Supporting Information

Zelezniak et al. 10.1073/pnas.1421834112

SI Methods

SCS Algorithm.

```
sc_scores = {}
for A in C:
 solutions = []
 do:
  species_set = minimize_donors_set(A, C, previously found solutions)
  solutions += species_set
 while species_set != ϕ
 for B in C:
  if A != B:
   sc_scores[(A, B)] = len([solutions containing B])/len(solutions)
```

The minimize_donors_set routine solves MILP problem where the objective is to minimize the number of donating species in community C while ensuring the growth of species A and satisfying all steady-state constraints as well as the uptake/secretion flux bounds. Binary constraints θ_s control the ON/OFF-state of member species. For $\theta_s = 0$, sum of all secretion fluxes for the species s are set to 0. Additional constraints ensuring biomass production for all ON-state species are also included. To enumerate all possible solutions, each time a new solution is found a new constraint blocking it from the further search space is added. Finally, we also ensure that vitamins are not used by any species as a carbon source by restricting their uptake ($v_{vit.uptake}$) to minimal requirement for growth (see *Methods*):

$$min \sum_{s \in C \setminus A} \theta_s$$

subject to:

$$S_s v_s = 0\,, \quad \forall s \in C$$

$$v_i^{lb} \le v_i \le v_i^{up}, \quad \forall i \in s$$

$$v_{A,growth} = 1$$

$$\sum_{S \in L} v_{s,secretion} - \gamma \cdot \theta_s \le 0,\ \forall S \in C,\ \theta_s \in \{0,1\}, \gamma > argmax(v)$$

$$v_{A,growth} - v_{A,\min\_growth} * \theta_s \ge 0\,, \quad \forall s \in C \setminus A$$

$$\sum_{s \in L} \theta_s < |\mathrm{L}|\,, \quad \forall L \in \{previously\ found\ solutions\}$$

$$-\varepsilon \le v_{vit.uptake} - v_{measured\ vit.uptake} * v_{vit.uptake} \le \varepsilon$$

MUS Algorithm.

```
mu_scores = {}
for A in C:
 solutions = []
 do:
  metabolites_set = minimize_received_metabolites_set(A, C, previously found solutions)
  solutions += metabolites_set
 while metabolites_set != ϕ
 for m in A.received_metabolites:
  mu_scores[(A, m)] = len([solutions containing m])/len(solutions)
```

The minimize_received_metabolites_set routine solves MILP problem analogous to the minimize_donors_set routine. Its objective is to find a minimal set of metabolites donated to species A by other community members. Here also steady-state constraints and uptake/secretion flux bounds have to be satisfied. Here too we introduce binary variable θ_m. $\theta_m = 1$ represents activation of uptake of metabolite m. All found solutions are excluded from the solution space by adding appropriate constraints. We also ensure that vitamins are not used by any species as a carbon source by restricting their uptake ($v_{vit.uptake}$) to minimal requirement for growth (see *Methods*):

$$\min \sum_{m \in \{metabolites \text{ from A}\}} \theta_m$$

subject to:

$$S_s v_s = 0, \quad \forall s \in C$$

$$v^{lb} \le v \le v^{up}$$

$$V_{A,growth} = 1$$

$$v_m - \gamma * \theta_m \le 0, \quad \forall m \in \{\text{metabolites uptakes from } A\}$$

$$\sum_{m \in L} \theta_m < |L|, \quad \forall L \in \{\text{previously found solutions}\}$$

$$-\varepsilon \le v_{vit.uptake} - v_{measured\ vit.uptake} * v_{vit.uptake} \le \varepsilon$$

MPS Algorithm. MPS is a binary value showing whether species B can produce metabolite m under a given nutritional environment.

```
mp_scores = {}
for B in C:
  for m in {metabolites produced by B}
    mp_scores[(B, m)] = maximize_metabolite_yield(m, C) >= 1
```

$$\max v_m, \quad m \in \{\text{secreted metabolites from } B\}$$

subject to:

$$S_s v_s = 0, \quad \forall s \in C$$

$$v^{lb} \le v \le v^{up}$$

$$-\varepsilon \le v_{vit.uptake} - v_{measured\ vit.uptake} * v_{vit.uptake} \le \varepsilon$$

Difference Between MIP and SMETANA Score. Whereas the MIP estimates the maximum number of nutritional components that a community can provide for itself (through interspecies metabolite exchanges), the SMETANA score quantifies the extent of interspecies exchanges. To account for the complexity of possible interspecies exchanges, due to metabolic plasticity, the SMETANA score is decomposed into three distinct factors: (*i*) SCS, which accounts for the plasticity at the level of community; (*ii*) MUS, which accounts for the plasticity at the level of nutritional requirements of member species; and (*iii*) MPS, which accounts for the by-product secretion capabilities of member species.

