

# Functional Alignment of Metabolic Networks

ARNON MAZZA,<sup>1</sup> ALLON WAGNER,<sup>1,2</sup> EYTAN RUPPIN,<sup>1,3,4</sup> and RODED SHARAN<sup>1</sup>

## ABSTRACT

**Network alignment has become a standard tool in comparative biology, allowing the inference of protein function, interaction, and orthology. However, current alignment techniques are based on topological properties of networks and do not take into account their functional implications. Here we propose, for the first time, an algorithm to align two metabolic networks by taking advantage of their coupled metabolic models. These models allow us to assess the functional implications of genes or reactions, captured by the metabolic fluxes that are altered following their deletion from the network. Such implications may spread far beyond the region of the network where the gene or reaction lies. We apply our algorithm to align metabolic networks from various organisms, ranging from bacteria to humans, showing that our alignment can reveal functional orthology relations that are missed by conventional topological alignments.**

**Key words:** network alignment, metabolic networks, genome-scale metabolic models.

## 1. INTRODUCTION

**W**ITH THE EVER-GROWING HIGH THROUGHPUT MEASUREMENTS of biological entities and relations, there is considerable interest in methods that go beyond sequence analysis to compare and contrast different species, conditions, or time points. Network alignment methods present a promising alternative as they are able to capture topological similarities or differences that cannot be gleaned from sequence alone, improving the prediction of protein function, interaction, and evolution (Sharan et al., 2005).

Network alignment was originally applied to metabolic pathways (Ogata et al., 2000) and soon thereafter to protein–protein interaction networks (Kelley et al., 2003). Over the last decade a plethora of methods were developed for the comparison of networks, including local alignment efforts (Sharan et al., 2005; Flannick et al., 2006) and global alignment methodologies (Zhenping et al., 2007; Singh et al., 2008). All these methods are based on comparing the topology of the networks in question.

A metabolic network can be modeled by a hypergraph, whose nodes represent metabolites, and hyperedges represent metabolic reactions. Many alignment algorithms transform this representation into a directed graph, where nodes represent reactions, or enzymes, and an edge is directed from reaction *A* to *B* if *A* produces a substrate of *B*. This transformation allows the application of generic network alignment

---

<sup>1</sup>Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv, Israel.

<sup>2</sup>Department of Electrical Engineering and Computer Science, University of California, Berkeley, Berkeley, California.

<sup>3</sup>The Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel.

<sup>4</sup>Center for Bioinformatics and Computational Biology and Department of Computer Science, University of Maryland, College Park, Maryland.

methods to metabolic networks, aiming to maximize a combined similarity measure that is based on enzyme homology and topology (Pinter et al., 2005; Li et al., 2008; Ay et al., 2011; Abaka et al., 2013).

Another common form of modeling metabolism is via a constraint-based model, which allows expressing the space of fluxes of metabolic reactions under steady state assumptions. Such models can add functional information on the networks being compared, which could be exploited for alignment computation. The approach of Baldan et al. (2012) aligns two metabolic networks by comparing their elementary flux modes (EFMs), defined as minimal sets of reactions that can operate at steady state (Schuster and Hilgetag, 1994). The concept of EFMs is also used by Ay and Kahveci (2010), where the similarity between two reaction sets is measured according to the impact incurred by their inhibition on the flux cone.

In this article we propose a novel algorithm to align metabolic networks by taking advantage of their constraint-based models. These models allow us to assess the functional implications of genes or reactions, captured by the metabolic fluxes that are altered following their deletion from the network. Such implications may spread far beyond the region of the network where the gene or reaction lies, enabling the discovery of functional orthology relations that cannot be gleaned from topology alone. In the context of a human network, finding these nonobvious relations may reveal novel proteins in some model species that are inferred to be functionally similar to a disease-causing protein and, hence, may allow new models for the disease in question.

The article is organized as follows: In section 2 we describe constraint-based modeling and present the metabolic network alignment problem. In section 3 we present our alignment algorithm. In section 4 we apply our algorithm to align metabolic networks from various organisms, ranging from bacteria to humans, and demonstrate its utility over topology-based approaches.

## 2. PRELIMINARIES

### 2.1. Metabolic modeling

A genome-scale metabolic model (GSMM) describes a metabolic network in terms of metabolites (nodes) and biochemical reactions (hyperedges) on these metabolites. A GSMM can be represented by a stoichiometric matrix  $S$ , whose rows correspond to metabolites and columns to reactions.

It is accepted to assume that in a living cell, the concentrations of all metabolites are kept constant over time, signifying that a metabolite's production rate equals its consumption rate; this is known as the steady-state assumption. Some of the metabolites are continuously being taken up from the environment (a.k.a., growth medium), while others are being secreted to it; a special type of reactions, called *exchange reactions*, take care of these types of transport.

