

Intracellular metabolic circuits shape inter-species microbial interactions

3

4

5

6 Ali R. Zomorodi¹ and Daniel Segre^{1,2,3*}

7

8 ¹Bioinformatics Program, Boston University, Boston, MA 02215

9 ²Department of Biology, Boston University, Boston, MA 02215

10 ³Department of Biomedical Engineering, Boston University, Boston, MA 02215

11

12

13 *Corresponding author:

14 Email: dsegre@bu.edu

15 Phone: 617-358-2301

16 Fax: 617-353-4814

17

18

19

20

21

22

23

24

25

26

27

Metabolite exchanges in microbial communities give rise to ecological interaction networks that influence ecosystem diversity and stability¹⁻⁶. These exchanges depend on complex intracellular pathways thus raising the question of whether ecological interactions are inferable from genomes. We address these questions by integrating genome-scale models of metabolism⁷, to compute the fitness of interacting microbes, with evolutionary game theory, which uses these fitness values to infer evolutionarily stable interactions in multi-species microbial “games”. After validating our approach using data on sucrose hydrolysis by *S. cerevisiae*, we performed over 80,000 *in silico* experiments to evaluate the rise of unidirectional and cross-feeding metabolic dependencies in populations of *Escherichia coli* secreting 189 amino acid pairs. We found that, despite the diversity of exchanged amino acids, most pairs conform to general patterns of inter-species interactions. However, several amino acid pairs deviate from these patterns due to pleiotropy and epistasis in metabolic pathways. To better understand the emergence of cross-feeding, we performed *in silico* invasion experiments and found possible evolutionary paths that could lead to such association. Overall, our study provides mechanistic insights into the rise of evolutionarily stable interdependencies, with important implications for biomedicine and microbiome engineering⁸⁻¹⁰.

Obligate dependencies among microorganisms, through the exchange of essential metabolites have been hypothesized to be ubiquitous in microbial ecosystems ¹. Similar interactions have also been engineered in laboratory systems, mainly based on genetically induced auxotrophies ^{3-6,11,12}. However, the evolutionary rise and maintenance of these interactions constitutes an unresolved puzzle, since genotypes that do not produce a given costly metabolite may have a selective advantage over producers. One theory, known as the Black Queen Hypothesis ¹³, suggests that metabolic dependencies could arise through adaptive gene loss: organisms that lose the capacity to produce a costly compound (non-producers) will have a selective advantage over organisms that produce and inevitably leak that compound (producers). This could give rise to an obligate dependency of non-producers on producers ¹³, or, in the case of more than one leaky function, to obligate cross-feeding (bidirectional dependency) ¹⁴. However, little is known about the conditions under which these dependencies would be established.

A limited number of theoretical studies have recently explored this question using ecological models ¹⁵⁻¹⁸, evolutionary game theory (see ¹⁹⁻²¹ for comprehensive reviews) and concepts from economics ²². While these approaches have provided valuable phenomenological insight into the general principles of metabolic interdependencies, they generally do not take into account the specific details of the organisms, pathways and molecules involved: behind the production, leakiness, and utilization of each metabolite, is a complex network of biochemical reactions, which may significantly vary from organism to organism, and across different conditions and exchanged metabolites. A powerful avenue to address this gap is the use of systems biology methods, such as stoichiometric models of metabolism ⁷. These models take into account the full metabolic circuitry of a cell and provide quantitative predictions of its growth capacity and metabolic fluxes. Recent work has started applying these approaches to model microbial communities (see ²³ for a recent review) and to study the evolution of adaptive diversification in long-term evolutionary experiments ²⁴. However, a systematic analysis of the underlying mechanisms and patterns of possible equilibrium states of inter-species interactions

as a function of the leakiness of different metabolites is still unexplored. Here we propose a new hybrid modeling approach that combines the theoretical insight of evolutionary game theory with the organism-specific detailed analysis of cell-wide metabolic networks. We demonstrate how this strategy allows us to map the landscape of possible inter-species interactions, for which genome-scale models provide unique mechanistic insight.

Our mechanistic evolutionary game theory approach uses estimates of microbial fitness (payoff), based on genome-scale metabolic models, (forming the “payoff matrix” of the game). These payoffs are subsequently utilized to compute the Nash equilibria of “microbial interaction games” (see Methods). A Nash equilibrium is a central concept in game theory, defined as a state where no player can increase its payoff by a unilateral change of strategy. These payoffs also allows us to determine which of the identified Nash equilibria are evolutionarily stable by modeling the dynamics of genotype frequencies ²⁵ (see Figure 1 and Methods). As a proof-of-concept of our approach, we modeled the yeast sucrose hydrolysis system²⁶: when growing on sucrose, *S. cerevisiae* produces the surface enzyme invertase (encoded by the *suc2* gene), hydrolyzing sucrose into glucose and fructose, part of which serve as a public good. Since invertase production is metabolically costly, a mutant strain, which has lost its *suc2* gene, may emerge (see Figure 2A). Our simulations (based on the *iAZ900* yeast metabolic model ²⁷) reproduced the three types of equilibria observed experimentally ²⁶, i.e. Prisoners’ Dilemma (non-producers dominate and the community collapses), Mutually Beneficial game (producers dominate), and Snowdrift game (producers and non-producers coexist) (see Figure 2B-2F, and supplementary text for details).

We next sought to characterize the landscape of possible ecological interactions in a large set of strains and exchanged metabolites. In particular, we explored the evolution of metabolic dependencies mediated by the leaky production of individual amino acids in *E. coli* and asked how the evolution of these dependencies varies across the 20 amino acids and different leakiness levels. Here, a prototrophic wild-

type (WT, producer), leaking a given amino acid could interact with a mutant strain (MT, non-producer) lacking the gene(s) for the biosynthesis of that amino acid (see Figure 2G). The *iJO1366* genome scale model of *E. coli*²⁸ was used to construct *in silico* producer and non-producer strains, and to explore what equilibria can be established for a given leakiness level. As shown in Figure 2H, Prisoner's Dilemma, Mutually Beneficial and Snowdrift outcomes are the major possible equilibria, similar to the yeast system (Figure 2B) (with the amino acid cost replacing the sucrose hydrolysis cost). However, a more complex pattern is observed here for the Mutually Beneficial region (see supplementary text for details), highlighting the organism- and product-specific nature of our model. In addition, one can observe that, for each amino acid, there is a threshold for leakiness level above which non-producer mutants (MT) dominate, leading to community collapse (i.e., Prisoner's Dilemma) (Figure 3H). Thus, we may expect the amino acids secretion levels in *E. coli* to lie below this threshold. Interestingly, the average of multiple published amino acid secretion datasets (see Figure 3J) displays a consistent trend of leakiness levels decreasing with increasing cost. To assess the evolutionary stability of above Nash equilibria we performed *in silico* invasion experiments, where a resident population of WT is invaded by a low frequency MT (and vice versa). This analysis showed that the equilibrium frequencies of WT and MT are independent of their initial frequencies (see Figure 3I). This had been theoretically¹⁴ and experimentally^{14,29} suggested to stem from the negative frequency dependence of fitness. In addition to recapitulating this pattern, our analysis provides a quantitative prediction of the selection coefficients of the 20 amino acids. As shown in Figure 3K, the predicted selection coefficients display reasonable agreement with previous experimental reports for *E. coli*²⁹. This establishes a nontrivial quantitative link between metabolic stoichiometry and important ecological parameters.

We further extended our analysis to map the landscape of ecological interactions in communities with two leaky amino acids. Under what conditions would the increased number of exchangeable metabolites give rise to more complex inter-species interdependencies, such as reciprocal exchanges? Four different genotypes

are possible in this case (see Figure 3A): a prototrophic genotype that produces and leaks two amino acids (i.e., a full producer, denoted as '11'), two partial producer mutants each lacking the gene(s) for the biosynthesis of one amino acid (denoted as '01' and '10'), and a no-producer mutant strain lacking the biosynthesis genes for both amino acids (denoted as '00'). In a community composed of all these four genotypes, different types of interactions are possible. Here we focus on pairwise interactions, such as cross-feeding, [01, 10] and unidirectional dependency, [00, 11], though higher-order interactions among three or four genotypes (e.g., [00, 01, 10] and [00, 01, 10, 11]) are possible as well. We next systematically computed all Nash equilibria, at varying leakiness levels, for 189 amino acid pairs, corresponding to all possible pairs of 20 amino acids (see Figure 3B) except for one, namely the (alanine, isoleucine) pair, because the 00 genotype for this pair is auxotrophic for a third amino acid, valine. (see supplementary Figures S2 to S7 for higher order interactions' equilibria).

As shown in Figure 3B, a wide spectrum of equilibria ensues across different amino acid pairs and across different leakiness levels (e.g., see Figures 4C and 4D). Interestingly, despite the diversity of strains and exchanged metabolites, a majority of pairs are found to conform to general ecological patterns: for example, 139 (out of 189, or 73.5%) of amino acid pairs display a qualitatively identical region of leakiness levels (green regions in Figures 4B,C), where a unidirectional association, [00, 11], and cross-feeding, [01, 10], simultaneously emerge as Nash equilibria. The leakiness levels leading to this mixed equilibrium are the ones for which the full producer (i.e., 11 genotype) can still sustain growth. We refer to this region as *sustainable leakiness region* (see Methods). This is analogous to the maximum leakiness level in the study of individual amino acid secretions (Figure 2H): any leakiness levels for the two amino acids that lie outside this region would be expected to lead to extinction of the wild-type.

Another feature common to several amino acid pairs is the existence of a region in the leakiness plane (shown by red in Figure 3C) where cross-feeding (i.e., [01, 10])

is the only viable association, since excessive leakiness makes the full producer, 11, non-viable. This region is contiguous to the green region (with [00, 11] and [01, 10] as Nash equilibria). One interesting aspect of this configuration is that cross-feeding could initially ensue in the green region, coexisting with [00, 11] genotypes, and gradually move towards the red region (by evolving increased leakiness through selective advantage relative to the ancestor 11), leading to obligate cross-feeding as the only viable association.

In addition to exploring the landscape of Nash equilibria across different leakiness levels, it is interesting to ask whether the details of the biochemical networks underlying the genome-scale metabolic model predictions matter, and whether they provide direct explanatory power at the ecological level. Notably, for 50 amino acid pairs, the region of sustainable leakiness levels does not conform to the general pattern described above. For some pairs this region is partitioned into a number of sub-regions each corresponding to a different equilibrium (e.g. see Figure 3D) while in some extreme cases this entire region corresponds to a single Nash equilibrium, e.g., for (arginine, glutamate) and (glycine, threonine) (see below).