Curated Models Used for Estimation of Reaction Directions. Manually reconstructed models for 16 different species were obtained from the ModelSEED resource (1): *Acinetobacter baylyi* ADP1 (2), *Escherichia coli* K-12 MG1655 (3), *Methanosarcina barkeri* Fusaro (4), *Bacilus subtilis* (1), *Lactococcus lactis* ssp. *lactis* IL1403 (5), *Bacillus subtilis* 168 (6), *Mycoplasma pneumoniae* M129 (7), *Saccharomyces cerevisiae* S288c (8), *Helicobacter pylori* 26695 (9), *Pseudomonas putida* KT2440 (10), *Escherichia coli* K-12 MG1655 (11), *Saccharomyces cerevisiae* S288c (12), *Pseudomonas aeruginosa* PA01 (13), *Saccharomyces cerevisiae* S288c (14), *Mycoplasma genitalium* G-37 (15), and *Staphylococcus aureus* N315 (16).

1. Henry CS, et al. (2010) High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat Biotechnol* 28(9):977–982.
2. Durot M, et al. (2008) Iterative reconstruction of a global metabolic model of Acinetobacter baylyi ADP1 using high-throughput growth phenotype and gene essentiality data. *BMC Syst Biol* 2:85.
3. Feist AM, 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.
4. Feist AM, Scholten JCM, Palsson BO, Brockman FJ, Ideker T (2006) Modeling methanogenesis with a genome-scale metabolic reconstruction of Methanosarcina barkeri. *Mol Syst Biol* 2: 0004.
5. Oliveira AP, Nielsen J, Förster J (2005) Modeling Lactococcus lactis using a genome-scale flux model. *BMC Microbiol* 5:39.
6. Henry CS, Zinner JF, Cohoon MP, Stevens RL (2009) iBsu1103: A new genome-scale metabolic model of Bacillus subtilis based on SEED annotations. *Genome Biol* 10(6):R69.

7. Yus E, et al. (2009) Impact of genome reduction on bacterial metabolism and its regulation. *Science* 326(5957):1263–1268.
8. Nookaew I, et al. (2008) The genome-scale metabolic model iIN800 of Saccharomyces cerevisiae and its validation: A scaffold to query lipid metabolism. *BMC Syst Biol* 2:71.
9. Thiele I, Vo TD, Price ND, Palsson BO (2005) Expanded metabolic reconstruction of Helicobacter pylori (iIT341 GSM/GPR): An in silico genome-scale characterization of single- and double-deletion mutants. *J Bacteriol* 187(16):5818–5830.
10. Nogales J, Palsson BO, Thiele I (2008) A genome-scale metabolic reconstruction of Pseudomonas putida KT2440: iJN746 as a cell factory. *BMC Syst Biol* 2:79.
11. Reed JL, Vo TD, Schilling CH, Palsson BO (2003) An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR). *Genome Biol* 4(9):R54.
12. Zomorrodi AR, Maranas CD (2010) Improving the iMM904 S. cerevisiae metabolic model using essentiality and synthetic lethality data. *BMC Syst Biol* 4:178.
13. Oberhardt MA, Puchałka J, Fryer KE, Martins dos Santos VA, Papin JA (2008) Genome-scale metabolic network analysis of the opportunistic pathogen Pseudomonas aeruginosa PAO1. *J Bacteriol* 190(8):2790–2803.
14. Duarte NC, Herrgård MJ, Palsson BO (2004) Reconstruction and validation of Saccharomyces cerevisiae iND750, a fully compartmentalized genome-scale metabolic model. *Genome Res* 14(7):1298–1309.
15. Suthers PF, et al. (2009) A genome-scale metabolic reconstruction of Mycoplasma genitalium, iPS189. *PLOS Comput Biol* 5(2):e1000285.
16. Becker SA, Palsson BO (2005) Genome-scale reconstruction of the metabolic network in Staphylococcus aureus N315: An initial draft to the two-dimensional annotation. *BMC Microbiol* 5:8.

