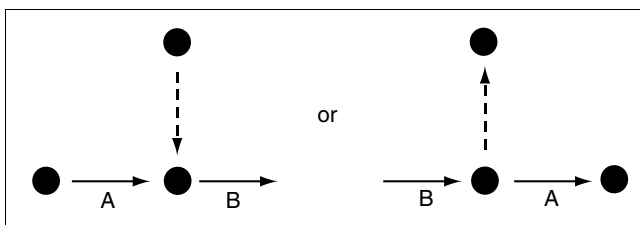
Cellular processes can only be fully understood by considering how the functions of proteins depend upon each other.

The relationship between two proteins can be symmetric - for example, when they mutually depend upon each other for their function within a protein complex. Proteins can also be

asymmetrically related. This occurs when the function of one protein (A) depends on another protein (B), but the function of protein B does not depend on A:  $A \rightarrow B$ . For example, in regulatory interactions, the function of the regulator depends on the presence of its target, but the target can often function without the regulator. Examples of asymmetrical relationships also exist in metabolism. For instance, multiple enzymes may produce the same substance (Figure 1), creating a situation in which the function of the proteins in the converging reaction fluxes (A) depends on the flux through B, but the function of B does not specifically depend on one of the converging fluxes. With the availability of accurate stoichiometric models of entire metabolic networks, it has become possible to infer symmetric and asymmetric coupling of reaction fluxes, not only at short metabolic distances, but throughout the complete network [1]. Asymmetrically coupled fluxes, when related to *in vivo* flux measures, do not exhibit a complete correlation (that is, symmetry) [2], and are much more frequent than the symmetric fully coupled fluxes (see below). Here we examine whether the asymmetric dependencies between proteins, as predicted from models of the complete metabolism of species at steady-state, are reflected in several genomic observables: which protein is expressed without the other, which is more essential than the other for survival or growth, which occurs in different genomes without the other and, finally, which is gained or lost without the other in evolution. To address these questions, we combined the dependencies of all reaction pairs in the metabolic networks of *Escherichia coli* [3] and *Saccharomyces cerevisiae* [4] with genome scale data sets for gene expression [5], gene essentiality [6,7], growth defects [8], and phylogenetic distribution [9].

## Results and discussion

Most coupled reaction pairs have an asymmetric dependency (that is, directional coupling): 82% in *Saccharomyces cerevisiae* [4] and 67% in the metabolic network of *Escherichia coli* [3] (see Materials and methods). As these asymmetric relations are so abundant in metabolism, we asked whether this



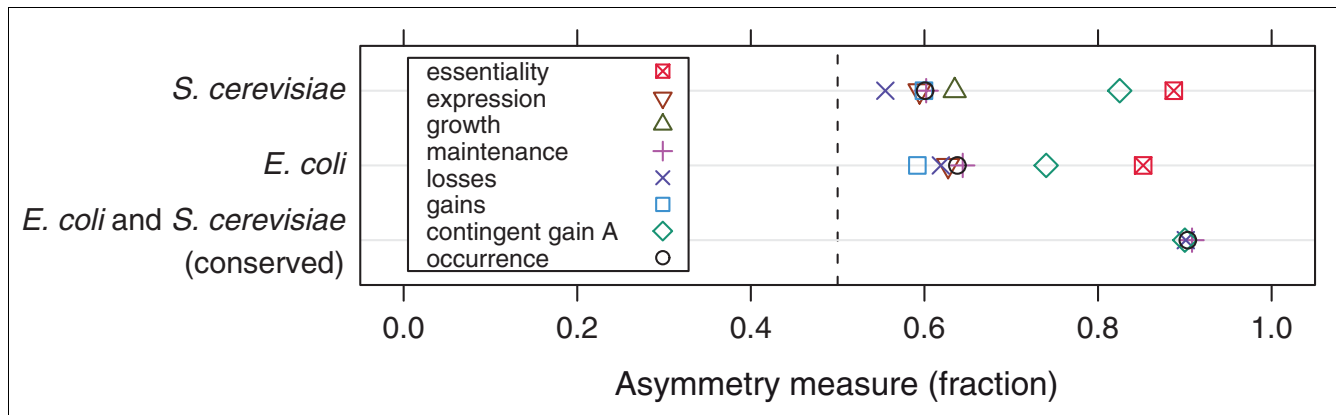
**Figure 1**  
Simple examples of asymmetric relationship between reactions A and B ( $A \rightarrow B$ ). Nodes and arrows indicate metabolites and metabolic reactions, respectively. At steady-state the activity (that is, carrying a flux) of reaction A depends on the activity of B, but the activity of B is independent of the activity of A, because there is an alternative converging or diverging flux (dashed arrows).

characteristic is also reflected in other system properties of the cell. Given an asymmetrically coupled reaction pair  $A \rightarrow B$  where A depends on B, but B does not depend on A (Figure 1), we expect that if one of the two reactions is inactive, it is most likely reaction A. To test this, we compared the asymmetric reaction pairs in the metabolic networks of *E. coli* and *S. cerevisiae* with four main types of genome scale data in which genes can be 'present' or 'absent'.