Every reaction is assigned with a flux, which measures the flow rate of compounds through the reaction. Flux capacities are naturally limited by availability of nutrients and capabilities of enzymatic activity. These presumptions are combined by applying a constraint-based modeling approach (CBM), representing mass balance, flux directionality, and capacity constraints on the space of possible fluxes in a metabolic network through a set of linear equations:

$$Sv=0 \tag{1}$$

$$v_{min} \leq v \leq v_{max} \tag{2}$$

where  $v$  is the flux vector (a.k.a., flux distribution) of the reactions in the model. Flux balance analysis (FBA) is then applied to study various properties of the model (Orth et al., 2010). FBA methods typically assume that the metabolic model attempts to optimize some objective function and use linear programming to compute the optimal solution of that function. A common biological optimization goal is the maximization of the amount of biomass that can be produced. The biomass function is an organism-dependent combination of metabolites that reflects its growth rate (Feist and Palsson, 2010). Often, the biomass optimal value may be achieved through many possible flux distributions, studied using a flux variability analysis (FVA) approach (Mahadevan and Schilling, 2003).

Another part of a GSMM is the gene-protein-reaction (GPR) associations, describing which genes and proteins catalyze which reactions, as well as the logical rules between the proteins required for catalyzation. This information allows simulating gene knockouts by inferring the affected reactions from the GPR and constraining the flux through them to 0. This approach was successfully used in numerous studies, for example, in distinguishing viable from lethal yeast gene deletion strains by testing whether the optimal biomass production rate was severely damaged under a given knockout (Mo et al., 2009).

## 2.2. Problem definition

In the metabolic network global alignment problem, one seeks a one-to-one correspondence between the reactions of two networks so that the similarity between matched reactions is maximized. Different variants of the problem can be derived depending on the definition of reaction similarity and on whether one-to-many or many-to-many relations are also allowed. Extant approaches to tackle this problem integrate sequence-based similarity data (on the genes catalyzing the reactions) as well as topology-based comparisons to construct a plausible matching.

In this article we address a problem that is similar in flavor, which we solve using a maximum matching approach. Specifically, we assume that the input consists of two GSMMs that are to be compared, representing two species. Our goal remains to align them so as to maximize reaction similarity. The crucial difference from standard alignment is that the similarity measure that we use relies on the input metabolic models. Briefly, we represent every metabolic reaction by the impacts that its deletion induces on the model's fluxes. The derived similarity measure is, hence, functional in nature. In our alignment we allow many-to-many relations. A detailed description of the algorithm appears next.

## 3. THE ALIGNMENT ALGORITHM

We devised a metabolic network alignment algorithm that takes as input two metabolic models (GSMMs) and outputs a many-to-many alignment of their reactions/genes. The algorithm proceeds in two phases. It is worth noting that while in this work we focused on evaluating reaction alignments, all steps of the algorithm are applicable to genes as well with minor adjustments.

### 3.1. Metabolite excretion similarity

For each reaction in a GSMM, we define a feature vector denoting the effect of its deletion on the species' ability to produce (or excrete) each of the metabolites in the model. To "delete" a reaction we constrain its flux to 0. To test the ability of the resulting model to produce a certain metabolite we apply linear programming to maximize the flux through a fictive reaction that excretes only that metabolite. We then record, per metabolite, the ratio between its maximal production rate under the knockout and the corresponding maximal rate in the wild-type (no knockout). We consider a metabolite to be affected by the deletion of a reaction if the obtained ratio is smaller than 99%, denoting some minimal effect that is not due to a numerical error. We call the resulting vector for a reaction its *excretion knockout profile*.

We exclude from the analysis reactions that are considered to be *dead ends*. These are defined as reactions that are unable to carry nonzero flux, even in the richest growth medium (with all exchange reactions open) and, thus, do not affect the computational model (Burgard et al., 2004).

To create a common feature set for comparing excretion profiles of two reactions in different species, only metabolites that are common to the two input models are examined; this limitation will be addressed in section 3.3. The similarity between two reactions, one per species, is then defined as the Jaccard index of their excretion profiles, that is, the number of metabolites that are jointly affected by the reactions over the total number of metabolites affected by them.

Given all pairwise similarities between the reactions of two models, we represent them using a weighted bipartite graph. In this graph, each side represents the reactions of a different species, and edges connect similar reactions, weighted by the corresponding similarity values. A maximum matching algorithm is then applied to find an alignment between the reactions. Precisely, we transform the similarities into distances (with the transformation  $d = 1 - s$ ) and apply the Hungarian method (Munkres, 1957), yielding a collection of reaction pairs with total minimum distance, or maximum similarity.

