

# Metabolic complementation in bacterial communities: necessary conditions and optimality

\*Matteo Mori<sup>1,2</sup>, \*Miguel Ponce-de-León<sup>1</sup>, Juli Peretó<sup>3</sup>, Francisco Montero<sup>1</sup>

1 - Departamento de Bioquímica y Biología Molecular I, Facultad de Ciencias Químicas, Universidad Complutense de Madrid, Madrid Spain

2 - Department of Physics, University of California at San Diego, La Jolla, CA, USA

3 - Department of Biochemistry and Molecular Biology and Institute for Integrative Systems Biology (I2SysBio), Universitat de València-CSIC, València, Spain

\* These authors contributed equally to this work

Correspondence: M Mori M Ponce-de-Leon, Departamento de Bioquímica y Biología Molecular I, Facultad de Ciencias Químicas, Universidad Complutense de Madrid, Ciudad Universitaria, Madrid 28045, Spain. E-mail: [mamori@ucsd.edu](mailto:mamori@ucsd.edu) or [migponce@ucm.es](mailto:migponce@ucm.es)

# Abstract

Bacterial communities may display metabolic complementation, in which different members of the association partially contribute to the same biosynthetic pathway. In this way, the end product of the pathway is synthesized by the community as a whole. However, the emergence and the benefits of such complementation are poorly understood. Herein we present a simple model to analyze the metabolic interactions among bacteria, including the host in the case of endosymbiotic bacteria. The model considers two cell populations, with both cell types encoding for the same linear biosynthetic pathway. We have found that, for metabolic complementation to emerge as an optimal strategy, both product inhibition and large permeabilities are needed. In the light of these results, we then consider the patterns found in the case of tryptophan biosynthesis in the endosymbiont consortium hosted by the aphid *Cinara cedri*. Using in-silico computed physicochemical properties of metabolites of this and other biosynthetic pathways, we verified that the splitting point of the pathway corresponds to the most permeable intermediate.

# Introduction

Species that coexists on time and space form complex networks of interactions which are shaped by abiotic and biotic factors (Faust and Raes, 2012; Seth and Taga, 2014). The association between different species can be analyzed from the perspective of the metabolic interactions. In this context, the possible nutritional interactions within different organisms can be grouped as follows: 1) competition for limiting nutrients; 2) syntrophy, *i.e.* the cooperation emerging as each of the partners gain by the metabolic reactions of the other, for example when one organism consumes a product of some other organism, allowing the producer to drive an originally energetically unfavorable metabolic process; 3) commensalism or nutrient cross-feeding, in which the presence of an organism that over-produces an essential nutrient enables auxotroph organisms to survive. While in the case of competition one of the organism will inevitably exclude the other (Hardin, 1960), in the cases of syntrophy and cross-feeding, the interaction will tend to stabilize the coexistence of both species, otherwise competitive. In particular, when the exchange of nutrients or precursors is bidirectional and thus, is beneficial for both partners, the nutritional interdependence will lead, in most of the cases, to a co-evolutionary process, the most striking example being the emergence of symbiotic associations (Douglas, 1998; Germerodt et al., 2016; Hansen et al., 2007).

In the context of the evolution of endosymbiotic bacteria the concept of functional complementation is widely used to describe the complex patterns of nutritional interactions between the members of the association. This concept has been applied at least with two different meanings. On one hand, it can be used in the sense of nutritional cross-feeding between the endosymbiont bacteria and its host, *i.e.* there are exchanges between the symbionts and/or the hosts of some essential compounds (*e.g.* amino acids, vitamins, etc.) (López-Sánchez et al., 2009; Macdonald et al., 2012; Russell et al., 2014; Wu et al., 2006). On the other hand, the concept is also used to refer to a more complex scenario, in which the patterns of gene loss and retention in each symbiont have lead to an inter-pathway genomic complementation (Van Leuven et al., 2014). This means that during the co-evolutionary process, a subset of genes coding a certain metabolic pathway are retained in one symbiont, while the other genes are retained in the other(s) member(s) of the association. In other words, the coding genes of one or more pathways end up distributed among different genomes. In the present paper we will use the second definition.

Events of metabolic complementation, as defined above, are commonly found between endosymbiont bacteria and their hosts, where some metabolic functions are distributed between them (Baumann et al., 2006; Rao et al., 2015; Zientz et al., 2004). For example, in different obligate endosymbiont hosted by sap feeding insects, transaminase activities, a critical step in the biosynthesis of amino acids, have been found to be lost (Baumann et al., 2006; Jiang et al., 2012; McCutcheon and von Dohlen, 2011; Rao et al., 2015; Zientz et al., 2004). In some of these cases, the transaminase activities have been found to be encoded in the host genome, indicating that the synthesis of amino acids requires the combined metabolic capabilities of both members of the association (Hansen and Moran, 2011; Russell et al., 2013). Moreover, it has been suggested that the host can regulate the production of amino acids by controlling the supply of key precursors, such as glutamine (Poliakov et al., 2011; Price et al., 2014; Russell et al., 2014). Other examples of metabolic complementation include the menaquinol biosynthetic pathway in the cockroach endosymbiont *Blattabacterium cuenoti*, where mevalonate, one of the precursors, is provided by the insect host, while the symbiont provides vitamins and other essential compounds (Ponce-de-Leon et al., 2013).

Furthermore, events of metabolic complementation have also been found between the members of endosymbiont consortia in different insects, *i.e.* certain metabolic capabilities are distributed among the different endosymbiotic bacteria living in the same host. A particularly interesting example is the case of the endosymbiont consortium of the cedar aphid *Cinara cedri*. In this system,

two different bacteria, *Buchnera aphidicola* BCc and *Ca. Serratia symbiotica* SCc (hereafter referred to as *S. symbiotica* SCc), coexist (Burke and Moran, 2011; Lamelas et al., 2011; Pérez-Brocal et al., 2006). In this system, most of the biosynthetic processes, *e.g.* biosynthesis of some amino acids, are being performed by only one member of the consortium, and thus each member has to rely on the other for specific nutrients. Other (non-essential) amino acids are, instead, only provided by the host, which also provides the critical transamination steps in the synthesis pathways of the symbionts. Tryptophan biosynthesis is a notable exception. This biosynthetic pathway is split in two parts, each operating in one member of the consortium. Therefore, in this case, the biosynthesis of tryptophan requires the presence of both endosymbionts. Surprisingly, a case of convergent evolution has been also found in the symbiotic system of the psyllid *Heteropsylla cubana* (Martínez-Cano et al., 2015). In this second case, the primary symbiont *Ca. Carsonella ruddii* encodes the first step of the pathway, whereas the secondary symbiont, which has lost almost all the genes for the biosynthesis of essential amino acids, still encodes the remaining genes for complementing the biosynthetic pathway (Sloan and Moran, 2012). More complex scenarios have been described in a cicada species of the genus *Tettigades* harboring endosymbionts. In this system it was found that the endosymbiont *Ca. Hodgkinia cicadicola* has evolved into two cytologically distinct species, partitioned into discrete cells, but which are metabolically interdependent (Van Leuven et al., 2014).

Most of the described complementation events have been identified through genomic analyses. However, these studies do not address the possible advantages or disadvantages of the observed metabolic design. Particularly, metabolic complementation presents certain biophysical problems regarding the splitting a metabolic pathway into different organisms. For example, there is the question of how the intermediate metabolites are exchanged between the endosymbiont and its host, or between the different members of a consortium, (*e.g.* *B. aphidicola* BCc and *S. symbiotica* SCc in the cedar aphid). This question becomes even more puzzling when considering that obligate endosymbionts have a very small repertory of genes coding for transporters (Charles et al., 2011). In addition, intermediate metabolites in biosynthetic pathway do not usually have associated transporters, which suggest diffusion as the most plausible mechanism for the exchanges with the surrounding environment.