For the first anomalous pair, (arginine, glutamate), Nash equilibria in the sustainable leakiness region includes [10, 11] and [00, 11], but not the cross-feeding state [01, 10] (see Figure 3E). Inspection of the biosynthesis pathway of arginine and glutamate in *E. coli* revealed that glutamate is required for the production of ornithine, which serves as an essential precursor for the biosynthesis of arginine. This implies that a mutant strain lacking the biosynthesis pathways for glutamate (i.e., strain 10) will not be able to synthesize and leak arginine, thus acting like a 00 genotype and preventing the occurrence of cross-feeding. This observation is consistent with a previous report on the inability of arginine and glutamate auxotrophic mutant strains to grow in a co-culture under the minimal medium ³. This effect is due to a pleiotropic metabolic gene (i.e. a gene whose modification affects more than one metabolic phenotype), and illustrates how core biochemistry can impact ecological interactions.

1

2 A more complex scenario occurs for the (glycine, threonine) pair, where cross-
 3 feeding, [01, 10], is the only Nash equilibrium that emerges in the sustainable
 4 leakiness region (see Figure 3F). What prevents [00, 11] from being a Nash
 5 equilibrium? One of the conditions for [00, 11] to be a Nash equilibrium is that, in
 6 presence of 11, the fitness of the non-producer (00) should be higher than any of the
 7 partial producers (01 and 10) (as one would intuitively expect, because 00 does not
 8 incur the production cost of the two amino acids). In this case, however, it turns out
 9 that 00 is less fit than 10 (*thr*⁻ mutant, see a sample payoff matrix in supplementary
 10 Figure S8) thereby preventing [00, 11] from being a Nash equilibrium. This
 11 anomalous effect is due to the fact that the concurrent removal of glycine and
 12 threonine biosynthesis genes (in 00) will lead to a reduction in the capacity to
 13 produce other essential biomass components (such as serine). Interestingly, this is a
 14 case of diminishing-return (or “negative”) epistasis, in which the effect of the double
 15 mutation (00) is less severe than expected based on the two single mutations (01
 16 and 10). Negative epistatic interactions preventing the appearance of [00, 11] as the
 17 Nash equilibrium can be observed for 34 other amino acid pairs (see supplementary
 18 Table S1). This pattern is supported by existing experimental reports showing that
 19 epistasis correlates negatively with the expected fitness of multiple “genome
 20 streamlining” mutations in *E. coli*, thereby causing diminishing returns³⁰. These
 21 results provide mechanistic insights into how epistatic interactions among
 22 intracellular pathways can affect ecological interactions, a feature that cannot be
 23 captured by abstract phenomenological models.

24

25 It is next interesting to ask whether our approach can shed light onto the possible
 26 paths towards the rise of different interactions, especially cross-feeding. In the
 27 landscapes of identified Nash equilibria (Figure 3B), cross feeding ([01, 10])
 28 emerges together with other equilibria (such as [00, 11], e.g. in the green region in
 29 Figure 3C). This raises the question of whether and under what conditions cross-
 30 feeding would be evolutionarily stable. By performing a number of targeted *in silico*
 31 invasion experiments (see Methods), we found that this depends strongly on the

initial frequencies of the different genotypes (shown in Figure 3A). In particular, our analysis demonstrates that cross-feeders (01 and 10) will go extinct if they invade the full producer (11) in presence of the non-producer genotype (00) (see Figure 4A). However, the cross-feeders can subsist if 00 invades at a later stage, e.g. if it arises from mutations in the partial producers (01 or 10) (see Figures 4B and 4C for two such mechanisms). An example of the latter scenario is the two-step process hypothesized in ¹⁴: first, the biosynthetic capacity for one amino acid is lost, e.g. resulting in a 01 genotype, which could equilibrate and co-exist with 11 (see Figures 2G-2I). In the second step, either 01 or 11 may lose their capacity for producing the second amino acid (because the other strain can compensate), giving rise to 00 and 10 genotypes, respectively (see Figure 4B). Here, we quantitatively explored this hypothesis by assessing the evolutionary dynamics for the second step, assuming that equilibration of the first step has already occurred. As shown in Figure 4B, the prototrophic (11) and no-producer (00) genotypes always survive in this competition, while cross-feeders evolutionarily emerge only at high leakiness levels. Further *in silico* invasion experiments demonstrate that established cross-feeding pairs tend to be resistant to invasion by non-producers (see supplementary Figure S10) and by prototrophs (see supplementary Figure S11). Thus, once established, obligate mutual metabolic exchange could be evolutionarily stable against invasion by other genotypes, even in a homogenous environment, consistent with previous experimental reports ³¹. Notably this stability against invasions is dependent on the specific metabolites exchanged and on the level of metabolic exchange (leakiness).

We demonstrated here that by adding new layers of details to abstract theoretical ecology models, we can reveal how intracellular molecular mechanisms (including pleiotropy and epistasis of metabolic enzyme genes) lead to the rise of ecological interactions. The analysis we presented spans over 80,000 *in silico* experiments (across 189 amino acid pairs and 441 leakiness level combinations), which is beyond the current experimental capabilities. This study provides a guideline for the design of future targeted experiments, e.g. built upon previously established synthetic communities ^{3-6,11}. For example, our computational results could be used

1 to suggest choices of metabolite pairs and ranges of engineered leakiness levels that
 2 lead to the establishment of a specific inter-dependency, e.g., unidirectional vs.,
 3 cross-feeding. In addition, our study offers a basis for better understanding of
 4 metabolic interdependencies in natural microbial communities, such as those in the
 5 human gut microbiota ⁸⁻¹⁰. From the perspective of biotechnological applications,
 6 our approach lays the foundation for proactively incorporating evolutionary
 7 concepts in the *de novo* design of synthetic microbial consortia that are resistant to
 8 invasion by competing strategies ²³.

9 **Methods**

10 **Background on evolutionary game theory.** Evolutionary game theory is the
 11 application of classical game theory to model the evolutionary dynamics of mixed
 12 populations. Modeling microbial communities with evolutionary game theory
 13 involves two steps: (i) Considering all pairwise interactions among genotypes and
 14 estimating the payoff (fitness) of a microbial player k upon interacting with a
 15 partner k' . Higher order interactions can be similarly defined; These estimated
 16 payoffs are represented in the form of a matrix, referred to as payoff matrix. (ii)
 17 Using the payoff matrix to identify the Nash equilibrium(s) – a fundamental concept
 18 in game theory defined as a state where no player has an incentive to unilaterally
 19 change its current strategy, because it cannot improve its payoff by doing so. An
 20 *evolutionary stable strategy* is a similar concept in evolutionary game theory: It is a
 21 Nash equilibrium, which is evolutionarily stable, i.e., natural selection alone is
 22 sufficient to prevent invasion by competing mutant strategies. Evolutionary stable
 23 strategies can be found by modeling the evolutionary dynamics of the game using
 24 the computed payoffs (see the following sections). In this text, by ‘evolutionary
 25 dynamics’ we mean how the relative genotype abundances (frequencies, or
 26 community structure) change over time following evolutionary game theory ²⁵, and
 27 microbial ecology literature ³². This aspect of evolutionary dynamics, as opposed to
 28 the simulation of *de novo* mutations and subsequent selection processes (which
 29 have been pursued in other studies ³³), is the main focus of our analysis.

Background on Flux Balance Analysis (FBA). FBA (described in detail elsewhere, e.g. ³⁴) is a linear optimization problem that uses genome-scale metabolic models to make quantitative predictions about the cell's growth rate, intracellular reaction fluxes and secretion rates of metabolites that are potentially excreted by the cell under a given condition. In the most common formulation, this is achieved by maximizing the flux of a pseudo-reaction called *biomass reaction* ($v_{biomass}$) whose reactants are precursors required for growth and whose flux is indicative of the cell's growth capacity. This is subject to constraints imposing the steady-state mass balance for each metabolite in the network (see Constraint 1 below) and lower and upper bounds on reactions fluxes to impose the reversibility of reactions and uptake and aeration conditions (see Constraint 2 below). The standard formulation of FBA is as follows:

maximize $v_{biomass}$

subject to

$$\sum_{j \in J} s_{ij} v_j = 0, \quad \forall i \in I, \quad (1)$$

$$LB_j \leq v_j \leq UB_j, \quad \forall j \in J, \quad (2)$$

where, I is the set of metabolites, J is the set of reactions, s_{ij} is the stoichiometric coefficient of metabolite i in reaction j (known from the metabolic model), LB_j and UB_j denote lower and upper bounds on flux of a reaction j (provided as inputs), respectively, and v_j is the flux of a reaction j (optimization variables).

Using genome-scale models and FBA to compute payoffs for interacting microbes. We used FBA to provide organism-specific and genomically-informed estimates of the payoffs upon specific pairwise interactions (see also Figure 1). For a given pair of genotypes k and k' , we solve a separate FBA problem for k and k' . Each FBA problem involves two new types of constraints added to the standard FBA: (i) The first type of constraints pertains to implementing *in silico* the gene deletions

that correspond to the genotype under consideration (see Equation 3 below). For example, if genotype k is auxotroph for lysine, this auxotrophy can be in principle induced by knocking out gene $lysA$, which codes for diaminopimelate decarboxylase (DAPC). In the FBA model, this gene deletion is simulated by setting $v_{DAPDC} = 0$. More complex gene-to-reaction mappings could lead to more complex set of constraints. In cases where several different choices are possible for genes whose deletion would induce a given auxotrophy, we select one. (ii) The second type of constraints (Equations 4 and 5) simulates the exchange of metabolites between the genotype under consideration and its partner(s). In our calculations, for any given exchanged metabolite (e.g., amino acid or amino acids pair), we systematically explore the effects of a range of possible leakiness levels. Thus in each *in silico* experiment, the leakiness levels of amino acids secreted by each genotype are fixed at a pre-specified value. Similarly, metabolites leaked by the partner are made available to this genotype through appropriate upper bounds on import fluxes. The optimal value of biomass flux obtained upon solving this FBA problem provides an estimate of the growth rate of each genotype in a given pairwise interaction, which we use as a proxy for its payoff. For example, the payoff of k when facing k' ($a_{kk'}$) is $v_{biomass}^k$ and the payoff of k' when facing k ($a_{k'k}$) is $v_{biomass}^{k'}$. The general form of the FBA formulation for each genotype is mathematically formulated as follows:

maximize $v_{biomass}$

subject to

$$\sum_{j \in J} s_{ij} v_j = 0, \quad \forall i \in I, \quad (1)$$

$$LB_j \leq v_j \leq UB_j, \quad \forall j \in J, \quad (2)$$

$$v_j = 0, \quad \forall j \in J^{mutant} \quad (3)$$

$$v_{EX_i(e)} \geq -u_i, \quad \forall i \in I^{auxotroph} \quad (4)$$

$$v_{EX_i(e)} \geq e_i, \quad \forall i \in I^{leaky} \quad (5)$$

where, $J^{mutation}$ denotes the set of reactions corresponding to specific gene mutations for the genotype under consideration. In addition, $I^{auxotroph} \subset I$

represents the set of metabolites that this genotype is auxotroph for, but that are provided by other genotypes, and $I^{leaky} \subset I$ is the set of leaky (secreted) metabolites by the genotype under consideration. $v_{EX_i(e)}$ denotes the flux of exchange reaction for a metabolite i . $u_i > 0$ and $e_i > 0$ denote the pre-specified net uptake and export flux of a metabolite i , respectively. Input parameters for this FBA problem include: (i) The list of metabolites that each genotype is leaking and those available for uptake from the partner genotype in a pairwise (or higher order) interaction, (ii) The leakiness level of exchanged metabolites, and (iii) the net uptake and export fluxes of exchanged metabolites (i.e., e_i and u_i) calculated based on the fixed leakiness levels and the specific interacting genotypes (see supplementary text for details). Constraint (3) sets to zero the flux of reactions corresponding to the specific gene mutations in the genotype under consideration. Constraints (4) allows for the uptake of metabolites available from partner genotype(s) in a pairwise or (higher-order) interaction. Constraint (5) requires the export of leaky metabolites at the pre-specified level e_i . The payoff of the genotype under consideration is set to the optimal value of the biomass flux, or to the death rate (a negative value), in the case of an infeasible problem. An infeasible FBA problem may occur due the lack of enough carbon source to satisfy maintenance ATP requirements in the model or due to imposing a high level of leakiness for a leaked metabolite. Imposed leakiness level causing this infeasibility are referred to ‘unsustainable leakiness levels’. The details of specific formulations for the presented case studies with *S. cerevisiae* and *E. coli* are given in the supplementary text. Additional environmental/strategic/genetic conditions can be incorporated through the addition of appropriately defined constraints. In addition, one can use a different objective function (e.g., the minimization of metabolic adjustment ³⁵) or other constraint-based community modeling tools e.g., ^{36,37} as an alternative.

