Supporting Information

Systems level analysis of the *Chlamydomonas reinhardtii* metabolic network reveals variability in evolutionary co-conservation

Amphun Chaiboonchoe[a†], Lila Ghamsari[b†], Bushra Dohai[a†], Patrick Ng[c†], Basel Khraiwesh[a], Ashish Jaiswal[a], Kenan Jijakli[a], Joseph Koussa[a], David R. Nelson[a], Hong Cai[a§], Xinping Yang[b], Roger L. Chang[d], Jason Papin[e*], Haiyuan Yu[c*], Santhanam Balaji[a,b,f*], Kourosh Salehi-Ashtiani[a,b*]

[a]Division of Science and Math, New York University Abu Dhabi and Center for Genomics and Systems Biology (CGSB), New York University Abu Dhabi Institute, Abu Dhabi, UAE
[b]Center for Cancer Systems Biology (CCSB) and Department of Cancer Biology, Dana-Farber Cancer Institute, and Department of Genetics, Harvard Medical School, Boston, MA, USA
[c]Department of Biological Statistics and Computational Biology and Weill
Institute for Cell and Molecular Biology, Cornell University, Ithaca, NY, USA
[d]Department of Systems Biology, Harvard Medical School, Boston, MA, USA
[e]Department of Biomedical Engineering, University of Virginia, Charlottesville, VA, USA
[f]MRC Laboratory of Molecular Biology, Cambridge, UK.

*Correspondence to: ksa3@nyu.edu, bsanthan@mrc-lmb.cam.ac.uk, haiyuan.yu@cornell.edu, papin@virginia.edu

[†]These authors contributed equally to this work
[§]Present address: BGI-Shenzhen, Shenzhen 518083, China

Fig. S1 Transformation of the Chlamydomonas Metabolic network to a gene-centric network.
Conceptual transformation of a metabolic pathway (where nodes are metabolites) to a protein centric
pathway (where nodes are gene products) is shown in (a), herein referred to as "type A transformation"
where edges (typically enzymes) are transformed to nodes. In the resulting pathway, edges are
connected to each other through metabolites. Following type A transformation of the Chlamydomonas
metabolic network [1, 2], the number of connections between genes (degree) and the number of genes with
the relevant degree is plotted for the resulting network in (b). The distribution suggests a non-linear
trend with many genes with "few" connections and few genes with "many" connections as observed in
most other biological networks like protein-protein interaction or gene regulatory networks. Currency
metabolites (see Network Transformation) were excluded from the transformation.
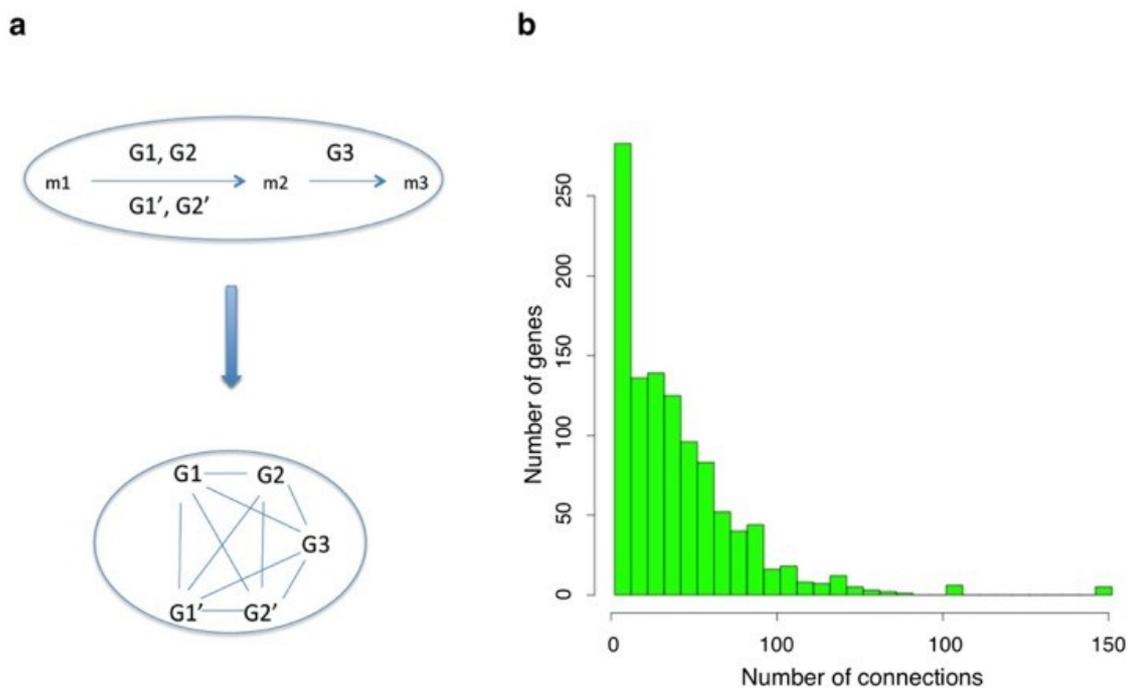
Fig. S2 The transformed, protein-centric metabolic network of *C. reinhardtii*. A graphical representation of the transformed *C. reinhardtii* metabolic network. Following removal of currency metabolites, the gene-reaction-metabolite associations described in the network were used to transform the genome-scale metabolic network of *C. reinhardtii* to a gene-centric network. There are altogether 1,081 gene products and 11,094 edges (links). This network displays a scale-free non-linear distribution between the number of genes and their connections (see Fig. S1). The colors of the nodes represent their connectivity degree (blue highest, dark-red lowest); the size of the nodes corresponds to the clustering co-efficient of the nodes; the size and color of the edges correspond to edge-betweenness (short or blue, low; long or red, high).
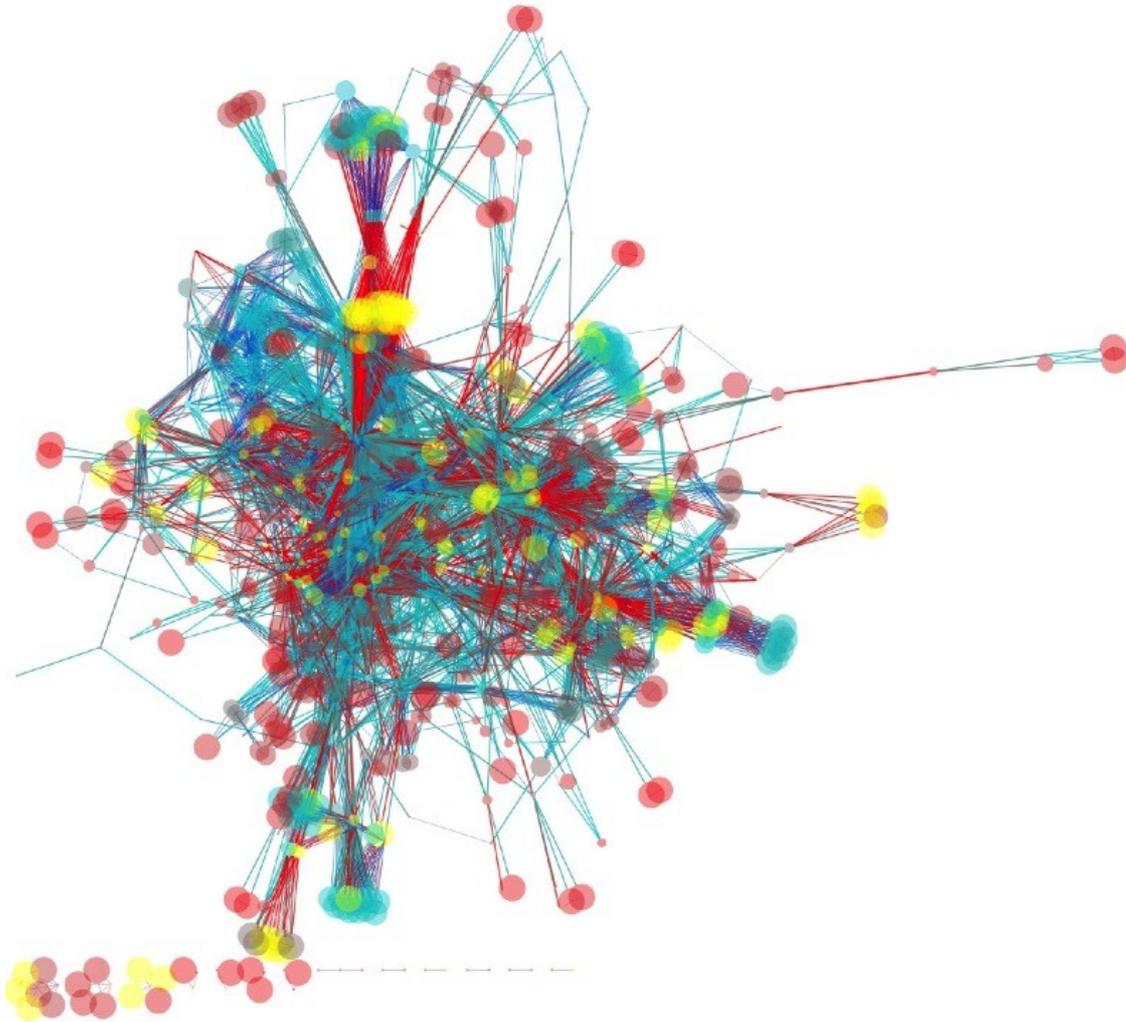
Fig. S3 GO term enrichment analyses of dynamically and statically co-conserved pairs. Statistically significant GO terms for Biological Process ontology with enrichment values of p< 0.1 are shown for static (red) or dynamic (green) pairs (a). Statistically significant GO terms for Molecular Function ontology with enrichment values of p< 0.1 are shown in (b) for static (red) or dynamic (green) pairs. The bars represent the enrichment score (y-axis) of the dynamic and static in different biological functions (x-axis).

