

1 **Full Title:** Managing Uncertainty in Metabolic Network Structure and Improving Predictions Using
2 EnsembleFBA

3
4 **Short Title:** Ensembles of Genome-Scale Metabolic Networks

5
6 Matthew B. Biggs¹ and Jason A. Papin^{1*}

7
8 1. Department of Biomedical Engineering, University of Virginia, Charlottesville, VA, USA

9
10 * Corresponding author

11 E-mail: papin@virginia.edu

12

1 **Abstract**

2 Genome-scale metabolic network reconstructions (GENREs) are repositories of knowledge about the
3 metabolic processes that occur in an organism. GENREs have been used to discover and interpret
4 metabolic functions, and to engineer novel network structures. A major barrier preventing more
5 widespread use of GENREs, particularly to study non-model organisms, is the extensive time required to
6 produce a high-quality GENRE. Many automated approaches have been developed which reduce this
7 time requirement, but automatically-reconstructed draft GENREs still require curation before useful
8 predictions can be made. We present a novel ensemble approach to the analysis of GENREs which
9 improves the predictive capabilities of draft GENREs and is compatible with many automated
10 reconstruction approaches. We refer to this new approach as Ensemble Flux Balance Analysis
11 (EnsembleFBA). We validate EnsembleFBA by predicting growth and gene essentiality in the model
12 organism *Pseudomonas aeruginosa* UCBPP-PA14. We demonstrate how EnsembleFBA can be included
13 in a systems biology workflow by predicting essential genes in six *Streptococcus* species and mapping the
14 essential genes to small molecule ligands from DrugBank. We found that some metabolic subsystems
15 contribute disproportionately to the set of predicted essential reactions in a way that is unique to each
16 *Streptococcus* species. These species-specific network structures lead to species-specific outcomes from
17 small molecule interactions. Through these analyses of *P. aeruginosa* and six *Streptococci*, we show that
18 ensembles increase the quality of predictions without drastically increasing reconstruction time, thus
19 making GENRE approaches more practical for applications which require predictions for many non-
20 model organisms. All of our functions and accompanying example code are available in an open online
21 repository.

22

23 **Author Summary**

24 Metabolism is the driving force behind all biological activity. Genome-scale metabolic network
25 reconstructions (GENREs) are representations of metabolic systems that can be analyzed mathematically
26 to make predictions about how a biochemical system will behave as well as to design biochemical

1 systems with new properties. GENREs have traditionally been reconstructed manually, which can require
2 extensive time and effort. Recent software solutions automate the process (drastically reducing the
3 required effort) but the resulting GENREs are of lower quality and produce less reliable predictions than
4 the manually-curated versions. We present a novel method (“EnsembleFBA”) which overcomes
5 uncertainties involved in automated reconstruction by pooling many different draft GENREs together into
6 an ensemble. We tested EnsembleFBA by predicting the growth and essential genes of the common
7 pathogen *Pseudomonas aeruginosa*. We found that when predicting growth or essential genes, ensembles
8 of GENREs achieved much better precision or captured many more essential genes than any of the
9 individual GENREs within the ensemble. By improving the predictions that can be made with
10 automatically-generated GENREs, we open the door to studying systems which would otherwise be
11 infeasible.

12

13 **Introduction**

14 Metabolism is the driving force behind the wondrous flurry of biological activity carpeting our planet. An
15 organism’s metabolism is determined by the metabolic enzymes encoded in its genome, the chemical
16 reactions catalyzed by those enzymes, and whether or not those enzymes are actively expressed [1]. The
17 simplest bacteria have hundreds of metabolic enzymes, while the most complex eukaryotes have
18 thousands. The products of these enzymatic reactions serve as substrates for other reactions, such that the
19 chemical transformations carried out in a cell can be represented as a vast network [2]. Mass and energy
20 flow through such networks, transforming environmental inputs into the building blocks of life. Every
21 species has a unique metabolic network driving its growth and interaction with the environment.

22

23 Genome-scale metabolic network reconstructions (GENREs) are formal representations of metabolic
24 networks [3]. GENREs serve as a comprehensive collection of metabolic knowledge about a particular
25 organism and they are amenable to mathematical analysis [4]. The process of reconstructing a GENRE
26 takes months to years, but the reconstruction process often leads to new discoveries [5]. Mathematical

1 analysis of GENREs gives insight into how particular metabolic pathways are used by an organism, what
2 substrates it can utilize, which of its genes are essential in a given environment, how a metabolic network
3 can be engineered to produce more of a desired product, or which enzymes within the network should be
4 targeted in order to halt growth in an organism [6–8]. The reconstruction and analysis of GENREs for
5 single species has greatly contributed to our understanding of microbes and our ability to engineer them.
6 Recently, analyses have been developed which predict metabolic interactions between microbes [9,10].
7 However, the application of these recent analyses has been greatly limited by the large investment in time
8 required to reconstruct a useful GENRE. Many microbial communities of interest consist of hundreds of
9 species [11,12]. It is decidedly impractical to spend decades manually curating hundreds of GENREs.

10
11 Many automated methods have been developed for rapidly reconstructing more accurate GENREs [13–
12 16]. We present a novel ensemble method that is complimentary to these existing automated methods
13 which we refer to as Ensemble Flux Balance Analysis (EnsembleFBA). EnsembleFBA pools predictions
14 from many draft GENREs in order to more reliably predict properties that arise from metabolic network
15 structure, such as nutrient utilization and gene essentiality (Fig 1). The primary benefits of this new
16 method are that it relies on automatically-generated GENREs (which can be generated in a matter of
17 minutes to hours) and yet produces more reliable predictions than individual GENREs within the
18 ensemble. We implement and discuss one possible way of generating useful ensembles, but emphasize
19 that other automated methods could be modified to generate useful ensembles.

20
21 **Fig 1. Alternative network structures can be analyzed collectively as an ensemble. A.** Gap filling a
22 network using the same media conditions, but in different orders, can lead to different network structures.
23 Here we display two networks for *P. aeruginosa* gap filled with two permutations of the same 10 minimal
24 media conditions. We highlight two unique reactions in each, and growth predictions which differ
25 between the two. **B.** An ensemble can be created by collecting many alternative network structures which
26 are all consistent with available data. Ensemble-level predictions are generated by treating the individual

1 network predictions like votes. We used three qualitatively different decision thresholds: the “any”
2 threshold requires that a single network predict growth; the “majority” threshold requires that a strict
3 majority predict growth; the “consensus” threshold requires all networks within the ensemble to be in
4 agreement. Note that the top five growth conditions result in the same prediction regardless of threshold,
5 while the bottom three conditions result in threshold-dependent outcomes.

6
7 We begin by discussing a common GENRE curation procedure known as gap filling. We demonstrate that
8 a global gap filling procedure does not perform any better than a sequential one. Instead, we introduce an
9 ensemble approach to pool the many possible network structures resulting from different sequences of the
10 input media conditions (Fig 1). We demonstrate that an ensemble reliably outperforms most of its
11 constituent GENREs in terms of predicting growth and gene essentiality. By tuning the stringency of the
12 voting threshold (e.g. requiring a majority of GENREs to agree vs. complete consensus) it is possible to
13 achieve greater precision or recall than any of the constituent GENREs. We show how additional steps to
14 increase the diversity among GENREs within the ensemble (e.g. reconstructing each member GENRE
15 using subsets of the available data) can further improve recall. Furthermore, we found that incorporating
16 negative growth information into our GENREs improved overall accuracy of the ensemble. We present
17 proof of concept of the use of ensembles by predicting carbon source utilization and gene essentiality in
18 *Pseudomonas aeruginosa*, a well-studied, clinically-relevant pathogen. We provide an example workflow
19 using EnsembleFBA by predicting gene essentiality in six *Streptococcus* species and mapping the
20 predicted essential genes to small molecules ligands in DrugBank. All of our data and code are available
21 in an online repository, including example scripts to make adoption of EnsembleFBA easy. Our ability to
22 make mechanistic predictions about complex cellular communities requires advances in the way we
23 leverage the data available to us, and the way we handle uncertainty. Ensemble FBA is a novel tool that
24 maintains the speed of automated reconstruction methods while improving predictions by intentionally
25 managing uncertainty in network structures.

26

1 **Results**

2

3 *Gap Filling Against Multiple Media Conditions in Different Orders Produces Different Network*

4 *Structures*

5 Gap filling is the process of identifying mismatches between computational predictions and experimental
6 results, and identifying changes to the network structure which will bring the computational predictions
7 into agreement with the experimental data. “Gaps” are missing reactions and can be filled by drawing
8 from a database of possible metabolic reactions. Given that there are usually many mismatches between
9 computational and experimental results, we demonstrate that simply changing the order in which
10 computational results are brought into agreement with experimental can result in different network
11 structures. For example, suppose that it is experimentally determined that a microbe can grow on glucose
12 minimal media and sucrose minimal media, but the computational predictions do not match. Gap filling
13 the GENRE against a representation of glucose minimal media first and sucrose minimal media second,
14 may result in a different network in the end than if sucrose minimal media were first. In practice, the
15 order of gap filling is arbitrary.