Another question is how the endosymbionts adapt their pathways to satisfy the needs of the host and of the other symbiotic endosymbionts. As the symbiotic relationships are established, bacteria overproduce nutrients needed by the host. The flux through the corresponding biosynthetic pathway

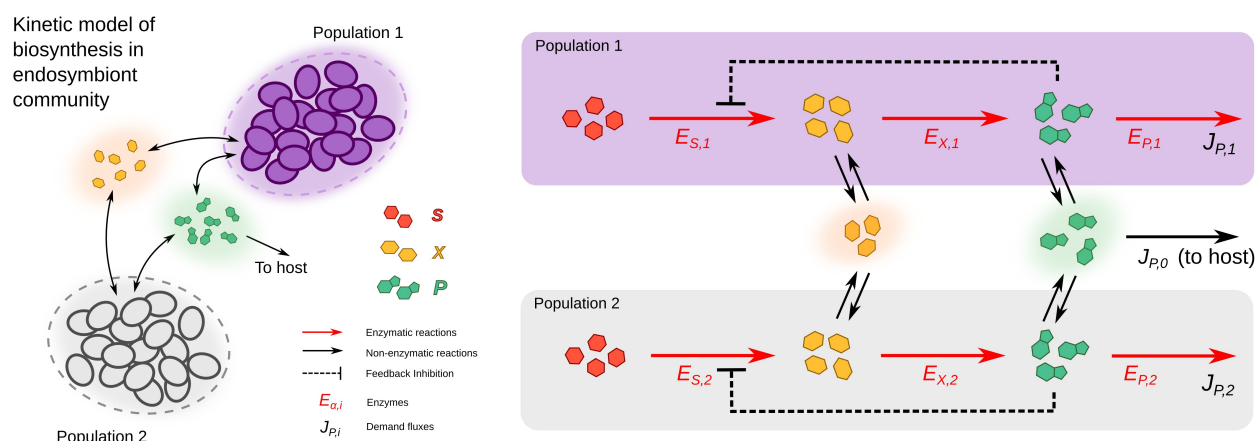
can be increased by acting on several properties of the enzymes (Kacser and Burns, 1973). The catalytic constant of the enzymes, and their affinities for the substrates, can be selected in order to yield larger fluxes. For instance, it has been shown that a selective pressure of improving the enzyme's efficiencies can affect lateral gene transfer in endosymbionts (Ringemann et al., 2006). A more straightforward way to increase reaction fluxes is to increase the enzyme levels. Although, metabolic processes are regulated at different levels (*e.g.* transcription), a common feature of obligate endosymbionts is the apparent absence of transcriptional regulatory mechanisms (Moran and Bennett, 2014; Wilcox et al., 2003). Therefore, enzyme levels can be increased either tuning the translation/transcription efficiency of the gene, or by modifying the gene copy number, inserting additional copies of the gene in the chromosome or using plasmids. For instance, many *Buchnera* strains, producing tryptophan for their hosts, possess multiple copies of the enzyme anthranilate synthase, which is considered to be a limiting step of the tryptophan biosynthetic pathway (Lai et al., 1994). The activity of the enzymes is also regulated by the presence of inhibitors or activators. In most biosynthetic pathways, flux is negatively regulated (through allosteric inhibition) by the final product of the pathway on the first reaction. One might also ask what is the role of feedback inhibition in systems like the *C. cedri* consortium, in which the metabolite inhibiting the pathway is not produced in the same cell. Does the division of labor operated by the members of the consortium effectively reduce the amount of allosteric inhibition and increase the production rate of tryptophan?

The very presence of metabolic complementation between symbionts and their hosts, and especially among the members of an endosymbiont consortium, rises many evolutionary questions: does metabolic complementation provides any selective advantage for the whole system? And, if it does, which is the evolutionary advantage provided by such strategy? Studies have been performed in order to elucidate if organisms exhibiting cross-feeding interactions do possess any selective advantage with respect to free-living organisms (Germerodt et al., 2016; Pfeiffer and Bonhoeffer, 2004). Microbial coexistence has been also explored by using thermodynamic and kinetic models (Großkopf and Soyer, 2016). At any rate, understanding the evolutionary and mechanistic properties of natural microbial communities might allow the design of synthetic ones with potential industrial interest (Großkopf and Soyer, 2014).

# The model

## A kinetic model of two parallel unbranched pathways with negative feedback inhibition

We developed a kinetic model to study metabolic complementation endosymbiont consortium of two different bacteria, when they are allowed to exchange intermediate metabolites of a hypothetical biosynthetic pathway. The model describes a consortium consisting of two different endosymbiotic bacterial species and their host (Fig 1). Each of the two bacterial populations is represented by a large number ( $N_i$  with  $i=1,2$ ) of identical cells, so that in practice it is sufficient to consider two different compartments, modeling a single cell of each bacterial strain, enclosed in a larger compartment, which allows for exchanges of metabolites between cells and the host. Both endosymbionts code for the same hypothetical biosynthetic pathway, which allows the conversion of a precursor  $S$  (produced by the bacteria), to an end product  $P$ , which is essential for any of the three members of the consortium. The biosynthetic pathway is a linear chain composed by two reactions (see Fig. 1B), with the precursor  $S$  being first converted into an intermediate metabolite  $X$ , and then  $X$  being converted to the final product  $P$ .



**Figure 1. Kinetic model of metabolic complementation.** Two cell types (purple and gray) can both potentially produce the product  $P$  starting from a substrate  $S$ , passing through an intermediate metabolite  $X$ . The enzymes can be freely allocated to the pathway's reaction and consumption of  $P$ . The intermediate ( $X$ ) and final ( $P$ ) metabolites can cross the cell membranes by diffusion. Both bacteria and the host need the final product to grow; leading to consumption rates  $J_{P,0}$ ,  $J_{P,1}$  and  $J_{P,2}$ . In the case of the *Buchnera/Serratia/Cinara* consortium, the pathway is completely split across the two populations ( $E_{S,2}$  and  $E_{X,1}$  are absent).

The intermediate metabolite  $X$  and the final product  $P$  can freely diffuse outside the cells (Fig. 1B). Moreover, since the product  $P$  is considered essential for both endosymbionts and its host, three demand fluxes are also introduced to represent the incorporation of  $P$  to the respective biomasses of the three member of the consortium. Finally, product inhibition has been suggest to be

present in some endosymbionts, *e.g.* anthranilate synthase of *B. aphidicola* shows a highly conserved allosteric binding site (Lai et al., 1994), albeit in other enzymes like prephenate dehydratase the inhibition could be mitigated by point mutations in the allosteric site (Jiménez et al., 2000). Thus, the model includes a negative feedback inhibition mechanism that regulates the production of  $P$ .