To account for different possible matches that are equally likely, we add a small random noise (a Gaussian function with parameters  $\mu = 0, \sigma = 0.02$ ) to the computed distances and recalculate the matching. We repeat this procedure four times, and keep only the stable matches, that is, the reaction pairs that are returned in all four repetitions.

### 3.2. Aggregating over multiple media

A growth medium in a GSMM is characterized by the set of exchange reactions that are allowed to carry incoming (negative) flux. The alignment thus far represented the reaction similarities computed under two

fixed media (one per model). Depending on the application, it is often desired to compare two metabolic models under a variety of media, exploring the metabolic spaces spanned by the different uptake constraints (Bilu et al., 2006; Guimerà et al., 2007). We restrict our computations to biologically relevant media, that is, media under which the species can plausibly grow. We define a medium to be *viable* if the biomass flux under this medium is at least 10% of the flux under the richest conditions.

To extend our comparison to different media, we repeat the similarity computation and alignment derivation in 100 random viable growth media. Each medium is randomly generated by allowing only a small fraction (25%) of the exchange reactions to carry inbound flux, in addition to enabling uptake reactions that are essential to survival, that is, reactions whose deletion reduces the biomass flux to less than 10% of the maximal one (for all the species we tested, the same essential reactions were found for all thresholds in the range 10–50%). The benefit of working with such poor media is that when only a small part of the network is activated, the deletion of a reaction has potentially more impact due to shortage in backups. In order to activate similar regions of the two metabolic networks, we limit the pool of exchange reactions that could be enabled to reactions that exist in both species (i.e., reactions that transport the same metabolite).

Applying the basic alignment algorithm in all media, we achieve 100 different sets of reaction matches. We gather all reaction pairs from all the matchings and score each pair by the percentage of matchings in which it appears. The result is a collection of aligned reaction pairs, each with its associated confidence score. This weighted collection comprises the output of the first phase of the algorithm.

### 3.3. Global flux similarity

The algorithm presented thus far restricts the feature space to the subset of the shared metabolites between the aligned species, which may be relatively small for distant species. To overcome this limitation, we computed per reaction  $r$  a second binary feature vector, called its *global knockout profile*, capturing those reactions through which the minimal (or maximal) flux increased (or decreased) due to the knockout of  $r$ . We compute these feature vectors for all the reactions in a given GSMM, resulting in a square matrix whose rows and columns represent reactions appearing in the same order. Given two species, our goal is to align their corresponding knockout matrices  $R_1$  and  $R_2$ .

Unlike the first phase of the algorithm, in this problem the features of the two matrices do not match, and thus it cannot be approached via straightforward Jaccard similarity computation. We therefore apply a simulated annealing approach, where the rows/columns of the smaller matrix  $R_1$  are initially mapped to a subset of similar size of the rows/columns in  $R_2$ , denoted  $R'_2$ . We score a potential mapping by the sum of the Jaccard indices over the rows plus the sum of the Jaccard indices over the columns. In each step we pick a random reaction in  $R'_2$  and compute the effect on the alignment score when switching it with each of the other reactions in  $R'_2$  or when removing it from the mapping and replacing with a reaction from  $R_2 \setminus R'_2$ . We examine the change in score for the best candidate: If the score increases, the step is accepted; otherwise, the step is accepted in some probability that decreases over the iterations. If a step is accepted, the corresponding rows and columns are updated in  $R'_2$ .

To narrow down this huge search space, we use the output of the first phase of the algorithm to determine the starting permutation: the many-to-many alignment of the first phase is transformed into a one-to-one alignment by traversing all pairs in order of decreasing score and accepting a pair if it does not collide with an existing one (we collect only pairs with confidence score >5%). These high-confidence reaction pairs serve as fixed anchors and are not allowed to switch in the optimization process.

We repeat the algorithm for the same set of 100 media and assign to each reaction pair a score reflecting the percentage of the media in which it was reported. To achieve the final alignment, we unify the outputs of the two phases, defining the confidence score of a reaction pair as the average score computed in each phase. We filter out pairs that raised only in the second phase with score less than 2%.

## 4. RESULTS

We implemented the alignment algorithm in Matlab. Linear optimizations were performed using the Tomlab Cplex optimization tool. The knockout simulations were executed using grid computing over five Intel Xeon X5650 servers with 24 cores each. A complete alignment (including both phases) between two species over 100 media was generated in approximately 48 hours. We applied the algorithm to align the

networks of several species pairs with varying evolutionary distances. We describe below our performance evaluation measures, the algorithms we compared and the alignment instances we processed.

#### 4.1. Performance evaluation

To evaluate our method and compare it to extant ones, we estimate the accuracy of the predicted reaction pairs with respect to a ground-truth set. The latter includes reaction pairs whose input and output metabolites are identical (determined based on name similarity). The evaluation is summarized in a precision-recall (PR) curve. For a given confidence threshold, *precision* is the percent of aligned pairs that are part of the ground-truth set, while *recall* is the percent of ground-truth pairs that occur in the alignment.