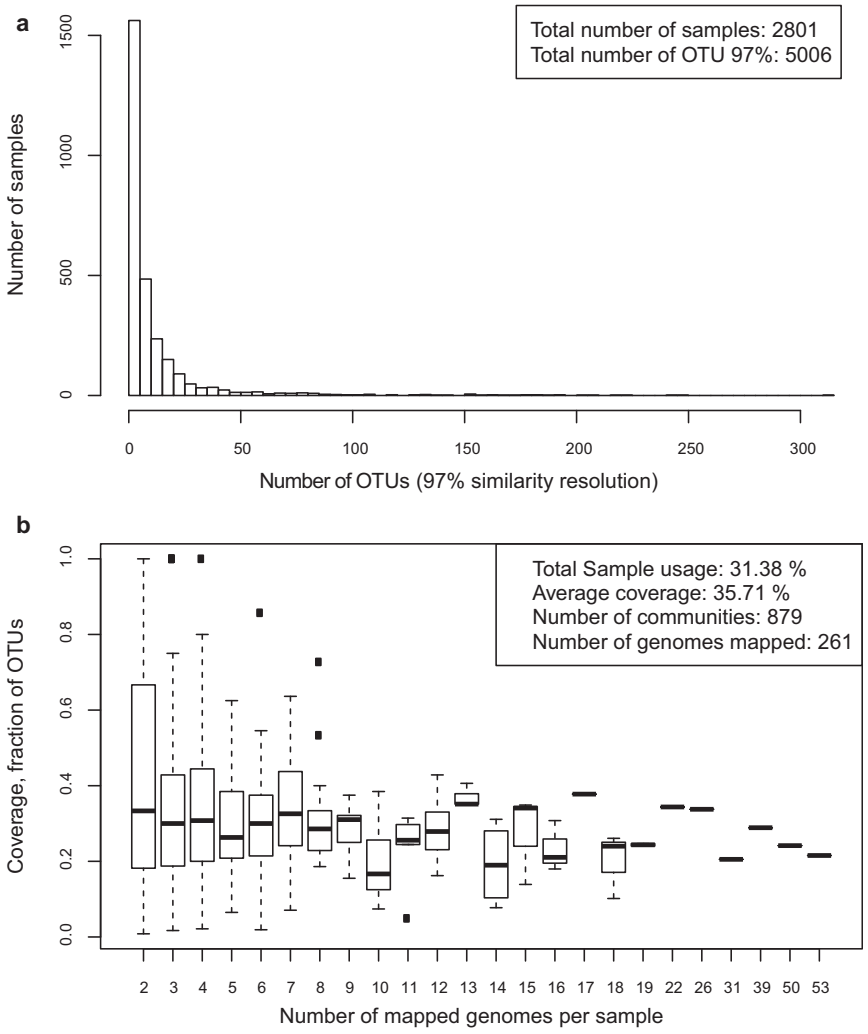

Fig. S1. Summary statistics of the OTU to genome mappings. (*A*) Distribution of OTUs across different sampling sites. (*B*) Sample coverage by mapped genomes.