We first assessed the asymmetry in the lethality [6,7] and condition-specific growth defects [8] of gene knockouts. In an  $A \rightarrow B$  situation, we expect that if only one of the two genes is essential or affects growth, this will be the B gene: in the absence of gene A, a flux may still flow through the reaction catalyzed by protein (gene) B, but without B, A cannot function. Indeed, we find that for 87% of the  $A \rightarrow B$  pairs, in which one of the genes is essential, B is the essential gene (Figure 2; McNemar test; *S. cerevisiae*,  $n = 417$ ; *E. coli*,  $n = 331$ ;  $p < 10^{-36}$ ). The result for the condition-specific growth defects of non-essential  $A \rightarrow B$  pairs is less pronounced, but still for 64% of the conditions, the loss of B causes a greater growth defect than the loss of gene A (Figure 2; two-sided Wilcoxon test; *S. cerevisiae*,  $n = 141$ ;  $p < 2 \times 10^{-3}$ ).

We also find a consistency of the asymmetric relations with gene expression patterns. Because gene A depends for its function on gene B, there should be few conditions where A is expressed without B, relative to situations where B is expressed without A. As expected, the B gene is expressed in 61% of the conditions where only one of two asymmetrically related genes is expressed (Figure 2; *S. cerevisiae*,  $n = 573$ ; *E. coli*,  $n = 1,166$ ;  $p < 10^{-6}$ ). In conclusion, these analyses show that asymmetric relations between metabolic enzymes are reflected in system properties of the specific organisms.

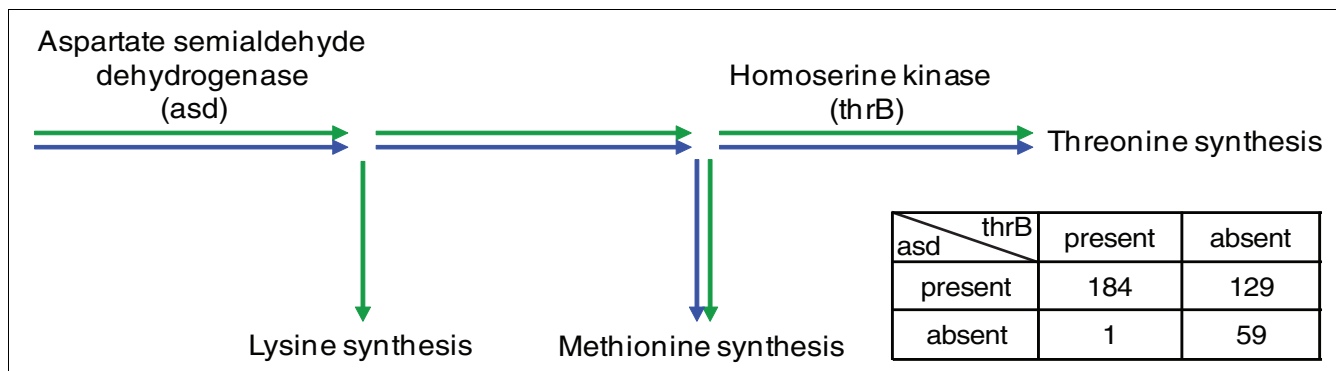
Next, we asked whether the asymmetric relations between enzymes are also reflected in evolution. Generally, functionally interacting proteins tend to co-occur across genomes [10,11]. This raises the question of whether the asymmetric relation of reactions is also reflected in the evolution of genomes. Although asymmetrically linked enzymes tend to co-occur [3], if only one of the two enzymes is absent from a genome, we expect this to be enzyme A: as A depends on the function of B, it will rarely be present in genomes where B is absent. To test this, we analyzed the phylogenetic distribution of all *E. coli* and *S. cerevisiae*  $A \rightarrow B$  pairs across 373 species [9]. Indeed, gene A is the absent gene in 62% of the species where one of the two genes is absent (Figure 2; two-sided Wilcoxon test; *E. coli*,  $n = 1,225$ ; *S. cerevisiae*,  $n = 2,242$ ;  $p \approx 0$ ). Besides asymmetry in the occurrence of genes in present day species, we also expect asymmetry in the gains and losses across evolutionary history. We inferred the occurrence of A and B in their ancestors by maximum parsimony [12]. In line with our expectations, gene A is more frequently lost (59%) in cases where a presence of both A and B in the ancestor was followed by a loss of either A or B (Figure 2; *E. coli*,  $n = 1,215$ ;