16

17 We implemented a custom gap fill algorithm based on the algorithms FASTGAPFILL and FastGapFilling
18 (see Materials and Methods) [17,18]. We used the Model SEED biochemistry database as our “universal”
19 reaction database from which to draw reactions for gap filling [13]. We used the Model SEED web
20 interface to automatically generate a draft GENRE for *Pseudomonas aeruginosa* UCBPP-PA14 (without
21 using the Model SEED gap filling feature). We gap filled this draft GENRE using 2, 5, 10, 15, 20, 25 and
22 30 media conditions that experimentally support growth, with 30 replicates for each. For example, we
23 selected five media conditions at random and selected two random permutations of these conditions (in
24 this case, there are 120 possible permutations). We gap filled in the order prescribed by the first
25 permutation and then in the order of the second, and compared the resulting networks. We repeated the
26 process 30 times, each time drawing a new set of five random media conditions and gap filling using two

1 random permutations of those five media conditions. We found that even with as few as two media
2 conditions, gap filling in a different order resulted in an average of 25 unique reactions per GENRE (Fig
3 2). As the number of media conditions increased, so did the average difference between the resulting
4 GENRES.

5
6 **Fig 2. Gap filling in different orders leads to different network structures.** Each error bar indicates an
7 empirical 95% confidence interval from 30 simulations. For a single simulation, a set of media conditions
8 was randomly selected (we simulated sets sizes of 2–30) and we gap filled a GENRE twice, using the
9 same media conditions but in different orders. We compared the resulting pair of GENREs and we found
10 that on average, GENREs within a pair contained an average of 25–35 unique reactions. The average
11 number of unique reactions increased with the number of media conditions used to gap fill.

12 13 *“Global” Gap Filling Provides No Advantages Over a Sequential Approach*

14 We hypothesized that rather than gap filling sequentially, perhaps a “global” gap fill approach would
15 result in more parsimonious, biologically-relevant solutions without the ambiguity associated with
16 changing the gap fill order. We extended our custom gap fill algorithm to identify a minimal set of
17 reactions which could be added to a GENRE to permit growth in multiple media conditions
18 simultaneously (see Materials and Methods). We started with the *P. aeruginosa* UCBPP-PA14 draft
19 GENRE from the Model SEED and repeated the 30 replicates from 2 to 30 media conditions as above,
20 but using the global gap fill approach that we developed (see Materials and Methods; Fig 3). We found
21 that this global approach did not identify solutions that were any more parsimonious (Fig 3A), and lead to
22 dramatic increases in solve times with increasing media conditions (Fig 2B). In order to determine
23 whether the global solution was any more “biologically-relevant”, we also compared the ability of the
24 global and sequential approaches to reconstruct a well-curated GENRE for *P. aeruginosa* UCBPP-PA14
25 called iPAU1129 (Bartell et al. In review). For each iteration (30 total), we removed 20% of reactions
26 from iPAU1129 and used the sequential and global approaches to gap fill from the universal database

1 using a random selection of five media conditions. The resulting networks were compared to the original
2 iPAU1129, under the assumption that the most biologically-relevant approach would most faithfully
3 reconstructed the curated GENRE, iPAU1129. We found no statistically significant difference between
4 the two approaches (Fig 3C) (p-value = 0.63 by two-sided, paired Wilcoxon signed rank test).

5
6 **Fig 3. Results of global gap filling approach are no more parsimonious or biologically relevant.** For
7 each number of media conditions, we reconstructed 30 pairs of GENREs. For each pair, a set of media
8 conditions was randomly selected and one GENRE was gap filled sequentially while the other was gap
9 filled using a global approach. **A.** We found that the GENREs resulting from the global approach were
10 slightly larger than those gap filled using the sequential approach, with an increasing size disparity as the
11 number of media conditions increased. **B.** The global gap fill approach required significantly more time to
12 run (note the log scale on the y-axis). The solve time increased quadratically with the number of media
13 conditions, such that with 30 media conditions the average solve time for the global approach was ~1000
14 times greater than the sequential approach. Error bars in panels A and B represent empirical 95%
15 confidence intervals. **C.** We compared the ability of the sequential and global approaches to replace
16 reactions removed from a manually-curated GENRE for *P. aeruginosa* UCBPP-PA14, iPAU1129. For
17 each replicate, we removed 20% of the reactions from iPAU1129 and applied the sequential and global
18 gap filling approaches with the same set of randomly selected media conditions. We compared the
19 reaction content of the gap filled GENREs with iPAU1129 using the Jaccard similarity metric. We found
20 that there was no difference between the sequential and global approaches in terms of recovering the
21 removed reactions (p-value = 0.63 by two-sided, paired Wilcoxon signed rank test). Box plots indicate
22 quartiles of the distributions.

23

24 ***Collecting Many Alternative Network Structures into an Ensemble Results in Improved Predictions***

1 Because the sequential gap filling approach produces different results depending on the order of gap
2 filling, we chose to maintain many possible structures resulting from random permutations of the input
3 media conditions rather than select a single GENRE structure for downstream analysis. Not knowing the
4 “true” network structure, we considered each different structure to be a “hypothesis” and analyzed them
5 collectively. For each of 2 to 30 training media conditions we produced 21 GENREs by randomizing the
6 gap fill order (Supplemental Fig 1). We then evaluated each GENRE individually by predicting growth or
7 no growth on 34 test media conditions (17 media conditions which experimentally supported growth and
8 17 which did not) using flux balance analysis (FBA). We found that each GENRE produced slightly
9 different growth predictions, resulting in some GENREs being more accurate than others (Fig 4). In order
10 to generate predictions using the ensemble, we treated each GENRE’s prediction as a single vote, and
11 pooled the votes using a threshold (Fig 1B). We tested three qualitatively different thresholds; “any”,
12 “majority”, and “consensus”. The “any” threshold simply requires that at least one GENRE predict
13 growth in a particular media condition. The “majority” threshold requires greater than half to predict
14 growth, and the “consensus” threshold requires all GENREs to predict growth. We evaluated the growth
15 predictions in terms of accuracy, precision, and recall (see Materials and Methods).

16

17 **Fig 4. Ensembles generated by gap filling against the same media conditions in different orders.**

18 Using 20 media conditions, we generated 21 GENREs, where each GENRE was gap filled using either: a
19 different order of the same input media conditions (“Order Only”), random weighting of reactions in
20 database and random subsets of reactions from draft (“Diverse”), or a diverse ensemble which also
21 included negative growth data through a trimming step (“Negative Growth Data”). We evaluated the
22 accuracy, precision, and recall of every individual GENRE and of the ensembles by predicting growth on
23 17 positive media conditions and 17 negative media conditions which were not used during gap filling.
24 The average of the individual GENREs is shown as black points with the maxima and minima as black
25 lines extended above and below. The ensemble predictions using the three different thresholds are shown
26 as red circles “any”, green triangles “majority”, and blue squares “consensus”. Note that there is less

1 ensemble diversity when differences result only from media condition ordering (maxima/minima of
2 “Order Only” compared to “Diverse” or “Negative Growth Data”). Adding additional diversity results in
3 GENREs with both greater and lower accuracy than the best and worst of “Order Only”. Addition of the
4 trimming step (“Negative Growth Data”) improves overall accuracy and precision by ~15%. In terms of
5 ensemble thresholds, the “majority” threshold tends to perform similarly to the average of the individual
6 GENREs. The “any” threshold achieves recall as good or better than the best individual GENREs. The
7 “consensus” threshold performs consistently well in terms of accuracy and precision if there is very little
8 diversity in the ensemble (“Order Only”).

9
10 We found that the “majority” threshold led the overall ensemble to achieve average accuracy with respect
11 to the individual GENREs, consistently outperforming the least accurate of the individual GENREs (Fig 4
12 “Order Only”). The “any” threshold decreased overall accuracy and precision to be worse than any
13 individual GENRE, but increased the recall to match the best individual GENREs. At the other extreme,
14 we found that the “consensus” threshold led to accuracy and precision that matched the very best
15 individual GENREs but diminished recall.

16

17 ***Increasing the Diversity of Network Structures Increases Recall***

18 While different gap fill order does result in different GENRE structures, the differences are relatively
19 small (tens of differences relative to hundreds of reactions overall). In order to span a greater range of
20 potential GENRE structures, we added random weights (drawn from a uniform distribution) to the
21 reactions in the gap filling step (see Materials and Methods). The rationale is that given two pathways of
22 slightly different length but the same biological function, random weights will occasionally favor the
23 longer pathway, thus exploring alternatives that would otherwise be unobserved given a strictly
24 parsimonious procedure. Additionally, each GENRE was reconstructed using a random subset of only
25 80% of the reactions from the draft GENRE from the Model SEED. Using this new procedure, we

1 reconstructed ensembles of 21 GENREs using 2 through 30 training media conditions (Supplemental Fig
2 2). We evaluated the accuracy by predicting growth on the same 34 test media conditions as before. The
3 resulting accuracy, precision and recall of the individual GENREs were essentially the same on average
4 (Fig 4 “Diverse”), but the distribution spanned a much greater range, both positively and negatively. In
5 this case, the “majority” threshold again achieved average behavior with respect to the individual
6 GENREs, outperforming the least accurate individual GENREs (Fig 4 “Diverse”). The “any” threshold
7 tended to achieve the best accuracy and precision, although not quite as good as the best individual
8 GENREs. However, the “any” threshold achieved the best recall, better than the best individual GENREs
9 and better than the recall achieved with a less diverse ensemble.