a

b

4
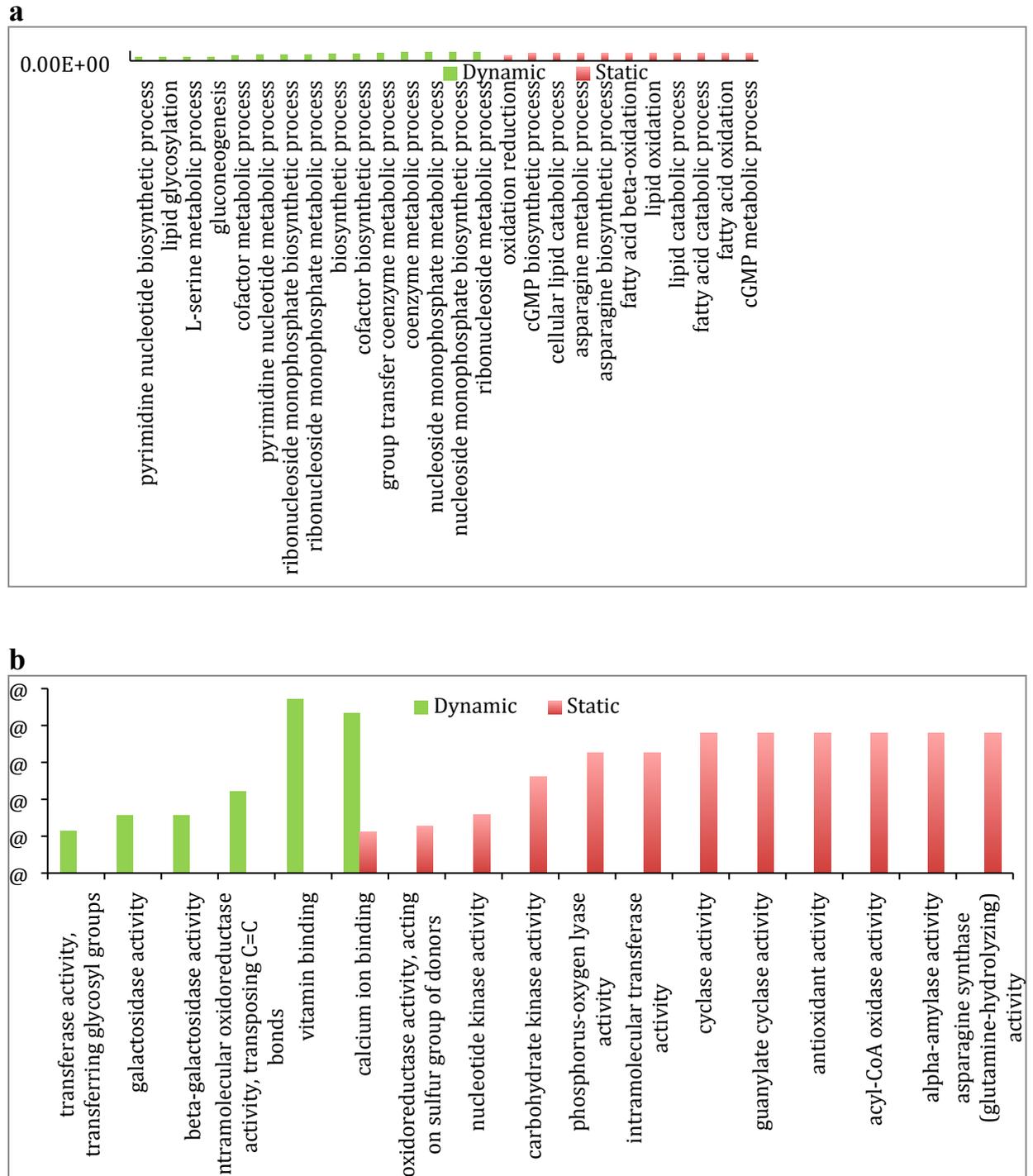
Fig. S4 The largest five co-conserved hubs in the network. The hubs in the network were defined as nodes forming the top 20% of highly connected genes. The five most connected hubs in the network are shown: Hubs 1-4 encode ferredoxins, hub 5 encodes an acyl-carrier protein (ACP2). Green edges are between nodes showing evidence of dynamic co-conservation, red edges are between statically co-conserved pairs.
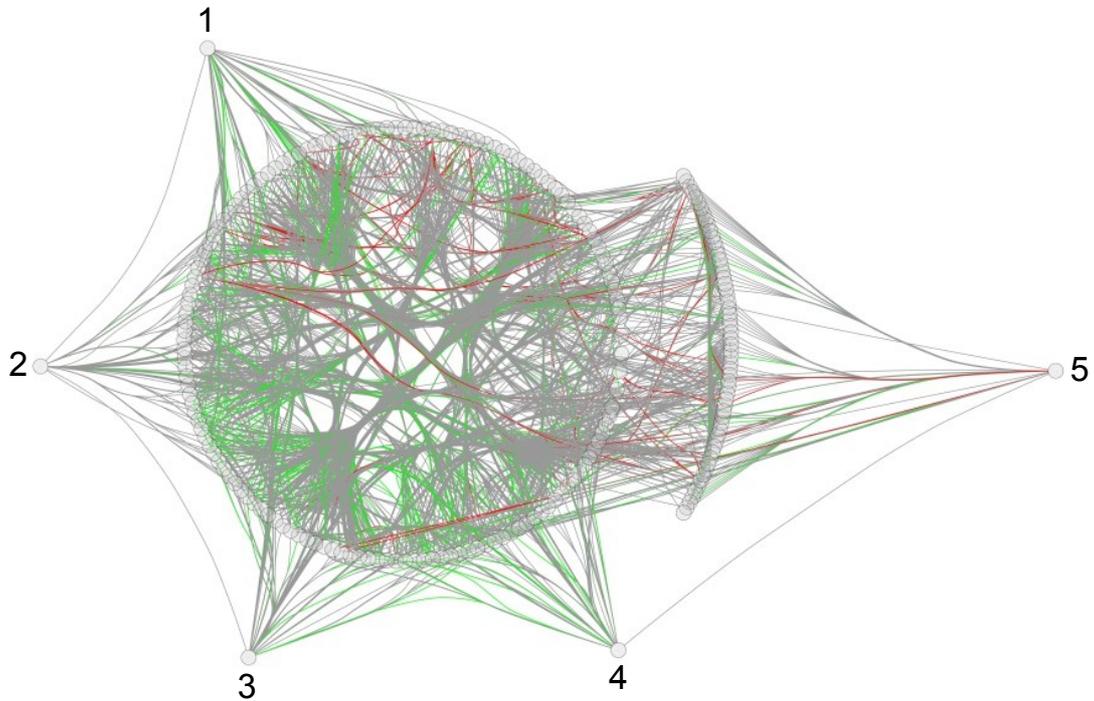
Fig. S5 Gene ontology (GO) and interolog analysis of *C. reinhardtii* with *S.cerevisiae* and *A.thaliana*. GO terms (Biological process) of dynamically co-conserved interologs of *C. reinhardtii/A. thaliana*, and interologs of *C. reinhardtii/S. cerevisiae* are shown in (a) and (b) respectively; GO terms of statically co-conserved interlogs of *C. reinhardtii/A. thaliana*, and interologs of *C.reinhardtii/S.cerevisiae* are shown in (c) and (d). BinGO (a biological network gene ontology tool) was used to visualize the GO terms that were statistically overrepresented (hypergeometric test) in the gene sets. Nodes highlighted in yellow correspond to GO terms with a *p*-value of 0.05 or smaller.