The two reactions in the biosynthetic pathway,  $S \rightarrow X$  and  $X \rightarrow P$ , are modeled using reversible Michaelis-Menten kinetics (Noor et al., 2013); we name the corresponding fluxes  $V_S$  and  $V_X$ , respectively. Moreover, in order to include the effect of product inhibition on the first step ( $S \rightarrow X$ ), a term representing such effect was introduced in the kinetic equation of the first reaction (Cornish-Bowden, 2013). There are diverse mechanisms of enzyme inhibition, including allosteric interactions, but in this approach we have analysed one of the simplest situations and the kinetics of the inhibition was considered, in a first place, to be uncompetitive, although the case of competitive inhibition is discussed later. Thus, the kinetic law of the first reaction  $S \rightarrow X$ , with flux  $V_S$ , is the following:

$$V_S = \frac{k_{cat}[E_S]}{1+[P]/K_I} \cdot \frac{([S]-[X]/K_{eq})/K_M^+}{1+[S]/K_M^++[X]/K_M^-} \quad (\text{Eq. 1})$$

where  $k_{cat}$  is the turnover number, or catalytic constant, of the reaction,  $K_M^+$  ( $K_M^-$ ) is the forward (backward) Michaelis constant,  $K_{eq}$  is the equilibrium constant, and  $K_I$  is the inhibition constant. It is worth to note that in the limit  $1/K_I \rightarrow 0$  the reversible Michaelis-Menten kinetic law, without inhibition, is recovered. The kinetic for the second reaction ( $X \rightarrow P$ , with flux  $V_X$ ) is analogous to (Eq. 1) but without the inhibition term ( $[P]/K_I$ ).

In order to allow the flux of metabolites between the members of the consortium, the metabolites  $X$  and  $P$  are allowed to be exchanged with the surrounding environment. Endosymbiotic bacteria usually lack transport systems (Charles et al., 2011); thus, the exchanges of  $X$  and  $P$  are modeled using first-order, non-saturating, reversible kinetic laws. The fluxes  $U_\alpha$  take the following form:

$$U_\alpha = D_\alpha([ \alpha ]_{in} - [ \alpha ]_{out}) , \quad \alpha = X, P \quad (\text{Eq. 2})$$

where  $D_\alpha$  are related to membrane permeability and diffusion constants of the metabolites (Laidler and Shuler, 1949). Note that the fluxes are assumed to be positive when the metabolites are excreted by the cells. On the other hand, if there were any kind of protein-mediated transport, the corresponding membrane proteins should also have been taken into account. Finally, the product  $P$



is assumed to be consumed during growth by the three members of the consortium, with consumption fluxes in each compartment. For the bacterial cells, we model the consumption of the metabolite  $P$  using an irreversible Michaelis-Menten reaction:

$$V_P = k_{cat} [E_P] \frac{[P]}{[P] + K_M} \quad (\text{Eq. 3})$$

The fluxes  $V_{P,i}$  (the index  $i=1,2$  corresponding to the cell type) are constrained to be larger than some fixed flux  $J_{P,i}$ . The extracellular product is instead consumed by the host. It is useful to normalize this rate to the size of the bacterial populations  $N = N_1 + N_2$ , so that the consumption rate equals  $d[P]_{out}/dt = -N J_{P,0}$ . The three parameters  $J_{P,0}$  and  $J_{P,i}$  thus set the demand of the metabolite  $P$  from the bacteria or the host.

Finally, the size of the populations of the two bacteria naturally influence both the total protein concentration and the magnitude of the transport fluxes between the cells and the host, and cannot be disregarded. It is convenient to introduce the normalized populations  $n_1 = N_1/N$  and  $n_2 = N_2/N$ , with  $N_1 + N_2 = N$ . We will therefore first show results obtained by fixing the cell relative populations to be  $n_1 = n_2 = 1/2$ , and then exploring the case in which the two population fractions differ. When the whole model is considered, the balance equations for all the metabolites of the system can be written as follows:

$$\begin{aligned} \frac{d[X_i]}{dt} &= V_{S,i} - V_{X,i} - U_{X,i}, \quad i=1,2 \\ \frac{d[P_i]}{dt} &= V_{X,i} - V_{P,i} - U_{P,i}, \quad i=1,2 \\ \frac{d[X_0]}{dt} &= N(n_1 U_{X,1} + n_2 U_{X,2}) \\ \frac{d[P_0]}{dt} &= N(n_1 U_{P,1} + n_2 U_{P,2} - J_{P,0}) \end{aligned} \quad (\text{Eq. 4})$$

where  $N$  corresponds to the total number of cells in both populations. It is worth to note that the equations for intracellular metabolites are symmetric for the two species, while the balance equations of the extracellular metabolites depend on the size of each symbiont population and, in the case of  $P_0$ , on the host demand flux  $J_{P,0}$ . The substrate  $S$  is a precursor produced by each of the two cell types, and its concentration is assumed to be fixed (Heinrich and Schuster, 1996). As a consequence, the balance equation of  $S$  is not included in the model. For the sake of simplicity, in the following we will set all kinetic parameters to one, unless stated otherwise. Also, we will set the



population fractions to  $n_1=n_2=1/2$  and the concentration of the substrate to  $[S]=10$ .

## The optimization criteria

Endosymbiont bacteria involved in nutritional interactions overproduce some essential products, such as essential amino acids or vitamins, and share them with their host (and in some cases with other endosymbionts) (Douglas, 1992). On the other hand, as discussed in the introduction, such over-production can be ensured by tuning the enzyme concentrations along the pathway, even if end-product inhibition may not allow to tune the flux easily (Baumann et al., 1999; Lai et al., 1994; Rouhbakhsh et al., 1997). Accordingly, improvements in the efficiency of over-productions, in terms of the economy on protein synthesis, should provide of some advantage to the whole system (Dekel and Alon, 2005; Flamholz et al., 2013; Goelzer and Fromion, 2011; Kafri et al., 2015; Mori et al., 2016; Ponce de León et al., 2008; Shachrai et al., 2010). Thus, in this work we will focus on the trade-off between the maximization of the flux through a pathway and the efficient allocation of the proteins. Using constraint-based optimization this trade-off can be modelled as two complementary optimization problems: (i) maximization of the flux subject to a constraint on the maximum amount of total enzyme concentration; and (ii) minimization of enzyme concentrations subject to a constraint (lower bound) on the flux of interest, which ensure that a certain demand is satisfied.

Herein, we will adopt the second formulation, *i.e.* enzyme minimization subject to a certain demand flux of the product  $P$ , although one can show that both formulations yield completely analogous results (see Supplementary Text S1). The optimization problem is defined in the following way. Given the demand fluxes  $J_{P,i}$ ,  $i=0, 1, 2$ :

$$V_{P,i} \geq J_{P,i} \quad \forall i=0, 1, 2 \quad (\text{Eq. 5})$$

we compute the metabolite and enzyme concentrations which minimize the total enzyme concentration (normalized by the total bacterial population), defined as follows:

$$C = \sum_{i=1,2} \sum_{\alpha=S,X,P} n_i [E_{\alpha,i}] \quad (\text{Eq. 6})$$

where the cost or objective function  $C$  correspond to the total amount of protein invested by both cell types in order to produce an amount of  $P$  enough to satisfy the demand required by the three members of the consortium. All the intermediate metabolites are subject to the steady-state condition, which is imposed by setting to zero the time derivatives in (Eq. 4). Accordingly, each solution will consist of a set of metabolites and enzymes concentrations, from which one can

compute the optimal fluxes using the kinetic laws. A complementary formulation of the optimization problem is possible, *i.e.* maximization of the flux subject to a constraint on the maximum amount of total enzyme concentration. This other approach can be shown to be completely equivalent to the “enzyme” minimization approach, as shown in Supplementary Note 1.