10

11 *Accounting for Negative Growth Conditions Greatly Improves Ensemble Accuracy, Precision and* 12 *Recall*

13 Our experimental growth data for *P. aeruginosa* UCBPP-PA14 included both positive (media conditions
14 which supported growth) and negative results (media conditions which did not support growth). We
15 formulated an optimization-based procedure which allowed us to incorporate information inherent in
16 negative growth conditions into our automated curation (see Materials and Methods). In brief, the
17 optimization problem identifies a minimal number of reactions to “trim” from a GENRE in order to
18 prevent growth on negative media conditions while maintaining growth on positive media conditions. As
19 before, we generated ensembles of 21 GENREs for 2 through 30 positive media conditions (Supplemental
20 Fig 3). We used random reaction weights, random subsets of 80% of the reactions from the draft GENRE
21 from Model SEED, and this time we selected 10 negative media conditions for each GENRE (distinct
22 from the negative conditions used to assess accuracy) and incorporated them using our trimming
23 procedure. We found that incorporation of the negative media conditions increased the accuracy and
24 precision of both the individual GENREs and the ensembles by ~15% (Fig 4 “Negative Growth Data”).
25 The “majority” threshold once again approximately tracked the average GENRE accuracy, precision and

1 recall. The “any” threshold achieved accuracy and recall that were often better than the top individual
2 GENREs, with recall exceeding that achieved previously.

3
4 ***Ensembles Achieve Greater Precision or Recall Than Best Individual GENREs When Predicting***
5 ***Essential Genes***

6 We evaluated the ability of ensembles to predict gene essentiality. We generated an ensemble of 51
7 GENREs, each created by gap filling with a random subset of 25 of the total 47 positive media conditions
8 (53%), 10 randomly-selected negative media conditions from the total of 40 (25%), and 1,210 randomly-
9 selected reactions from a total of 1,512 in the draft GENRE generated by Model SEED (80%). Genes
10 were associated with reactions based on the assigned gene-protein-reaction (GPR) relationships from the
11 draft GENRE. We used an *in silico* representation of CF sputum medium and predicted gene essentiality
12 by removing reactions associated with each gene in turn (according to the GPR logic) and running FBA.
13 We compared the resulting gene essentiality predictions with experimental results [19]. We found that the
14 “majority” threshold resulted in better accuracy and recall than the average of individual GENREs, and
15 drastic improvement over the worst GENREs (Fig 5). The “consensus” threshold resulted in a ~20%
16 increase in precision over the best individual GENRE and greater than 100% increase over the worst
17 individual GENRE. Unsurprisingly, the increased precision of the “consensus” threshold comes at the
18 cost of reduced recall. The “any” threshold resulted in lower precision but a ~40% increase in recall over
19 the best individual GENRE and a ~170% increase over the worst individual GENREs.

20
21 **Fig 5. Ensembles outperform individual GENREs when predicting gene essentiality.** We generated
22 an ensemble of 51 GENREs by gap filling against 25 randomly-selected positive growth conditions, 10
23 negative growth conditions, and 80% of the reactions from the Model SEED draft network. We predicted
24 gene essentiality in CF sputum medium and compared the predictions to *in vitro* gene essentiality data.
25 We found that the “consensus” threshold (blue squares) achieved a ~20% increase in precision over the
26 best individual GENRE and a ~100% increase in precision over the worst individual GENRE. Similarly,

1 the “any” threshold (red circles) achieved a ~40% increase in recall over the best individual GENRE and
2 a ~170% increase over the worst. Note the threshold-dependent tradeoff between precision and recall.

3
4 ***Increasing Ensemble Size Improves Predictions for Small Ensembles***
5 Using the same ensemble of 51 GENREs from above, we examined the effect of ensemble size on
6 predicting essential genes. We sampled with replacement 10,000 small ensembles from among the 51
7 GENREs for ensemble sizes of 2 through 51. We evaluated the accuracy, precision and recall against the
8 same gene essentiality data set using a “majority” threshold. We found that smaller ensembles were less
9 accurate, less precise, and more variable than larger ensembles (Figs 6A and 6B). Increasing size
10 improved predictions but with diminishing benefits as the ensemble grew larger. Interestingly, with this
11 “majority” threshold, average recall increased initially, but diminishes again as the ensemble grows larger
12 (Fig 6C).

13
14 **Fig 6. Increasing ensemble size improves performance initially.** Using the same ensemble of 51
15 GENREs, we used bootstrap sampling to simulate 10,000 ensembles of sizes 2 through 51. We evaluated
16 the performance of each sampled ensemble in terms of accuracy (A), precision (B), and recall (C) on the
17 gene essentiality predictions using the “majority” threshold. We found that accuracy and precision
18 increased sharply until around 15 GENREs, at which point gains were less pronounced with additional
19 GENREs. This result suggests that increasing ensemble size does not infinitely improve ensemble
20 performance.

21
22 ***Common Reactions in Ensemble Are Consistent with Manually-Curated Reconstruction***

23 In order to characterize the way gap filling distributes reactions throughout the ensemble, we generated an
24 ensemble of 100 GENREs (Fig 7A). Each GENRE was reconstructed using a randomly-selected 80% of
25 the reactions in iPAU1129, and then sequentially gap filled from the independent, universal reaction

1 database using 25 random positive growth conditions. We found that before gap filling, the “correct”
2 reactions from iPAU1129 were initially distributed in a bell curve throughout the ensemble (Fig 7B). The
3 vast majority of “correct” reactions were found in 50 or more of the GENREs, and in 80 GENREs on
4 average. In contrast, the “incorrect” reactions (those added by gap filling but which were not in the
5 original iPAU1129) were distributed sporadically, with the majority being found in 10 or fewer GENREs.
6 After the gap filling step, 65 “correct” reactions were found to have been added to every GENRE,
7 suggesting a core set of “correct” reactions that were required for biomass production in any condition.
8 We observed that the most common reactions (found in 50 or more GENREs) were overwhelmingly
9 “correct” reactions from iPAU1129 (Fig 7C). All of these most common reactions (both “correct” and
10 “incorrect”) were involved in the production of biomass components, particularly amino acids.

11
12 **Fig 7. Common reactions in ensemble are consistent with manually-curated reconstruction. A.** We
13 generated an ensemble of 100 GENREs using a randomly-selected 80% of the reactions in iPAU1129,
14 and then sequentially gap filled from the universal reaction database using 25 randomly-select positive
15 growth conditions. **B.** The distribution of reactions throughout the ensemble is displayed as the proportion
16 of reactions (y-axis) which are found in a given number of GENREs (x-axis). “Correct” reactions are
17 those which are found in the manually-curated iPAU1129, while “incorrect” reactions are those which
18 are added during gap filling but not found in iPAU1129. We observed that there is a common set of
19 reactions which were found in all 100 GENREs. The majority of this common set are “correct” (88
20 reactions) while 23 are “incorrect”. **C.** The common reactions (found in 50 or more GENREs) consist of a
21 greater proportion of “correct” reactions. “Incorrect” reactions tend to be uncommon.

22
23 ***Identifying Small Molecules Which Interact with Unique Streptococcus Species***
24 We demonstrate how EnsembleFBA can be implemented in a systems biology workflow. We selected six
25 species from the genus *Streptococcus* which all have growth phenotype data available through a previous
26 study [20]. We reconstructed an ensemble for each species: *Streptococcus mitis*, *Streptococcus*

1 *gallolyticus*, *Streptococcus oralis*, *Streptococcus equinus*, *Streptococcus pneumoniae* and *Streptococcus*
2 *vestibularis* (Fig 8A). For each species, we generated a draft GENRE using the Model SEED online
3 interface. We generated a diverse ensemble of 21 GENREs from each Model SEED draft, and gap filled
4 each member GENRE using 25 random growth conditions specific to that species. We mapped all genes
5 (translated to protein sequences) from each *Streptococcus* species to small molecule protein binding
6 sequences from DrugBank using NCBI standalone BLASTP and an e-value threshold of 0.001 [21,22].
7 For all potential gene targets, we used the ensembles to predict gene essentiality using a “majority”
8 threshold in rich media.
9 We found 261 small molecules in DrugBank that potentially bind to the products of 169 essential genes
10 (evenly distributed throughout the six species). Many of these small molecules (113) interact with an
11 essential gene in only one of the species, while 44 were predicted to target conserved essential genes in all
12 six species (Fig 8B). *S. equinus* was predicted to have the most essential genes interact with unique small
13 molecules while *S. pneumoniae* was not predicted to have essential genes interact with any unique small
14 molecules (Fig 8C). As an example of a conserved small molecule interaction, DB04083 (N'-Pyridoxyl-
15 Lysine-5'-Monophosphate) is predicted to interact with essential aspartate aminotransferases in all six
16 species. Alternatively, DB03222 (2'-Deoxyadenosine 5'-Triphosphate) is only predicted to interact with
17 an essential ribonucleotide reductase in *S. gallolyticus*.