a

b

branched chain family amino acid biosynthetic process

branched chain family amino acid metabolic process

cellular aromatic biosynthetic process

carboxylic acid metabolic process

oxoacid metabolic process

cellular amino acid biosynthetic process

cellular aromatic compound biosynthetic process

carboxylic acid biosynthetic process

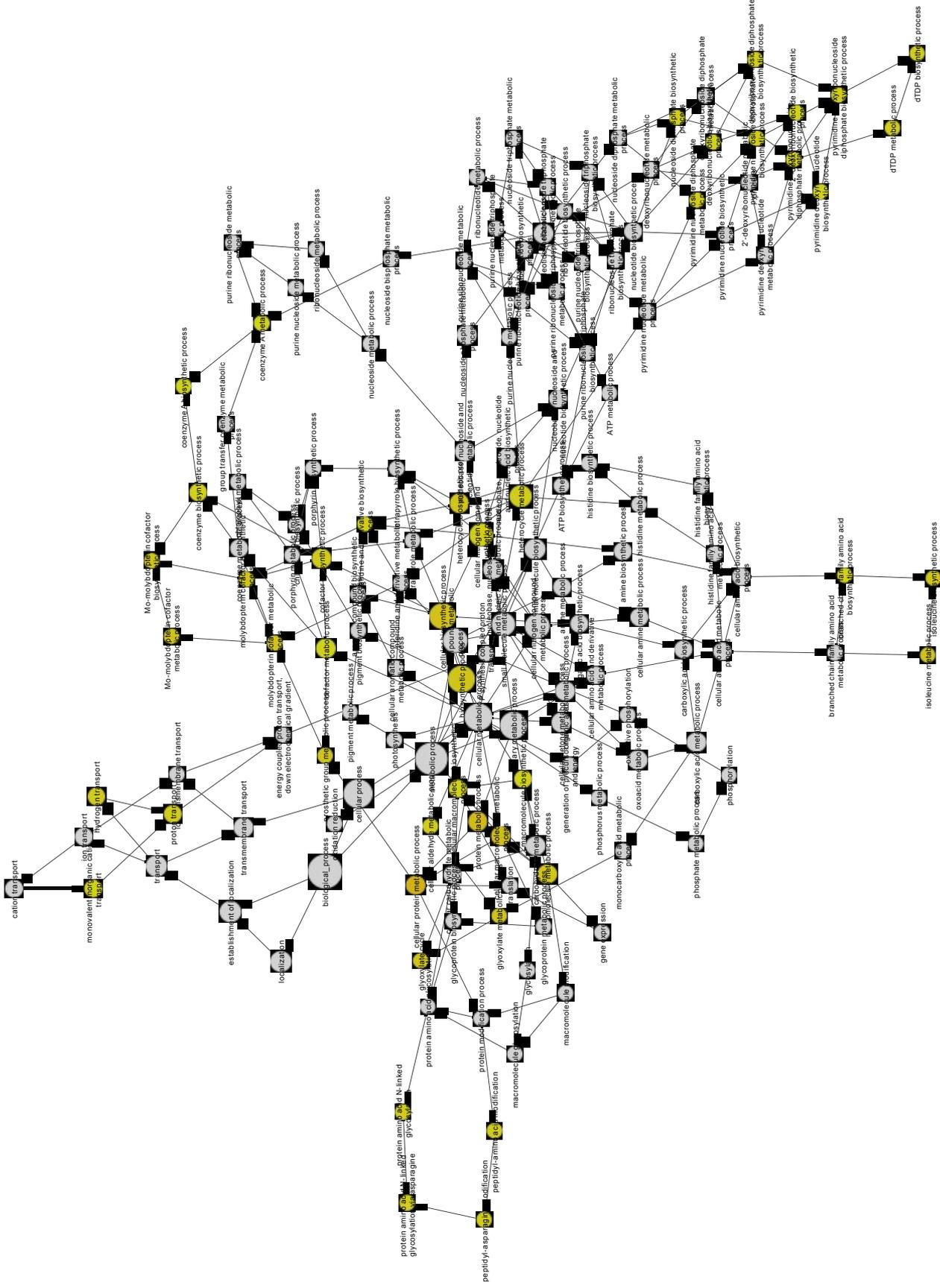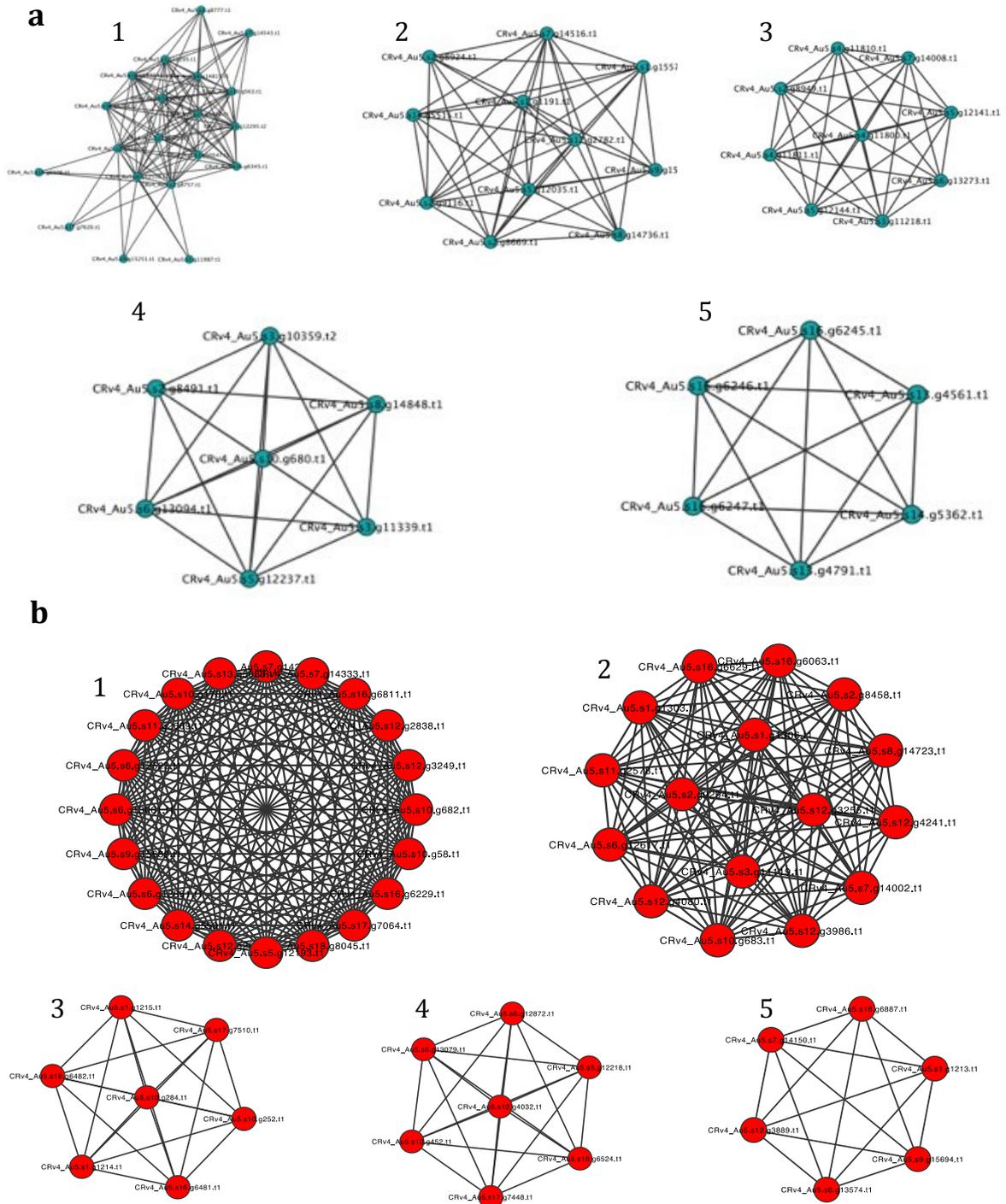organic acid metabolic process

amine biosynthetic process

cellular amino acid metabolic process

organic acid and derivative metabolic process

cellular ketone metabolic process

cellular nitrogen compound biosynthetic process

small molecule biosynthetic process

organic acid biosynthetic process

transmembrane transport

transport

amine metabolic process

cellular nitrogen compound metabolic process

cellular biosynthetic process

biosynthetic process

cellular metabolic process

small molecule metabolic process

cellular process

establishment of localization

nitrogen compound metabolic process

primary metabolic process

cellular macromolecule biosynthetic

metabolic process

oxidation reduction

biological_process

localization

gene expression

macromolecule biosynthetic process

macromolecule metabolic process

cellular macromolecule metabolic process

cellular macromolecule metabolic

alcohol metabolic process

monosaccharide metabolic process

translation

cellular carbohydrate metabolic

protein metabolic process

cellular protein metabolic process

alcohol biosynthetic process

glycoprotein metabolic process

carbohydrate metabolic process

cellular carbohydrate metabolic process

glycoprotein biosynthetic process

hexose metabolic process

cellular protein metabolic process

glycoprotein biosynthetic process

polyol metabolic process

glucose metabolic process

glycerol metabolic process

glycosylation

cellular macromolecule glycosylation

macromolecule modification

protein amino acid glycosylation

protein modification process

macromolecule glycosylation

protein modification process

protein amino acid N-linked glycosylation

peptidyl-amino acid modification

protein amino acid N-linked glycosylation via asparagine

peptidyl-asparagine modification

7

c

d

9

Fig. S6 Network module analysis using MCODE. The top 5 modules identified based on MCODE scores are represented from (a) 21 total modules for statically co-conserved gene pairs and (b) 41 total modules of dynamically co-conserved gene pairs. The regions of the dynamic and static sub networks where these modules were detected (Fig. 2 in the main text) can be found based on the numbers associated with each module.

Fig. S7 Light and dark expressed genes (a) and interactions (b). Dark and light expressed genes have differential topological and functional characteristics in the network.

Fig. S8 Predicted flux values from the *in silico* Chlamydomonas model. Flux balance analyses were carried out for maximization of biomass under dark with acetate (blue bars) and light without acetate (red bars) conditions. Biomass yield is given on the y-axis.