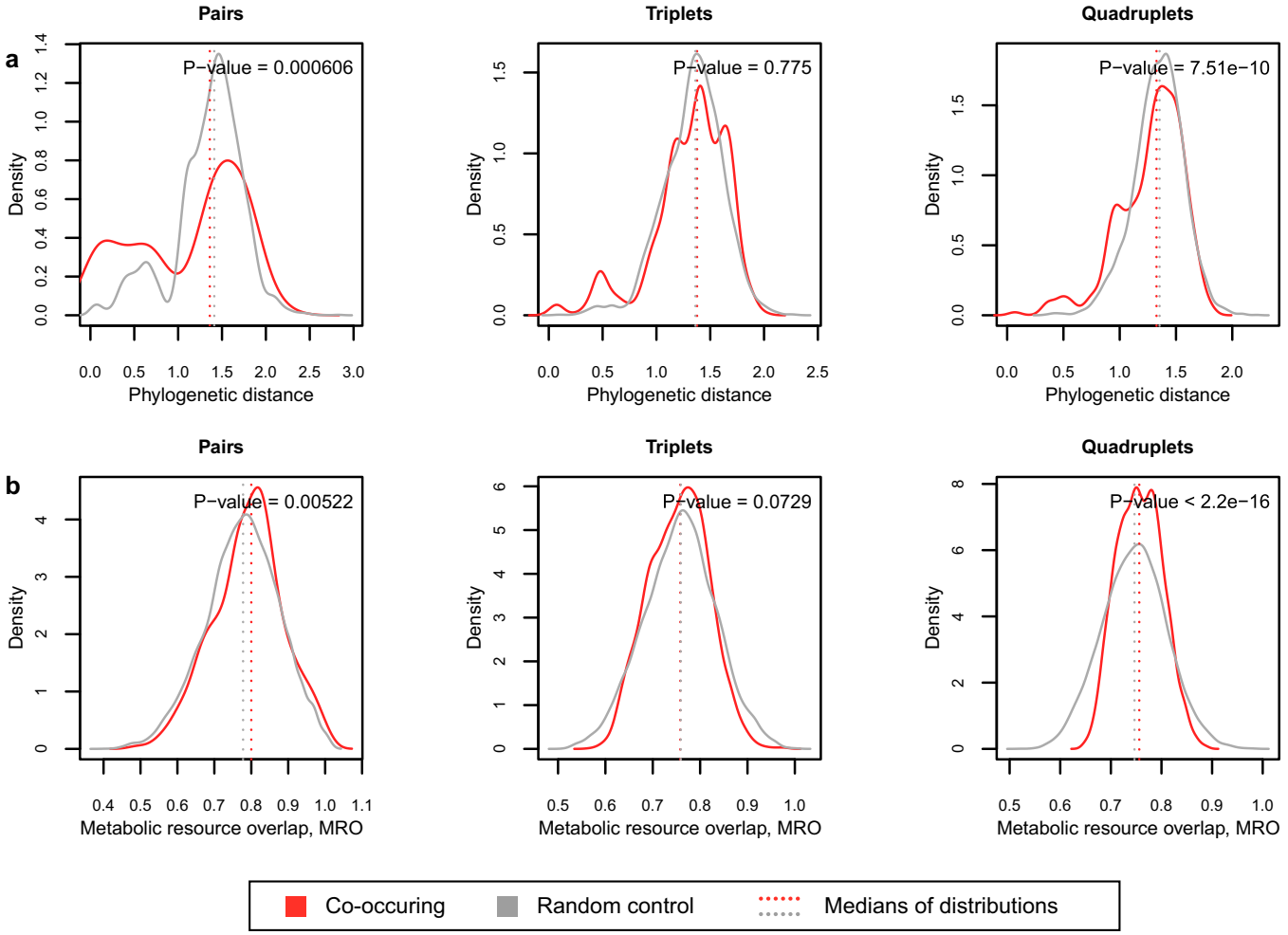

Fig. S2. Competition metrics of co-occurring subcommunities compared with random assemblies. Neither phylogenetic distance (*A*) nor MRO (*B*) can discern co-occurring subcommunities from random assemblies. Results for the random control are based on simulations of 10,000 groups randomly assembled from the same species pool.

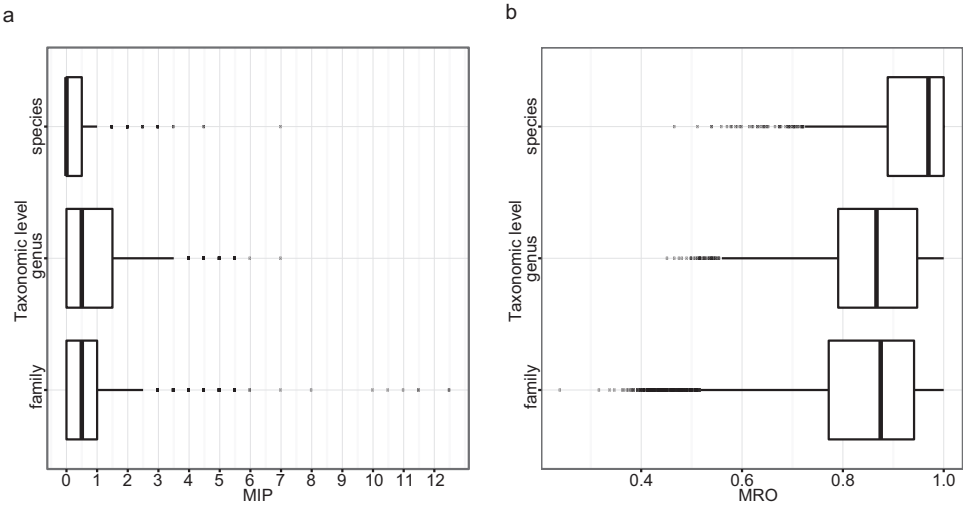

Fig. S3. MIP (*A*) and MRO (*B*) distribution for the species pairs belonging to the same taxonomic level.