Optimization-driven automated identification of the Nash equilibria. Upon constructing the payoff matrix, as described above, one can identify the Nash equilibrium(s) of the game. We developed an efficient optimization-based

1 procedure called *NashEq Finder* to automate the identification of all pure strategy
 2 Nash equilibria of a game given its payoff matrix. NashEq Finder is an integer linear
 3 program (ILP), which relies on binary variables to decide on whether or not a
 4 particular entry of the payoff matrix satisfies the conditions of a Nash equilibrium
 5 (see supplementary text for a detailed formulation). This algorithm is able to
 6 identify all possible Nash equilibria of a game with any number of players and thus
 7 can be reliably used for the rapid identification of the equilibrium states of complex
 8 communities.

9

10 **Modeling the evolutionary dynamics at a genome-scale resolution.** Following
 11 standard approaches in evolutionary game theory we model evolutionary dynamics
 12 using the Replicator Equation ²⁵. In particular, to take into account interactions
 13 higher than pairwise, we used an extended form of the classical Replicator equation
 14 ²⁵. This equation predicts the changes in the relative abundance (frequencies) of
 15 genotypes over time according to their reproductive fitness under the assumption of
 16 a roughly constant population size ²⁵ (see supplementary text for the difference
 17 between this equation and multi-species dynamics flux balance analysis ³⁸). Our
 18 formulation, following ³⁹, can be expressed as follows:

$$\frac{dx_k}{dt} = (f_k(\mathbf{x}) - \phi(\mathbf{x}))x_k, \quad k = 1, 2, \dots, K, \quad (6)$$

$$f_k(\mathbf{x}) = \sum_{k'=1}^K a_{kk'}x_{k'} + \sum_{k'=1}^K \sum_{k''=1}^K a_{kk'k''}x_{k'}x_{k''} + \dots, \quad k = 1, 2, \dots, K, \quad (7)$$

$$\phi(\mathbf{x}) = \sum_{k'=1}^K f_{k'}(\mathbf{x})x_{k'}. \quad (8)$$

19 Here, K denotes the set of community members and $\mathbf{x} = [x_1, x_2, \dots, x_K]^T$ is the
 20 composition of the community with x_k being the frequency of genotype k . $f_k(\mathbf{x})$ is
 21 the average reproductive fitness of genotype k that depends not just on other
 22 genotypes it may encounter but also on their frequencies. $a_{kk'}$ and $a_{kk'k''}$ denote the
 23 payoffs of genotype k when encountering another genotype k' in a two-player game

(i.e. pairwise interaction) or two other genotypes k' and k'' in a three-player game, respectively. Finally, $\phi(x)$ is the average fitness of the entire community.

3

Here we use the replicator equation to perform targeted *in silico* invasion experiments, in which a newly emerged low frequency genotype invades an existing resident genotype. While such *in silico* experiments are based only on evolutionary dynamics (modeled through the Replicator equation), future developments could take into account additional aspects such as eco-evolutionary feedback^{32,40-42}.

9

10 References

- 11 1 Zelezniak, A. *et al.* Metabolic dependencies drive species co-occurrence in
12 diverse microbial communities. *Proc Natl Acad Sci U S A* **112**, 6449-6454,
13 doi:10.1073/pnas.1421834112 (2015).
- 14 2 Freilich, S. *et al.* Competitive and cooperative metabolic interactions in
15 bacterial communities. *Nat Commun* **2**, 589, doi:10.1038/ncomms1597
16 (2011).
- 17 3 Wintermute, E. H. & Silver, P. A. Emergent cooperation in microbial
18 metabolism. *Mol Syst Biol* **6**, 407, doi:10.1038/msb.2010.66 (2010).
- 19 4 Mee, M. T., Collins, J. J., Church, G. M. & Wang, H. H. Syntrophic exchange in
20 synthetic microbial communities. *Proc Natl Acad Sci U S A* **111**, E2149-2156,
21 doi:10.1073/pnas.1405641111 (2014).
- 22 5 Shou, W., Ram, S. & Vilar, J. M. Synthetic cooperation in engineered yeast
23 populations. *Proc Natl Acad Sci U S A* **104**, 1877-1882,
24 doi:10.1073/pnas.0610575104 (2007).
- 25 6 Hoek, T. A. *et al.* Resource Availability Modulates the Cooperative and
26 Competitive Nature of a Microbial Cross-Feeding Mutualism. *PLoS Biol* **14**,
27 e1002540, doi:10.1371/journal.pbio.1002540 (2016).
- 28 7 Becker, S. A. *et al.* Quantitative prediction of cellular metabolism with
29 constraint-based models: the COBRA Toolbox. *Nat Protoc* **2**, 727-738,
30 doi:10.1038/nprot.2007.99 (2007).
- 31 8 Rakoff-Nahoum, S., Foster, K. R. & Comstock, L. E. The evolution of
32 cooperation within the gut microbiota. *Nature* **533**, 255-259,
33 doi:10.1038/nature17626 (2016).
- 34 9 Waldor, M. K. *et al.* Where next for microbiome research? *PLoS Biol* **13**,
35 e1002050, doi:10.1371/journal.pbio.1002050 (2015).
- 36 10 Magnúsdóttir, S. *et al.* Generation of genome-scale metabolic reconstructions
37 for 773 members of the human gut microbiota. *Nat Biotechnol* **35**, 81-89,
38 doi:10.1038/nbt.3703 (2017).

- 1 11 Harcombe, W. Novel cooperation experimentally evolved between species.
2 *Evolution* **64**, 2166-2172, doi:10.1111/j.1558-5646.2010.00959.x (2010).
- 3 12 Zhang, X. & Reed, J. L. Adaptive evolution of synthetic cooperating
4 communities improves growth performance. *PLoS One* **9**, e108297,
5 doi:10.1371/journal.pone.0108297 (2014).
- 6 13 Morris, J. J., Lenski, R. E. & Zinser, E. R. The Black Queen Hypothesis:
7 evolution of dependencies through adaptive gene loss. *MBio* **3**,
8 doi:10.1128/mBio.00036-12 (2012).
- 9 14 Morris, J. J. Black Queen evolution: the role of leakiness in structuring
10 microbial communities. *Trends Genet* **31**, 475-482,
11 doi:10.1016/j.tig.2015.05.004 (2015).
- 12 15 Oliveira, N. M., Niehus, R. & Foster, K. R. Evolutionary limits to cooperation in
13 microbial communities. *Proc Natl Acad Sci U S A* **111**, 17941-17946,
14 doi:10.1073/pnas.1412673111 (2014).
- 15 16 Mas, A., Jamshidi, S., Lagadeuc, Y., Eveillard, D. & Vandenkoornhuys, P.
16 Beyond the Black Queen Hypothesis. *ISME J*, doi:10.1038/ismej.2016.22
17 (2016).
- 18 17 Estrela, S., Morris, J. J. & Kerr, B. Private benefits and metabolic conflicts
19 shape the emergence of microbial interdependencies. *Environ Microbiol* **18**,
20 1415-1427, doi:10.1111/1462-2920.13028 (2016).
- 21 18 Germerodt, S. *et al.* Pervasive Selection for Cooperative Cross-Feeding in
22 Bacterial Communities. *PLoS Comput Biol* **12**, e1004986,
23 doi:10.1371/journal.pcbi.1004986 (2016).
- 24 19 Frey, E. Evolutionary game theory: Theoretical concepts and applications to
25 microbial communities. *Physica a-Statistical Mechanics and Its Applications*
26 **389**, 4265-4298, doi:10.1016/j.physa.2010.02.047 (2010).
- 27 20 Hummert, S. *et al.* Evolutionary game theory: cells as players. *Mol Biosyst* **10**,
28 3044-3065, doi:10.1039/c3mb70602h (2014).
- 29 21 Blaser, M. J. & Kirschner, D. The equilibria that allow bacterial persistence in
30 human hosts. *Nature* **449**, 843-849, doi:10.1038/nature06198 (2007).
- 31 22 Tasoff, J., Mee, M. T. & Wang, H. H. An Economic Framework of Microbial
32 Trade. *PLoS One* **10**, e0132907, doi:10.1371/journal.pone.0132907 (2015).
- 33 23 Zomorodi, A. R. & Segrè, D. Synthetic Ecology of Microbes: Mathematical
34 Models and Applications. *J Mol Biol*, doi:10.1016/j.jmb.2015.10.019 (2015).
- 35 24 Großkopf, T. *et al.* Metabolic modelling in a dynamic evolutionary framework
36 predicts adaptive diversification of bacteria in a long-term evolution
37 experiment. *BMC Evol Biol* **16**, 163, doi:10.1186/s12862-016-0733-x (2016).
- 38 25 Nowak, M. A. *Evolutionary dynamics : exploring the equations of life*. (Belknap
39 Press of Harvard University Press, 2006).
- 40 26 Gore, J., Youk, H. & van Oudenaarden, A. Snowdrift game dynamics and
41 facultative cheating in yeast. *Nature* **459**, 253-256, doi:10.1038/nature07921
42 (2009).
- 43 27 Zomorodi, A. R. & Maranas, C. D. Improving the iMM904 *S. cerevisiae*
44 metabolic model using essentiality and synthetic lethality data. *BMC Syst Biol*
45 **4**, 178, doi:10.1186/1752-0509-4-178 (2010).

1 28 Orth, J. D. *et al.* A comprehensive genome-scale reconstruction of Escherichia
2 coli metabolism--2011. *Mol Syst Biol* **7**, 535, doi:10.1038/msb.2011.65
3 (2011).

4 29 D'Souza, G. & Kost, C. Experimental Evolution of Metabolic Dependency in
5 Bacteria. *PLoS Genet* **12**, e1006364, doi:10.1371/journal.pgen.1006364
6 (2016).