**Figure 2**  
Asymmetrically linked reaction pairs (A→B) related to asymmetry in gene essentiality, growth defects, gene expression and genome evolution. The fraction ( $f_{0/1} = n_{0/1}/(n_{0/1} + n_{1/0})$ ) where only B is essential in rich medium (essentiality) or has an effect on the growth across conditions (growth), where only B is expressed across conditions (expression), where only B is present across species (occurrence), where only B is present after gain, loss or maintenance over evolutionary lineages, and where A is contingently gained over evolutionary lineages (contingent gain A) is averaged over all reaction pairs (see Materials and methods). For conserved pairs there is no relevant result on gain, because too few (n = 2) events were found.

*S. cerevisiae*, n = 1,423;  $p < 10^{-7}$ ). Gene B is more often gained (60%) in cases where an absence of both A and B in the ancestor was followed by a gain of either A or B (*E. coli*, n = 605; *S. cerevisiae*, n = 1,449;  $p < 10^{-6}$ ). It is also expected that a gain of A depends on the presence of B (contingent evolution [13]). Indeed, a gain of gene A occurs more often when B is present (78%; *E. coli*, n = 824; *S. cerevisiae*, n = 1,472;  $p \approx 0$ ) than when B is absent (see Materials and methods). Finally, there are also situations where a presence of only one gene in the ancestor is maintained along the evolutionary lineage (that is, neither of genes A or B were gained or lost). As expected, maintenance of A absent and B present was found more frequently than the reverse (62%; *E. coli*, n = 1,223; *S. cerevisiae*, n = 2,230;  $p \approx 0$ ).

Although the various genomic and phylogenetic properties correlate significantly with the asymmetric relationships in the metabolic networks of *E. coli* or *S. cerevisiae*, exceptions remain where gene A is present while gene B is not. How can this be explained? For phylogenetic presence/absence patterns, one explanation for these irregularities is species-specific differences in metabolism. For example, the large scale replacement of amino-acid biosynthetic pathways by amino acid importers in *Thermofilum pendens* [14] has led to a situation where aspartate semialdehyde dehydrogenase (asd), one of the basal enzymes for amino-acid synthesis, is absent while homoserine kinase (thrB), which depends on asd, is still present (Figure 3). To examine such cases with unexpected phylogenetic occurrence systematically, we listed all asym-



**Figure 3**  
The asymmetric relationship between asd and thrB, two proteins conserved between *E. coli* (green) and *S. cerevisiae* (blue), is reflected in their asymmetric phylogenetic distributions. The activity of asd does not depend on thrB while the activity of thrB does depend on asd. Although in most cases both enzymes are present or absent together (243), thrB is more frequently absent while asd is present (129) than vice versa (1). The exception to the pattern comes from *Thermofilum pendens*, a species that has lost a large number of amino acid biosynthetic pathways, and imports most of its amino acids [14]. Note that a second asymmetric reaction pair between asd and the initial enzyme in the lysine synthesis pathway, present in *E. coli*, is not conserved in *S. cerevisiae*.

metrically dependent reaction pairs that lost gene A but not gene B in at least five monophyletic species (the expected pattern), and also lost gene B but not gene A in at least five monophyletic species (the unexpected pattern). Species with both genes present or both genes absent were allowed in both partitions (Additional data file 1). Some of these cases indeed reflect a change of metabolism, such as ubiquinone synthesis, which, in a species like *S. cerevisiae*, depends on the tryptophan biosynthesis pathway, while in *Homo sapiens* tryptophane is part of the diet and tryptophan biosynthesis has been lost but ubiquinone synthesis has been conserved. In most cases of unexpected loss, however, B has been replaced by a non-orthologous functional equivalent. Thus, the metabolic dependency of reaction A on B as identified in our reference metabolism may have remained intact, but the protein catalyzing B has changed. We also found cases of multiple functional specificities in orthologous group A, corresponding to a different substrate specificity of A in the species where B was lost, relative to the reference species *E. coli* or *S. cerevisiae* (Additional data file 1).

Even when genes and reactions are conserved across evolution, the nature of their relation can vary among species, as it depends on the overall functional and metabolic capabilities of the organism. Such variations could reduce the extent of asymmetry in the phylogenetic distribution. If this is the case, we expect to find a stronger correlation for genes with a conserved asymmetric dependency between the distantly related species *E. coli* and *S. cerevisiae* (see Figure 3 for an example). Indeed, we find a stronger correlation between the asymmetry in metabolism and the asymmetry in genomic occurrence across present day species and ancestral states if we consider reaction pairs with a conserved asymmetric relationship ( $n = 16$ ) between the two studied networks (approximately 90%; Figure 2). Nevertheless, this set of conserved reactions has few exceptions to the predicted asymmetry which, like the exceptions above, can be explained by differences in the metabolism between species (Additional data file 2).