Fig. S9 Metabolic locations for correlated downregulation (a) and upregulation (b) of flux and gene expression levels. Numbers correspond to the number of genes expressed in each of the subcellular compartments for each set.
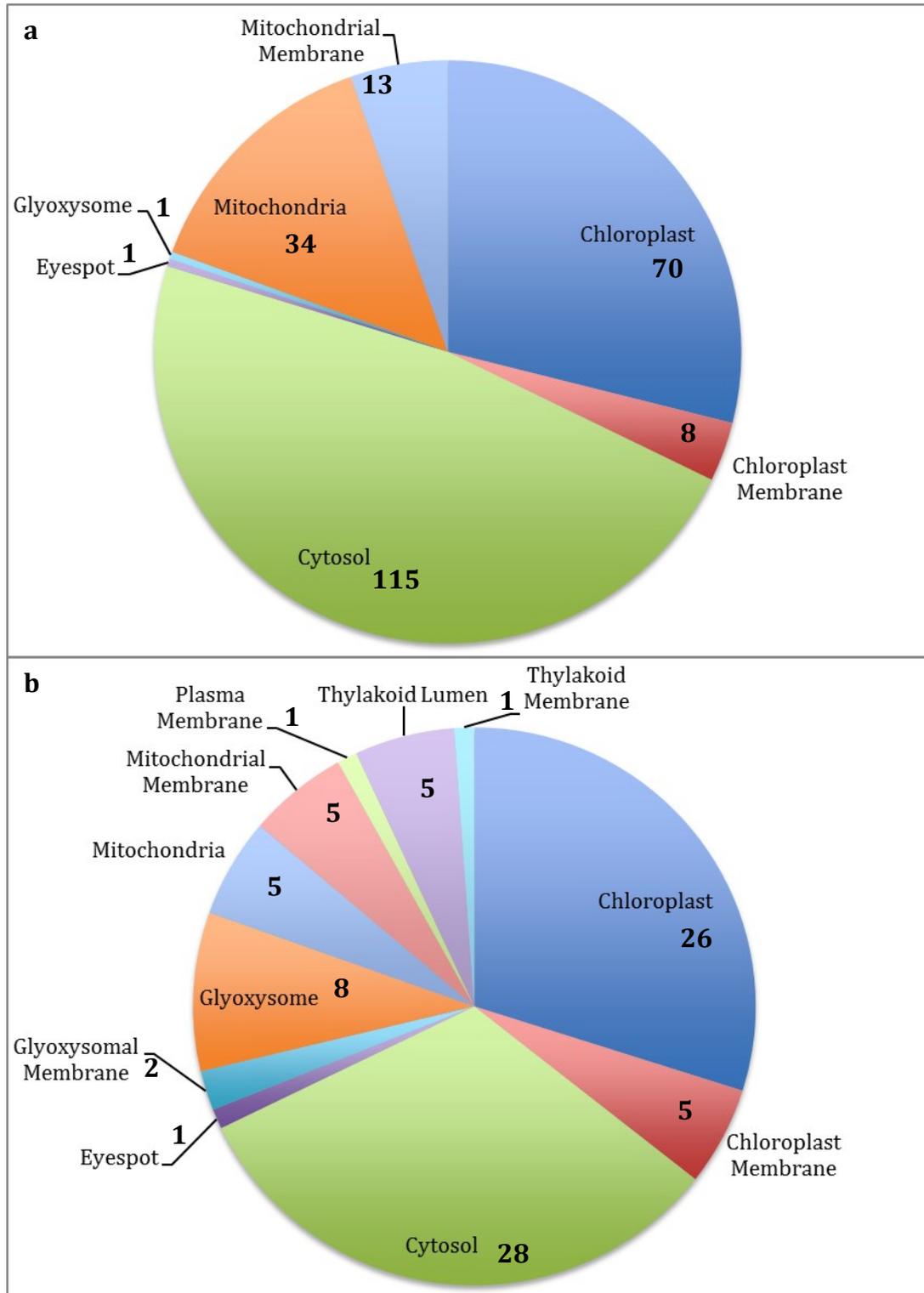
Fig. S10 Distribution of synthetic interactions with respect to their severity and growth condition.
Synthetic interactions were identified under growth in dark with acetate (DA) and light with no acetate
(LNA), the identified interactions were then binned based on the extent that biomass is reduced (in
comparison to the wild-type model).