18
19 To better understand the differences between the metabolic networks which underpin these small
20 molecule screen results, we predicted reaction essentiality in rich media for all six species using a
21 “majority” threshold. We found that several metabolic subsystems were enriched among essential
22 reactions beyond what would be expected from random chance (Fig 8D). Some subsystems were enriched
23 in all six species, such as Peptidoglycan biosynthesis, indicating that these reactions related to cell wall
24 biosynthesis are disproportionately essential in all six species. Other subsystems were enriched among
25 essential reactions in a unique species. For example, *S. mitis* is predicted to have a greater proportion of
26 essential reactions related to Amino acid metabolism than other species, perhaps indicating that *S. mitis*

1 has less redundancy in those pathways than the other six species. Essential reactions related to Butanoate
2 metabolism were most enriched in *S. pneumoniae*, while essential reactions in Lysine degradation were
3 most enriched in *S. equinus*. Interestingly, reactions associated with core metabolic functions (e.g. Amino
4 acid biosynthesis, Valine and leucine biosynthesis, Phenylalanine biosynthesis) were not equally enriched
5 among essential reactions for all species.

6

7 **Fig 8. EnsembleFBA predicts unique essential genes targets of small molecules in six *Streptococcus***

8 **species. A.** We reconstructed ensembles of 21 GENREs for six *Streptococcus* species based on draft

9 GENREs from the Model SEED and 25 growth conditions. From among all genes within all six species,

10 we identified with potential binding interactions with small molecules from the DrugBank database, and

11 used the ensembles to predict the essentiality of those genes. We found 261 small molecules with

12 potential binding to essential gene products. **B.** Many small molecules interact with an essential gene in

13 only one species, while a core set of 44 small molecules interact with essential genes in all six species. **C.**

14 Distribution of small molecule interactions with essential genes, unique and conserved among the six

15 species. Note that 44 small molecules interact with essential genes in all six species. *S. equinus* is

16 predicted to have essential genes uniquely interact with 43 small molecules, while *S. pneumoniae* is

17 predicted to not have any essential genes which interact with unique small molecules. **D.** Subsystem

18 enrichment among essential reactions by species. We predicted reaction essentiality for all six species in

19 rich media and then calculated a p-value indicating the likelihood of observing each subsystem among the

20 essential reactions given the total number of reactions associated with that subsystem. For clarity we

21 display the $-\log(\text{p-value})$, where darker colors indicate greater enrichment (i.e. a disproportionate number

22 of reactions in that subsystem are predicted to be essential). Note that some subsystems are enriched

23 among essential reactions in all six species (e.g. Peptidoglycan biosynthesis) while others are uniquely

24 enriched in a specific species (e.g. Phenylalanine biosynthesis in *S. mitis*).

25

26 **Discussion**

1 Genome-scale metabolic network reconstructions (GENREs) have been used for decades to assemble
2 information about an organism’s metabolism, to formally analyze that information, and in so doing, to
3 make predictions about that organism’s behavior in unobserved or unobservable contexts. A major barrier
4 preventing more widespread use of GENREs, particularly in non-model organisms, is the extensive time
5 and effort required to produce a high-quality GENRE. Many automated approaches have been developed
6 which reduce this time requirement (e.g. Model SEED, GLOBUS, CoReCo, RAVEN) [13–15]. We
7 demonstrate that gap filling—although our results apply to many automated curation approaches—can
8 lead to many potential GENRE structures depending on the ordering of the input data. Rather than
9 arbitrarily selecting a single GENRE from among many possible networks (which are all reasonably
10 consistent with the available data), we found that collecting many GENREs into an ensemble improved
11 the predictions that could be made. We call this approach “EnsembleFBA” and emphasize that ensembles
12 are a useful tool for dealing with uncertainty in network structure. We demonstrated how ensemble
13 diversity impacts predictions. We show that EnsembleFBA correctly identifies many more essential genes
14 in the model organism *P. aeruginosa* UCBPP-PA14 than the best individual GENREs. We showcase how
15 EnsembleFBA can be utilized in a systems biology workflow by predicting how small molecules interact
16 with different essential genes in six *Streptococcus* species. Ensembles increase the quality of predictions
17 without incurring months of manual curation effort, thus making GENRE approaches more practical for
18 applications which require predictions for many non-model organisms. We have provided code to
19 facilitate the creation and analysis of ensembles of GENREs.

20
21 Gap filling is a common step during the GENRE curation process, both for manually- and automatically-
22 curated GENREs [5]. We used a linear (rather than binary) gap filling algorithm to expand GENREs so
23 that they are capable of producing biomass *in silico* on growth media which supports growth of the
24 organism *in vitro*. Gap filling algorithms suggest parsimonious reaction sets from some “universal”
25 biochemical database which, if added to a GENRE, will allow growth in the new environment
26 [6,17,18,23]. Often, multiple reaction sets can enable growth, so some heuristics are needed to select a

1 final solution. Sometimes gene homology metrics are used to select a solution, such that genes which
2 catalyze the suggested reactions are compared to the current genome, and the reaction set with the best
3 matches in the current genome are selected as the final solution. When validated, these solutions can lead
4 to re-annotation of the genome [6]. During automated curation, there is less opportunity for extensive
5 validation, and so the first or the most parsimonious solution is selected. As we demonstrated, the order of
6 gap filling can change the final outcome, thus producing GENREs with different structures from the exact
7 same input data (Fig 2). Under these circumstances, it is difficult to know which solution is most correct
8 without additional data.

9
10 A possible way around this issue of gap fill order is to remove the sequential nature of gap filling entirely
11 and use a global gap filling approach. We demonstrate that not only is such a global approach much
12 slower (quadratic increases in solution time as growth media conditions are added), but the solutions are
13 no more parsimonious or biologically relevant (Fig 3). Alternatively, we found that two additional
14 innovations improved the predictions that could be obtained from automatically-generated GENREs: the
15 addition of negative growth conditions and the collection of multiple GENREs into ensembles.

16
17 Negative growth conditions have not been extensively incorporated into GENRE curation. To our
18 knowledge, only one group has developed an approach for removing reactions in order to prevent growth
19 in specific conditions [23]. In that case, the reactions were not removed from the GENRE, but rather,
20 prevented from carrying flux under particular conditions. This approach was supported by a biological
21 justification that certain enzymes may not be functional under certain conditions [23]. Our approach is
22 different in that it seeks to produce a single GENRE structure that is consistent with all available data,
23 positive and negative. We achieved this by removing a minimal reaction set to simultaneously prevent
24 growth in negative conditions and allow growth in positive growth conditions (see Materials and
25 Methods). By utilizing this untapped source of information, we found that average GENRE accuracy

1 increased by ~15% (Fig 4). Automatically incorporating negative growth conditions is a little-explored
2 area that has the potential to make better use of growth screening data.

3
4 Ensembles have been used for many years in the machine learning community to leverage the strengths of
5 many different models to improve predictions [24,25]. Ensembles have been used previously to analyze
6 GENREs from a kinetic standpoint [26]. Because kinetic parameters are usually unknown for an entire
7 genome-scale network, ensembles of kinetic parameters are generated such that all parameter sets lead to
8 the same steady state [26]. In this way, ensembles can represent the space of allowable kinetic parameters.
9 Our approach to generating ensembles is different in that we attempt to represent the space of allowable
10 GENRE structures rather than kinetic parameters.

11
12 Ensembles provide a significant advantage over individual GENREs by tuning for specific results with
13 defined decision thresholds (Figs 4 and 5). Consistently, by using the “any” threshold, recall can be made
14 to equal or exceed the best individual GENREs. This result makes sense, considering that different
15 network structures will result in different growth or gene essentiality predictions. By accepting any
16 essential gene prediction from among the constituent GENREs, we cast a wider net and capture many
17 more of the true essential genes and growth conditions. The fact that many individual GENREs contribute
18 unique but true predictions suggests that each GENRE recapitulates elements of the “true” network
19 structure (Fig 7). Similarly, by using the “majority” threshold, the ensemble predictions perform like the
20 average GENRE (Figs 4 and 5). By requiring a majority of GENREs to agree, the ensemble guards
21 against poor predictions and, in most cases, outperforms the worst individual GENREs. Finally, if
22 precision is the overall goal, a “consensus” threshold provides confidence that the majority of positive
23 predictions are true positives (Fig 5).