In order to highlight the patterns of proteome allocation between the cell populations, we introduce the measure of enzyme asymmetries  $A_\alpha$ , defined for each enzyme  $\alpha$  as follows:

$$A_\alpha = \frac{[E_{\alpha,1}] - [E_{\alpha,2}]}{[E_{\alpha,1}] + [E_{\alpha,2}]}, \alpha = S, X, P \quad (\text{Eq. 7})$$

This quantity reflects the relative difference in the concentration of each enzyme between the two different endosymbiont. Thus,  $A_\alpha$  will be zero if the enzyme concentrations in the two cells are the same, while it will equal either 1 or -1 if one cell type completely lacks the corresponding enzyme. In particular, if the two type of cells have the same sets of kinetic parameters, one would expect that the protein asymmetries would be equal to 0, unless some metabolic interaction emerges.

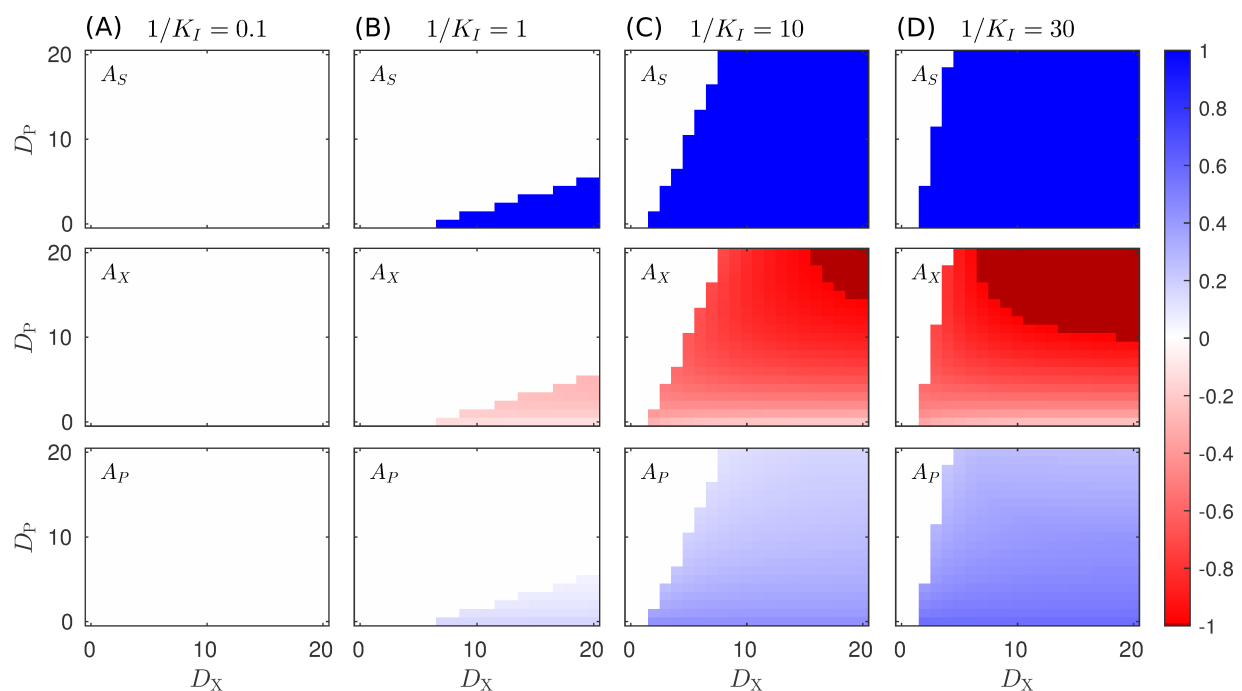
## Results

### Emergence of metabolic complementation

Using the model and the optimization criteria introduced in previous sections, we now turn on the analysis of the optimal proteome allocation obtained, for different scenarios. As a starting point, we will set the host demand flux to zero,  $J_{P,0}=0$ . As explained in Supplementary Note S1, since the total enzyme concentrations is proportional to the demand fluxes, the sum of the demand fluxes  $J_{P,1}+J_{P,2}$  is not relevant for the optimal strategy of proteome allocation, and only their ratio  $J_{P,1}/J_{P,2}$  is (as long as  $J_{P,0}=0$ ). We will therefore choose  $J_{P,i}=1$  for both the two cell types. (The results for different values of  $J_{P,0}=0$  and  $J_{P,1}/J_{P,2}$  are shown below.)

We show in Fig. 2 the optimal protein asymmetries obtained by varying the value of the diffusion constants,  $D_X$  and  $D_P$ , for different values of the inhibition constant  $K_I$ . Given the symmetry of the problem, *i.e.* the reactions in the two cell types have identical kinetic parameters, we will only consider the solutions with  $A_S \geq 0$ . In the extreme case where  $D_X=D_P=0$ , the two cells cannot interact in any way and no complementation can arise. However, when the metabolites are allowed to permeate, the two cell populations are in principle allowed to interact, and an asymmetric solution corresponding to metabolic complementation may emerge. If the effect of the product inhibition is too low (e.g.  $1/K_I \leq 0.1$ ), then the optimal solution is always characterized by

a symmetric enzyme allocation, i.e.  $A_\alpha=0$ , for all enzymes involved (Fig 2A), and for all values of the permeability constants  $D_X$  and  $D_P$ . Conversely, when the inhibition gets larger and larger (e.g.  $1/K_I \geq 1$ ), asymmetric solutions starts to emerge as the optimal strategy, depending on the values of the diffusion constants  $D_X$  and  $D_P$  (Fig 2B-D). For a given value of  $1/K_I$ , the diffusion constants define a transition line, which divides the plane into two zones, corresponding to the range of parameter for which the symmetric or asymmetric solution is optimal.

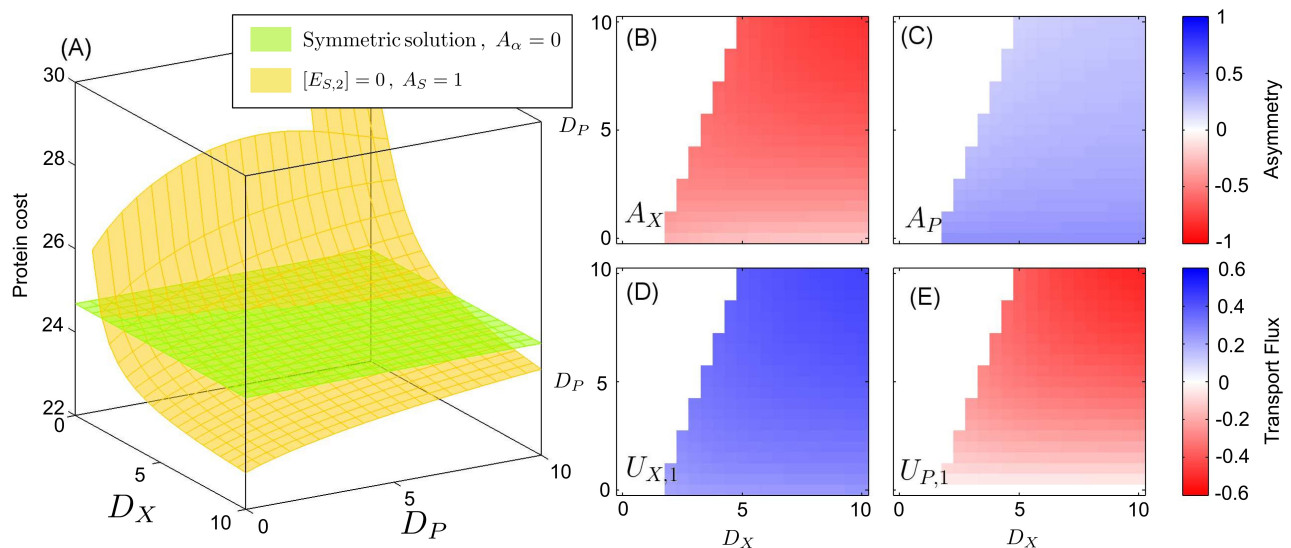


**Figure 2. Enzyme asymmetries  $A_\alpha$  for optimal proteome allocation solutions obtained in different scenarios.** The figure shows the optimal solution in terms of the enzyme asymmetries as a function the diffusion constants  $D_X$  and  $D_P$ , and for different values of the inhibition constant, in the case of uncompetitive inhibition of  $P$  on the enzyme  $E_S$  (see Eq. 1). Subplots on each column, labeled A-D correspond to the solution obtained for a different values of  $1/K_I$ , whereas rows represents the different enzymes of the pathway. The darker color in the  $1/K_I=10$  and  $1/K_I=30$  subplot columns highlight the region in which  $[E_{S,2}]=[E_{X,1}]=0$ , and the pathway is completely split between the two cell types. The color scale corresponds to the asymmetry value of the different enzymes.