Synthetic Interactions
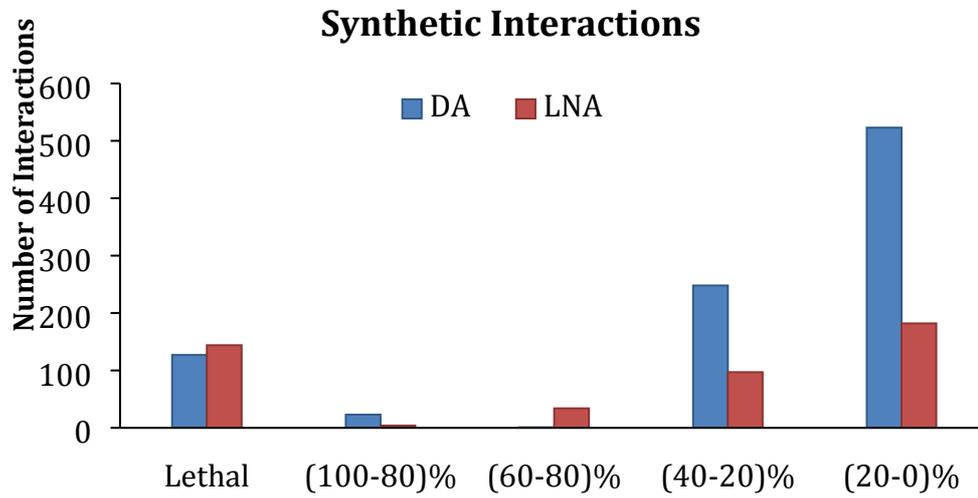
Fig. S11 Subnetworks of synthetic lethal interactions under dark with acetate (a) and light with out acetate (b). Synthetic lethal pairs were used to generate the shown subnetworks. The blue box represents Carbohydrate metabolic process (GO: biological process) and N-Glycan Biosynthesis (KEGG pathway). The green circle represents Electron transport (GO: biological process) and Valine, leucine and isoleucine degradation (KEGG pathway).
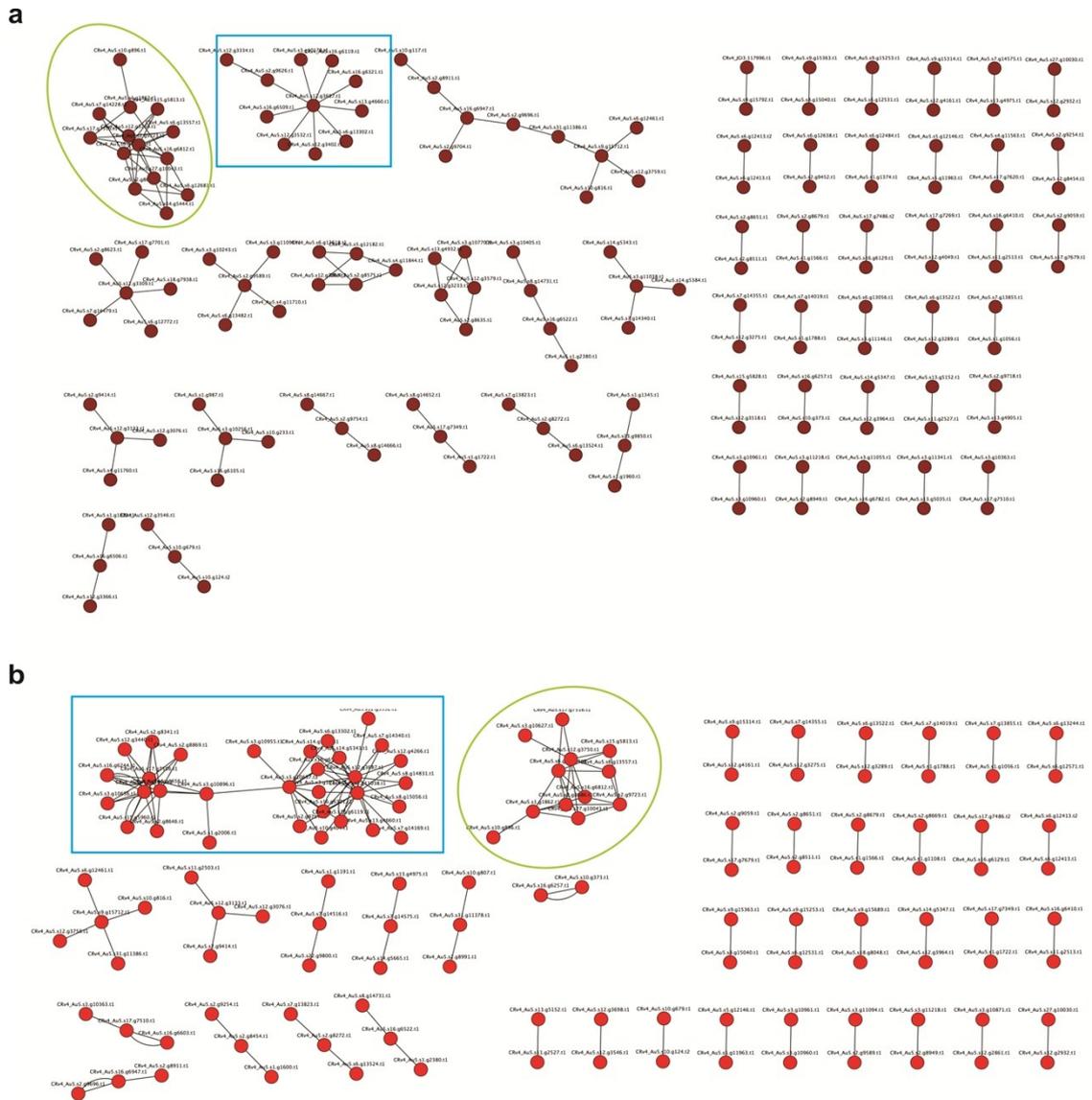
Fig. S12 KEGG term enrichment between two conditions (light and dark) for synthetic lethal interactions. The bars represent the enrichment score (y-axis) of the dark with acetate (red) and light with no acetate (blue) in different KEGG pathways (x-axis).
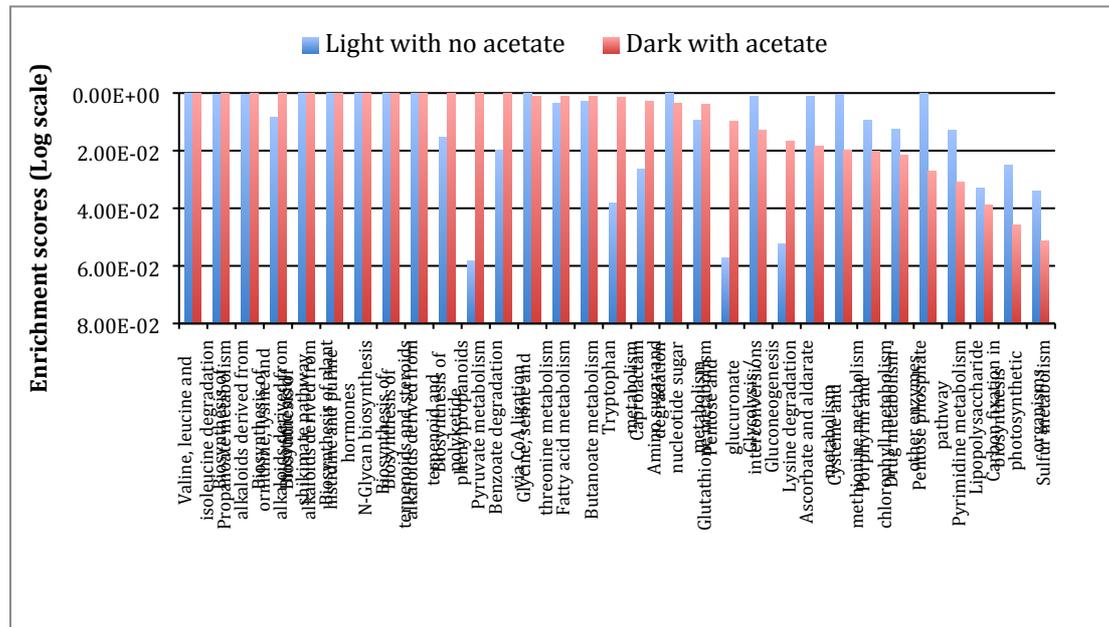
Fig. S13 The pathways of the reactions of co-set1 in set 4 that has zero average profile distance (Table S18) in purine metabolism. The figure shows the purine metabolism as obtained from KEGG. The reactions XNDH (via xanthine dehydrogenase) and URO (via urate oxidase) are highlighted in red rectangular indicated with numbers 1 and 2, respectively. In XNDH xanthine dehydrogenase dehydrogenases the xanthine into Urate, in which the latter is oxidized into 5-Hydroxyisourate in the reaction URO.
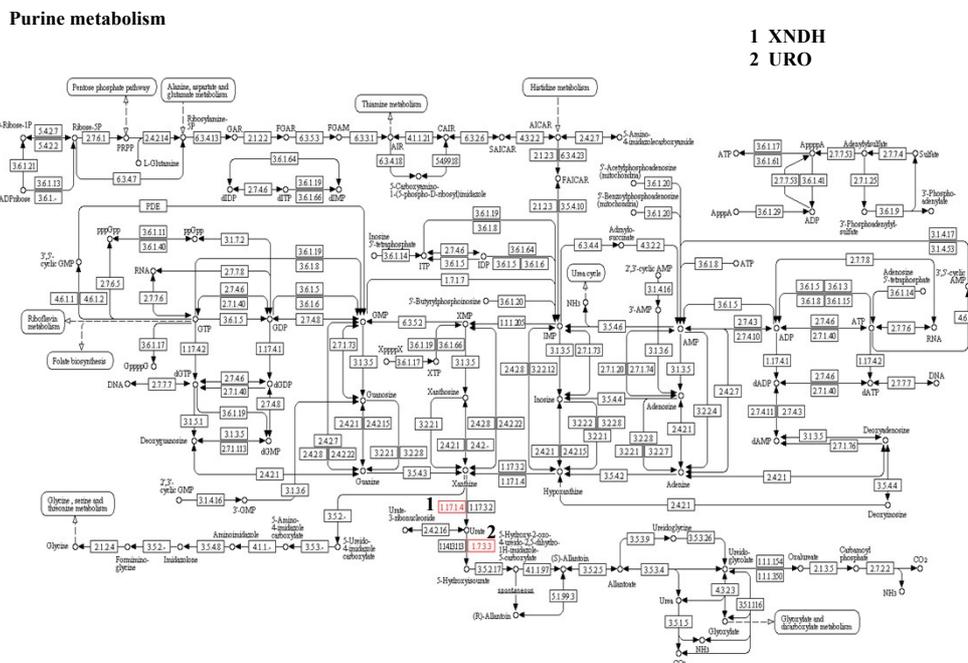
Fig. S14 The pathways in the n-glycan biosynthesis from KEGG of reactions in co-set5 in set 2, reactions of co-set4 set 1, and the reactions associated with lethal gene pair CRv4_Au5.s16.g6321.t1 and CRv4_Au5.s12.g3687.t1 with average profile distance 3.6 with average profile distance of 3.2,1.5 and 3.6, respectively. In this figure the red rectangles, indicated with number 2,6, and 7, correspond to reactions of co-set5 in set 2 (Suppl. Data Table S18), where the reaction DOLPTG1 is catalyzed via glycotransferase to activate a nucleotide sugar and converts it into a nucleophillic glycosyl acceptor. The reactions of GLNACT (via N-acetylglucosaminyltransferase) and G12MT2 (via ALG11 mannosyltransferase with a DXD motif) produce the sequon Asn-X-Ser/Thr through DOLPTG1 (via glycosyltransferase). The blue rectangles, indicated with numbers 2,4, and 5, correspond to reactions of co-set4 in set 1 (Suppl. Data Table S17), where both GLNACT and DOLPMT (via dolichyl-phosphate beta-D-mannosyltransferase) are needed as precursor reactions for the production of the Asn-X-Ser/Thr sequon through DOLPMT2 (via dolichyl-phosphate-mannose-glycolipid alpha-mannosyltransferase).The pink rectangles, indicated with numbers 1 and 3, correspond to the reactions associated with the lethal gene pair CRv4_Au5.s16.g6321.t1 and CRv4_Au5.s12.g3687 (Suppl. Data Table S12).CRv4_Au5.s16.g6321.t1 associates the reaction GLNACPT (via UDP-GlcNAc:dolichyl-phosphate N-acetylglucosaminephosphotransferase) in purine metabolism) and CRv4_Au5.s12.g3687.t1 associates the reaction AGAKI (via D-xylose isomerase in the fructose and mannose metabolism). The metabolisms in which both genes are involved are related through the production of mannose-GDP.


1	GLCNACPT
2	GLNACT
3	AGAKI
4	DOLPMT
5	DOLPMT2
6	G12MT2
7	DOLPGT1

16
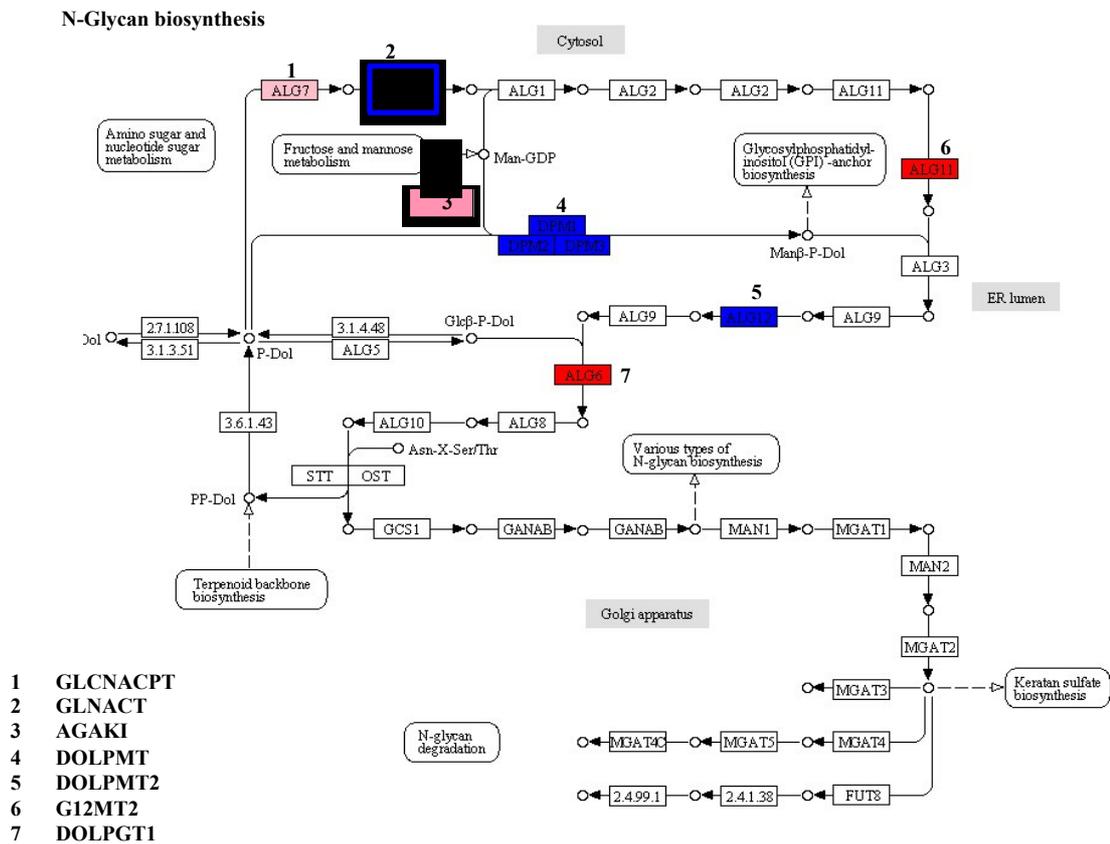
Fig. S15 The pathways of the reactions of co-set1 in set 6 that has 3.2 average profile distance (Table S17) in fatty acid biosynthesis. The figure shows the pathways of the fatty acid biosynthesis as obtained from KEGG. The red rectangles correspond to the reaction 3OAR40 and the reaction 3HAD120, indicated with numbers 1 and 2, respectively. In this biosynthesis, 3HAD120 produces dodecanoic acid through the enzyme (3R)-3-Hydroxybutanoyl-[acyl-carrier-protein] hydro-lyase, and 3OAR40 produces the Butyryl-acp through 3-oxoacyl-[acyl-carrier-protein] reductase (n-C4:0).