We compared our algorithm to two state-of-the-art network alignment methods. The first, IsoRankN (Liao et al., 2009), is a leading approach for aligning protein–protein and other molecular interaction networks. This algorithm performs topological-based alignment and is able to exploit prior information on node similarity (e.g., sequence based). To create the input for IsoRankN, we constructed a reaction graph per model, in which a node represents a reaction and an undirected (unweighted) edge connects two reactions if one of them produces a substrate of the other (ignoring very abundant metabolites that “contribute” to that graph more than 150 edges). The prior node similarity scores were computed based on the EC number categorization. Specifically, for two reactions, one from each network, the similarity score was set as the Jaccard index between the sets of EC numbers associated with the genes that catalyze the reactions. Finally, we experimented with two values for the parameter  $\alpha$  that balances between the prior information and network-derived match scores— $\alpha=0.5$ , which gives equal weight to both, and  $\alpha=0.99$ , which emphasizes the topology-based score.

We also aimed to compare our algorithm to a recent metabolic network alignment method, CAMPways (Abaka et al., 2013), whose code was readily available. However, when applying this method to any of the alignment instances described below, it did not finish processing any of them (nor produced a progress log) within a 96-hour time frame, even when restricting attention to the cytosol compartment, which is the only subnetwork of significant size we were able to neatly extract from the models. Hence, we could not report its results.

#### 4.2. Aligning similar models

As a basic validation of our approach we applied it to align the yeast metabolic network of Mo et al. (2009) with itself (randomly permuting the reactions). The model contains 1577 reactions over 1228 metabolites (Fig. 1). With a precision of 85%, out of 1024 non-deadend reactions 875 were correctly aligned (to themselves, 85% recall), compared to 690 aligned reactions (67% recall) by the first phase only. The PR curve is displayed in Figure 2a. Notably, this recall further increases to 95% when considering in the ground-truth only the 919 reactions whose knockout has some observable effect in either the excretion or the global knockout profile (Fig. 1). This gap between the number of non-deadend reactions and the number of reactions having nonempty profiles can be explained by the existence of alternative pathways in the metabolic network and may indicate that higher-order deletions may improve the recall.

The predicted matches also contain 702 nonidentical reaction pairs (over 235 distinct reactions) with score at least 2%. While these pairs represent different reactions, they admit perfect similarity in some media and, thus, are indistinguishable from their real matches in those media (22% are neighboring reactions,  $p=e^{-210}$  by a hypergeometric test). Notably, the majority (215) of these 235 reactions also appear in the predicted matching with their real match.

We compared the performance of our approach to the IsoRankN algorithm, executed over the induced reaction graph (4610 edges). Given two reactions, their prior similarity score was computed as the overlap percentage between the sets of EC numbers corresponding to their associated genes in the model. The results for the two parameter choices (0.5 and 0.99) are shown in Figure 2b. To allow a fair comparison, these plots relate to the complete set of reactions, including also deadend ones. Evidently, our algorithm outperforms both competitors, producing higher-quality alignments.

As a second validation test, we aligned successive versions of the *E. coli* metabolic model that were published by the same lab, iAF1260 (Feist et al., 2007) and iJO1366 (Orth et al., 2011) (2382 and 2583 reactions, respectively, see Fig. 1). These models have 2114 (non-deadend) reaction pairs in common, identified by comparing their internal codes. Notably, with a precision of 85%, 77% of the ground-truth pairs are discovered by our algorithm, compared to only 51% when applying only the first phase of the