When the condition favors the asymmetric strategy, in general all the enzymes exhibit certain degree of asymmetry. In particular in the case of the enzyme  $E_S$ , the asymmetry only takes the extreme values,  $A_S=0$  or  $|A_S|=1$ , whereas for the other two enzyme the degree of asymmetry depends on the diffusion constants. Interestingly, we find that large values of  $D_X$  promote the asymmetry, while for large values of  $D_P$  the symmetry is restored. In other words, asymmetric solutions arise when  $D_X/D_P$  is large enough. Summarizing, asymmetric proteome allocation becomes an optimal strategy in the presence of by product inhibition and when the product diffuses

slower than the intermediate metabolite  $X$ . If  $P$  diffuses too fast, the concentrations within both cell types will be roughly the same, and therefore the inhibition cannot be by-passed. Conversely, if  $P$  diffuses much slower than  $X$ , its concentration on the cell coding for the first enzyme (i.e.  $E_S$ ) will be low and its contribution as an inhibitor will be negligible. As shown in Fig 2, in the regions where asymmetry emerges as the optimal solution, we have the following relations for enzyme asymmetries:  $A_S=1$ ,  $A_X<0$  and  $A_P>0$ . Thus, in order to have a deeper insight on the relation between diffusion and the sudden change from symmetric to asymmetric solution, we focus now on single value of the inhibition constant ( $1/K_I=10$ ). We can individually study the symmetric and asymmetric solutions by imposing that  $A_S=0$  (in the symmetric case) or  $A_S=1$  (in the asymmetric case). Then, for these two cases, the optimal proteome allocation was computed for different values of the diffusion constants,  $D_X$  and  $D_P$ .

Fig 3A shows that the protein cost of the asymmetric strategy grows when the metabolites cannot easily diffuse across the membranes, i.e. when the values of  $D_X$  and  $D_P$  decrease (yellow surface/lines). On the other hand, the asymmetric strategy becomes optimal for large values of the diffusion rates. For large values of the diffusion constants we see that the asymmetric solution is optimal, with the  $A_X$  enzyme asymmetry approaching the value -1 (Fig. 3B). In this conditions, the pathway is completely splitted between the two cells, with one enzyme ( $E_S$ ) only present in the first cell (Fig. 3B), and the second enzyme ( $E_X$ ) only present in the other cell (Fig. 3C). We see in Fig. 3D that the asymmetric condition,  $A_S=1$ , is always accompanied by an excretion of the  $X$  metabolite from the first cell, which is gathered by the second cell  $U_{X,1}=-U_{X,2}>0$ . Interestingly, the exchange of  $X$  (Fig. 3D) still takes place even in the case in which  $P$  cannot cross the cell membrane, i.e.  $D_P=0$  (Fig. 3E). In this condition, the second population becomes a cheater, and takes advantage of the leakage of  $X$  from the first population. Nevertheless, if the diffusion constant  $D_P$  is increased, the second cell starts sharing the product  $P$  with the first cell, thus establishing a metabolic complementation relationship. Finally, if  $D_P$  is too large, the concentrations of the  $P$  metabolites are very similar in both cells, so that there is no way to reduce the effect of inhibition by partitioning the metabolic route. Therefore, the cell switches from metabolic complementation to an “neutralist” strategy, with completely symmetric enzyme allocation and no exchange of metabolites between the two cells types.



**Figure 3. Optimal solutions (minimum proteins) in the case of uncompetitive inhibition.** The figure shows protein costs plotted as a function of  $D_X$  and  $D_P$  for a fixed value of the inhibition  $1/K_I=10$ . (A) The cost function (that is, the total protein cost) is shown in the z-axis; the green surface, corresponding to the symmetric solution, is obtained by imposing that the two cells have identical enzyme allocation, while the yellow one is obtained imposing that one of the two cells lacks the  $E_S$  enzyme. (B) proteome asymmetry of the  $E_X$  enzyme; (C) Proteome asymmetry for the  $E_P$  enzyme; (D) Optimal transport flux  $U_{X,1}$  of the  $X$  metabolite from cell 1 to cell 2; (E) Optimal transport flux  $U_{P,1}$  of the  $P$  metabolite from cell 1 to cell 2 (the negative value of the flux means that  $P$  is actually shuttled from cell 2 to cell 1). The color scale of subplots B-C and D-E, corresponds to the asymmetry value of the different enzymes and the sense of the transportation fluxes, respectively.

## Assessing the impact of alternative inhibition mechanisms and different product demands

In this section we extend our results to more general settings. First, we check that our results do not depend qualitatively on the inhibition kinetics. Instead of a non-competitive mechanism (Eq. 1), we use a competitive mechanism (Cornish-Bowden, 2013) so that the concentration of the product metabolites affects the saturation of the enzyme, and not the turnover rate. Although this mechanism is more complicated than the uncompetitive case, due to the interplay between the different concentrations of the  $S$ ,  $X$  and  $P$  metabolites, the results obtained are analogous to the simpler case of uncompetitive inhibition (see Supplementary Fig. S1). This suggests that while product inhibition is a necessary condition for the emergence of metabolic complementation, it does not depend on specific inhibition mechanisms, and the metabolic complementation strategy is favored by strong feedback inhibition.

Then, we probed the effect of different demand fluxes, both to the host ( $J_{P,0}$ ) and to the bacterial species ( $J_{P,1}$  and  $J_{P,2}$ ). As one increases the host demand flux  $J_{P,0}$ , the qualitative pattern of the different optimal strategies is only slightly affected by the presence of a flux of

products to the host (see Supplementary Fig. S2), even though the total amount of protein to be allocated clearly increases with the demand flux. Similarly, we consider the case where the two bacterial populations have different demands for the product  $P$ , for instance because of a different biomass composition. We see that an asymmetry in the demands for  $P$  can promote, even more, the asymmetry in the proteome allocation (see Supplementary Fig. S3); for large enough  $1/K_I$  and demand flux asymmetry  $\rho_J = (J_{P,1} - J_{P,2}) / (J_{P,1} + J_{P,2})$ , the optimal solution is characterized by a complete splitting of the pathway within the two cells,  $[E_{S,1}] = [E_{X,2}] = 0$ .

A similar effect is found when the two bacterial species have different populations, that is,  $n_1 \neq n_2$ . As  $n_1$  is moved away from  $1/2$  (with  $n_2 = 1 - n_1$ ), the protein asymmetry is generally enhanced, with larger asymmetries  $A_X$  (see Supplementary Fig. S4). All these results show that the complementation patterns described by the model are robust against variation in the demand fluxes and the relative population sizes.