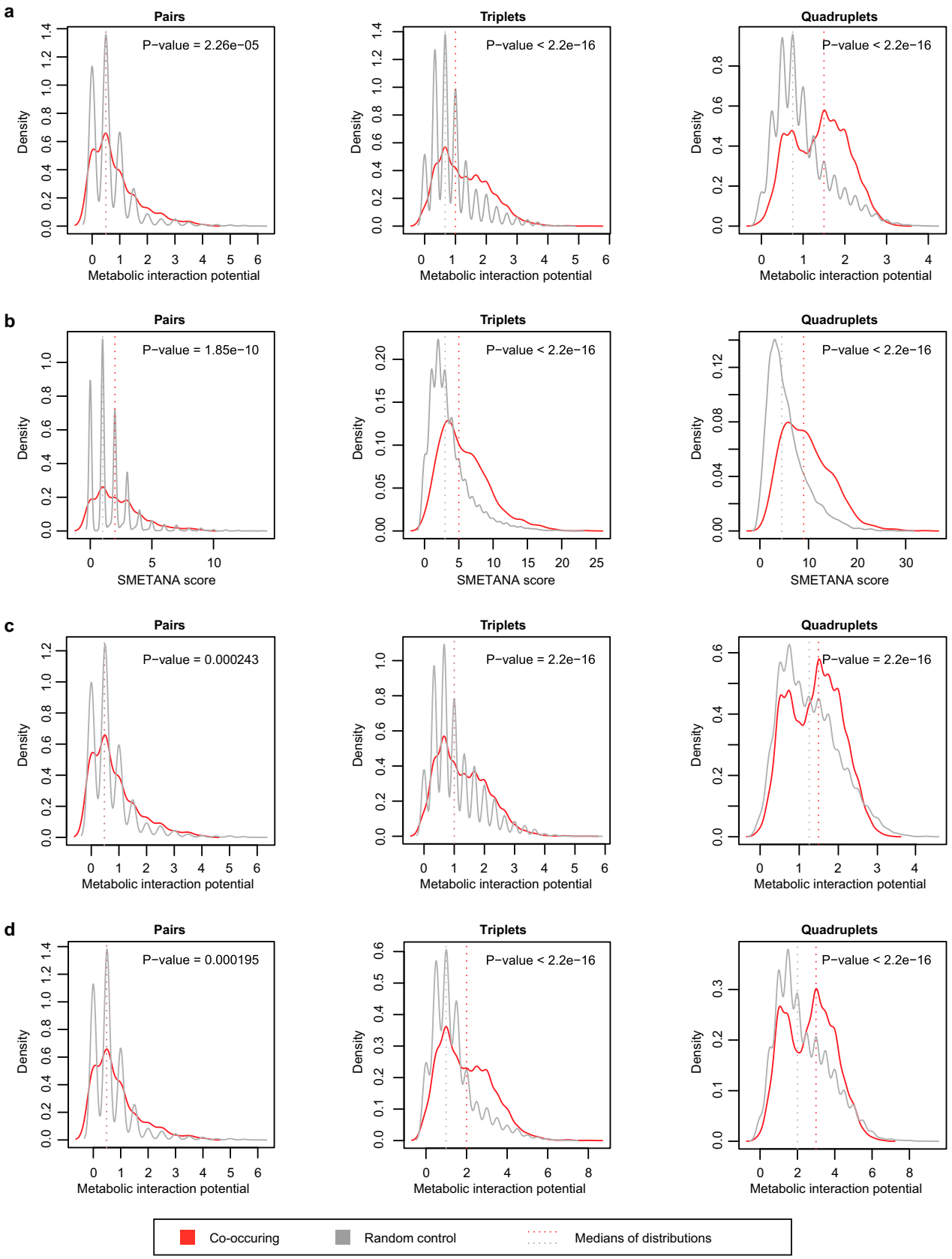

Fig. S4. Cooperation metrics of co-occurring communities compared with random assemblies. Co-occurring communities feature higher interaction potential (*A*) and higher metabolic coupling (*B*) than non-co-occurring groups. Results for the random control are based on simulations of 10,000 groups randomly assembled from the same species pool. Choosing these random assemblies from habitat-filtered (*C*) or rich habitats (*D*) preserves the distinction of co-occurring groups. Habitats for which the descriptions lacked the words "water" or "rock" were considered as rich. Habitats with the annotation "sewer" were, however, retained as rich.