Fatty acid biosynthesis

1 3OAR40
2 3HAD120
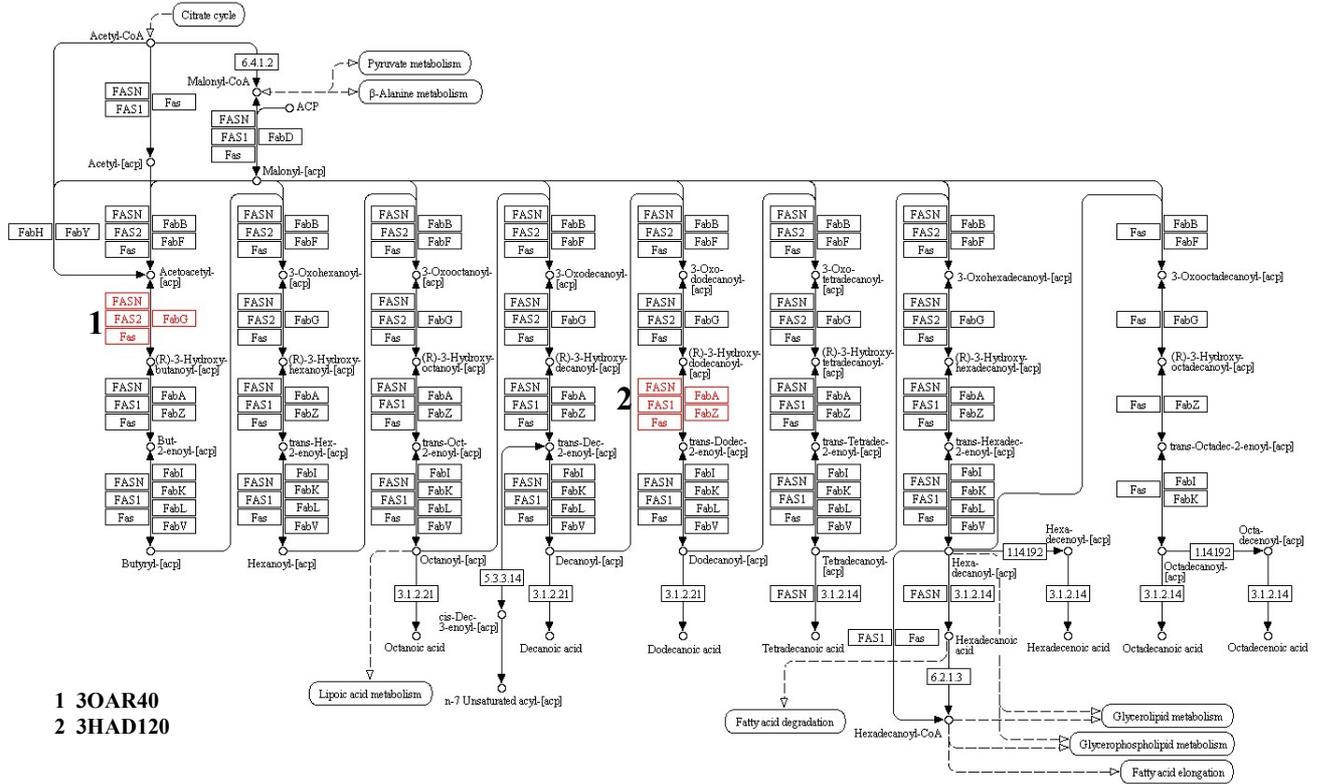
Fig. S16 The pathways of the reactions associated with lethal gene pair CRv4_Au5.s16.g6244.t1 and CRv4_Au5.s17.g7556.t1 (Table S12), that has zero average profile distance. In this figure, the red rectangles show the reactions associated with the genes CRv4_Au5.s16.g6244 (indicated with number 1) and CRv4_Au5.s17.g7556.t1 (indicated with number 2) in steroid biosynthesis and pentose phosphate pathway, respectively, as obtained from KEGG. The gene CRv4_Au5.s17.g7556.t1 is used to catalyze the reaction glucose-6-phosphate-1-dehydrogenase (g6p-A) which converts D-Glucose 6-phophate into D-Glucono-1,5-lactone 6-phosphate in the pentose phosphate pathway. The other gene, CRv4_Au5.s16.g6244.t1, is used in steroids biosynthesis.


Steroid biosynthesis

Pentose Phosphate pathway

1 CRv4_Au5.s16.g6244.t1
2 CRv4_Au5.s17.g7556.t1
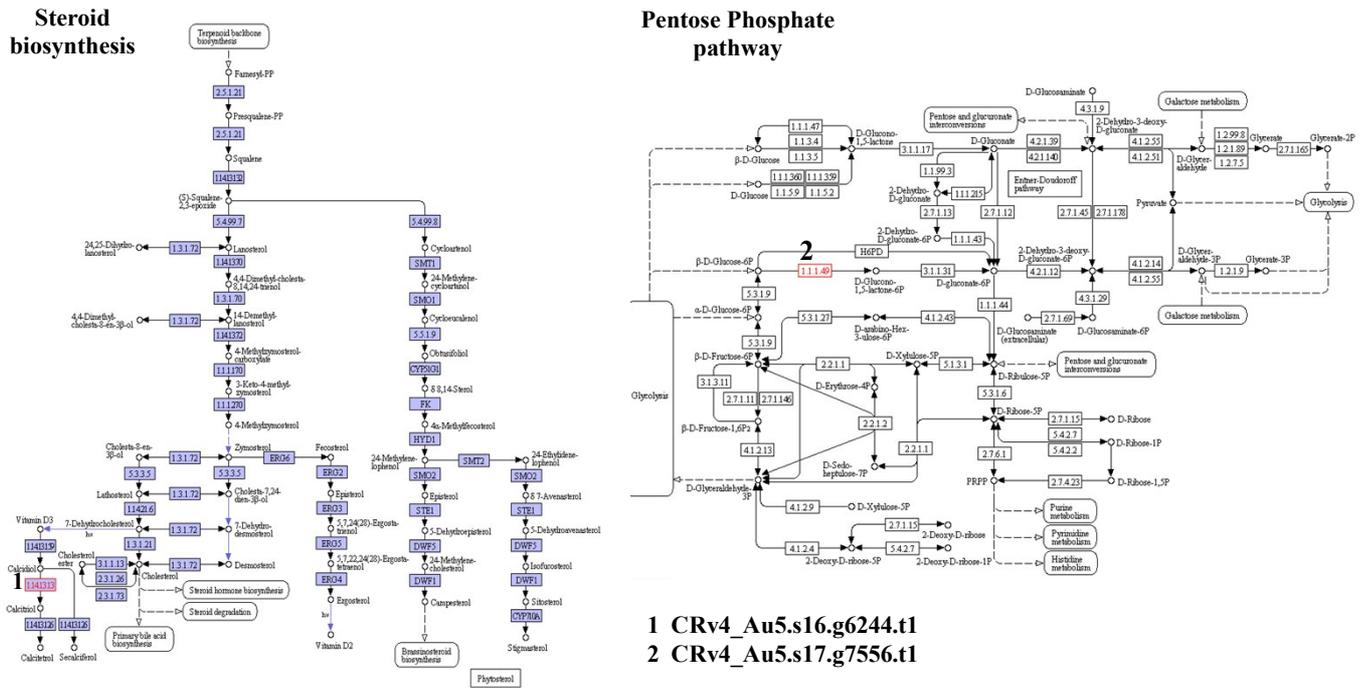
Fig. S17 The pathways of the reactions associated with lethal gene pair CRv4_Au5.s16.g6244.t1 and CRv4_Au5.s17.g7556.t1 (Table S13) that has 3.4 average profile distance. In this figure the red rectangles show reactions associated with lethal gene CRv4_Au5.s9.g15712.t1 (indicated with number 1) and CRv4_Au5.s10.g816.t1 (indicated with number 2) that are utilized in the Glycine, Serine and Threonine metabolism and Glyoxylate and Dicarboxylate metabolism, respectively obtained from KEGG. In the glycine, serine and threonine metabolism, the glycine is converted into L-serine by the reaction glycine hydroxymethyltransferase, specifically associated with CRv4_Au5.s9.g15712.t1 gene. CRv4_Au5.s10.g816.t1 is associated with glycine synthase for the conversion of glycine into ammonia in the glyoxylate and dicarboxylate metabolism