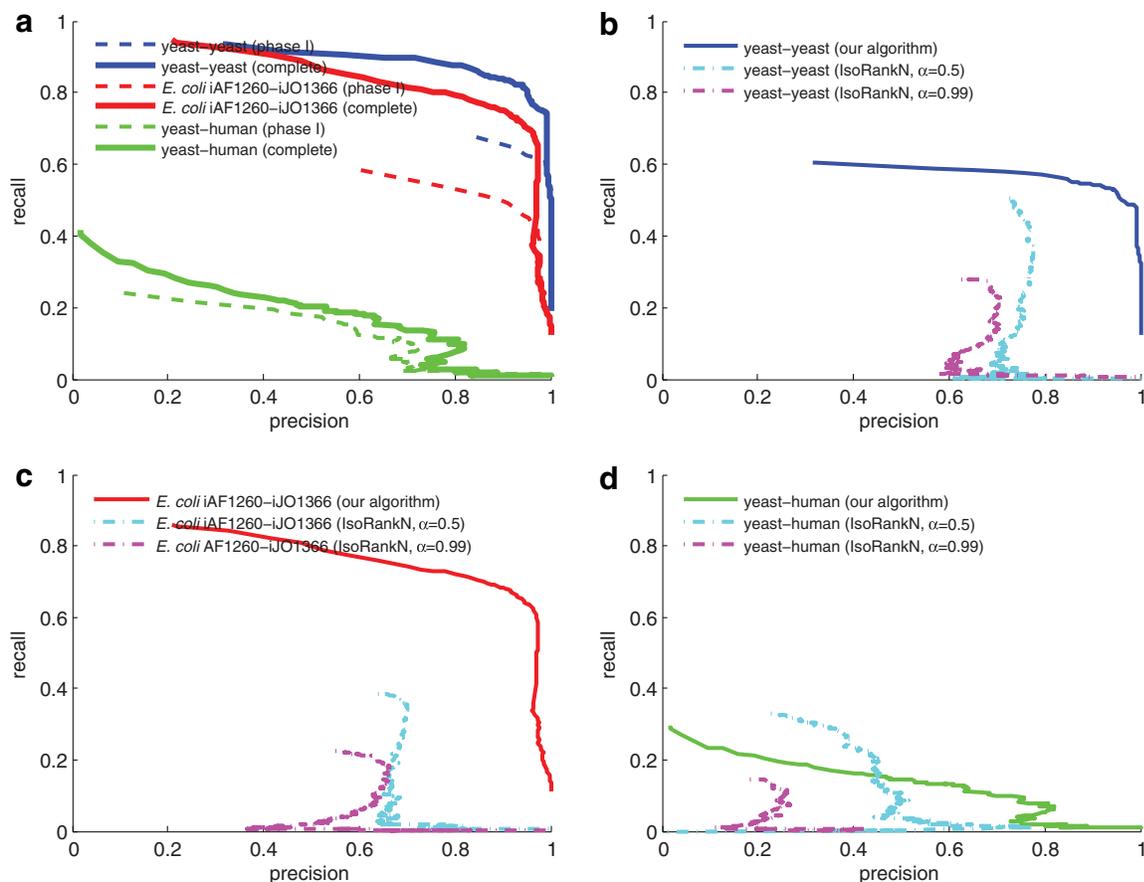
<i>measure/model</i>	yeast	human	<i>E. coli</i> iAF1260	<i>E. coli</i> iJO1366
#metabolites	1228	2766	1668	1805
#shared metabolites	663	663	1621	1621
#reactions	1577	3788	2382	2583
#non-deadend reactions	1024	2515	2159	2351
#reactions with nonempty excretion knockout profile	722	1072	1860	2337
#reactions with nonempty excretion knockout profile over the shared metabolites	657	766	1707	2323
#reactions with nonempty global knockout profile	912	1599	2159	2351
#reactions with observable knockout effect	919	1603	2159	2351

**FIG. 1.** Model summary and reaction knockout statistics. This table displays statistics for each pair of aligned models, including the number of metabolites in each model (all or shared between the models) and the number of reactions (all or non-dead-end only). To assess the potential of the two alignment phases, the number of reactions having some observable knockout effect is shown: (1) when considering the excretion profile of a reaction, (2) when considering only the part of that profile that corresponds to metabolites shared between the aligned species, (3) when examining the global profile, and (4) when accounting both (1) and (3).

algorithm (Fig. 2a). The maximal recall achieved was nearly 95%, leaving only 114 true matches that were not revealed in any confidence score. It is worth noting that the good performance in this case study demonstrates the power of our framework when most reactions have informative knockout profiles (Fig. 1). Figure 2c demonstrates a clear advantage of our algorithm over the topology-based IsoRankN. This result also shows that our approach is robust to the inherent incompleteness of metabolic networks, as the aligned models have a large core of common reactions but also considerable differences.

#### 4.3. Aligning the yeast and human models

To test our algorithm on distant species, we applied it to align GSMs of yeast and human. We used the yeast iMM904 model (Mo et al., 2009) (1577 reactions, 1024 non-deadend; 1228 metabolites) and the human Recon1 model (Duarte et al., 2007) (3788 reactions, 2515 non-deadend; 2766 metabolites). Our alignment algorithm was limited to 919 yeast reactions (58% of total reactions) and 1603 human reactions (42%) with nonempty excretion or global profiles; 663 common metabolites between the models were identified based on name similarity (Fig. 1). Figure 2a displays the PR curve of the predictions with respect to a ground-truth set of 421 non-deadend reaction pairs. For a precision of 60%, our algorithm achieves a recall of 18%, an improvement of 3% over the recall achieved by applying the first phase only. The relatively inferior performance compared to the alignments computed in the previous section can be related to the small number of human reactions that have nonempty excretion knockout profiles (20% when considering only the shared metabolites). This in turn results in a poor starting point for the simulated