7 30 D'Souza, G., Waschina, S., Kaleta, C. & Kost, C. Plasticity and epistasis strongly
8 affect bacterial fitness after losing multiple metabolic genes. *Evolution* **69**,
9 1244-1254, doi:10.1111/evo.12640 (2015).

10 31 Pande, S. *et al.* Fitness and stability of obligate cross-feeding interactions that
11 emerge upon gene loss in bacteria. *ISME J* **8**, 953-962,
12 doi:10.1038/ismej.2013.211 (2014).

13 32 Sanchez, A. & Gore, J. feedback between population and evolutionary
14 dynamics determines the fate of social microbial populations. *PLoS Biol* **11**,
15 e1001547, doi:10.1371/journal.pbio.1001547 (2013).

16 33 Elena, S. F. & Lenski, R. E. Evolution experiments with microorganisms: the
17 dynamics and genetic bases of adaptation. *Nat Rev Genet* **4**, 457-469,
18 doi:10.1038/nrg1088 (2003).

19 34 Orth, J. D., Thiele, I. & Palsson, B. What is flux balance analysis? *Nat Biotechnol*
20 **28**, 245-248, doi:10.1038/nbt.1614 (2010).

21 35 Segrè, D., Vitkup, D. & Church, G. M. Analysis of optimality in natural and
22 perturbed metabolic networks. *Proc Natl Acad Sci U S A* **99**, 15112-15117,
23 doi:10.1073/pnas.232349399 (2002).

24 36 Stolyar, S. *et al.* Metabolic modeling of a mutualistic microbial community.
25 *Mol Syst Biol* **3**, 92, doi:10.1038/msb4100131 (2007).

26 37 Zomorodi, A. R. & Maranas, C. D. OptCom: a multi-level optimization
27 framework for the metabolic modeling and analysis of microbial
28 communities. *PLoS Comput Biol* **8**, e1002363,
29 doi:10.1371/journal.pcbi.1002363 (2012).

30 38 Zhuang, K. *et al.* Genome-scale dynamic modeling of the competition between
31 Rhodoferax and Geobacter in anoxic subsurface environments. *ISME J* **5**, 305-
32 316, doi:10.1038/ismej.2010.117 (2011).

33 39 Cressman, R. & Tao, Y. The replicator equation and other game dynamics.
34 *Proc Natl Acad Sci U S A* **111 Suppl 3**, 10810-10817,
35 doi:10.1073/pnas.1400823111 (2014).

36 40 Post, D. M. & Palkovacs, E. P. Eco-evolutionary feedbacks in community and
37 ecosystem ecology: interactions between the ecological theatre and the
38 evolutionary play. *Philos Trans R Soc Lond B Biol Sci* **364**, 1629-1640,
39 doi:10.1098/rstb.2009.0012 (2009).

40 41 Rauch, J., Kondev, J. & Sanchez, A. Cooperators trade off ecological resilience
41 and evolutionary stability in public goods games. *J R Soc Interface* **14**,
42 doi:10.1098/rsif.2016.0967 (2017).

43 42 Widder, S. *et al.* Challenges in microbial ecology: building predictive
44 understanding of community function and dynamics. *ISME J* **10**, 2557-2568,
45 doi:10.1038/ismej.2016.45 (2016).

- 1 43 Paczia, N. *et al.* Extensive exometabolome analysis reveals extended overflow
2 metabolism in various microorganisms. *Microb Cell Fact* **11**, 122,
3 doi:10.1186/1475-2859-11-122 (2012).
4 44 Valle, J. *et al.* The amino acid valine is secreted in continuous-flow bacterial
5 biofilms. *J Bacteriol* **190**, 264-274, doi:10.1128/JB.01405-07 (2008).
6
7

8 **Figures' Legends**

9 **Figure 1. Outline of the proposed genome-driven evolutionary game theory**
10 **approach.** Annotated genomes of community members are used to construct
11 genome-scale metabolic models. For each possible pair of genotypes in the
12 community, constraint-based analysis tools for metabolic models such as Flux
13 Balance Analysis ³⁴ are used to estimate the fitness (or “payoff”) of each genotype as
14 they engage in a specific metabolite-mediated interaction. These payoffs form the
15 payoff matrix of the game. Based on this payoff matrix we identify all Nash equilibria
16 of the game, using a newly developed automated tool (NashEq Finder, see
17 supplementary text). The payoff matrix also allows us to model evolutionary
18 dynamics (i.e., how genotype frequencies change over time) ²⁵ and to determine
19 which of the identified Nash equilibria are evolutionarily stable (see Methods).
20

21 **Figure 2. Metabolic dependencies in populations of *S. cerevisiae* growing on**
22 **sucrose and in populations of *E. coli* secreting an amino acid.** (A) Metabolic
23 interactions between producer (wild-type, WT) and non-producer (mutant, MT)
24 genotypes of *S. cerevisiae* growing on sucrose ²⁶. Here, *e* represents the percentage
25 of glucose/fructose that diffuses away and serves as a public good. (B) Nash
26 equilibria and (C) the equilibrium frequency of WT for the community shown in (A)
27 as a function of the capture efficiency of the glucose/fructose and the invertase
28 production cost (modeled by changing the stoichiometric coefficient of ATP in the
29 sucrose hydrolysis reaction, indicated as *x*). An alternative *in silico* formulation of
30 the energetic cost of invertase production that reproduces exactly the setup used in
31 the experiment by Gore et al ²⁶ (based on histidine auxotrophy) proved to be

1 qualitatively equivalent to the analysis presented here (see supplementary text for
2 details). The equilibrium frequency of WT was obtained from *in silico* invasion
3 experiments (see Methods) for two cases of a small fraction of MT invading a
4 resident population of WT and vice versa demonstrated that the equilibrium
5 frequency of WT is the same in both cases. (D) Metabolic interactions between WT
6 and MT when additional glucose is provided in the growth medium (see
7 supplementary text for details). (E) Nash equilibria and (F) the equilibrium
8 frequency of WT in the presence of glucose in the growth medium. The entire
9 Snowdrift game region and part of the Mutually Beneficial region in (B) are replaced
10 by the Prisoner's Dilemma game, in (E), which is consistent with previous reports ²⁶
11 and serves as an additional verification of our modeling approach. This is because in
12 the presence of glucose, MT is less dependent on WT thereby increasing the average
13 fitness of MT. (G) Possible genotypes in populations of *E. coli* leaking an amino acid
14 include a prototrophic wild-type strain (WT) self-synthesizing a leaky amino acid
15 and a mutant strain (MT) lacking the gene(s) for the biosynthesis of this amino acid.
16 (H) The identified Nash equilibria for various leakiness levels (as a percentage of an
17 *in silico* determined maximum: see supplementary text) across all 20 amino acids.
18 Amino acids are shown here by using their standard three-letter code in the order of
19 increasing *in silico* growth cost (see also supplementary Figure S1). (I) The
20 equilibrium frequency of WT as a function of the leakiness level and amino acid
21 type. *In silico* invasion experiments for two cases of MT invading WT and WT
22 invading MT revealed that the equilibrium frequencies are insensitive to the initial
23 frequencies. (J) Experimentally reported leakiness levels of amino acids averaged
24 over three different datasets ^{43,44}. Values in each dataset were normalized to their
25 maximum (see supplementary Table S1 for values of data). Error bars show
26 standard deviation over the three datasets. (K) Predicted selection coefficients
27 across all amino acids and leakiness levels vs. the experimentally reported ones for
28 *E. coli* ²⁹. Empty circles show the average predicted selection coefficient for each
29 amino acid. The black and blue lines show the fitted lines to the predicted and
30 experimental values ²⁹, respectively. The equation for the fitted line to experimental
31 data was inferred from a graph of reference ²⁹.

1

2 **Figure 3. Equilibrium metabolic dependencies in populations of *E. coli* with**

3 **two leaky amino acids.** (A) Genotypes involved include a prototrophic strain self-

4 synthesizing two leaky amino acids (i.e., 11), two single-mutant strains each lacking

5 the gene(s) for the biosynthesis of one amino acid but synthesizing and leaking the

6 other (i.e., 01 and 10), and a mutant strain lacking the genes for the biosynthesis of

7 both leaky amino acids (i.e., 00). Here, '1' and '0' denote the presence or absence of

8 biosynthesis pathways (genes) for an amino acid, respectively. (B) The identified

9 Nash equilibria of two-player games (i.e., pairwise interactions) for all amino acids

10 pair across different leakiness levels, zoomed in for four selected pairs including (C)

11 (lysine, isoleucine), (D) (glutamate and leucine), (E) (arginine, glutamate) and (F)

12 (glycine, threonine). A sample payoff matrix of the game is shown in (C) for a

13 leakiness level of 15% for both lysine and isoleucine. Non-viable equilibria signify

14 associations between genotypes that are non-viable leading to community collapse.

15 Nash equilibria of three- and four-player games for a selected number of amino acid

16 pairs are also given in supplementary Figures S2-S7. Metabolic maps in (E) And (F)

17 show the metabolic pathways involved in the synthesis of the respective amino

18 acids. Sample payoff matrices in the region of sustainable leakiness levels in (E) and

19 (F) are given in supplementary Figure S8.

20

21 **Figure 4. Impact of the initial genotype frequencies on the evolutionary**

22 **emergence of metabolic dependencies in populations of *E. coli* with two leaky**

23 **amino acids.** Here, we have shown the results of targeted *in silico* invasion

24 experiments for a representative amino acid pair (lysine, isoleucine) (see also

25 Figure 3c). (A) 00, 01 and 10 simultaneously originate from 11 (i.e., wild-type)

26 through genome streamlining and invade an existing population of 11 genotypes.

27 (B) A small population of 10 and 00 invade a resident population of the 11 and 01.

28 This simulates the second step of a two-step process for the loss of the leaky

29 functions hypothesized in ¹⁴ (see the main text and also supplementary Figure S9).

30 (C) A small population of 01 and 10 invades a resident population of 11. This models

31 an alternative scenario for the two-step loss of leaky functions leading to stable

1 cross-feeders: Two partial producer mutant genotypes (01 and 10) originate from
2 11, followed by the rise of the 00 genotype from 01 and/or 10. As shown here,
3 cross-feeders can evolutionarily stabilize and co-exist with 11 genotypes in the first
4 step. Further analysis showed that cross-feeders are also resistant to invasion by 00
5 genotypes arising in the second step (see supplementary Figure S10). Dynamic plots
6 in (A)–(C) show the sample evolutionary dynamics of the system for selected equal
7 leakiness levels for both lysine and isoleucine. Pie charts show the equilibrium
8 frequencies of each genotype starting from the initial genotype frequencies shown
9 in each panel. These equilibrium frequencies are given only for the sustainable
10 leakiness region (green region in Figure 3C).