Glycine, Serine, and Threonine metabolism
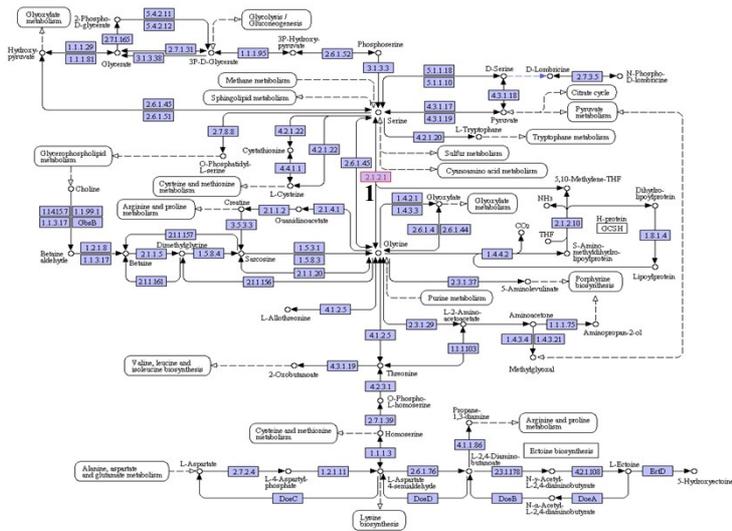
Glyoxylate and Dicarboxylate metabolism
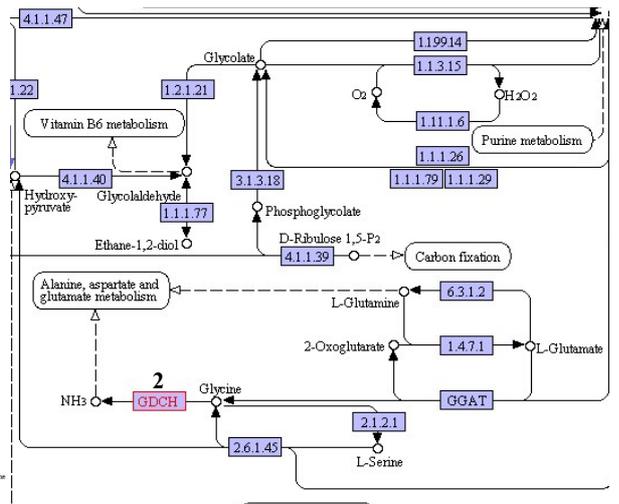
1 **CRv4_Au5.s9.g15712.t1**
2 **CRv4_Au5.s10.g816.t1**

Supplementary Methods: S1-S6

Method S1: Evolutionary affinity assignments. BLASTP searches were done for each protein sequence of each gene in the network against proteomes of available and completely sequenced eukaryotic genomes. These proteome sequences were obtained from SUPERFAMILY database (version: genome_sequence_29-Aug-2010.sql). The top ten hits for each genome (e-value cutoff 0.001) were identified and BLASTCLUST (http://www.ncbi.nlm.nih.gov/IEB/ToolBox/C_DOC/lxr/source/doc/blast/blastclust.html) was done on these hits to identify homology clusters (sequence identity of 30% and length cut-off of 70%) [3]. Each genome was further categorized into one of the thirteen eukaryotic lineages. Evolutionary conservation of *C. reinhardtii* genes was defined on the basis of lineages present in each cluster was identified by BLASTCLUST. Using this approach, each metabolic gene in the network was determined to be conserved/non-conserved in each of the major eukaryotic lineages e.g., fungi, animals, plants, etc.

Method S2: Co-conservation analyses. *Mutual information index (MI) and evolutionary profile distance (PD) calculations* – Mutual information index [4] for neighboring gene pairs *X, Y,* were calculated as follows:

MI (X, Y) = H(X)+H(Y)-H(X,Y)

Where H(N) = Σ(-(frequency of affinity index)*log(frequency of affinity index)); H(X,Y) is the frequency of co-occurrence affinity index and affinity index values are 0 or 1.

Evolutionary profile distance (PD) was obtained by measuring Euclidean distance between each neighboring gene in the real or randomized network as follows:

PD (X, Y) = √ (Σ over N's (value of lineage N of Gene1 – Value of Lineage N of Gene 2)^2), where N = 1 to 13). This definition is a variant of profile similarity used for gene expression clustering and analysis[5].

Profile randomization – For the randomization processes, the phylogenetic profile for each gene is randomized such that the binary affinity values of "1" and "0" were retained but their positional occurrences were varied. Thus the generated randomized phylogenetic profiles evaluated MI and PD values for each pair of genes in the network. We note that for these analyses, the positions of genes in the network were not randomized to evaluate the significance as evolution information content comes from phylogenetic profiles.

Network randomization and p-value calculations - For the MI and PD calculations, significant p-values were estimated by using random networks. These random networks were generated from the original network by maintaining the number of partners per gene while rewiring randomly their partners. We generated 1,000 such random networks and for every pair of genes in each network we calculated MI and PD values then performed randomization of profiles as described (above) and evaluated how often

the difference in the trend between randomized profiles and original profiles for these random networks was found to be equal or higher for the real network.

Method S3: Statistical analysis of synthetic interactions-Kolmogrov-Smirnov test. To check the normality of the profile distance distribution within the synthetically interacting gene sets, we performed a Kolmogorov-Smirnov (KS) test measuring the maximum absolute difference between our data and standard normal distribution. The standard normal distribution was obtained from profile distances between all the 1,086 genes in the network. The KS test determines if the null hypothesis, that the data comes from a standard normal distribution, or the alternative, that it does not come from such a distribution, applies. A result of h = 1 is returned if the test rejects the null hypothesis or 0 otherwise. The decision is transformed into a significance level (p-value) directly compared to a chosen threshold, 5%; if p-value was less than 5%, then the difference is significant, otherwise it is not considered. As illustrated in Table S11, the KS test revealed that the synthetic interactions distribution does not come from a standard normal distribution. A significant difference was observed between the random pairs in the network and the synthetic interactions profile distances at both conditions. These results show that the evolutionary affinities of the genes involved in the synthetic interactions at both conditions LNA and DA differ from overall pairwise distance distributions of the network.

Method S4: Statistical analysis of synthetic interactions and co-sets-Hypergeometric test. In our analysis we performed a hypergeometric test on the profile distance data of the synthetic interactions genes. Three success values were chosen; greater or equal to 1, greater or equal to 2, and greater or equal to 3 for both conditions. The reason behind these choices is that most of the profile distance values lie above 1. The test provides a statistical measure to examine if the synthetic interaction distances are enriched for larger than random network values or not. As illustrated in Table S15, probabilities at dark conditions were higher than those at light condition in most cases. Similarly, statistically significant enrichment was observed in co-sets (see main text for details of these enrichment analyses).

Method S5: Transformed metabolic networks and paired-ortholog analyses. To examine if the found dynamic and static pairs occur in Arabidopsis and yeast networks. The latest version of a yeast metabolic network (http://sourceforge.net/projects/yeast/files/yeast_7.11.zip) was used to generate a gene-centric network. The yeast gene-centric network consisted of 117,559 edges (connections) between 903 metabolic gene products, with an average connectivity of ~260, and clustering coefficient of 0.83. The network has 2 connected components. We transformed the *Arabidopsis* metabolic network, reported by Mintz-oron et al., 2011, which consisted of 1,363 unique reactions and 1,078 metabolites [6]. The Arabidopsis gene-centric network consisted of 369,479 edges (connections) between 1,792 metabolic gene products with an average connectivity of ~73, and clustering coefficient of 0.50. The network has 5 connected components.

The ortholog annotation of genes from multiple genomes between *C. reinhardtii*, *A. thalina* and *S. cerevisiae* were derived from the 11th version of OMA (Orthologus Matrix) Database[7] (http://omabrawer.org). The Chlamydomonas IDs are from mapping of Augustus 5 (with BLAST) to Phytozome version 8 (http://www.phytozome.net/).

Gene ontology analysis was used to identify enrichment, the conservation or rewiring of functional categories of the orthologue genes that had a common ancestor between *C. reinhardtii* to *A. thalina* and *S. cerevisiae.* The entire interologs static and dynamic pairs found between *C. reinhardtii*/*S. cerevisiae* and *C. reinhardtii*/*A. thalina* were used to examine the functional categories. We identified the multiple GO enriched functional categories that conserved and varied between the three model organisms as shown in Figure 4 and Figure S5 A-D. Over-representation of GO terms in a set of genes was determined by using the Biological Networks Gene Ontology tool (BiNGO) (http://ww.psb.ugent.be/cbd/papers/BiNGO/) [8]. BiNGO retrieved the relevant GO annotation then tested for significance using the Hypergeometric test with no corrected for multiple testing. Statistically, hypergeometric distribution is a measure of the probability that describes the number of successes in a sequence of n draws from a finite population without replacement. GO categories that were significantly over-represented among the differentially expressed proteins were identified.

Method S6: Correlated and anti-correlated analyses of expression and flux. The crosstalk between metabolic reactions and gene regulation are essential and fundamental to the living cell. The cell adjusts gene regulatory network and shifts metabolic phenotypic states according to environmental changes. In this segment, differential expression of mRNA under two conditions (dark with acetate and light without acetate) was compared with metabolic flux levels based on the reactions as shown in Figure S8. The reactions were optimized for biomass production under the corresponding light (no acetate) and dark with acetate growth conditions. We considered the collective expression of all genes in reactions rather than individual genes; we computed the total expression levels of all genes from dark (with acetate) to light (without acetate) conditions.

The expression patterns and regulation of a number of enzymes that catalyze the same reaction are often different [9]. To explore the agreement between changes of gene expression and predicted fluxes under light or dark conditions, we compared these results to the reaction's predicted fluxes by flux balance analysis (FBA) to demonstrate the concordances of transcriptional and flux changes (Fig. S8). We further characterized these agreements based on the compartments that the reactions occur (Fig. S9).