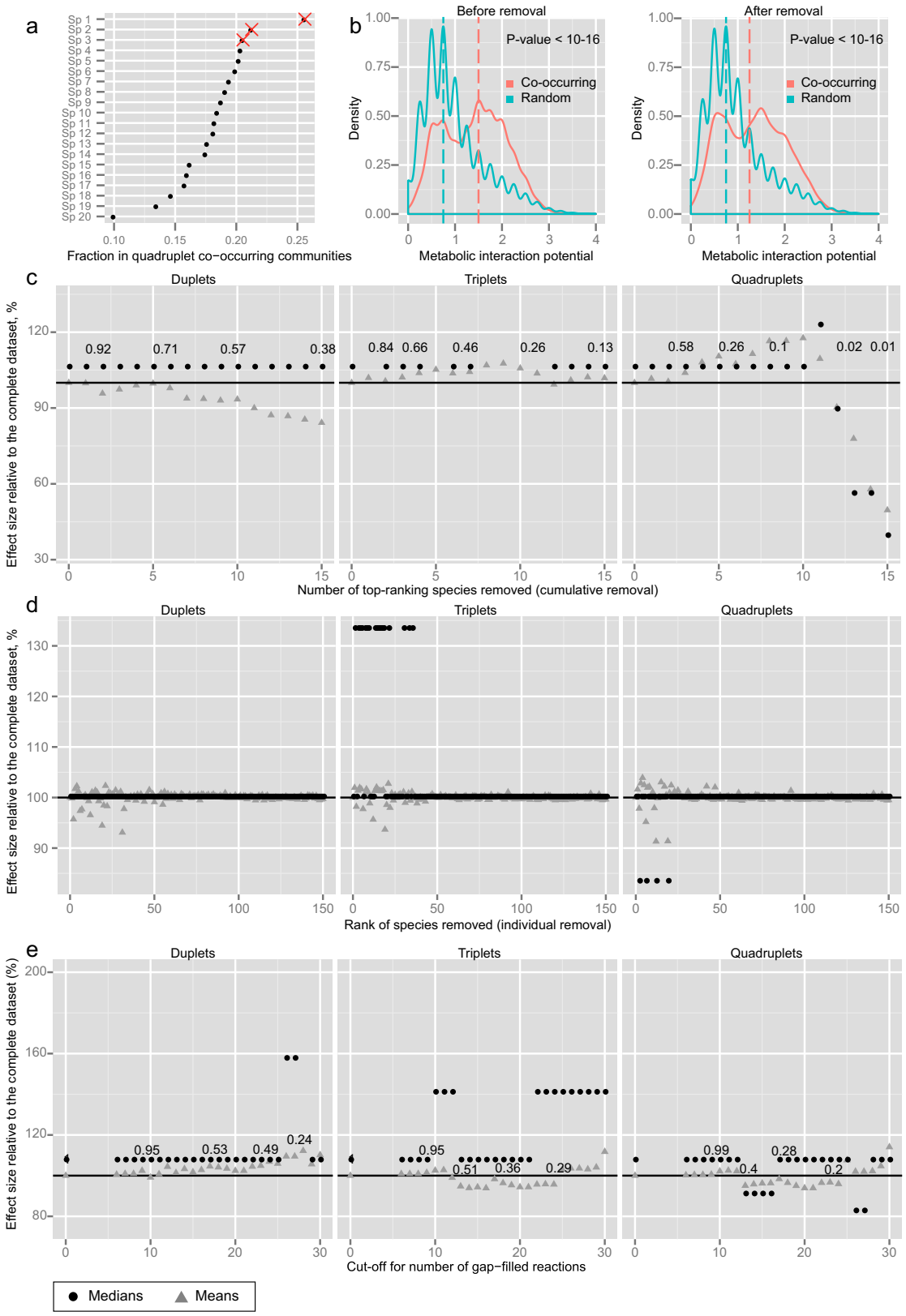

Fig. S5. Robustness of MIP values of co-occurring subcommunities toward species composition and gap-filled reactions in metabolic models. (*A* and *B*) Removal of the three most frequent species (found in 52% of all co-occurring subcommunities) retains the contrast between the co-occurring subcommunities (red density plots) and random assemblies (gray density plots). (*C* and *D*) Cumulative (*C*) or individual (*D*) removal of top-ranking co-occurring species (Table S3) retains the contrast between co-occurring subcommunities and random assemblies. Shown are the effect sizes following species removal [log ratio of the medians (circles) or means (triangles)]. (*E*) The use of subsets of models with different numbers of gap-filled reactions does not affect the MIP effect size discriminating co-occurring subcommunities from random assemblies. Numbers in the plot show the fraction of remaining subcommunities, after removing those containing species with fewer than *n* gap-filled reactions (*x* axis). We here note that the distribution of the number of gap-filled reactions in species forming co-occurring subcommunities is similar to that for the non-co-occurring subcommunities ($P = 0.66$).

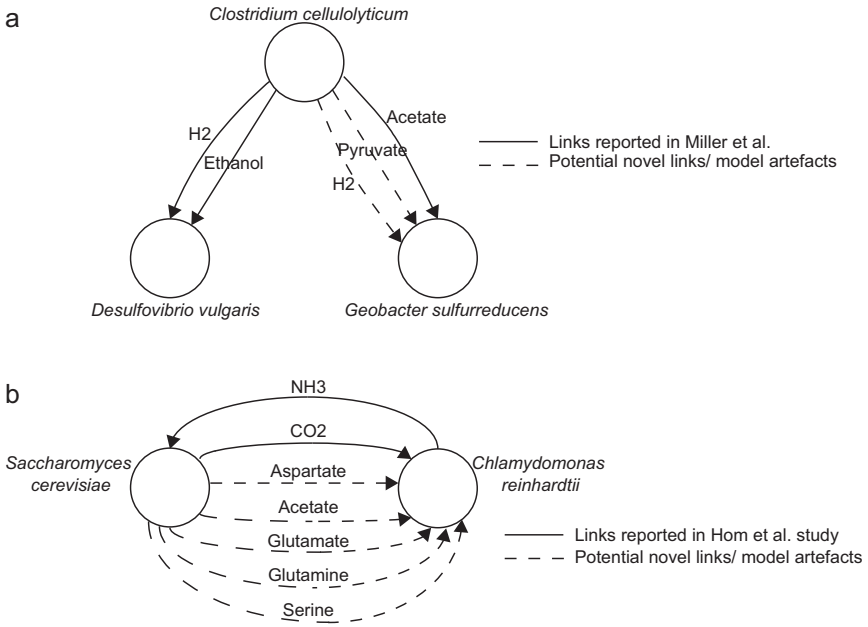