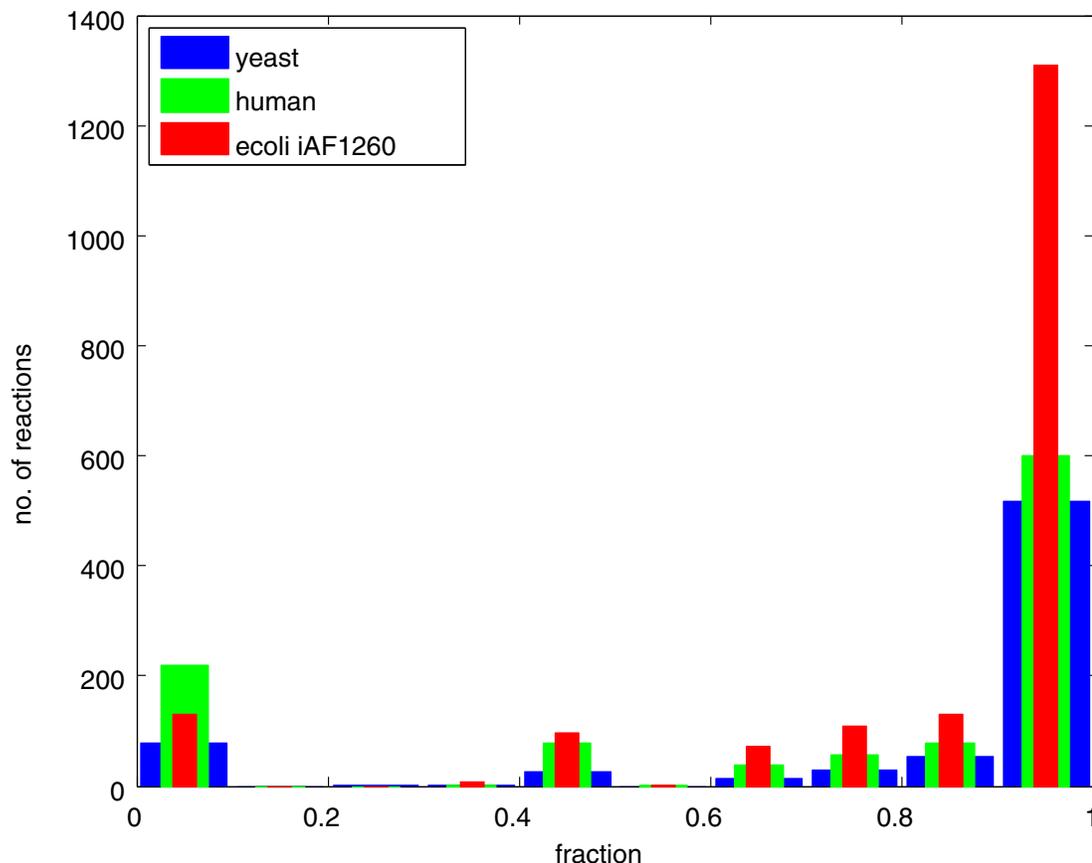


**FIG. 2.** Performance evaluation and comparison to IsoRankN. **(a)** Precision-recall curves for our functional alignment algorithm (phase I only and the complete algorithm) over the three case studies: yeast vs. yeast, *E. coli* iAF1260 vs. *E. coli* iJO1366, and yeast vs. human. The recall is measured with respect to the number of non-dead-end ground-truth reaction pairs between the metabolic networks. **(b–d)** These subplots compare the performance of our (complete) algorithm and the IsoRankN algorithm over the three case studies. Here the complete ground-truth set is considered, including matches of dead-end reactions.

annealing phase, which gets stuck in distant local optima. In comparison to IsoRankN (Fig. 2d), the PR plot of our algorithm dominates that of the topology-based variant ( $\alpha=0.99$ ). The second variant, which balances sequence and topology information ( $\alpha=0.5$ ) has higher recall for low precision values (up to 45%), but its recall drops to almost zero for higher precision values (above 55%), while our algorithm maintains relatively stable recall values even in this range (13% recall at 60% precision, with respect to all 587 ground-truth pairs, including deadend reactions).

To systematically evaluate the quality of the predicted matches that are not in the ground-truth alignment, we calculated the functional similarity between the gene sets catalyzing the reactions in each of these pairs. We defined the functional similarity between two genes as the maximum semantic similarity (Resnik, 1999) between their annotated gene ontology terms (The Gene Ontology Consortium, 2000). We extended this definition to gene sets by defining their similarity as the maximum value obtained over any two members of these sets. Using these definitions, we computed the average functional similarity between 201 predicted reaction pairs with a confidence score of at least 5% (and at least one associated gene in both models). The average score was 5.7, corresponding to an empirical  $p$ -value of 0.005 (permutation test).