Having established that asymmetric dependencies derived from the metabolic networks are reflected in both species-specific system properties and evolution, we asked whether this correlation could simply be an effect of local network topology rather than the complete metabolism. We defined network distance between two reactions in the network as the minimal number of metabolites that separate them. For all the genomic properties studied, we find in most cases that the asymmetry is actually more pronounced at larger (non-trivial) network distances ( $d \geq 4$ ), with a fraction ranging from 56% to 99% (Additional data file 3). This shows that the asymmetric dependencies are not simply an effect of local network topology.

## Conclusion

We show here that the relationships between proteins that arise from their functional dependencies can have an important influence on other elements of the biological system. The analysis of relationships between genes has so far focused on symmetric relations, including correlated and anticorrelated phylogenetic distributions of genes, and on higher order logic [10,11,15,16]. Our findings underline the relevance of asymmetric binary relationships between proteins, such as those that can be inferred from metabolic networks, to explain the evolution and functioning of the system. We demonstrate that asymmetric flux relations between enzymes are more abundant than symmetric relations. Furthermore, we show that this asymmetry is reflected in gene expression, gene essentiality and the evolution of genomes, even for proteins at large metabolic distances. Our results suggest a potential to predict asymmetric functional relations between proteins on the basis of genomic data.

## Materials and methods

### Flux coupling analysis

Flux coupling [1] between reactions within the genome-scale metabolic networks of *E. coli* K12 (*iJR904* GSM/GPR) [17] and *S. cerevisiae* *iLL672* [18] was based on two recent studies [3,4]. Flux coupling relies on minimization and maximization of flux ratios ( $R_{\min}$  = lowest possible  $v_A/v_B$  ratio and  $R_{\max}$  = highest possible  $v_A/v_B$ ) to determine the dependency between reaction A and B within the network (at steady-state [19]), given mass-balance constraints and flux capacity constraints (range of possible flux values; see also [1] for details).

In this study we mainly investigated the most abundant type of flux coupling, referred to as directional coupling (asymmetric dependency): the activity (flux) of one reaction (A) implies the activity of the other (B), but not necessarily the reverse ( $A \rightarrow B$ ,  $R_{\min} = 0$  and  $R_{\max} = \text{finite value}$ ). These reactions are coupled, but may not always operate together. In contrast, in fully coupled pairs (symmetric dependency) the activity of one reaction implies the activity of the other and *vice versa* ( $R_{\min} = R_{\max} = \text{finite value}$ ). Calculations were done without assuming a constant biomass composition to avoid coupling of a large set of fluxes to the biomass reaction. All biomass components were allowed to be drained independently of one another (see [1,2] for details). Directional coupling between reactions was computed at a condition where all external nutrients were allowed for uptake and secretion (via capacity constraints on the exchange fluxes with environment) [3,4].

### Network distance

Network distances ( $d$ ) were calculated by representing the network as a directed graph consisting of nodes (metabolites) and edges (reactions), and applying a shortest path algorithm. Distances correspond to the minimal number of nodes that separate any two reactions in the network. To increase

the functional relevance of network distance, we removed the most highly connected nodes, including ATP, ADP, AMP, CO<sub>2</sub>, CoA, glutamate, H, NAD, NADP, NADH, NADPH, H<sub>2</sub>O, NH<sub>3</sub>, phosphate, and pyrophosphate [20].

We grouped directionally coupled pairs (A→B) into two network distance groups - close network distance ( $d < 4$ ) and non-trivial distance ( $d \leq 4$ ) - to investigate whether the identified asymmetric relations are independent of network distance. Our conclusions are not affected by the exact distance cutoff between small and large network distance (Additional data file 3).

### Gene essentiality

Essentiality data for *S. cerevisiae* was obtained from the MIPS (Munich Information Center for Protein Sequences) database [7] (gene disruption table, 14-11-2005). Only essentiality information that referred to an original publication was retained, that is, database entries with a PubMed ID. If a gene was classified as both essential and non-essential by different sources, we assigned essentiality according to a majority rule and if no decision was possible, we marked the gene as ambiguous. For *E. coli*, we used the gene essentiality determined by Gerdes *et al.* [6]. We analyzed the essentiality on the level of reactions, using the gene-reaction associations as defined in each metabolic model. Reactions can be catalyzed by complexes of multiple enzymes (subunits linked by 'AND' in the model). Only if all subunits of an enzyme complex were essential did we consider the reaction essential. Conversely, only if all subunits were non-essential was the reaction considered non-essential. Otherwise, reactions were discarded. Reactions can also be catalyzed by iso-enzymes (linked by 'OR' in the model). If the individual iso-enzymes are classified as non-essential in single knockout experiments, it is still possible that the reaction is essential, because the loss of one iso-enzyme can be compensated by the other iso-enzymes. For this reason, we did not consider reactions with iso-enzymes. We summarized the combinations of essentiality and non-essentiality of all directionally coupled reactions in a  $2 \times 2$  contingency table and tested for its symmetry by a McNemar test as implemented in R [21].