Fig. S6. Predicted metabolic interactions accurately capture experimental results (*A*) in a three-species community reported by Miller et al. (1) and (*B*) a yeast–algal community reported by Hom and Murray (2). Models for three species community were obtained from Zomorrodi and Maranas (3). Published species-level models of *C. reinhardtii* (4) and the *S. cerevisiae* (5) were used for reconstructing the yeast–algal community model. NO_3^- and H_2S were used in simulations instead of NO_2^- and SO_4^{2-} to maintain compatibility with the species models. Glucose uptake reaction in the *C. reinhardtii* model was blocked as suggested in the experimental study (2). Dotted arrows mark potential novel interactions, or possibly model artifacts [e.g., predicted pyruvate link in the community (*A*)].

1. Miller LD, et al. (2010) Establishment and metabolic analysis of a model microbial community for understanding trophic and electron accepting interactions of subsurface anaerobic environments. *BMC Microbiol* 10:149.
2. Hom EF, Murray AW (2014) Plant-fungal ecology. Niche engineering demonstrates a latent capacity for fungal-algal mutualism. *Science* 345(6192):94–98.
3. Zomorrodi AR, Maranas CD (2012) OptCom: A multi-level optimization framework for the metabolic modeling and analysis of microbial communities. *PLOS Comput Biol* 8(2):e1002363.
4. Dal'Molin CGD, Quek LE, Palfreyman RW, Nielsen LK (2011) AlgaGEM—A genome-scale metabolic reconstruction of algae based on the Chlamydomonas reinhardtii genome. *BMC Genomics* 12(Suppl 4):S5.
5. Zomorrodi AR, Maranas CD (2010) Improving the iMM904 S. cerevisiae metabolic model using essentiality and synthetic lethality data. *BMC Syst Biol* 4:178.