24
25 We observed that ensemble performance is limited by the quality of the GENREs which form the
26 ensemble. The choice of decision threshold (“any”, “majority”, or “consensus”) did not consistently

1 improve overall accuracy of the ensemble. However, by improving the individual GENREs using
2 negative information, the overall ensemble accuracy improved dramatically (Fig 4). Also, it should be
3 noted that the computational burden required by ensembles will always be greater than the burden of a
4 single GENRE. For all the examples in this study, computational burden scales linearly with the number
5 of GENREs in the ensemble (ensemble of size N GENREs will require N times longer to calculate FBA
6 solutions) which is a modest expectation in practice. Other applications, like predicting species
7 interactions, would not scale linearly if all possible pairs of GENREs between two ensembles were
8 simulated.

9
10 Increasing ensemble diversity impacted ensemble recall, but did not have an obvious effect on overall
11 accuracy. Some degree of diversity is required in order to gain any advantage through an ensemble
12 representation. In the “Order Only” ensemble (generated simply by changing the order of gap filling; Fig
13 4) there were only small differences between any of the GENREs so it was difficult to improve on the
14 best GENRE. By injecting greater diversity through random weights and random subsets of the data, we
15 observed much greater variation in individual GENRE performance (both positively and negatively), but
16 the average accuracy was the same as the low diversity ensemble (Fig 4). The advantage of diversity is in
17 casting a wide net and thus improving ensemble recall, particularly when combined with an “any”
18 decision threshold. In practice, the choice to increase diversity or not will depend on the goals of the
19 analysis. If the goal is to generate many candidate essential genes or media conditions, then more
20 diversity will be advantageous. If the goal is to generate fewer, more confident predictions, then
21 minimizing diversity will be most effective.

22
23 EnsembleFBA is easily integrated into systems biology workflows. As an example, a current challenge in
24 systems biology is to identify species-specific drug targets so that therapies will not disrupt the healthy
25 microbiome structure [27,28]. We reconstructed ensembles for six *Streptococcus* species by gap filling
26 with growth phenotype data, we predicted essential genes and mapped those genes to potential small

1 molecule binding partners within a matter of hours, and can have more confidence in the quality of the
2 gene essentiality predictions than if we were to work with single GENREs for each species (Fig 8). The
3 process scales well with the number of species, such that 12 or 100 species would not take significantly
4 longer than six, and the quality of the predictions is maintained with scale. It is interesting to note that
5 among *Streptococcus* species, there are generally small molecules which can be selected to uniquely
6 interact with essential genes in a single species, and other small molecules which interact with conserved
7 essential genes (Fig 8C). The observed interactions between essential genes and small molecule ligands
8 are species-specific because of differences in network structure which lead to some metabolic subsystems
9 being disproportionately represented among essential reactions (Fig 8D). In the search for species-specific
10 drug targets, it is important to consider, not only the presence or absence of a particular gene, but also the
11 role of that gene in the broader network context, and improved systems biology tools such as
12 EnsembleFBA can help to elucidate that context with greater confidence.

13
14 Gap filling is not the only GENRE reconstruction approach that produces many possible solutions.
15 Likelihood-based gap filling produces a distribution of possible annotations for each gene in a genome,
16 assigning a probability to each [14,16,29]. Network structure is then based on maximizing the likelihood
17 over all possible solutions. Ensembles could be generated easily using this type of framework by
18 sampling many alternative solutions around the maximum likelihood. Indeed, it may be beneficial to
19 create an ensemble using GENREs reconstructed using several different methods. We suggest that there
20 are many possible ways to generate ensembles such that they will allow researchers to generate better
21 predictions about under-studied organisms.

22
23 Finally, we foresee ensembles playing an important role beyond improving predictions, for example, in
24 experimental design and model reconciliation. Within a diverse ensemble, many possible network
25 structures are represented, and it is expected that some structures will be closer to the truth than others.
26 We suggest that ensembles can be leveraged to design an optimal series of experiments to weed out the

1 most incorrect network structures. For instance, such an approach could select the most differentiating
2 carbon sources to experimentally test, or the most differentiating essential genes. Ensemble-guided
3 experimental design could save time and experimental resources. Model reconciliation is another field
4 that could benefit from ensembles [8,30]. Given GENREs for two different species, reconciliation is the
5 process of removing systematic differences from the two GENREs so that any differences which remain
6 are due to biology alone. Systematic differences often result from arbitrary choices during the process of
7 reconstruction. Ensembles could be used to automate the reconciliation process by representing the space
8 of possible GENREs for each species and the reconciled versions would be the two models from the two
9 spaces that are most similar to each other. Thus, ensembles have potential to improve other tasks than
10 prediction, including experimental design and mapping the space of GENRE structures for tasks like
11 reconciliation.

12

13 **Materials and Methods**

14

15 *Code and Data Availability*

16 All data, Matlab (Natick, MA, USA) implementations of algorithms, Matlab simulation scripts, results
17 files and figure generation scripts are publically available in our online repository:

18 <https://github.com/mbi2gs/ensembleFBA>

19

20 *Data Sources*

21 All biochemical reference data was obtained from the Model SEED database
22 (<https://github.com/ModelSEED/ModelSEEDDatabase>). The metabolic reaction and compound databases
23 were parsed and formatted for use in Matlab using a custom Python script available in our repository
24 (“format_SEED_data.py”). A draft network for *P. aeruginosa* UCBPP-PA14 was automatically generated
25 using the Model SEED web service (<http://modelseed.org/genomes/>). Similarly, draft networks were
26 generated for *Streptococcus mitis* ATCC 6249, *Streptococcus gallolyticus* ICDDR-B-NRC-S3,

1 *Streptococcus oralis* ATCC 49296, *Streptococcus equinus* AG46, *Streptococcus pneumoniae* (PATRIC
2 ID 1313.5731), and *Streptococcus vestibularis* 22-06 S6.

3
4 Representations of media conditions (including minimal media and cystic fibrosis sputum medium), and
5 biomass representations were drawn from previous GENRE analyses of *Pseudomonas aeruginosa*
6 [31,32].

7
8 *P. aeruginosa* PA14 essential genes in cystic fibrosis sputum medium were experimentally identified
9 previously [19].

10
11 A manually curated, and thoroughly validated GENRE of *P. aeruginosa* UCBPP-PA14 called iPAU1129
12 was developed previously (Bartell et al. In review), along with Biolog growth screen data for *P.*
13 *aeruginosa* UCBPP-PA14 indicating many media conditions in which this strain will and will not grow.

14
15 Growth phenotype data for six *Streptococcus* species was obtained from the file “Supplementary Data 1”
16 of [20]. Small molecule amino acid binding target sequences were downloaded from the DrugBank
17 website (<http://www.drugbank.ca/>) [21]. After identifying homologous genes to the target sequences
18 using BLASTP [22], we used a custom python script to parse the results for input into Matlab
19 (“listPossibleTargets.py”, available in repository).

20
21 ***Linear Gap Filling***

22 We implemented a linear (as opposed to binary) gap filling algorithm in Matlab, based on the algorithms
23 FASTGAPFILL and FastGapFilling [17,18]. We used the Gurobi solver version 6.0.5 for all optimization
24 tasks (Gurobi, Houston, TX, USA). To begin, we provide the algorithm with a universal database of
25 metabolic reactions U, a universal database of exchange reactions X, a biomass reaction, and a set of
26 growth conditions formatted as lower bounds on exchange reactions. The algorithm identifies a set of

1 reactions from U and X that allow flux through the biomass reaction under all growth conditions. The
 2 algorithm is implemented as a linear program (LP) that minimizes the sum of the absolute value of all
 3 fluxes through U and X. The optimization problem takes the form:

4

5 minimize $\text{sum}(r_{u,z_u}) + \text{sum}(r_{x,z_x})$

6 s.t. $Uv + Xw = 0$ (1)

7 $lb_{u,i} \leq v_i \leq ub_{u,i}$ for all $i \in [1, N_u]$ (2)

8 $lb_{x,i} \leq w_i \leq ub_{x,i}$ for all $i \in [1, N_x]$...

9 $-z_{u,i} \leq v_i \leq z_{u,i}$ for all $i \in [1, N_u]$ (3)

10 $-z_{x,i} \leq w_i \leq z_{x,i}$ for all $i \in [1, N_x]$...

11 $z_i \geq 0$ for all $i \in [1, N_u + N_x]$ (4)

12 $v_{\text{biomass,gc=j}} \geq 0.05$ for all $j \in [1, N_{gc}]$ (5)

13 $z_{u,i} \geq C_{z_u,i}$ for all $i \in [1, N_u]$ (6)

14 $z_{x,i} \geq C_{z_x,i}$ for all $i \in [1, N_x]$...

15

16 Where:

17 U is the universal reaction library (as a stoichiometric matrix)

18 X is the universal exchange library (same metabolites as U)

19 N_u and N_x are the number of reactions in U and X, respectively

20 N_{gc} is the number of growth conditions

21 v is the vector of fluxes through U

22 w is the vector of fluxes through X

23 $lb_{u,i}$, $ub_{u,i}$, $lb_{x,i}$, and $ub_{x,i}$ are the lower and upper bounds on v_i and w_i , respectively

24 $v_{\text{biomass,gc=j}}$ is the flux through the biomass reaction under growth condition j

25 $C_{z_u,i}$ and $C_{z_x,i}$ are variables that can force reactions from U and X to be included

1 z is a continuous variable which acts as a constraint on the absolute values of elements in v and
2 w.