Figure 1

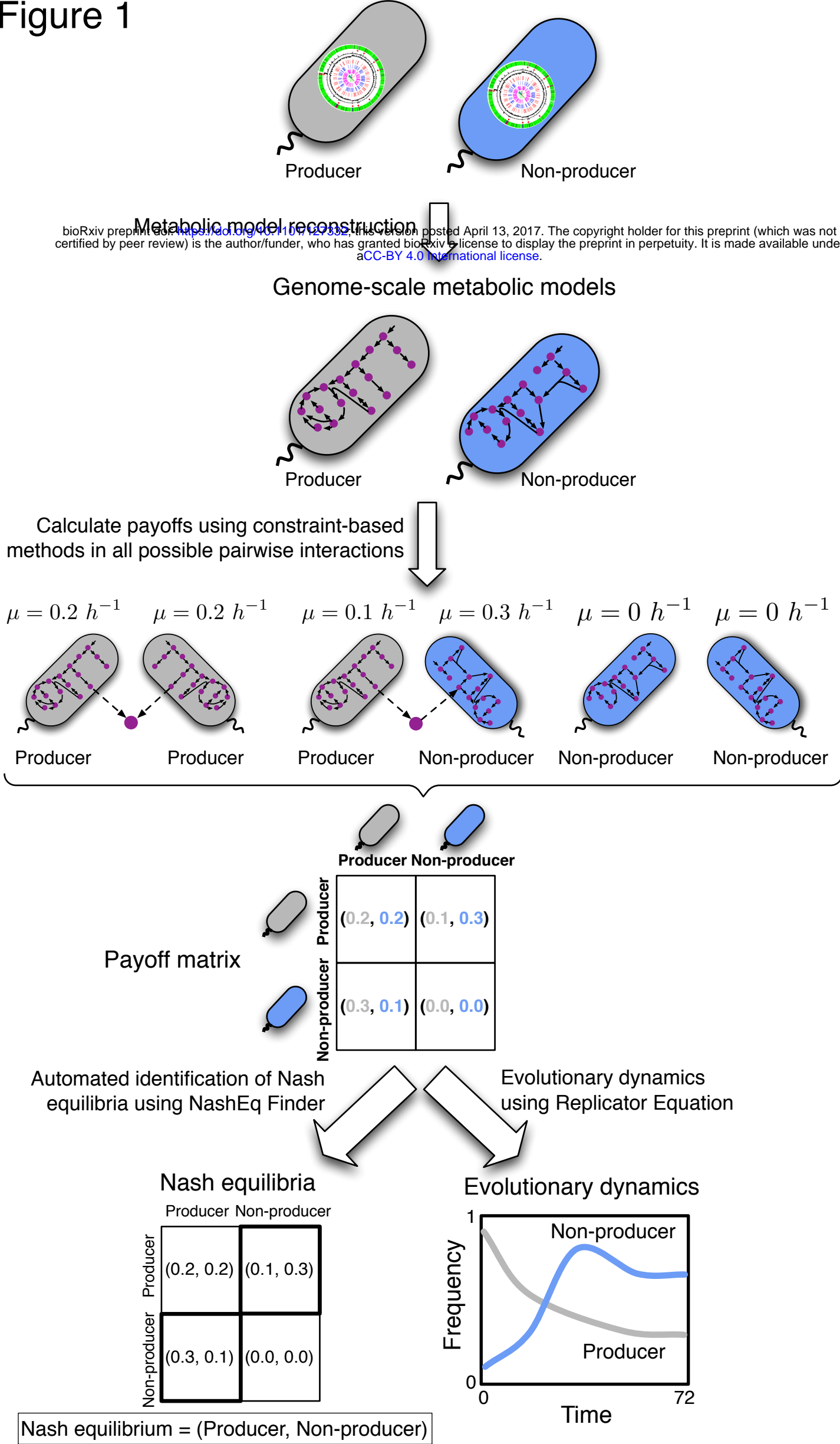
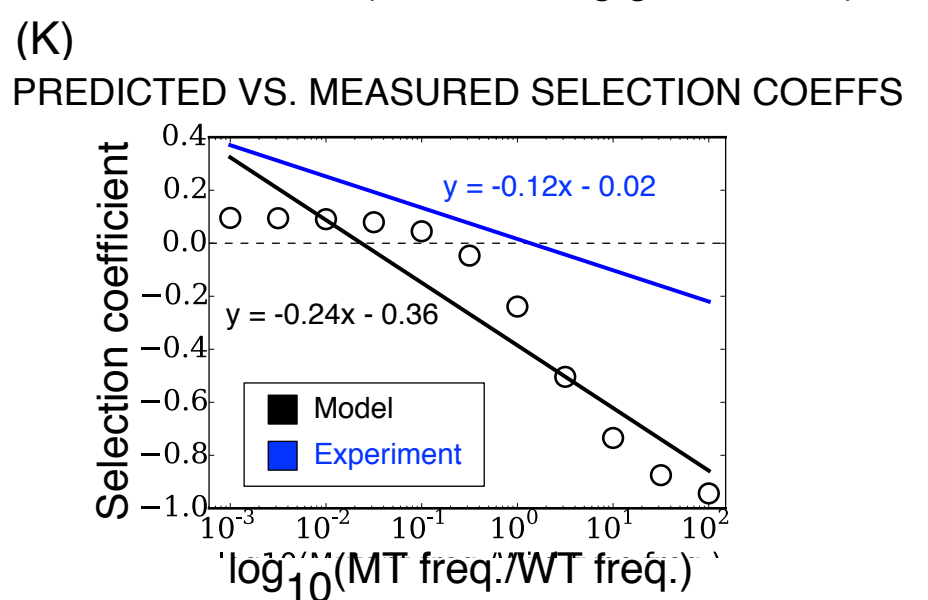
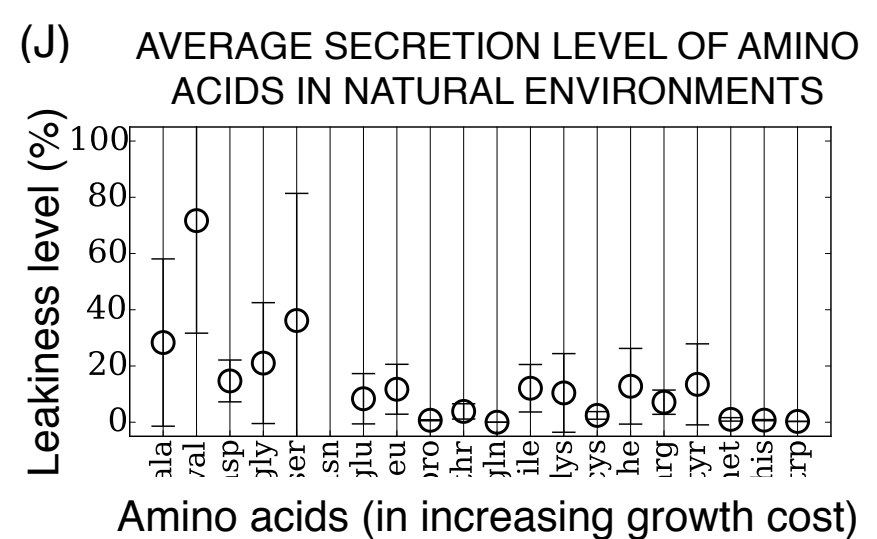
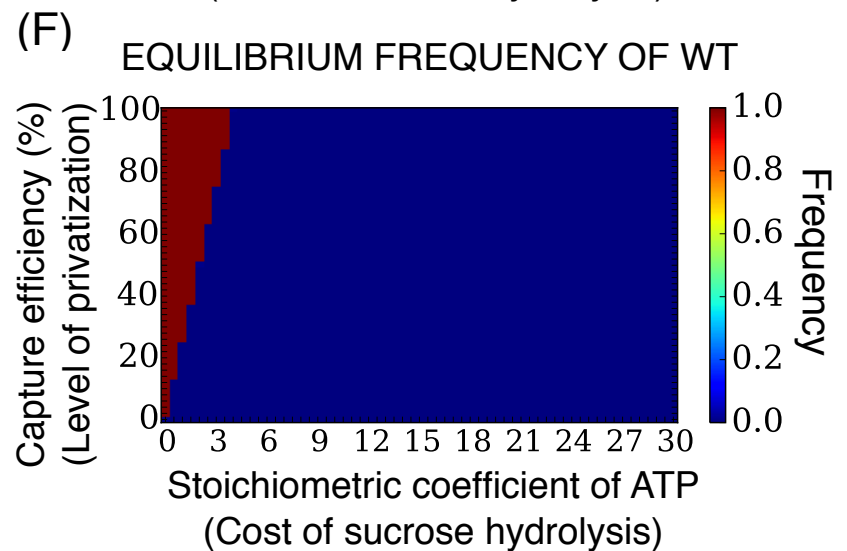
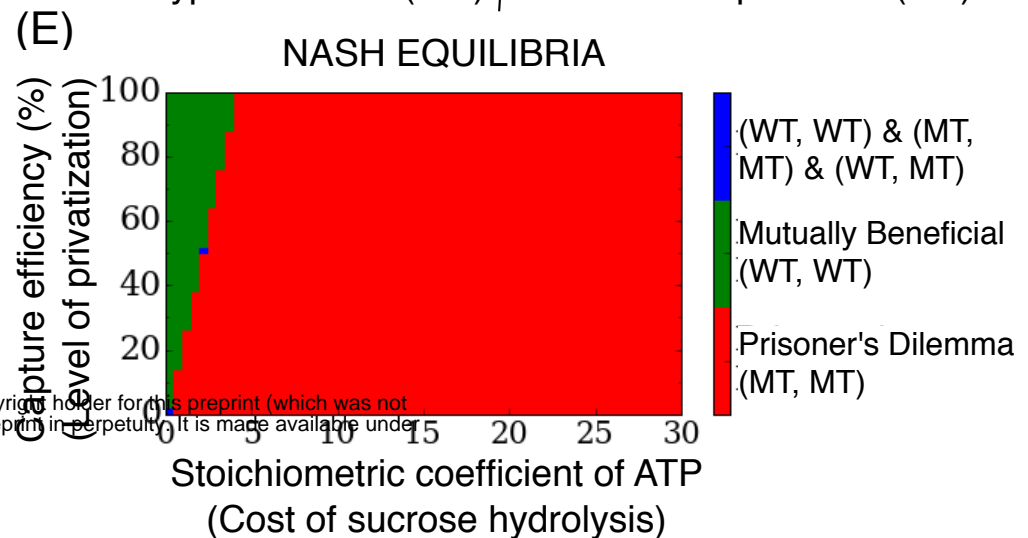
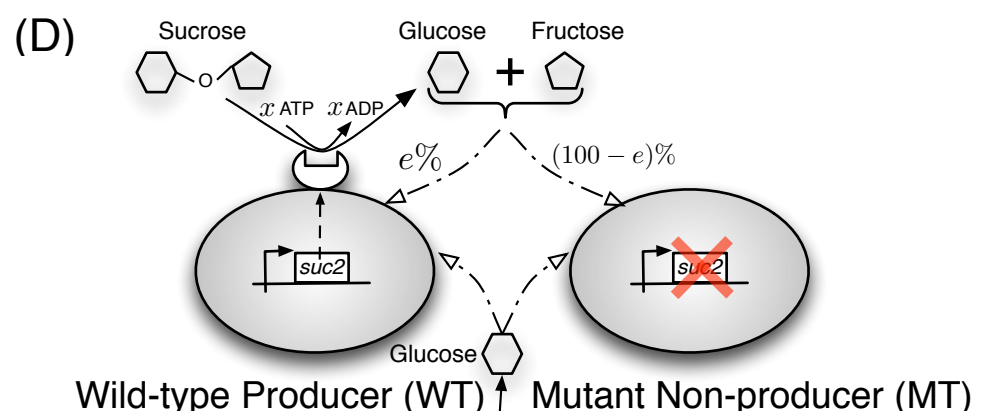
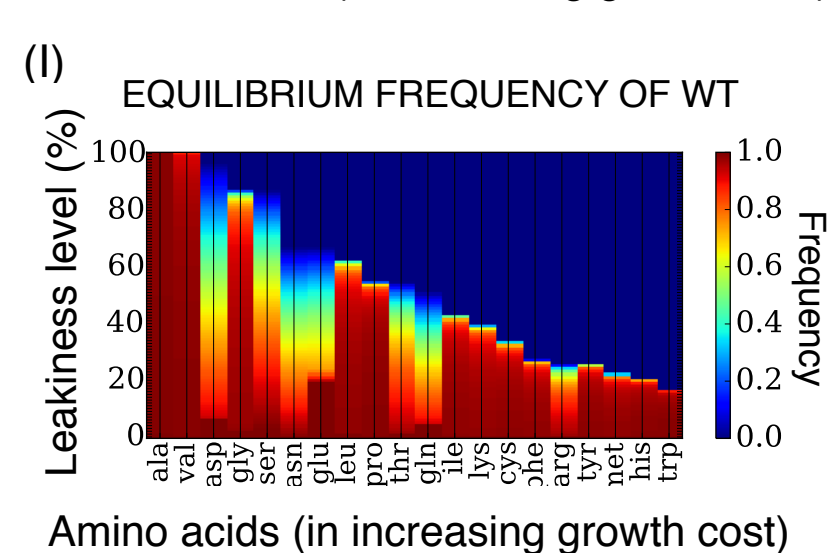
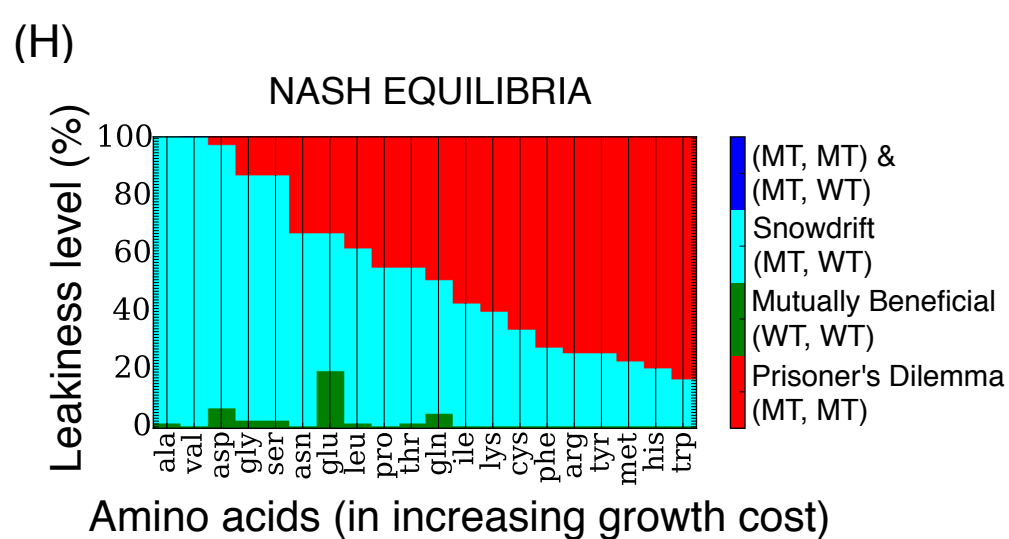
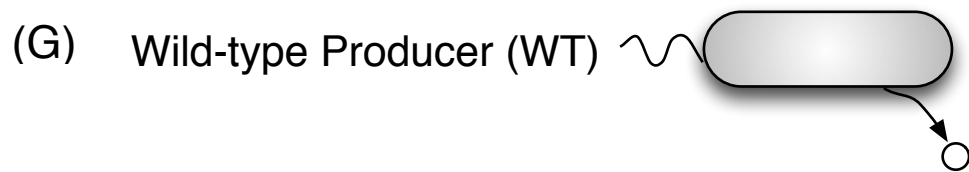
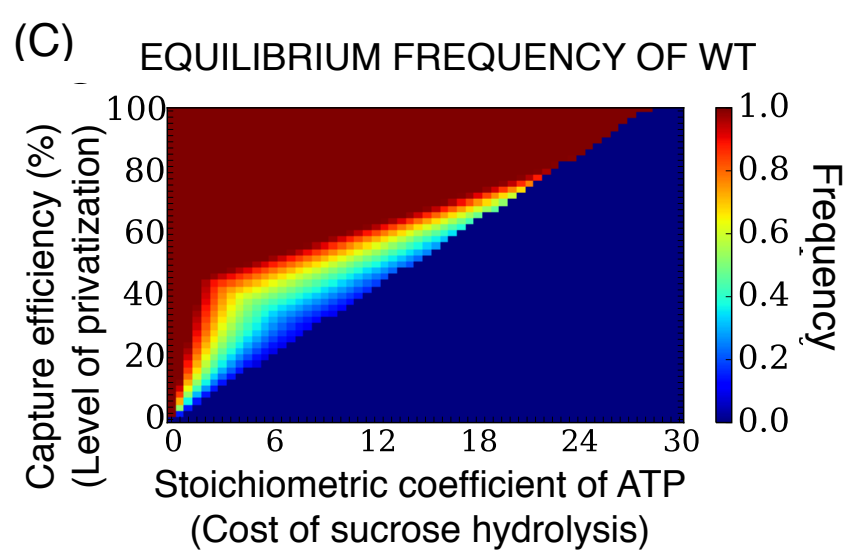
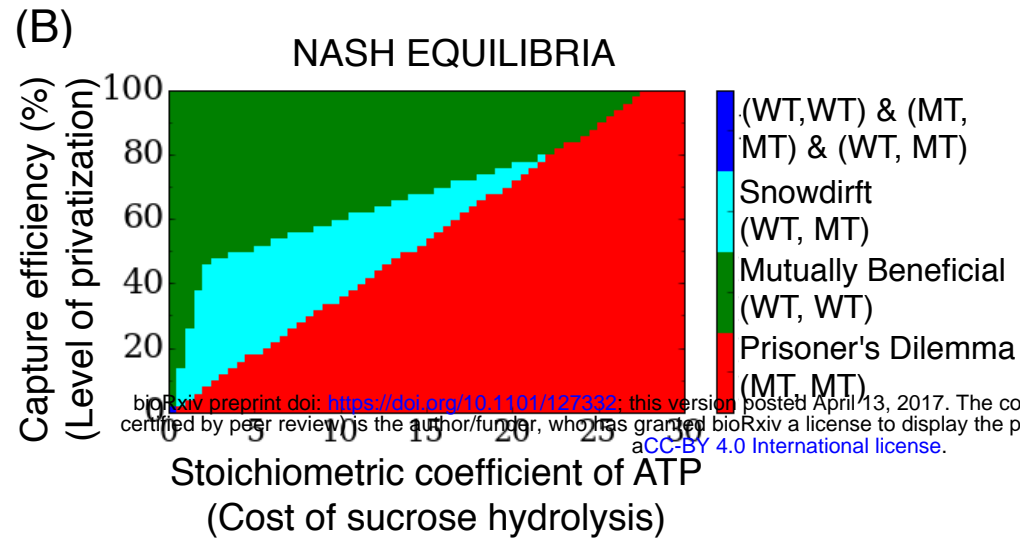
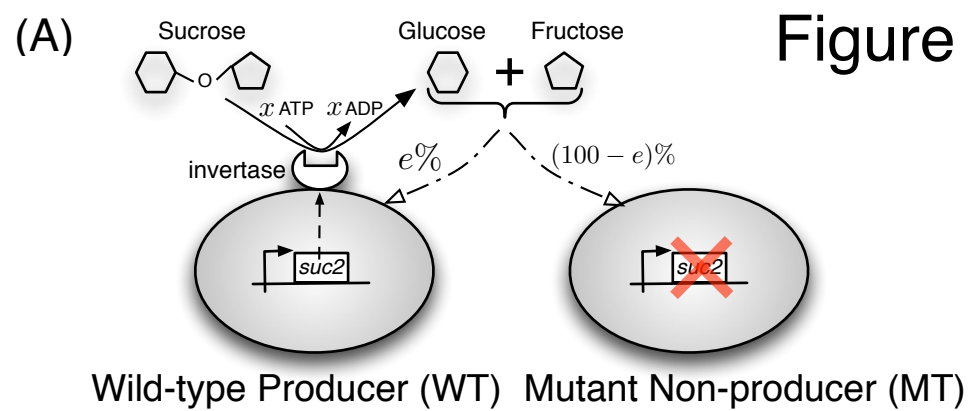