### Growth defects of gene knockouts

We used the condition-specific growth data of Hillenmeyer *et al.* [8] restricted to measurements at generation 5 of homozygous strains (12 conditions including dropouts of adenine, arginine, isoleucine, lysine, threonine, tryptophan, or tyrosine, as well as YP glycerol, minimal, sorbitol, synthetic complete media). We used the empirical  $p$ -values published by Hillenmeyer and co-workers [8] to derive binary profiles of significant (1) and insignificant (0) growth defects. To obtain unique  $p$ -values for every gene and condition, we calculated the geometric mean over batches, pools and scanners. A growth defect was considered significant if this average  $p$ -value was  $< 10^{-3}$ . The mapping from gene to reaction level was done in the same way as for the essentiality data (see above).

Subsequently, for each reaction pair A→B with a corresponding pair of growth effect profiles we calculated the fraction ( $f_{o/1}$ ) of conditions in which reaction A showed no growth effect while reaction B did ( $n_{o/1}$ ), relative to the total number of conditions in which only one of the reactions showed a growth effect ( $n_{o/1} + n_{1/o}$ ). We tested the distribution of these fractions against the null-hypothesis that there is no bias, that is, no asymmetry ( $H_0: f_{o/1} = 0.5$ ), with the two-sided one-sample Wilcoxon test as implemented in R [21]. We averaged the calculated fractions over all pairs. For this and all other datasets, our results were qualitatively the same if we summarized the distribution as the mean or as the fraction of reaction pairs with a  $f_{o/1} > 0.5$ .

### Gene expression

The expression data were based on 13 studies with 327 conditions for *S. cerevisiae* and 12 studies with 420 conditions for *E. coli* (Additional data file 4). These data were obtained from the Gene Expression Omnibus (GEO) [5] at the National Center for Biotechnology Information (NCBI). Presence (expressed)/absence (not expressed) calls were made using the BioConductor affy package [22]. For each experimental condition, the presence/absence calls of individual genes were translated into 'presence/absence calls' of reactions based on the gene-reaction associations. Reactions that were catalyzed by multiple enzymes (iso-enzymes or subunits; see above) were considered present if at least one of the iso-enzymes or all subunits of enzyme complexes were present. For each reaction pair A→B with a corresponding pair of expression profiles, we calculated the fraction ( $f_{o/1}$ ) of conditions in which reaction A is absent while reaction B is present ( $n_{o/1}$ ) relative to the total number of conditions in which only one of the reactions is present ( $n_{o/1} + n_{1/o}$ ). We tested the distribution of these fractions against the null-hypothesis that there is no bias - that is, no asymmetry ( $H_0: f_{o/1} = 0.5$ ) - with the two-sided one-sample Wilcoxon test as implemented in R [21].

### Reaction-level phylogenetic profiles and ancestral state reconstruction

We constructed phylogenetic profiles that denote the presence and absence of enzymes across 373 species according to the STRING 7.0 orthologous groups [9]. To explore the presence and absence of reactions across species, we mapped the enzyme orthology information to the reactions-level using the gene-reaction associations. In situations of iso-enzymes, we considered the reaction present in a species if at least one iso-enzyme was present. If a reaction was catalyzed by an enzyme that had multiple subunits, it was considered present in a species only if all these subunits were encoded in the genome. For each reaction pair A→B with a corresponding pair of 'reaction-level' phylogenetic profiles, we calculated the fraction ( $f_{o/1}$ ) of genomes in which reaction A is absent while reaction B is present ( $n_{o/1}$ ) relative to the total number of genomes in which exactly one of the reactions is present ( $n_{o/1} + n_{1/o}$ ). We tested the distribution of these fractions against the null-

hypothesis that there is no bias - that is, no asymmetry ( $H_0: f_{0/1} = 0.5$ ) - with the two-sided one-sample Wilcoxon test as implemented in R [21].

We inferred the most parsimonious ancestral presence/absence states of A and B using a phylogenetic tree of all 373 species included in this analysis (this tree contained some multifurcations to account for uncertainties [9]) and PAUP [12]. The tree was manually rooted at the trifurcation of eukaryotes, Eubacteria and Archaea. All results were based on a gain/loss cost ratio of 2/1 [23] and a delayed transition assumption ('DELTRAN'). Importantly, varying the parameters did not affect our conclusions.