3 r is an optional weight on z which can be randomized

4
5 In order to incorporate genome annotations from a specific organism, we force the inclusion of all
6 associated reactions from those annotations using the C_{zu} variables. Note that unlike a binary
7 optimization, the LP minimizing the sum of the absolute flux values through U and X does not
8 necessarily result in a solution with the fewest reactions, but rather the solution which requires the
9 minimum sum of the absolute values of the fluxes through it. The LP here can be extended to utilize
10 multiple growth conditions simultaneously (global approach) by duplicating the U and X matrices, once
11 for each growth condition, but minimizing a single set of z variables across all conditions. To gap fill
12 using multiple growth conditions sequentially, we gap fill using the first growth condition, incorporate the
13 solution into the GENRE, then repeat the process for all growth conditions. Our Matlab function
14 “expand()” implements this optimization problem.

15

16 ***Incorporating Negative Growth Conditions by Trimming Reactions***

17 We implemented a binary optimization problem to trim minimal reactions from a GENRE in order to
18 prevent growth under negative growth conditions while simultaneously maintaining growth in the
19 positive growth conditions. As input to the algorithm, we provide a GENRE, and a set of both positive
20 and negative growth conditions. We chose to run FBA first to identify mismatches between the
21 computational predictions and the *in silico* data (negative growth conditions that erroneously predicted to
22 support growth *in silico*). Having identified those, we then ran FBA on all the positive growth conditions
23 to identify the top five with flux distributions most similar to the flux distribution of the negative growth
24 condition. The GENRE and the selected growth conditions are passed to the trimming problem, which
25 takes the form:

26

$$\begin{aligned}
 &1 \quad \text{maximize } \text{sum}(r_u y_u) + \text{sum}(r_x y_x) \\
 &2 \quad \text{s.t.} \quad Uv + Xw = 0 \quad (1) \\
 &3 \quad y_{u,i} \text{lb}_{u,i} \leq v_i \leq y_{u,i} \text{ub}_{u,i} \quad \text{for all } i \in [1, N_u] \quad (2) \\
 &4 \quad y_{x,i} \text{lb}_{x,i} \leq w_i \leq y_{x,i} \text{ub}_{x,i} \quad \text{for all } i \in [1, N_x] \quad \dots \\
 &5 \quad v_{\text{biomass,gc=j}} \geq 0.05 \quad \text{for all } j \in [1, N_{\text{gc}}] \quad (3) \\
 &6 \quad \max(v_{\text{biomass,ngc=j}}) = 0 \quad \text{for all } j \in [1, N_{\text{ngc}}] \quad (4) \\
 &7 \quad y_u \in \{0, 1\}
 \end{aligned}$$

8

9 Where:

10 U is the set of reactions from the GENRE (as a stoichiometric matrix)

11 X is the set of exchange reactions from the GENRE (same metabolites as U)

12 N_u and N_x are the number of reactions in U and X , respectively

13 v is the vector of fluxes through U

14 w is the vector of fluxes through X

15 $\text{lb}_{u,i}$, $\text{ub}_{u,i}$, $\text{lb}_{x,i}$, and $\text{ub}_{x,i}$ are the lower and upper bounds on v_i and w_i , respectively

16 $v_{\text{biomass,gc=j}}$ is the flux through the biomass reaction under growth condition j

17 y is a set of binary variables which determine inclusion of reactions in the network

18 r is an optional weight on y which can be randomized

19

20 The term $\max(v_{\text{biomass,ngc=j}}) = 0$ requires that the maximum possible flux through the biomass reaction for

21 the negative growth conditions is constrained to be zero. In order to implement this constraint, we took

22 advantage of duality theory as has been done previously [7]. Specifically, the optimal objective value of

23 the dual of a linear program will equal the optimal value of the primal. By constraining the primal and

24 dual objectives to equal each other, we can ensure that the flux through the biomass objective is

25 maximized. We can replace the term $\max(v_{\text{biomass,ngc=j}}) = 0$ with the following constraints:

26

$$1 \quad v_{\text{biomass,ngc=j}} = \lambda_{\text{ub}} \text{ub } y - \lambda_{\text{lb}} \text{lb } y \quad (4.1a)$$

$$2 \quad S^T \lambda_{\text{mets}} + \lambda_{\text{ub}} - \lambda_{\text{lb}} = c \quad (4.2)$$

$$3 \quad \lambda_{\text{ub}}, \lambda_{\text{lb}} \geq 0 \quad (4.3)$$

4

5 Where λ_{mets} , λ_{ub} and λ_{lb} are the dual vectors associated with the metabolites, upper and lower bounds of
6 the primal problem. Note that the terms $\lambda_{\text{ub}} \text{ub } y$ and $\lambda_{\text{lb}} \text{lb } y$ are quadratic, requiring a multiplication of the
7 binary inclusion variable y with the dual variables. Because y is a binary variable, in this case the
8 quadratic constraints can be converted to linear constraints through the use of additional variables:

9

$$10 \quad t_{\text{ub}} \leq L_{\lambda} y_i \quad (4.4)$$

$$11 \quad t_{\text{ub}} \geq 0 \quad (4.5)$$

$$12 \quad t_{\text{ub}} \leq \lambda_{\text{lb},i} \text{ub}_i \quad (4.6)$$

$$13 \quad t_{\text{ub}} \geq \lambda_{\text{lb},i} \text{ub}_i - L_{\lambda} (1 - y_i) \quad (4.7)$$

14

15 Where t_{ub} is a stand-in for the product $\lambda_{\text{ub}} \text{ub } y$ and L is a large number greater than or equal to the upper
16 bound on $\lambda_{\text{ub}} \text{ub } y$ (e.g. 1000). Similar constraints would be produced for the product $\lambda_{\text{lb}} \text{lb}$. The quadratic
17 constraint above can then be replaced by a linear constraint:

$$18 \quad v_{\text{biomass,ngc=j}} = t_{\text{ub}} \text{ub} - t_{\text{lb}} \text{lb} \quad (4.1b)$$

19

20 Our Matlab function “trim_active()” implements this optimization problem.

21

22 ***Iterative Approach to Reconstructing GENREs Consistent with All Growth Screening Data***

23 We implemented an iterative algorithm to integrate the LP expansion step with the binary trimming step.
24 The algorithm first applies the expansion step to produce a GENRE that is capable of growing in all
25 positive growth conditions. Next, the algorithm checks for negative growth conditions that allow for
26 biomass flux and for any that do, applies the trim step as described above. The algorithm iterates between

1 the expand and trim steps until either a completely consistent GENRE structure is identified, or it reaches
2 a maximum attempts limit. A single attempt is completed if the GENRE structure is not yet consistent
3 with the input growth conditions but stops making progress (stuck in a local optimum). In this case, a
4 random reaction is removed from the GENRE and the search is re-initiated. If the maximum attempts
5 limit is reached, the algorithm removes any negative growth conditions that are inconsistent with the
6 positive growth conditions, and returns the final GENRE. This iterative algorithm is implemented in our
7 Matlab function “build_network()”.

8

9 *Predicting Growth and Essential Genes*

10 Growth media were simulated by setting the lower bounds on exchange reactions for the appropriate
11 nutrients to negative values. The uptake of carbon source(s) limited the final flux through biomass.
12 “Growth” was determined by maximizing flux through the biomass objective. We predicted “growth” if a
13 positive, non-zero flux could be achieved through biomass. Gene knock-outs were simulated by
14 generating a new GENRE which was missing the reactions dependent on the knocked-out gene. The
15 reaction-gene dependence was determined by evaluating the binary logic of the GPRs provided by Model
16 SEED. Our custom script to evaluate GPR logic is “simulateGeneDeletion()”.

17

18 We evaluated the growth predictions in terms of accuracy $(TP + TN) / (TP + FP + TN + FN)$, precision
19 $(TP / TP + FP)$, and recall $(TP / TP + FN)$ where TP = number of true positives, FP = the number of false
20 positives, TN = the number of true negatives and FN = the number of false negatives. Precision indicates
21 the number of positive predictions which are true positives. Recall indicates the number of positive events
22 which were correctly predicted by the method.