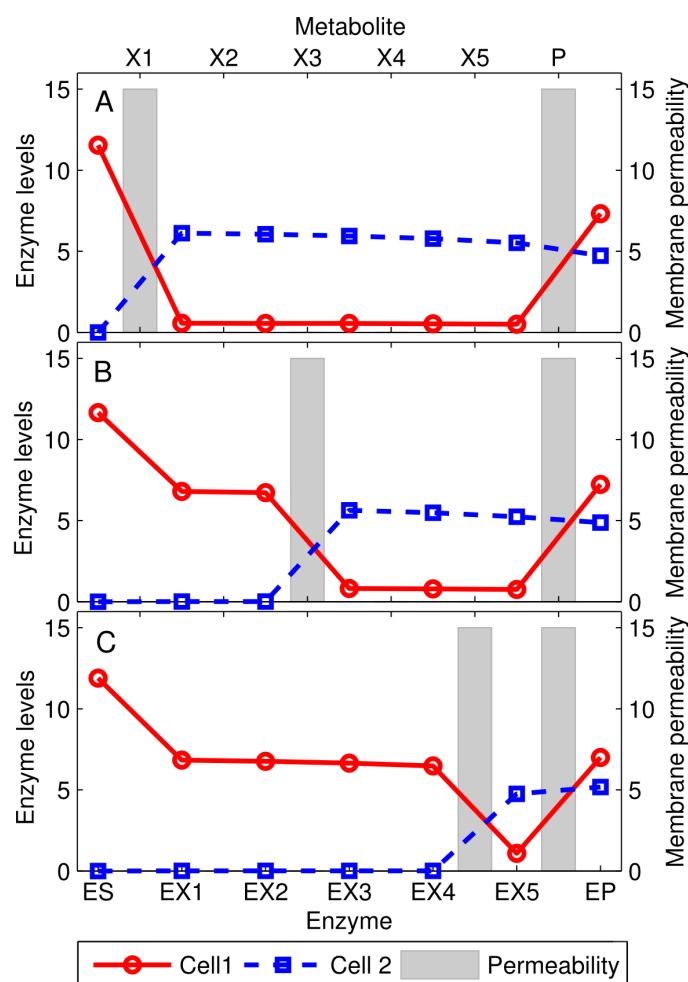
## Sensitivity analysis

The number of kinetic parameters in the model is quite large, and it is very difficult to probe the effect of perturbations for each different combination of parameters. Therefore, we performed a sensitivity analysis, on the model with uncompetitive inhibition, by randomly perturbing the parameters, and studying how the perturbation of each parameter correlates with the insurgence of the asymmetric optimal solution (see Materials and Methods for details). We see that the parameters which promote the enzyme asymmetry most are, as expected, the inverse inhibition constant  $1/K_I$  (reflecting the amount of inhibition), the concentration of the  $S$  metabolite, and the diffusion constant  $D_X$  of the  $X$  metabolite (Supplementary Table S1). We also find that the equilibrium constant of the  $S \rightarrow X$  reaction also positively affects the asymmetry, albeit slightly; this is because increased concentrations of the  $X$  metabolite allows for larger shuttling fluxes across the two cells. On the other hand, the diffusion constant  $D_P$  of the  $P$  metabolite is found to promote the symmetry between the two cells, decreasing the average value of  $A_S$ .

## Extended model with multiple intermediates

In the previous section we gave an in depth characterization of a model in which the biosynthetic pathway is composed by two reactions. However, real biosynthetic pathways are usually composed by a larger number of reactions. In order to study how the permeability affects the emergence of a complementation, an extended version of the model was developed, which keeps the fundamental features of its simplest version, while adding more intermediates in the pathway. The extended

version of the model was obtained by introducing five intermediate metabolite,  $X_1, \dots, X_5$ , along with five enzymes  $EX_1, \dots, EX_5$ . Each enzyme  $EX_i$ , with the index  $i$  ranging between 1 and 5, catalyzes the reversible conversion of  $X_i$  into  $X_{i+1}$  or, in the case of  $EX_5$ , the product metabolite  $P$ . This extended model allows us to study how a long biosynthetic pathway is differently partitioned within the two cells for different permeabilities of the intermediate metabolites.



**Figure 4. Solutions of the extended model with 5 intermediate metabolites  $X_1, \dots, X_5$ .** The points represent the optimal enzyme levels of the two cells, while the gray bars are the membrane permeabilities ( $D_i$ ) for the intermediate metabolites. In each panel, the  $P$  metabolite and one of the intermediate metabolites  $X_i$  ( $i = 1, 3$  and  $5$  in panels A, B and C, respectively) is allowed to permeate through the membrane ( $D=15$ ), while all others were not ( $D=10^{-3}$ ). In these conditions, the pathway is split in correspondence of the permeable intermediate metabolite. Other settings:  $K_{eq}=4/3$  for all reversible reactions (all but the one catalyzed by the  $E_P$  enzyme). Inhibition constant:  $K_I=1/10$ ; the concentration of  $[S]=20$ .

In the extreme case in which only one intermediate metabolite can permeate through the cell membranes, we find that the general behavior of the extended model is totally consistent with the results obtained using the simplified version (see “Emergence of metabolic complementation”



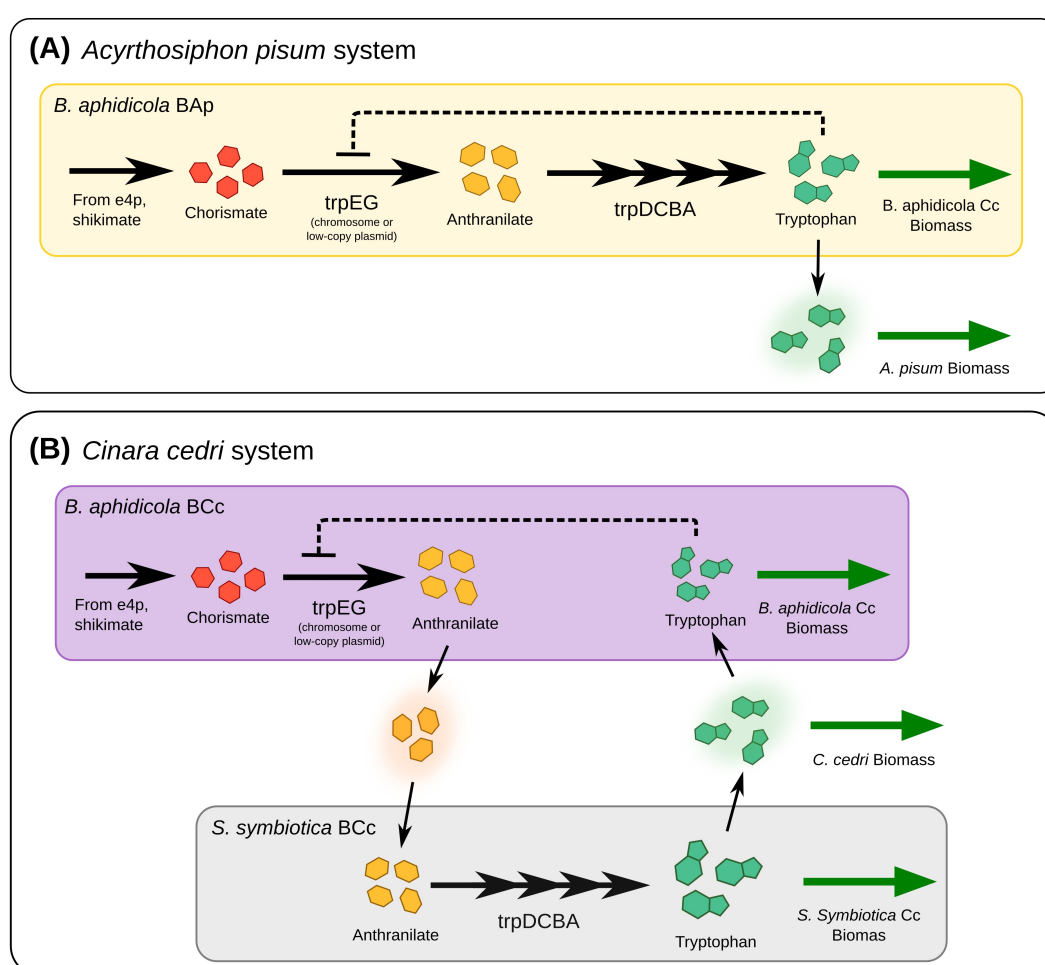
section). As shown in Fig. 4, the pathway is split in correspondence of the most permeable intermediate metabolite. The first cell converts the substrate  $S$  into the intermediate metabolite with large permeability, which in turn is (mostly) shuttled to the other cell and converted into the product  $P$ , which is then shared between the two bacteria. A wide range of behaviors are obtained when different, more complex, scenarios are considered (see Supplementary Fig. S5). For example, when all the intermediate metabolites can permeate across the membranes, the optimal solution still exhibits metabolite complementation; However, enzymes levels change smoothly along the pathway with many metabolite being exchanged, and therefore the qualitative pattern lacks a unique breaking point.