Figure 2



(A)

Figure 3

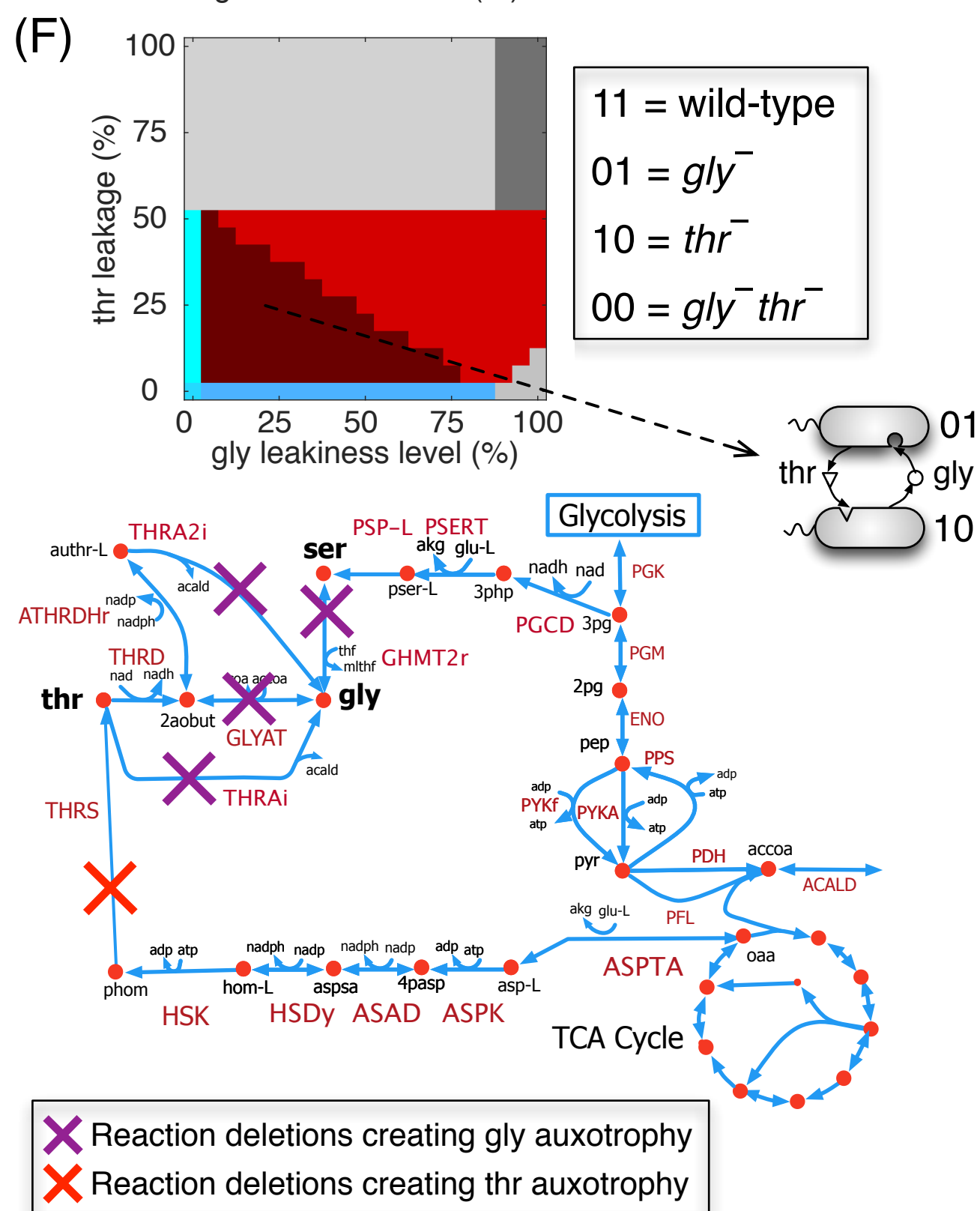
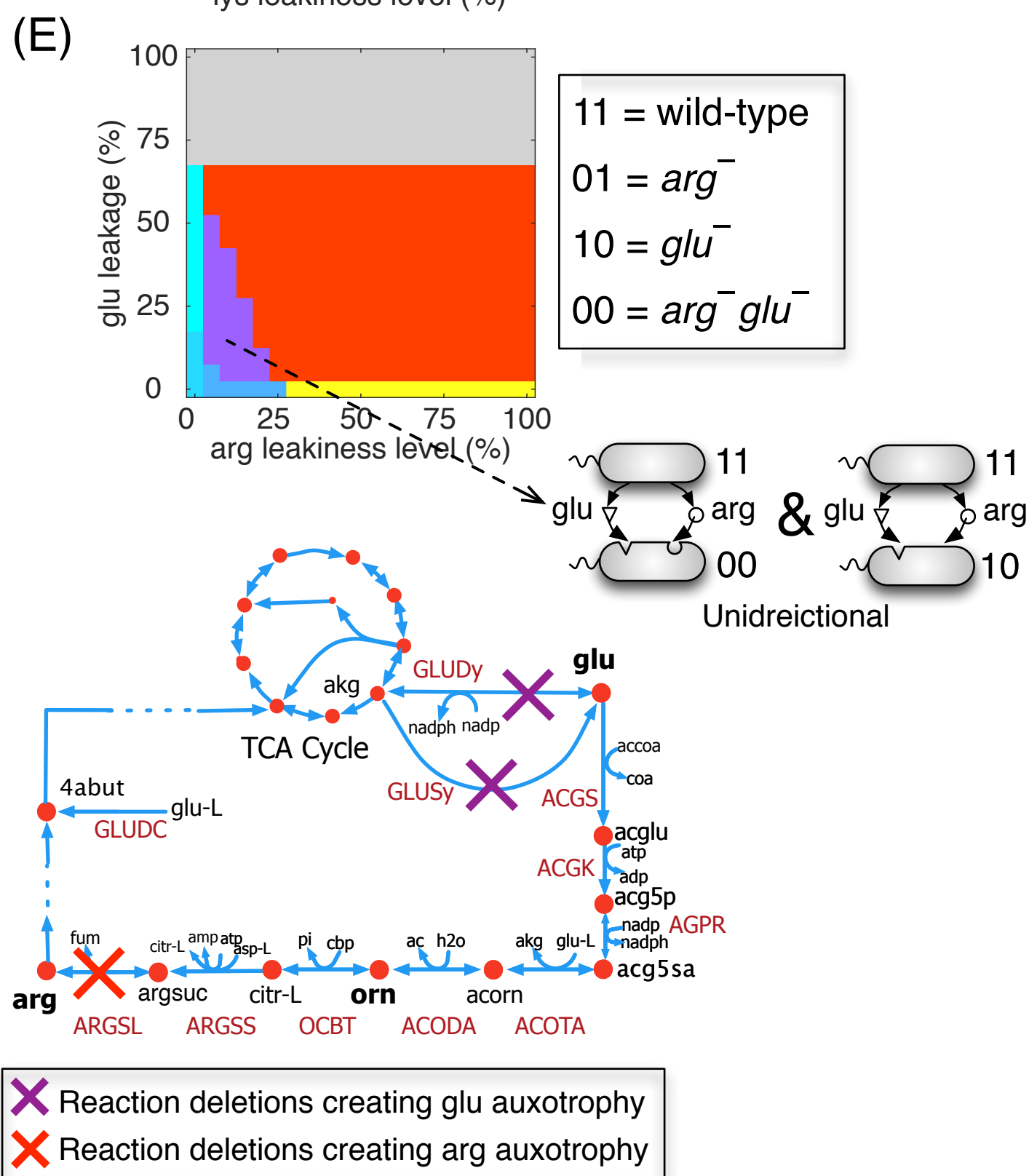
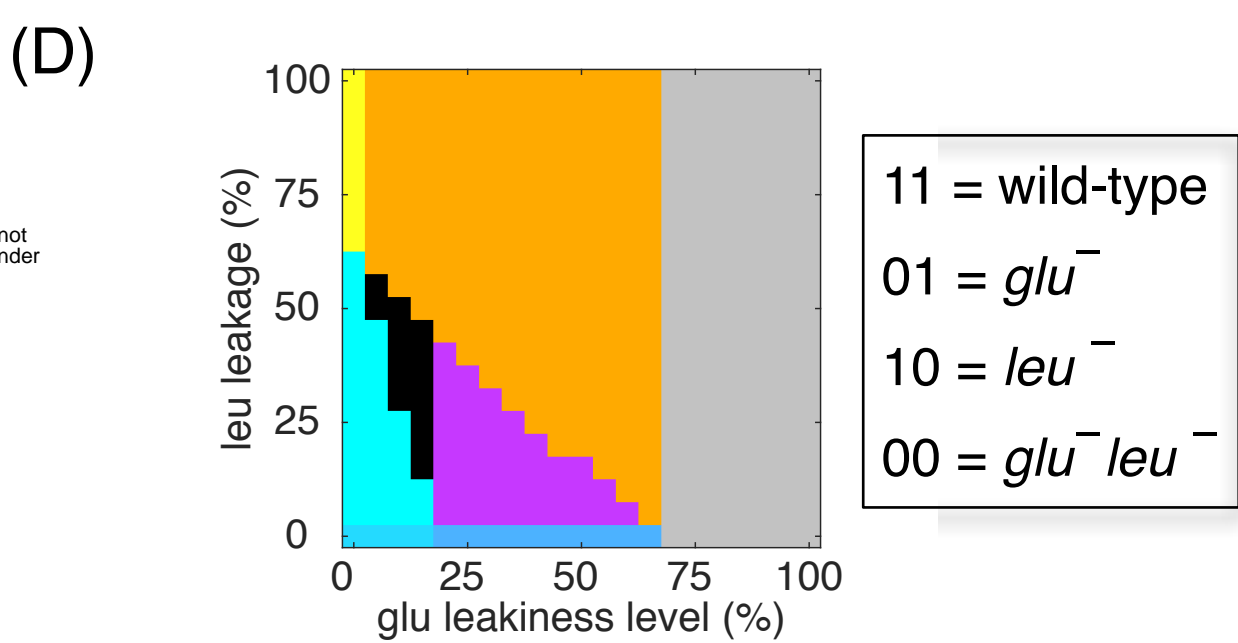
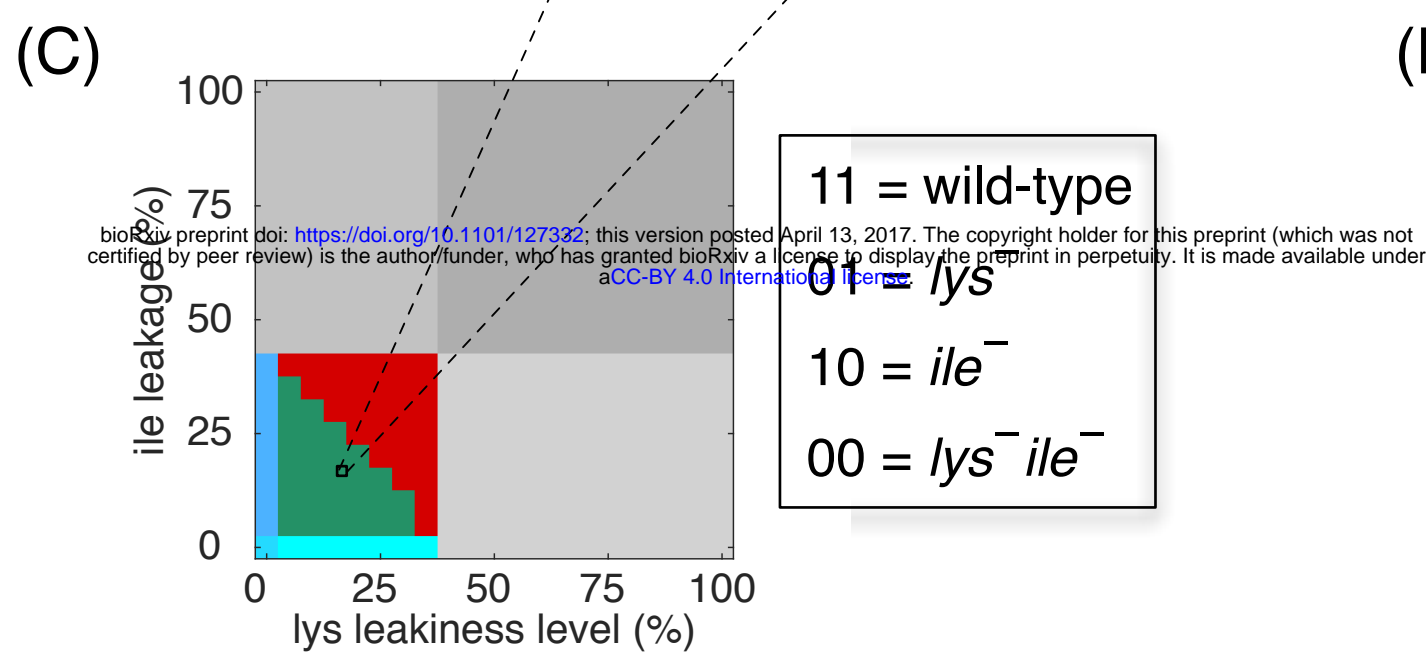