To demonstrate the utility of our approach in identifying nonobvious orthology relations, we first computed how many yeast or human reactions have long-range impacts. To this end, we counted for each perturbed reaction the ratio between the number of affected metabolites that are not part of the reaction and the total number of metabolites affected by that reaction. Figure 3 shows that the majority of the reactions that have nonempty excretion knockout profiles also have long-range effects in the network. To test



**FIG. 3.** Long-range knockout effects. For different fraction values  $t$  (in bins of size 0.1), shown is the number of reactions for which a fraction of  $t$  of the affected metabolites after a knockout of the reaction is not part of it.

whether distant functional alignments may suggest nonobvious disease models, we looked for a human reaction that (i) is associated with a disease from the OMIM database (Amberger et al., 2009), and (ii) has a ground-truth yeast match that is distant in the reaction graph from the suggested match in our alignment. As an example, our method functionally aligned the human reaction catalyzed by the enzyme cystathionine- $\beta$ -synthase (CBS) with the exchange reaction that imports sulfate into the yeast cell. CBS deficiency in humans leads to a severe disease due to disruption of sulfur metabolism, homocystinuria, in which the inability of CBS to convert homocysteine leads to its excessive accumulation in the blood and urine. CBS is part of the transsulfuration pathway, which FBA analysis suggests is used for homocysteine degradation to ultimately (and indirectly) increase the availability of sulfur to the cell. It is thus plausible that some of the phenotypes of CBS deficiency can be modeled by blocking sulfate uptake. Our algorithm detected this functional alignment even though yeast has a close sequence ortholog to the human CBS enzyme and the reaction catalyzed by that enzyme is topologically distant from sulfate exchange.

Moreover, this coupling correctly reflects pathologies associated with CBS deficiency in humans: the alignment is due to the similar effects incurred by the deletion of the yeast/human reaction over a set of eight sulfur-containing metabolites, all of them derived from cysteine or glutathione; the latter is an antioxidant with key cellular functions, which is thought to be produced in considerable quantities by the transsulfuration pathway through the intermediate cysteine (Mosharov et al., 2000). As this process depends on CBS, glutathione deficiency may partly account for homocystinuria's symptoms (cf. Robert et al., 2005).

## 5. CONCLUSIONS

We presented a model-based alignment strategy to align metabolic networks. Our strategy employs GSMMs to compute the functional implications of metabolic reactions, thereby aligning them. We applied

our strategy to align different metabolic models, demonstrating its utility over topological approaches. Importantly, our method is applicable to current-large scale metabolic models, unlike most of the leading tools, which are practical for aligning only a single pathway at a time, requiring prior knowledge on the model division to distinct pathways. Our approach could also be integrated into a multistep alignment framework that exploits data of three types: topology, sequence, and functional impacts.

One limitation of our approach is its applicability only to reactions whose deletion has some observable effect. In some models, many of the reactions do not exhibit an effect when deleted in isolation, suggesting that better results can be obtained if extending the functional profiles to knockouts of higher order.

Another limitation, computational in nature, lies in our attempt to find a complete alignment in the second phase of the algorithm (all reactions of the smaller model must be aligned). As distant species such as yeasts and humans are likely to have only a small core of orthologous reactions, this procedure may be affected by irrelevant noise. A future challenge is to avoid this noise, applying the simulated annealing procedure to different subsets of the reactions.

### ACKNOWLEDGMENTS

A.M. and A.W. were supported in part by a fellowship from the Edmond J. Safra Center for Bioinformatics at Tel-Aviv University. R.S. was supported by a research grant from the Israel Science Foundation (grant no. 241/11).

### AUTHOR DISCLOSURE STATEMENT

No competing financial interests exist.