Note S1: Co-sets examples for dark and light conditions. Based on Figure 6, co-set1 in set 4 (Table S18) can be extracted as an example of a co-set with an average profile distance of 0, which signifies that the genes associated with the co-set have similar profile distances. The reactions in this co-set, namely URO (via urate oxidase) and XNDH (via xanthine dehydrogenase) were found to be conserved in Amoebazoa, Fungi, Metazoa, Stramenopiles, and Verdiplantae, and they are involved in purine metabolism. Specifically, in XNDH, the xanthine dehydrogenase dehydrogenases the xanthine into Urate, which is oxidized into 5-Hyroxyisourate by urate oxidase and the URO reaction. In *C. reinhardtii*, a urate oxidase type II enzyme was found to be activated early in response to nitrogen starvation [10]. Generally however, URO involves the oxidation of uric acid to form allantoin [11], and XNDH involves the oxidation of hypoxanthine to xanthine and xanthine to uric acid during purine catabolism [12] (Fig. S13).

On the other hand, co-set 5 in set 2 (Table S18) serves as an example of a co-set with a long average profile distance (average distance of co-set 5 is 3.2). Although the genes of this co-set differ evolutionarily due to their long profile distance, their reactions share the same biological function; which is N-glycan biosynthesis (DOLPGT1, GLCNACT, G12MT2) with the exception of (DM_asnglcnacglcnacman-LPAREN-man-RPAREN-man) being a demand reaction. In the n-glycan biosynthesis, the first step is the transferase reaction of Uridine diphosphate N-acetylglucosamine and the lipid P-Dol (dolichol phosphate) to generate a protein bound N-glycan in the cytoplasmic side of the endoplasmic reticulum's (ER's) membrane. This precursor is modified by a series of reactions in the ER and the Golgi, starting with its catalysis by glycotransferase (DOLPTG1) to activate a nucleotide sugar and converts it into a nucleophillic glycosyl acceptor. This ensures proper folding of the nascent protein through enabling its interaction with ER-resident chaperones, such as calnexin and calreticulin. The reactions of GLNACT (N-acetylglucosaminyltransferase) and G12MT2 (via ALG11 mannosyltransferase with a DXD motif) produce the sequon Asn-X-Ser/Thr through DOLPTG1 (glycosyltransferase). Found in many glycosyltransferases that utilize nucleotide sugars, the DXD motif is thought to be involved in the binding of a manganese ion that is required for association of the enzymes with nucleotide sugar substrates. GLCNACT is associated with the CRv4_Au5.s13.g4777.t1 or CRv4_Au5.s16.g6452.t1 genes, and G12MT2 is associated with CRv4_Au5.s14.g5384.t1 (glycosyl transferase). DOLPGT1, and GLCNACT are conserved in all 13 linages except Aloveolata and Choanoflgellida, while the G12MT2 is not conserved in any of the 13 lineages (Fig. S14).

The genes associated with coupled reactions of co-set4 of set 1 have a low profile distance (1.5), which means that they differ slightly in terms of evolutionary affinities. Their genes were conserved in all 13 eukaryotic lineages except Amoebozoa and Euglenzoa. The reactions of this co-set are DOLPMT2 (dolichyl-phosphate-mannose-glycolipid-alpha-mannosyltransferase), GLCNACT, and DOLPMT, and they are also involved in the N-Glycan biosynthesis pathway, the same pathway as the previous

example. Both GLNACT and DOLPMT are needed as precursor reactions for the production of the Asn-X-Ser/Thr sequon through DOLPMT2 (Fig. S14).

An average profile distance of 3.2 was found between genes associated with reactions of co-set 1 in set 6 (Table S17). The coupled reactions 3HAD120 and 3OAR40 are involved in fatty acid biosynthesis (Fig. S14). Where 3HAD120 produces dodecanoic acid through the enzyme (3R)-3-Hydroxybutanoyl-[acyl-carrier-protein] hydro-lyase, and 3OAR40 produces the Butyryl-acp through 3-oxoacyl-[acyl-carrier-protein] reductase (n-C4:0). The reaction 3HAD120 is associated with seven genes; five of them are conserved in Ameobazoa, Heptophceae, and Viridiplantae. The 3OAR40 reaction is carried out through the same seven genes with the addition of CRv4_Au5.s3.g10590.t1. CRv4_Au5.s3.g10590.t1 is conserved in Rhodophyta and Stramenopiles and codes for an enzyme that resembles an enoyl reductase, which catalyzes the second reduction step in fatty acid synthesis. This enzyme is also predicted to belong to the SDRs family of diverse oxidoreductases that have a single domain with structurally conserved Rossmann fold (alpha/beta folding pattern with a central beta-sheet) NAD(P)(H) binding region and a structurally diverse C-terminal region.

Note S2: Synthetic interactions examples for light and dark conditions. From Fig. 6, the largest profile distance was 3.6 and the shortest profile distance was zero. Both values exist frequently in lethal interactions and synthetic sick interactions (80% to 100 % growth reduction).

The longest profile distance in a lethally interacting pair and under light condition was found to be 3.6. This value is because one of the genes is conserved in all lineages while the other is not conserved in any of the 13 eukaryotic lineages. For example the gene CRv4_Au5.s16.g6321.t1 is conserved in all lineages while CRv4_Au5.s12.g3687.t1 is not conserved in any of the lineages. From biological aspect, CRv4_Au5.s16.g6321.t1 is involved in the N-glycan metabolism through the UDP-GlcNAc:dolichyl-phosphate N-acetylglucosaminephosphotransferase reaction, while the other gene (CRv4_Au5.s12.g3687.t1) is utilized in the fructose and mannose metabolism as D-xylose isomerase. The metabolisms in which both genes are involved are related through the production of mannose-GDP (Fig. S14).

In contrast, the shortest profile distance among the lethal gene pairs was zero. When the profile distance is zero, the evolutionary profile of the involved genes is the same. This means they are both either conserved in same lineages or are both conserved in none of the thirteen eukaryotic lineages. For example CRv4_Au5.s17.g7556.t1 (glucose-6-phosphate-1 dehydrogenase) and CRv4_Au5.s16.g6244.t1 (cytochrome P450) are conserved in none of the thirteen lineages and are utilized in the Pentose phosphate pathway and the Biosynthesis of steroids respectively. The gene CRv4_Au5.s17.g7556.t1 is used to catalyze the reaction glucose-6-phosphate-1-dehydrogenase (g6p-A) which converts D-Glucose 6-phophate into D-Glucono-1,5-lactone 6-phosphate in the pentose phosphate pathway. The other gene, CRv4_Au5.s16.g6244.t1, is used in steroids biosynthesis specifically for the oxidation of calcidiol into calcitriol (Fig. S16).

Under dark growth, the lethal gene pair, CRv4_Au5.s9.g15712.t1 and CRv4_Au5.s10.g816.t1, has a large profile distance value of 3.4. The first gene is conserved in all lineages while other one is not conserved in any other lineages. Biologically, CRv4_Au5.s9.g15712.t1 is involved in glycine, serine and threonine metabolism, while CRv4_Au5.s10.g816.t1 is involved in Nitrogen metabolism. In the glycine, serine and threonine metabolism, the glycine is converted into L-serine by the reaction glycine hydroxymethyltransferase, specifically associated with C Rv4_Au5.s9.g15712.t1 gene. CRv4_Au5.s10.g816.t1 is associated with glycine synthase for the conversion of glycine into ammonia in the nitrogen metabolism (Fig. S17).

References

1.   R. L. Chang, L. Ghamsari, A. Manichaikul, E. F. Hom, S. Balaji, W. Fu, Y. Shen, T. Hao, B. O. Palsson, K. Salehi-Ashtiani and J. A. Papin, *Mol Syst Biol*, 2011, **7**, 518.
2.   H. Yu, Y. Xia, V. Trifonov and M. Gerstein, *Genome Biol*, 2006, **7**, R55.
3.   S. F. Altschul, W. Gish, W. Miller, E. W. Myers and D. J. Lipman, *Journal of molecular biology*, 1990, **215**, 403-410.
4.   S. Balaji, M. M. Babu and L. Aravind, *Journal of molecular biology*, 2007, **372**, 1108-1122.
5.   M. B. Eisen, P. T. Spellman, P. O. Brown and D. Botstein, *Proceedings of the National Academy of Sciences of the United States of America*, 1998, **95**, 14863-14868.
6.   S. Mintz-Oron, S. Meir, S. Malitsky, E. Ruppin, A. Aharoni and T. Shlomi, *Proceedings of the National Academy of Sciences*, 2012, **109**, 339-344.
7.   A. M. Altenhoff, A. Schneider, G. H. Gonnet and C. Dessimoz, *Nucleic Acids Research*, 2011, **39**, D289-D294.
8.   S. Maere, K. Heymans and M. Kuiper, *Bioinformatics*, 2005, **21**, 3448-3449.
9.   Y. Bilu, T. Shlomi, N. Barkai and E. Ruppin, *PLoS Comput Biol*, 2006, **2**, e106.
10.   F. Merchán, H. van den Ende, E. Fernández and C. F. Beck, *Planta*, 2001, **213**, 309-317.
11.   K. Motojima, S. Kanaya and S. Goto, *Journal of Biological Chemistry*, 1988, **263**, 16677-16681.
12.   I. Kimiyoshi, A. Yoshihiro, N. Kumi, M. Shinsei, H. Tatsuo, S. Osamu, S. Nobuyoshi and N. Takeshi, *Gene*, 1993, **133**, 279-284.