We examined for each reaction pair A→B the following situations: type i, both reactions are absent in the ancestor and one is gained in the descendent; type ii, both reactions are present in the ancestor and one is lost in the descendent; type iii, the presence of exactly one of the reactions is maintained, that is, no change of state occurs. We calculated the fraction ( $f_{0/1}$ ) where B was gained ( $n_{0/1}$ , type i) and where A was lost ( $n_{0/1}$ , type ii) or maintained ( $n_{0/1}$ , type iii) relative to the total number of instances of that type (that is,  $n_{0/1} + n_{1/0}$ ). We tested the distribution of these fractions (over all AB pairs) against the null-hypothesis as mentioned above.

To analyze contingent gain of A, we determined for all gain events of A whether B was already present in the ancestor or not. The fraction of gains in presence of B (over all AB pairs) was tested against the null hypothesis that a gain of A is independent of the presence of B (that is,  $H_0: f_{\text{gain}} \text{ of A in presence of B} = 0.5$ ).

### Conserved directionally coupled reaction pairs

We considered a reaction to be conserved between *S. cerevisiae* and *E. coli* if it was catalyzed by orthologous enzymes. In the case of iso-enzymes we required that at least one orthologous enzyme was present in both organisms. For reactions catalyzed by enzyme complexes, we required that orthologs of all subunits were present in both organisms. The deviation of the asymmetry in gene gain, loss and maintenance was tested as discussed in the section 'Reaction-level phylogenetic profiles and ancestral state reconstruction'.

The absolute number of conserved directionally coupled pairs is limited ( $n = 16$ ) because conservation of directional coupling required: both genes of a pair to be present in *S. cerevisiae* and *E. coli*; the type of coupling to be conserved; and the directionality (A→B) to be conserved.

### Authors' contributions

BD, RN, PK, MH conceived and designed the study. RN, BD and PK performed the analyses. RN, PK, MH and BD wrote the manuscript.

### Additional data files

The following additional data are available with the online version of this paper. Additional data file 1 is a table listing asymmetrically dependent reaction pairs A→B for which the independent gene B was lost while gene A was retained ('AB = 10') and vice versa ('AB = 01'), both in at least five species. Additional data file 2 is a figure that shows an exception to the predicted genomics occurrence of two enzymes. Additional data file 3 is a figure that shows asymmetrically linked reaction pairs (A→B) related to asymmetry in gene essentiality, growth defects, gene expression and phylogenetic distribution for which the pairs are categorized according to network distance cutoffs. Additional data file 4 contains two tables listing *Saccharomyces cerevisiae* [24-34] and *Escherichia coli* [35-44] expression datasets.

### Acknowledgements

We thank Balázs Papp, Berend Snel and Bas Teusink for suggestions on the manuscript and we thank the anonymous reviewers for their useful comments. This work was supported by: The BioRange programme of The Netherlands Bioinformatics Centre (NBIC), supported by a BSIK grant through The Netherlands Genomics Initiative (NGI); The Kluiver Centre for Genomics of Industrial Fermentation; The European Union's 6th Framework Program, contract number LSHB-CT-2005-019067 (EPITEM); The Dutch Science Foundation (NWO) Horizon Project 050-71-058.

### References

- Burgard AP, Nikolaev EV, Schilling CH, Maranas CD: **Flux coupling analysis of genome-scale network reconstructions.** *Genome Res* 2004, **14**:301-312.
- Notebaart RA, Teusink B, Siezen RJ, Papp B: **Co-regulation of metabolic genes is better explained by flux coupling than by network distance.** *PLoS Comput Biol* 2008, **4**:e26.
- Pal C, Papp B, Lercher MJ: **Adaptive evolution of bacterial metabolic networks by horizontal gene transfer.** *Nat Genet* 2005, **37**:1372-1375.
- Bundy JG, Papp B, Harmston R, Browne RA, Clayson EM, Burton N, Reece RJ, Oliver SG, Brindle KM: **Evaluation of predicted network modules in yeast metabolism using NMR-based metabolite profiling.** *Genome Res* 2007, **17**:510-519.
- Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, Evangelista C, Kim IF, Soboleva A, Tomashevsky M, Edgar R: **NCBI GEO: mining tens of millions of expression profiles - database and tools update.** *Nucleic Acids Res* 2007, **35**:D760-765.
- Gerdes SY, Scholle MD, Campbell JW, Balazsi G, Ravasz E, Daugherty MD, Somera AL, Kyrpides NC, Anderson I, Gelfand MS, Bhattacharya A, Kapatral V, D'Souza M, Baev MV, Grechkin Y, Mseeh F, Fonstein MY, Overbeek R, Barabasi AL, Oltvai ZN, Osterman AL: **Experimental determination and system level analysis of essential genes in Escherichia coli MGI655.** *J Bacteriol* 2003, **185**:5673-5684.
- Mewes HW, Amid C, Arnold R, Frishman D, Guldener U, Mannhaupt G, Munsterkotter M, Pagel P, Strack N, Stumpflen V, Warfsmann J, Ruepp A: **MIPS: analysis and annotation of proteins from whole genomes.** *Nucleic Acids Res* 2004, **32**:D41-44.
- Hillenmeyer ME, Fung E, Wildenhain J, Pierce SE, Hoon S, Lee W, Proctor M, St Onge RP, Tyers M, Koller D, Altman RB, Davis RW, Nislow C, Giaever G: **The chemical genomic portrait of yeast: uncovering a phenotype for all genes.** *Science* 2008, **320**:362-365.
- von Mering C, Jensen LJ, Kuhn M, Chaffron S, Doerks T, Kruger B, Snel B, Bork P: **STRING 7 - recent developments in the integration and prediction of protein interactions.** *Nucleic Acids Res* 2007, **35**:D358-362.
- Marcotte EM, Pellegrini M, Ng HL, Rice DW, Yeates TO, Eisenberg D: **Detecting protein function and protein-protein interactions from genome sequences.** *Science* 1999, **285**:751-753.
- Huynen MA, Bork P: **Measuring genome evolution.** *Proc Natl*