### **Predicting Biological cases: The case of the tryptophan biosynthesis**

In this section we discuss the well documented case of the complementation of the tryptophan biosynthetic pathway between the members of an endosymbiotic consortium in relation with the results here obtained (Gosalbes et al., 2008; Sloan and Moran, 2012). This event of metabolic complementation has been reported to be convergent in two different insects systems and evolutionary unrelated. On one hand, the complementation of the tryptophan biosynthesis has been described in the endosymbiotic consortium composed by the P-endosymbiont *B. aphidicola* BCc and the co-primary endosymbiont, *S. symbiotica* SCc, hosted by the cedar aphid (Gosalbes et al., 2008). On the other hand, the complementation has also been described in the system of the psyllid *Heteropsylla cubana*, where the coding genes for the tryptophan pathway are distributed among the genomes of two endosymbionts, *Ca. Carsonella ruddii* and a tentatively identified secondary symbiont (Sloan and Moran, 2012), thus, exhibiting the same pattern than the one described for the case of the aphid *C. cedri* (Martínez-Cano et al., 2015). It is worth to note that although aphids and psyllids belongs to the same suborder Sternorrhyncha, their corresponding endosymbionts belong to different bacteria lineages, indicating the independence of the symbiotic events (Baumann et al., 2006). This convergence suggests the existence of some evolutionary advantage of the complementation strategy. In the following we will focus on the case of *C. cedri*, because this system has been extensively characterized (Burke and Moran, 2011; Gosalbes et al., 2008; Lamelas et al., 2011; Pérez-Brocal et al., 2006).

The tryptophan biosynthetic pathway involves five enzymes, encoded by seven genes (*trpA-F*), which allows the conversion of chorismate to tryptophan (Crawford, 1989) (see Fig. 5). Moreover, as in the case of most amino acid biosynthesis, the first step (catalyzed by anthranilate synthase) is subject to allosteric feedback inhibition by tryptophan the end product of the pathway (Kwak et al.,

1999; Rouhakhsh et al., 1996, 1997; Spraggon et al., 2001). Previous studies on different *B. aphidicola* strains have showed that all the key residues in the allosteric site are conserved in the *trpE* sequences (Lai et al., 1994; Rouhakhsh et al., 1996). However, since the *trpE* gene of *B. aphidicola* BCc was not considered in the mentioned study, we have performed a multi-sequence analysis to compare this sequence with its homologs in other eleven *B. aphidicola* strains, and in *E. coli* and *S. marcescens*. Accordingly, we have found that the key residues P21 and S40, as well the motif LLES, purportedly involved in allosteric feedback inhibition (Caligiuri and Bauerle, 1991; Kwak et al., 1999; Tang et al., 2001), are even better conserved in *B. aphidicola* BCc than in the other *B. aphidicola* strains (see Supplementary Figure S6, and Materials and Methods sections for more details), indicating that a product inhibition mechanism could be still active.



**Figure 5. Tryptophan biosynthetic pathway in two different endosymbiont-host systems.** Most Buchnera strains can synthesize tryptophan from chorismate derived from central metabolites. The first step is catalyzed by anthranilate synthase, encoded in the *trpEG* genes, then, anthranilate is converted into tryptophan by four additional enzymatic steps, encoded in the *trpDCBA* operon. In A) the structure of the pathway in the system *B. aphidicola* and the aphid *A. pisum* is presented. Feedback inhibition of tryptophan on anthranilate synthase is present. (B) In the aphid *C. cedri*, the primary (*B. aphidicola* BCc) and the co-primary (*S. symbiotica* SCc) endosymbionts share the production of tryptophan: *trpEG* genes are only present in Buchnera, while the *trpDCBA* genes only in Serratia.

If there is an evolutionary pressure to increase the production of tryptophan, then the inhibiting effect of tryptophan accumulation can be overcome by producing large amounts of the anthranilate synthase enzyme (Lai et al., 1994). In fact, in several *B. aphidicola* strains, *trpEG* genes are amplified, either with multiple copy of the genes in the chromosome, or with multiple-copy plasmids, suggesting an adaptation to overproduce tryptophan for hosts (Baumann et al., 2006; Rouhbakhsh et al., 1997). However, increasing protein levels to overcome the inhibition effect, may represent an large cost in terms of protein synthesis. According to our model, the metabolic complementation strategy can be used overcome feedback inhibition, while reducing the total amount of protein required to supply enough product. Indeed, in the *C. cedri* consortium the genes coding for the tryptophan biosynthetic pathway are distributed among the two different endosymbionts: the first two genes, *trpEG* encoding the anthranilate synthase activity, are hosted in a plasmid in the P-symbiont *B. aphidicola* BCc, whereas the gene products necessary to catalyze the synthesis of tryptophan from anthranilate (*trpDCBA*) are encoded by the genome of the co-primary *S. symbiotica* SCc, which in turn, lacks the *trpEG* genes (Gosalbes et al., 2008; Lamelas et al., 2011). Thus, we argue that the presence of metabolic complementation in these bacterial endosymbionts could be interpreted as an evolutionary innovation to reduce the protein cost required to supply tryptophan to all the members of the consortium.

Furthermore, the model also predicts that when a metabolic complementation is an optimal strategy, the point where the pathway should be split will correspond to the most permeable metabolite. To contrast this prediction to the present case of study, we have collected structural data of all the intermediate metabolite of the tryptophan biosynthesis pathway, and we used different physicochemical properties and rule-based estimators to evaluate membrane permeability (see the Material and Methods section), including the molecular weights and the solubility of the compounds. Using these parameters we identify plausible candidate metabolites for becoming breaking points in the tryptophan biosynthetic pathway. As seen in Table 1, among all the intermediate metabolites in this pathway, both anthranilate and indole appear as most plausible “breaking points”, while the other intermediates are less likely to cross membranes at an appreciable rate. Indole is an important signaling molecule in organisms such as *E. coli* (Martino et al., 2003), which is known to diffuse easily through cell membranes (Piñero-Fernandez et al., 2011). This raises the question: why the pathway breaking point is anthranilate instead of indole? An answer may lie in the structure of the tryptophan synthase complex (Dunn, 2012; Miles, 2013). The complex is a tetramer  $\alpha_2\beta_2$ . The  $\alpha$  subunit, encoded by *trpA* gene, catalyzes the synthesis of

indole from indole-3-glycerol-phosphate. Then, indole is directly channeled to the  $\beta$  subunit, encoded by *trpB*, where it is used to synthesize tryptophan. Interestingly, it has been shown experimentally that if the  $\alpha$  subunit is absent, the catalytic activity of tryptophan synthase is reduced by 1-2 orders of magnitude (Kirschner et al., 1991). Therefore, keeping both *trpA* and *trpB* in the same compartment appears to be a constraint and thus, indole is not a feasible candidate for the breaking point, since it is not allowed to diffuse freely in the cell. Anthranilate is thus the intermediate most likely to diffuse through cell membranes.