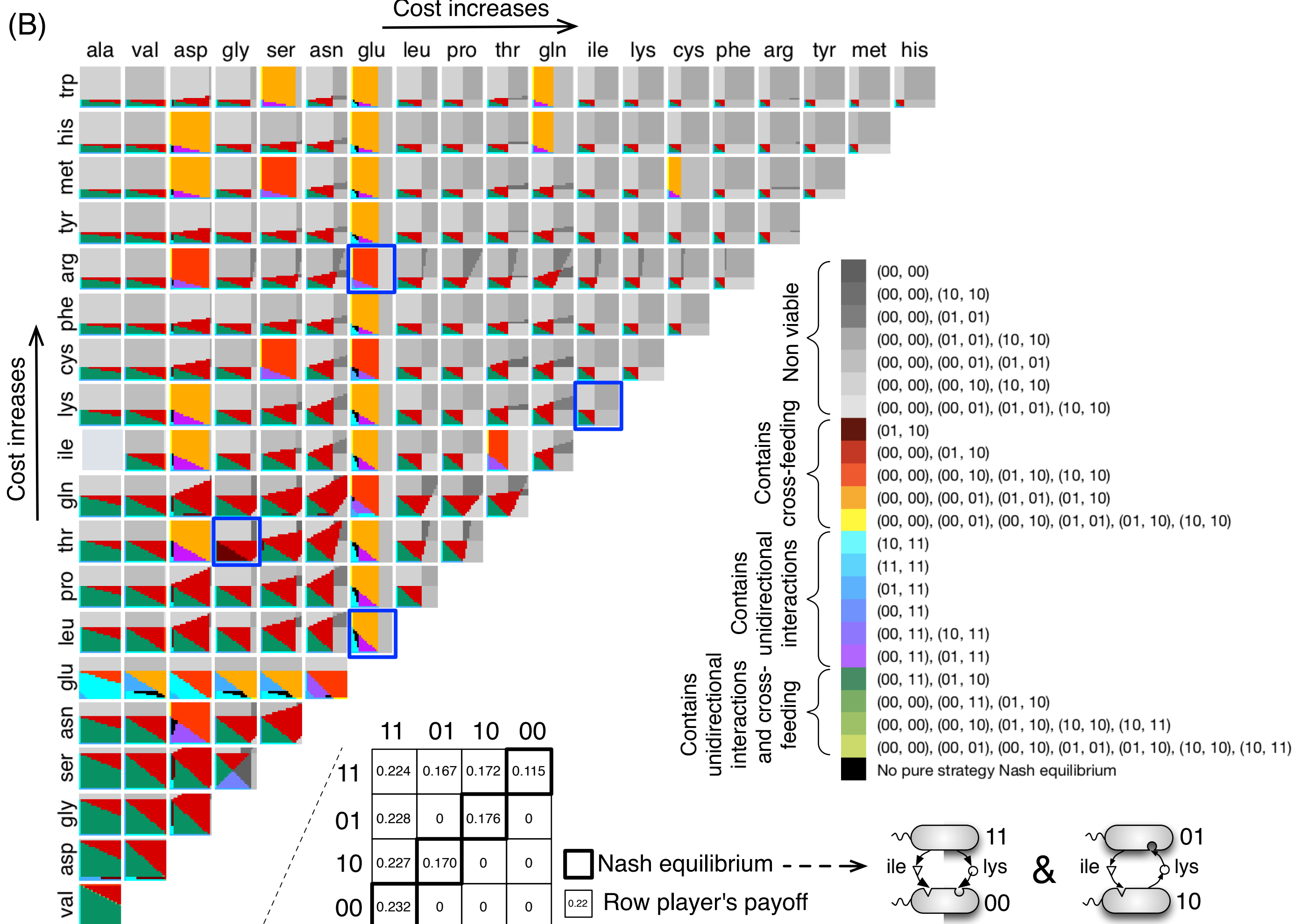
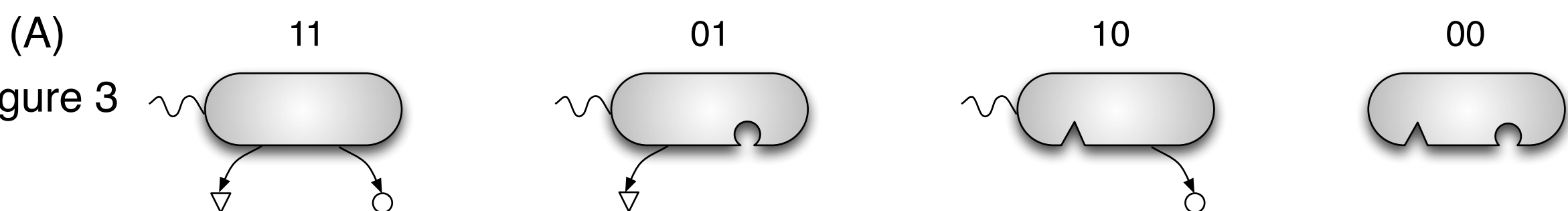
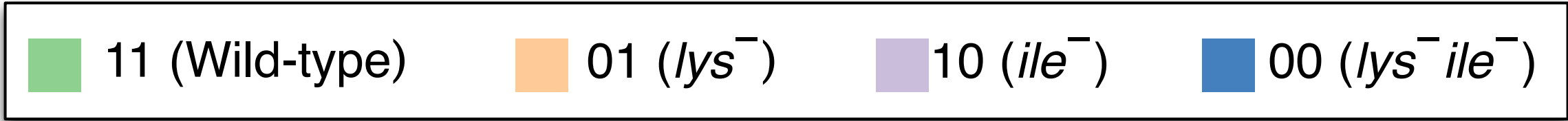
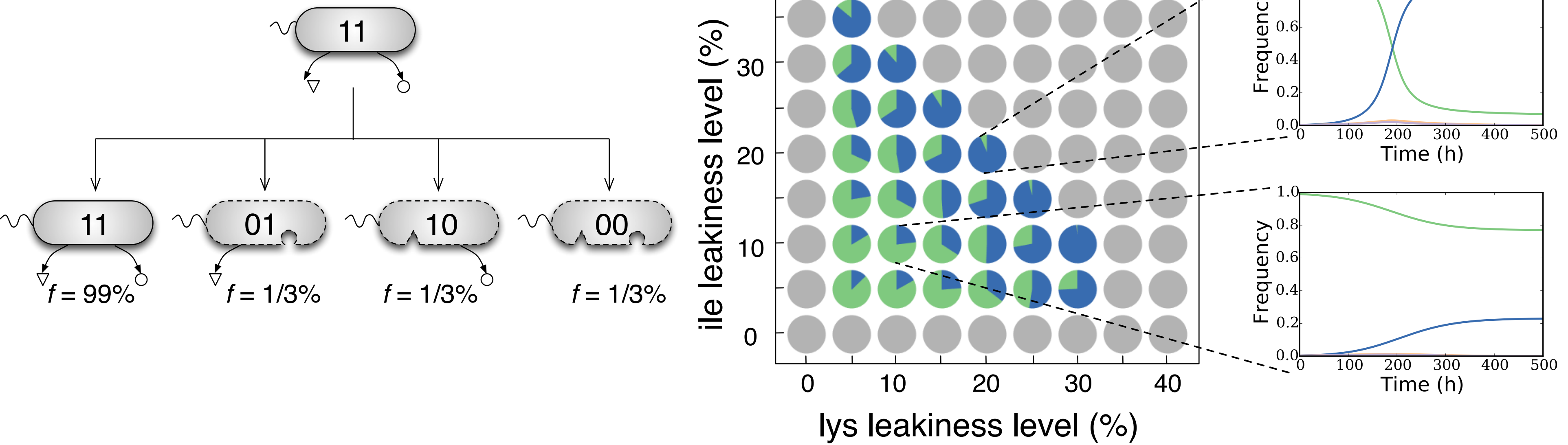