- Acad Sci USA* 1998, **95**:5849-5856.
12. Swofford DL: *PAUP\*: Phylogenetic Analysis Using Parsimony (\*and Other Methods)*. Version 4 Sunderland, Massachusetts: Sinauer Associates; 2003.
  13. Barker D, Pagel M: **Predicting functional gene links from phylogenetic-statistical analyses of whole genomes.** *PLoS Comput Biol* 2005, **1**:e3.
  14. Anderson I, Rodriguez J, Susanti D, Porat I, Reich C, Ulrich LE, Elkins JG, Mavromatis K, Lykidis A, Kim E, Thompson LS, Nolan M, Land M, Copeland A, Lapidus A, Lucas S, Detter C, Zhulin IB, Olsen GJ, Whitman W, Mukhopadhyay B, Bristow J, Kyrpides N: **Genome sequence of *Thermofilum pendens* reveals an exceptional loss of biosynthetic pathways without genome reduction.** *J Bacteriol* 2008, **190**:2957-2965.
  15. Bowers PM, Cokus SJ, Eisenberg D, Yeates TO: **Use of logic relationships to decipher protein network organization.** *Science* 2004, **306**:2246-2249.
  16. Morett E, Korbel JO, Rajan E, Saab-Rincon G, Olvera L, Olvera M, Schmidt S, Snel B, Bork P: **Systematic discovery of analogous enzymes in thiamin biosynthesis.** *Nat Biotechnol* 2003, **21**:790-795.
  17. Reed JL, Vo TD, Schilling CH, Palsson BO: **An expanded genome-scale model of *Escherichia coli* K-12 (ijR904 GSM/GPR).** *Genome Biol* 2003, **4**:R54.
  18. Kuepfer L, Sauer U, Blank LM: **Metabolic functions of duplicate genes in *Saccharomyces cerevisiae*.** *Genome Res* 2005, **15**:1421-1430.
  19. Varma A, Palsson BO: **Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110.** *Appl Environ Microbiol* 1994, **60**:3724-3731.
  20. Kharchenko P, Church GM, Vitkup D: **Expression dynamics of a cellular metabolic network.** *Mol Syst Biol* 2005, **1**:2005.0016.
  21. Team RdC: *R: A Language and Environment for Statistical Computing* Vienna, Austria: R Foundation for Statistical Computing; 2007.
  22. **Bioconductor affy** [<http://www.bioconductor.org/packages/2.0/bioc/html/affy.html>]
  23. Snel B, Bork P, Huynen MA: **Genomes in flux: the evolution of archaeal and proteobacterial gene content.** *Genome Res* 2002, **12**:17-25.
  24. Yarragudi A, Parfrey LW, Morse RH: **Genome-wide analysis of transcriptional dependence and probable target sites for Abf1 and Rap1 in *Saccharomyces cerevisiae*.** *Nucleic Acids Res* 2007, **35**:193-202.
  25. Singh J, Kumar D, Ramakrishnan N, Singhal V, Jervis J, Garst JF, Slaughter SM, DeSantis AM, Potts M, Helm RF: **Transcriptional response of *Saccharomyces cerevisiae* to desiccation and rehydration.** *Appl Environ Microbiol* 2005, **71**:8752-8763.
  26. Sabet N, Volo S, Yu C, Madigan JP, Morse RH: **Genome-wide analysis of the relationship between transcriptional regulation by Rpd3p and the histone H3 and H4 amino termini in budding yeast.** *Mol Cell Biol* 2004, **24**:8823-8833.
  27. Hochwagen A, Wrobel G, Cartron M, Demougin P, Niederhauser-Wiederkehr C, Boselli MG, Primig M, Amon A: **Novel response to microtubule perturbation in meiosis.** *Mol Cell Biol* 2005, **25**:4767-4781.
  28. Schawalder SB, Kabani M, Howald I, Choudhury U, Werner M, Shore D: **Growth-regulated recruitment of the essential yeast ribosomal protein gene activator Ifh1.** *Nature* 2004, **432**:1058-1061.
  29. Pitkanen JP, Torma A, Aiff S, Huopaniemi L, Mattila P, Renkonen R: **Excess mannose limits the growth of phosphomannose isomerase PMI40 deletion strain of *Saccharomyces cerevisiae*.** *J Biol Chem* 2004, **279**:55737-55743.