23

24 *Predicting Small Molecule Interactions*

1 We downloaded the Drug Target Sequences for small molecules in FASTA format from DrugBank [21].
2 Using NCBI standalone BLASTP and an e-value cutoff of 0.001, we identified homologous sequences in
3 all six *Streptococcus* proteomes [22].
4

5 ***Metabolic Subsystem Enrichment***

6 We downloaded KEGG subsystem annotations for the reactions in the Model SEED database
7 (“KEGG.pathways.tsv”). After predicting essential reactions for each *Streptococcus* species, we used the
8 hypergeometric distribution to calculate the probability of drawing k essential reactions and finding that x
9 or more are annotated with subsystem j , from a population of size M reactions, of which N are annotated
10 with subsystem j .
11

12 ***Computational Resources***

13 The majority of our reconstructions and simulations were performed on a 64-bit Dell Precision T3600
14 Desktop computer with 32 GB RAM and eight 3.6 GHz Intel Xeon CPUs, running Windows 7.
15 Incorporating negative growth information often lead to longer reconstruction times (sometimes 2 hours
16 per GENRE) due to the binary optimization step. To accelerate the reconstruction time while
17 incorporating negative growth information, we used the University of Virginia’s High Performance
18 Computing Cluster.
19

20 ***Generating Ensembles and Making Predictions Using Our Software***

21 Our Matlab scripts for generating an ensemble (using the gap filling approach described in this work) and
22 for analyzing an ensemble are freely available in a github repository (see Code and Data Availability).
23 The Gurobi solver is required, in addition to our Matlab scripts. We have also included a tutorial script to
24 guide the user through the necessary steps to generate and analyze an ensemble (“test_eFBA.mat”).
25

26 **Acknowledgements**

1 The authors would like to thank Jennifer Bartell for providing a copy of iPAU1129 and the accompanying
2 carbon source utilization data.

3

4 **References**

- 5 1. Caspi R, Altman T, Dale JM, Dreher K, Fulcher CA, Gilham F, et al. The MetaCyc database of
6 metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases.
7 *Nucleic Acids Res.* 2010;38: D473–D479. doi:10.1093/nar/gkp875
- 8 2. Jeong H, Tombor B, Albert R, Oltvai ZN, Barabási AL. The large-scale organization of metabolic
9 networks. *Nature.* 2000;407: 651–4. doi:10.1038/35036627
- 10 3. Oberhardt MA, Palsson BØ, Papin JA. Applications of genome-scale metabolic reconstructions.
11 *Mol Syst Biol.* 2009;5:320–334. doi:10.1038/msb.2009.77
- 12 4. McCloskey D, Palsson BØ, Feist AM. Basic and applied uses of genome-scale metabolic network
13 reconstructions of *Escherichia coli*. *Mol Syst Biol.* 2013;9: 661. doi:10.1038/msb.2013.18
- 14 5. Thiele I, Palsson BØ. A protocol for generating a high-quality genome-scale metabolic
15 reconstruction. *Nat Protoc.* 2010;5: 93–121. doi:10.1038/nprot.2009.203
- 16 6. Reed JL, Patel TR, Chen KH, Joyce AR, Applebee MK, Herring CD, et al. Systems approach to
17 refining genome annotation. *Proc Natl Acad Sci U S A.* 2006;103: 17480–17484.
18 doi:10.1073/pnas.0603364103
- 19 7. Burgard AP, Pharkya P, Maranas CD. OptKnock: A Bilevel Programming Framework for
20 Identifying Gene Knockout Strategies for Microbial Strain Optimization. *Biotechnol Bioeng.*
21 2003;84: 647–657. doi:10.1002/bit.10803
- 22 8. Bartell JA, Yen P, Varga JJ, Goldberg JB, Papin JA. Comparative metabolic systems analysis of

- 1 pathogenic *Burkholderia*. *J Bacteriol.* 2014;196: 210–226. doi:10.1128/JB.00997-13
- 2 9. Stolyar S, Van Dien S, Hillesland KL, Pinel N, Lie TJ, Leigh JA, et al. Metabolic modeling of a
3 mutualistic microbial community. *Mol Syst Biol.* 2007;3: 92. doi:10.1038/msb4100131
- 4 10. Zomorodi AR, Maranas CD. OptCom: a multi-level optimization framework for the metabolic
5 modeling and analysis of microbial communities. *PLoS Comput Biol.* 2012;8: e1002363.
6 doi:10.1371/journal.pcbi.1002363
- 7 11. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human
8 microbiome. *Nature.* 2012;486: 207–214. doi:10.1038/nature11234
- 9 12. Lundberg DS, Lebeis SL, Paredes SH, Yourstone S, Gehring J, Malfatti S, et al. Defining the core
10 *Arabidopsis thaliana* root microbiome. *Nature.* 2012;488: 86–90. doi:10.1038/nature11237
- 11 13. Henry CS, DeJongh M, Best AA, Frybarger PM, Linsay B, Stevens RL. High-throughput
12 generation, optimization and analysis of genome-scale metabolic models. *Nat Biotechnol.*
13 2010;28: 977–982. doi:10.1038/nbt.1672
- 14 14. Plata G, Fuhrer T, Hsiao T-L, Sauer U, Vitkup D. Global probabilistic annotation of metabolic
15 networks enables enzyme discovery. *Nat Chem Biol.* 2012;8(10): 848–854.
16 doi:10.1038/nchembio.1063
- 17 15. Agren R, Liu L, Shoaie S, Vongsangnak W, Nookaew I, Nielsen J. The RAVEN Toolbox and Its
18 Use for Generating a Genome-scale Metabolic Model for *Penicillium chrysogenum*. *PLoS Comput*
19 *Biol.* 2013;9(3): e1002980. doi:10.1371/journal.pcbi.1002980
- 20 16. Pitkänen E, Jouhten P, Hou J, Syed MF, Blomberg P, Kludas J, et al. Comparative genome-scale
21 reconstruction of gapless metabolic networks for present and ancestral species. *PLoS Comput Biol.*
22 2014;10: e1003465. doi:10.1371/journal.pcbi.1003465

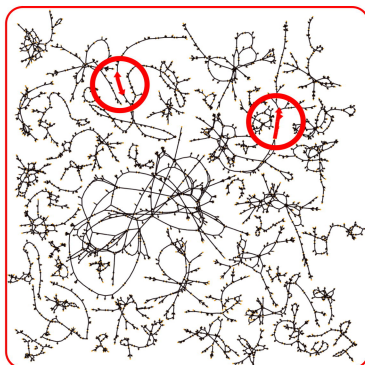
- 1 17. Thiele I, Vlassis N, Fleming RMT. FASTGAPFILL: efficient gap filling in metabolic networks.
2 *Bioinformatics*. 2014;30: 2529–2531. doi:10.1093/bioinformatics/btu321
- 3 18. Latendresse M. Efficiently gap-filling reaction networks. *BMC Bioinformatics*. 2014;15: 225.
4 doi:10.1186/1471-2105-15-225
- 5 19. Turner KH, Wessel AK, Palmer GC, Murray JL, Whiteley M. Essential genome of *Pseudomonas*
6 *aeruginosa* in cystic fibrosis sputum. *Proc Natl Acad Sci*. 2015;112: 4110–4115.
7 doi:10.1073/pnas.1419677112
- 8 20. Plata G, Henry CS, Vitkup D. Long-term phenotypic evolution of bacteria. *Nature*.
9 2014;517(7534): 369–372. doi:10.1038/nature13827
- 10 21. Wishart DS. DrugBank: a comprehensive resource for in silico drug discovery and exploration.
11 *Nucleic Acids Res*. 2006;34: D668–D672. doi:10.1093/nar/gkj067
- 12 22. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+:
13 architecture and applications. *BMC Bioinformatics*. 2009;10: 421. doi:10.1186/1471-2105-10-421
- 14 23. Kumar VS, Maranas CD. GrowMatch: An automated method for reconciling *in silico/in vivo*
15 growth predictions. *PLoS Comput Biol*. 2009;5(3): e1000308. doi:10.1371/journal.pcbi.1000308
- 16 24. Opitz D, Maclin R. Popular Ensemble Methods: An Empirical Study. *J Artif Intell Res*. 1999;11:
17 169–198.
- 18 25. Breiman L. Random Forests. *Mach Learn*. 2001;45: 5–32. doi:10.1023/A:1010933404324
- 19 26. Tran LM, Rizk ML, Liao JC. Ensemble Modeling of Metabolic Networks. *Biophys J*. 2008;95:
20 5606–5617. doi:10.1529/biophysj.108.135442
- 21 27. Round JL, Mazmanian SK. The gut microbiota shapes intestinal immune responses during health

- 1 and disease. *Nat Rev Immunol*. 2009;9: 313–323. doi:10.1038/nri2515
- 2 28. McCaughey LC, Ritchie ND, Douce GR, Evans TJ, Walker D. Efficacy of species-specific protein
3 antibiotics in a murine model of acute *Pseudomonas aeruginosa* lung infection. *Sci Rep*. 2016;6:
4 30201. doi:10.1038/srep30201
- 5 29. Benedict MN, Mundy MB, Henry CS, Chia N, Price ND. Likelihood-based gene annotations for
6 gap filling and quality assessment in genome-scale metabolic models. *PLoS Comput Biol*.
7 2014;10(10): e1003882. doi:10.1371/journal.pcbi.1003882
- 8 30. Oberhardt MA, Puchařka J, Martins dos Santos VAP, Papin JA. Reconciliation of genome-scale
9 metabolic reconstructions for comparative systems analysis. *PLoS Comput Biol*. 2011;7(3):
10 e1001116. doi:10.1371/journal.pcbi.1001116
- 11 31. Oberhardt MA, Goldberg JB, Hogardt M, Papin JA. Metabolic Network Analysis of *Pseudomonas*
12 *aeruginosa* during Chronic Cystic Fibrosis Lung Infection. *J Bacteriol*. 2010;192: 5534–5548.
13 doi:10.1128/JB.00900-10
- 14 32. Oberhardt MA, Puchařka J, Fryer KE, Martins dos Santos VAP, Papin JA. Genome-scale
15 metabolic network analysis of the opportunistic pathogen *Pseudomonas aeruginosa* PAO1. *J*
16 *Bacteriol*. 2008;190(8): 2790–2803. doi:10.1128/JB.01583-07