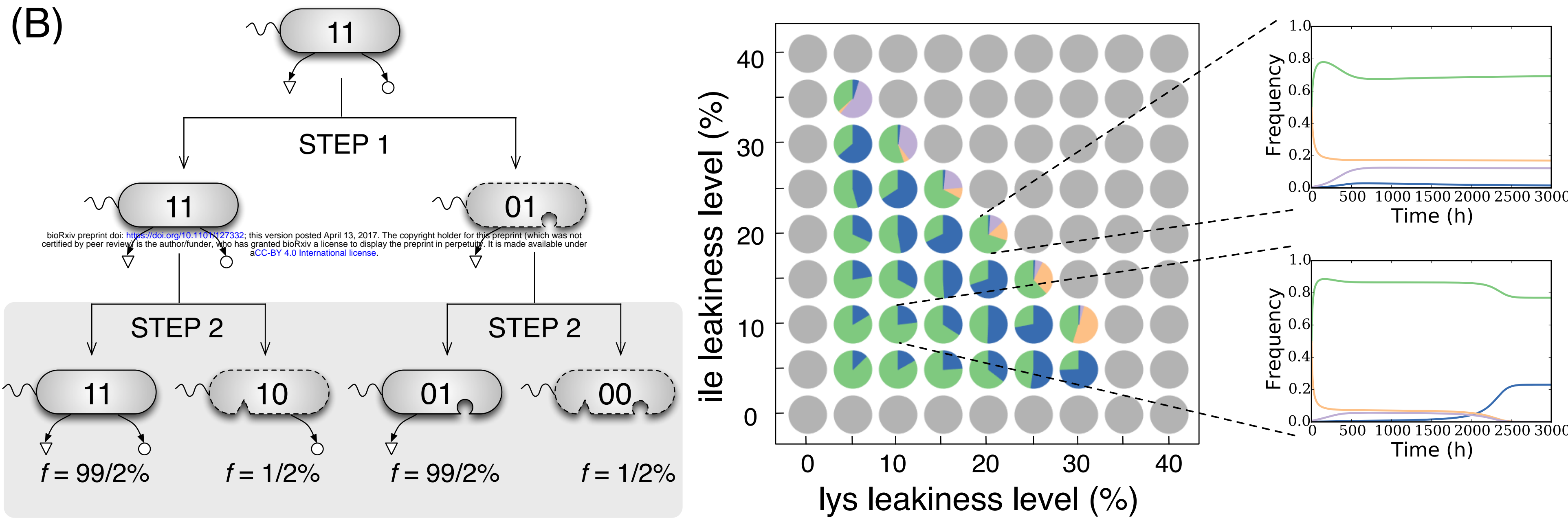
Figure 4



(A)



(B)



(C)