Table S1. Summary statistics of the number of co-occurring subcommunities and OTU to genome mappings

OTUs/sites		Pairs	Triplets	Quadruplets	Total
Co-occurring subcommunities identified, FDR <0.01		381	3,322	3,518	7,221
No. of possible subcommunities observed in samples		2,379	21,570	77,664	10,1613
No. of OTUs with 97% sequence identity	5,006				
No. of OTUs mapped with 95% sequence identity and appearing at least three times among samples	536				
No. of sampling sites in which mapped genomes were present	1,297				

Table S2. List of compounds assumed to be present in all environments

Compound ID	Compound name synonyms
cpd04098	Arsenite
cpd00971	Na+, sodium
cpd00254	Mg(2+), Mg, Mg2+, magnesium
cpd00009	Orthophosphoric acid, phosphoric acid, phosphate, orthophosphate
cpd00209	Nitric acid, nitrate
cpd00074	Sulfur, precipitated, S, sulfur
cpd01012	Cd2+, cadmium
cpd00528	N2, nitrogen
cpd00058	Copper, Cu+, Cu(I), Cu1+, copper1, Cu(II), Cu2+, copper2
cpd00099	Hydrochloride, hydrogen chloride, hydrochloric acid, chloride ion, Cl–, HCl, chloride
cpd00063	Ca(2+), Ca^{2+}, calcium
cpd10515	Iron(2+), ferrous ion, Fe(II), Fe2+
cpd00048	SLF, sulfuric acid, sulfate
cpd00012	PPi, diphosphate, pyrophosphoric acid, pyrophosphate
cpd00034	Zn(II), Zn2+, zinc
cpd00007	dioxygen, O2, oxygen
cpd10516	fe3, Iron(3+), ferric ion, Fe(III), Fe3+
cpd00011	Carbon dioxide, CO_2
cpd00075	Nitrite
cpd00244	Ni2+, nickel
cpd00067	H+
cpd00149	Co2+, cobalt
cpd04097	Pb2+, Pb, lead
cpd00030	Mn(III), Mn(II), Mn2+, manganese
cpd00205	K+, potassium
cpd00001	OH–, HO–, water, H_2O

Table S3. List of the top 20 frequent species mapping to co-occurring subcommunities

Name	Model_seed ID	Fraction
Corynebacterium diphtheriae NCTC 13129	Seed257309.4.51162	0.206938776
Rothia mucilaginosa DY-18	Seed680646.3.51162	0.15755102
Streptococcus sanguinis SK36	Seed388919.8.51162	0.156054422
Streptococcus mitis B6	Seed365659.3.51162	0.14952381
Staphylococcus lugdunensis HKU09-01	Seed698737.3.51162	0.147210884
Streptococcus pyogenes NZ131	Seed471876.6.51162	0.144489796
Staphylococcus aureus subsp. *aureus* USA300_TCH1516	Seed451516.9.51162	0.143129252
Methylobacterium radiotolerans JCM 2831	Seed426355.14.51162	0.142585034
Streptococcus gordonii str. Challis substr. CH1	Seed467705.9.51162	0.140952381
Acinetobacter baumannii AYE	Seed509173.8.51162	0.139455782
Corynebacterium aurimucosum ATCC 700975	Seed548476.3.51162	0.139455782
Anaerococcus prevotii DSM 20548	Seed525919.6.51162	0.138231293
Neisseria meningitidis alpha14	Seed662598.3.51162	0.136734694
Neisseria meningitidis FAM18	Seed272831.7.51162	0.135782313
Streptococcus pneumoniae D39	Seed373153.27.51162	0.132244898
Kocuria rhizophila DC2201	Seed378753.5.51162	0.13170068
Veillonella parvula DSM 2008	Seed479436.4.51162	0.123809524
Staphylococcus saprophyticus subsp. *saprophyticus* ATCC 15305	Seed342451.4.51162	0.116326531
Lactobacillus crispatus ST1	Seed748671.3.51162	0.113877551
Propionibacterium acnes SK137	Seed553199.9.51162	0.094829932

Note that the species mapping is subject to the pool of genome-sequenced species/strains against which mapping is performed.