  30. Ronald J, Akey JM, Whittle J, Smith EN, Yvert G, Kruglyak L: **Simultaneous genotyping, gene-expression measurement, and detection of allele-specific expression with oligonucleotide arrays.** *Genome Res* 2005, **15**:284-291.
  31. Takagi Y, Masuda CA, Chang WH, Komori H, Wang D, Hunter T, Joazeiro CA, Kornberg RD: **Ubiquitin ligase activity of TFIID and the transcriptional response to DNA damage.** *Mol Cell* 2005, **18**:237-243.
  32. Guan Q, Zheng W, Tang S, Liu X, Zinkel RA, Tsui KW, Yandell BS, Culbertson MR: **Impact of nonsense-mediated mRNA decay on the global expression profile of budding yeast.** *PLoS Genet* 2006, **2**:e203.
  33. Kresnowati MT, van Winden WA, Almering MJ, ten Pierick A, Ras C, Knijnenburg TA, Daran-Lapujade P, Pronk JT, Heijnen JJ, Daran JM: **When transcriptome meets metabolome: fast cellular responses of yeast to sudden relief of glucose limitation.** *Mol Syst Biol* 2006, **2**:49.
  34. Yu C, Palumbo MJ, Lawrence CE, Morse RH: **Contribution of the histone H3 and H4 amino termini to Gcn4p- and Gcn5p-mediated transcription in yeast.** *J Biol Chem* 2006, **281**:9755-9764.
  35. Dong T, Kirchhof MG, Schellhorn HE: **RpoS regulation of gene expression during exponential growth of *Escherichia coli* K12.** *Mol Genet Genomics* 2008, **279**:267-277.
  36. Zoetendal EG, Smith AH, Sundset MA, Mackie RI: **The BaeSR two-component regulatory system mediates resistance to condensed tannins in *Escherichia coli*.** *Appl Environ Microbiol* 2008, **74**:535-539.
  37. Wang L, Li J, March JC, Valdes JJ, Bentley WE: **luxS-dependent gene regulation in *Escherichia coli* K-12 revealed by genomic expression profiling.** *J Bacteriol* 2005, **187**:8350-8360.
  38. Lee J, Page R, Garcia-Contreras R, Palermino JM, Zhang XS, Doshi O, Wood TK, Peti W: **Structure and function of the *Escherichia coli* protein YmgB: a protein critical for biofilm formation and acid-resistance.** *J Mol Biol* 2007, **373**:11-26.
  39. Lee J, Jayaraman A, Wood TK: **Indole is an inter-species biofilm signal mediated by SdiA.** *BMC Microbiol* 2007, **7**:42.
  40. Reigstad CS, Hultgren SJ, Gordon JL: **Functional genomic studies of uropathogenic *Escherichia coli* and host urothelial cells when intracellular bacterial communities are assembled.** *J Biol Chem* 2007, **282**:21259-21267.
  41. Faith JJ, Hayete B, Thaden JT, Mogno I, Wierzbowski J, Cottarel G, Kasif S, Collins JJ, Gardner TS: **Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles.** *PLoS Biol* 2007, **5**:e8.
  42. Hayes ET, Wilks JC, Sanfilippo P, Yohannes E, Tate DP, Jones BD, Radmacher MD, BonDurant SS, Slonczewski JL: **Oxygen limitation modulates pH regulation of catabolism and hydrogenases, multidrug transporters, and envelope composition in *Escherichia coli* K-12.** *BMC Microbiol* 2006, **6**:89.
  43. Maurer LM, Yohannes E, BonDurant SS, Radmacher M, Slonczewski JL: **pH regulates genes for flagellar motility, catabolism, and oxidative stress in *Escherichia coli* K-12.** *J Bacteriol* 2005, **187**:304-319.
  44. Covert MW, Knight EM, Reed JL, Herrgard MJ, Palsson BO: **Integrating high-throughput and computational data elucidates bacterial networks.** *Nature* 2004, **429**:92-96.