### REFERENCES

- Abaka, G., Biyikoglu, T., and Erten, C. 2013. CAMPways: constrained alignment framework for the comparative analysis of a pair of metabolic pathways. *Bioinformatics* 29, i145–i153.
- Amberger, J., Bocchini, C.A., Scott, A.F., et al. 2009. McKusick's online Mendelian inheritance in man (OMIM). *Nucleic Acids Res.* 37, D793–D796.
- Ay, F., and Kahveci, T. 2010. Functional similarities of reaction sets in metabolic pathways. Proceedings of the First ACM International Conference on Bioinformatics and Computational Biology—BCB'10. pp. 102–111.
- Ay, F., Kellis, M., and Kahveci, T. 2011. SubMAP: aligning metabolic pathways with subnetwork mappings. *J. Comput. Biol.* 18, 219–235.
- Baldan, P., Cocco, N., and Simeoni, M. 2012. Comparison of metabolic pathways by considering potential uxes. BioPPN2012—3rd International Workshop on Biological Processes and Petri Nets, vol. 85, pp. 2–17.
- Bilu, Y., Shlomi, T., Barkai, N., et al. 2006. Conservation of expression and sequence of metabolic genes is reected by activity across metabolic states. *PLoS Comput. Biol.* 2, e106.
- Burgard, A.P., Nikolaev, E.V., Schilling, C.H., et al. 2004. Flux coupling analysis of genome-scale metabolic network reconstructions. *Genome Res.* 14, 301–312.
- Duarte, N.C., Becker, S.A., Jamshidi, N., et al. 2007. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc. Natl. Acad. Sci. USA* 104, 1777–1782.
- Feist, A.M., and Palsson, B.O. 2010. The biomass objective function. *Curr. Opin. Microbiol.* 13, 344–349.
- Feist, A.M., Henry, C.S., Reed, J.L., et al. 2007. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol. Syst. Biol.* 3, 121.
- Flannick, J., Novak, A., Srinivasan, B.S., et al. 2006. Graemlin: general and robust alignment of multiple large interaction networks. *Genome Res.* 16, 1169–1181.
- Guimerà, R., Sales-Pardo, M., and Amaral, L.A.N. 2007. A network-based method for target selection in metabolic networks. *Bioinformatics* 23, 1616–1622.
- Kelley, B.P., Sharan, R., Karp, R.M., et al. 2003. Conserved pathways within bacteria and yeast as revealed by global protein network alignment. *Proc. Natl. Acad. Sci. USA* 100, 11394–11399.
- Li, Y., de Ridder, D., de Groot, M.J., et al. 2008. Metabolic pathway alignment (M-Pal) reveals diversity and alternatives in conserved networks, 273–285. In *Advances in Bioinformatics and Computational Biology*, vol. 6. Imperial College Press, London.

- Liao, C.S., Lu, K., Baym, M., et al. 2009. IsoRankN: spectral methods for global alignment of multiple protein networks. *Bioinformatics* 25, i253–i258.
- Mahadevan, R., and Schilling, C. 2003. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab. Eng.* 5, 264–276.
- Mo, M.L., Palsson, B.O., and Herrgård, M.J. 2009. Connecting extracellular metabolomic measurements to intracellular ux states in yeast. *BMC Syst. Biol.* 3, 37.
- Mosharov, E., Cranford, M.R., and Banerjee, R. 2000. The quantitatively important relationship between homocysteine metabolism and glutathione synthesis by the transsulfuration pathway and its regulation by redox changes. *Biochemistry* 39, 13005–13011.
- Munkres, J. 1957. Algorithms for the assignment and transportation problems. *J. Soc. Ind. Appl. Math.* 5, 32–38.
- Ogata, H., Fujibuchi, W., Goto, S., et al. 2000. A heuristic graph comparison algorithm and its application to detect functionally related enzyme clusters. *Nucleic Acids Res.* 28, 4021–4028.
- Orth, J.D., Conrad, T.M., Na, J., et al. 2011. A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism—2011. *Mol. Syst. Biol.* 7, 535.
- Orth, J.D., Thiele, I., and Palsson, B.O. 2010. What is ux balance analysis? *Nat. Biotechnol.* 28, 245–248.
- Pinter, R.Y., Rokhlenko, O., Yeager-Lotem, E., et al. 2005. Alignment of metabolic pathways. *Bioinformatics* 21, 3401–3408.
- Resnik, P. 1999. Semantic similarity in a taxonomy: an information-based measure and its application to problems of ambiguity in natural language. *J. Artif. Intell. Res.* 11, 95–130.
- Robert, K., Nehmé, J., Bourdon, E., et al. 2005. Cystathionine  $\beta$  synthase deficiency promotes oxidative stress, fibrosis, and steatosis in mice liver. *Gastroenterology* 128, 1405–1415.
- Schuster, S., and Hilgetag, C. 1994. On elementary ux modes in biochemical reaction systems at steady state. *J. Biol. Syst.* 2, 165–182.
- Sharan, R., Suthram, S., Kelley, R.M., et al. 2005. Conserved patterns of protein interaction in multiple species. *Proc. Natl. Acad. Sci. U. S. A.* 102, 1974–1979.
- Singh, R., Xu, J., and Berger, B. 2008. Global alignment of multiple protein interaction networks with application to functional orthology detection. *Proc. Natl. Acad. Sci. USA* 105, 12763–12768.
- The Gene Ontology Consortium. 2000. Gene ontology: tool for the unification of biology. *Nat. Genet.* 25, 25–29.
- Zhenping, L., Zhang, S., Wang, Y., et al. 2007. Alignment of molecular networks by integer quadratic programming. *Bioinformatics* 23, 1631–1639.

Address correspondence to:  
Prof. Roded Sharan  
Blavatnik School of Computer Science  
Tel Aviv University  
Ramat Aviv  
Tel Aviv 69978  
Israel

E-mail: roded@post.tau.ac.il