17

A*P. aeruginosa* GENRE 1

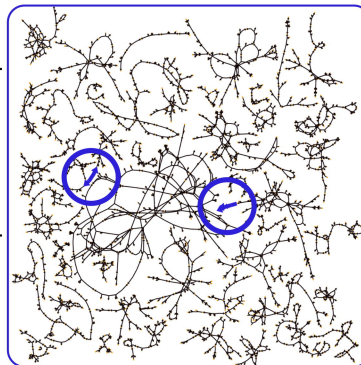
1489 Reactions (30 unique)

Unique
ReactionsMaltose glucohydrolase
GMP reductase
...Growth
Predictions

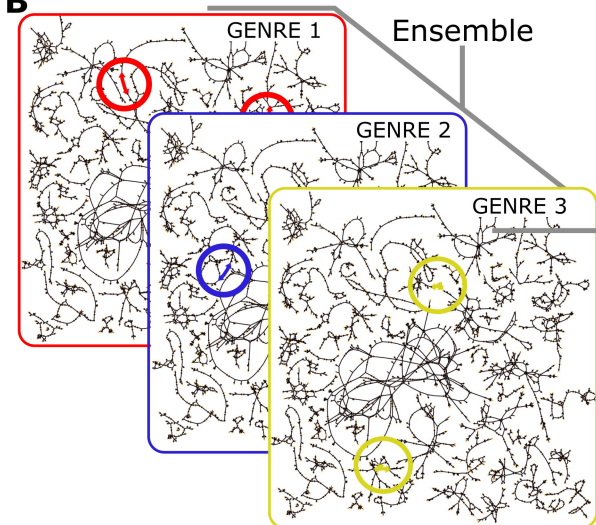
D-Galacturonate	✗
Glyoxalate	✓
Tartrate	✗
Amylotriose	✓
...	

P. aeruginosa GENRE 2

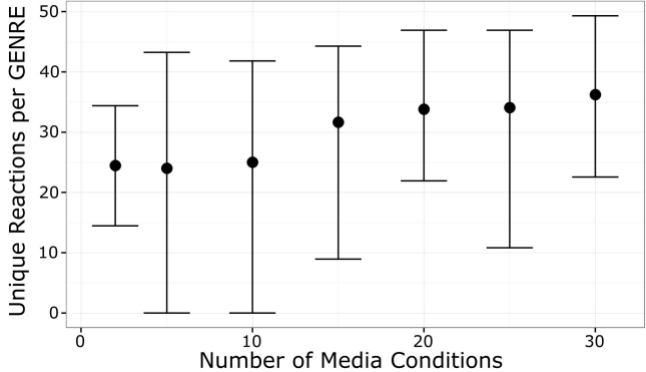
1484 Reactions (25 unique)

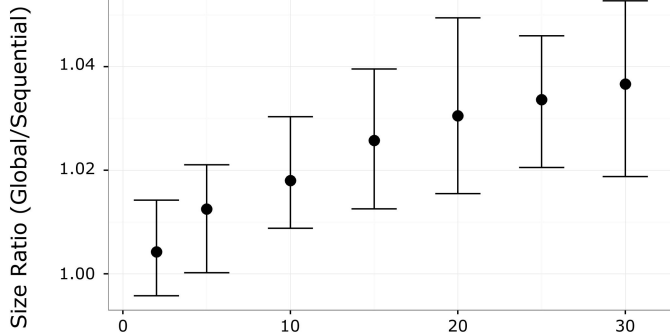
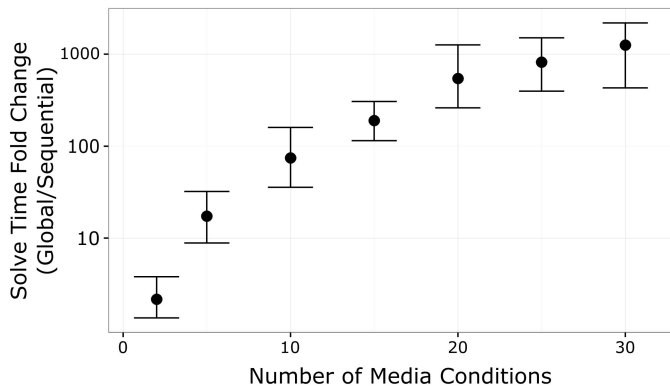
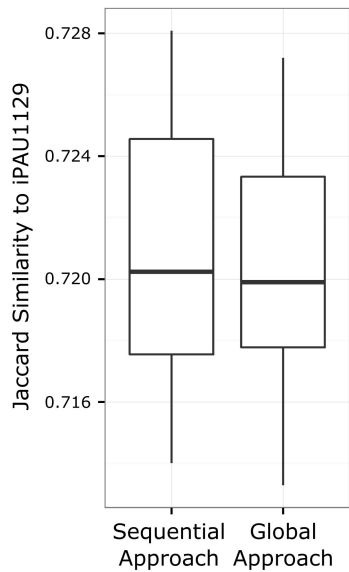
Unique
ReactionsFormamide amidohydrolase
Hydroxyproline oxidase
...Growth
Predictions

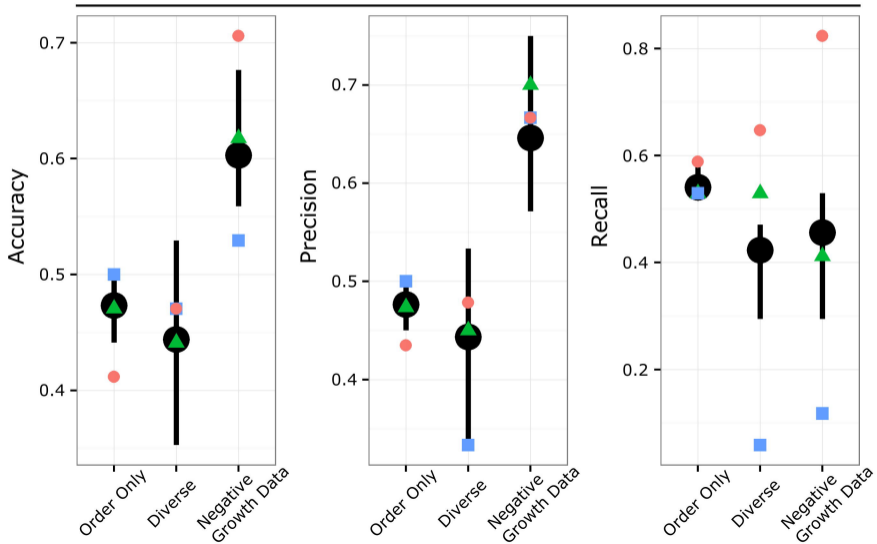
D-Galacturonate	✗
Glyoxalate	✓
Tartrate	✓
Amylotriose	✗
...	

B

Carbon Sources	GENRE Growth Predictions			Ensemble Growth Predictions		
	1	2	3	Any	Major	Cons.
Acetate	✓	✓	✓	✓	✓	✓
Citrate	✓	✓	✓	✓	✓	✓
D-Alanine	✗	✗	✗	✗	✗	✗
D-Galacturonate	✗	✗	✗	✗	✗	✗
Glyoxalate	✓	✓	✓	✓	✓	✓
Tartrate	✗	✓	✗	✓	✗	✗
Amylotriose	✓	✗	✗	✓	✗	✗
D-Mucic acid	✗	✓	✓	✓	✓	✗



A**B****C**

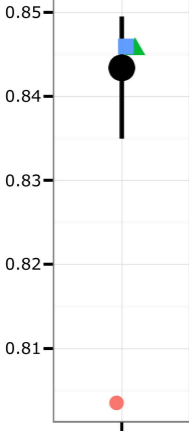
**Ensemble
Thresholds**Any ●
Majority ▲
Consensus ■**Individual
GENRES**Max. |
Mean ●
Min. |

**Ensemble
Thresholds**Any
Majority
Consensus**Individual
GENRES**

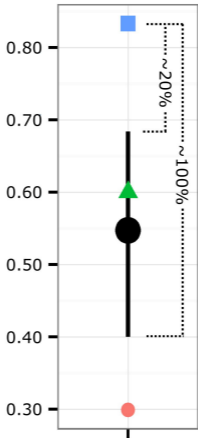
Max.

Mean

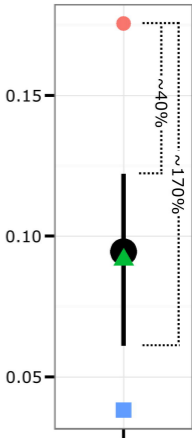
Min.



Accuracy



Precision



Recall