**Table 1. Physicochemical properties and rule-based estimators of the permeability for the intermediate metabolites of the tryptophan biosynthetic pathway.** For each metabolite the molecular weight, an octanol/water log ratio estimator (AlogP), the topological polar surface area (TPSA), and two rule-based classifiers (1PR and 3PR) are shown (see Estimation of metabolites permeability in Material and Methods section for details).

| Step | Metabolite                                   | MW      | AlogP | TPSA (Å <sup>2</sup> ) | 1PR | 3PR |
|------|--|---------|-------|------------------------|-----|-----|
| 1    | Chorismate                                   | 224,17  | 1,5   | 109,72                 | MH  | H   |
| 2    | Anthranilate                                 | 136,13  | 1,9   | 66,15                  | MH  | H   |
| 3    | N-(5-phosphoribosyl)-anthranilate            | 346,21  | -0,7  | 174,27                 | ML  | ML  |
| 4    | 1-(o-carboxyphenylamino)-1-deoxyribose-5-P   | 349,05  | -0,8  | 182,11                 | ML  | ML  |
| 5    | (1S,2R)-1-C-(indol-3-yl)glycerol 3-phosphate | 285,193 | -1,4  | 128,67                 | ML  | MH  |
| 6    | Indole                                       | 117,15  | 2,1   | 15,79                  | H   | H   |
| 7    | L-tryptophan                                 | 204,228 | -0,4  | 78,98                  | MH  | ML  |

\* Abbreviations: high (H); medium high (MH); medium low (ML)

Finally, we will briefly discuss the case of the transaminase activities losses in different lineages of *B. aphidicola* (Hansen and Moran, 2011; Macdonald et al., 2012; Poliakov et al., 2011), as well other endosymbionts, such as *Tremblaya* and *Portiera* (Baumann et al., 2006; McCutcheon and von Dohlen, 2011; Zelezniak et al., 2015; Zientz et al., 2004). Transaminase activities are a critical step in the biosynthesis of amino acids; transamination corresponds to the last step in the production of branched-chain amino acids (BCAs) and aromatic amino acids. Moreover, all BCAs share the same transaminase enzyme (EC 2.6.1.42), and the same happens for the aromatic amino acids (*i.e.* phenylalanine and tyrosine) (with enzyme EC 2.6.1.57). In several lineages of endosymbionts, such as most strains of *Buchnera*, these transamination steps are lost, while the rest of the pathway is retained. Transaminases are expressed in the host cells (Hansen and Moran, 2011; Russell et al., 2013) originally with a catabolic function but now recruited to perform the last step in a biosynthetic pathway. As the in the case of tryptophan, the synthesis of these amino amino acids is commonly regulated at different levels, which usually include product feedback inhibition (White et al., 2012).

The above mentioned complementation does not involve two different endosymbionts, but is instead a metabolic relationship between the endosymbiont and its host. Still, this case includes all the main features of the model discussed in this work, *i.e.* a metabolite demand, feedback inhibition, and the absence of transporters. Thus, in order to test if the breaking point of these pathway corresponds to the most permeable metabolite, we have computed the permeability estimators for the intermediates in the biosynthesis of the branched amino acid as well the aromatic amino acid. The results show that in all the cases, the breaking point correspond to the most permeable metabolite (see Supplementary Table S2), thus suggesting that the complementation may have emerged due to an evolutionary pressure on the consortium.

## Discussion

The emergence of metabolic complementation in endosymbiotic consortia is a complex phenomena poorly understood from a theoretical point of view. For this reason we have developed a simple model to analyzed the possible advantages (or disadvantages) of metabolic complementation, as well as the necessary conditions that may drive its emergence. Under the model assumptions, we have found that for metabolic complementation to emerge, two necessary conditions should be fulfilled: first, the pathway must exhibit product inhibition, and the inhibition constant should be strong enough so that the over-production requires a high increase in the enzyme concentration. Second, some of the intermediate metabolites of the pathway should be permeable enough to diffuse across cell membrane so that it can reach the other cell type. Thus, an “indifferent, or “neutralist” strategy, with both cell populations allocating the same enzymes, is favored if the metabolites cannot easily permeate through the cell membranes, and/or if there is no substantial product inhibition on the first step of the pathway. On the other hand, if the intermediate metabolites can permeate easily through the membrane, and if product inhibition is present, metabolic complementation strategies allow to reduce the global enzyme burden. Moreover, when analyzing the case of tryptophan biosynthesis in the endosymbiotic consortium of *C. cedri*, we have found that the split point corresponds to the most permeable (and not channeled) metabolite. Thus, metabolite permeability is a plausible predictor of the splitting point of complementation events.

### Evolution of metabolic complementation

The evolution of metabolic complementation in bacterial communities is most likely shaped by different forces, many of them may acting simultaneously (Tan et al., 2015). On one side, endosymbionts are subject to a process of irreversible gene loss and genome shrinkage, known as

the Muller's ratchet, due to its confinement to intracellular life and the bottleneck caused by its small population size (Moran, 1996). Thus, under this scenario, the gene losses with no detrimental effect on the system could be fixed at the population level by the effect of genetic drift. In this case, metabolic complementation would not necessarily confer a selective advantage, and would be a consequence of random gene losses.

Instead, a selective explanation focuses on the cost of protein synthesis, which affects both the microbial cells and the host (which also has to provide nutrients to the endosymbionts). Therefore, hosts whose endosymbionts are able to sustain a given demand of certain product while reducing the protein cost will exhibit an improvement in their fitness (Shachrai et al., 2010). The same can apply to the fitness of microbial communities as a whole; interestingly, experimental results on synthetic cross-feeding bacteria has showed that the interactions that emerge in this scenarios confer a significant fitness advantage to the genotypes involved in the interaction, and thus, stabilize the coexistence of cross-feeding organisms (Germerodt et al., 2016; Pande et al., 2014).

Another evolutionary mechanism for the emergence of metabolic complementation is given by the Black Queen Hypothesis (BQH). According to the BQH, whenever a metabolic function is redundant among the member of a microbial community, and the final product of the process leaks or escape from the producer cells, becoming a public good, the rest of the members can take advantage of this “common good” (Morris et al., 2012). Thus the BQH suggest that the loss of a leaky function will have some selective advantage at the level of each microbial species within the community; eventually, most of the members of the community will loss the genes coding for the function, until the production of public goods is just sufficient to support the community. This is not necessarily in contrast with the previous hypothesized fitness advantage for cross-feeders communities, since the loss of redundant functions may also increase the fitness of the system as a whole; the minimization of the total protein amounts might be a good approximation of the final state of a BQH-like dynamics of gene losses, where “intermediate goods” are available to all community members.

## **Application to the design of synthetic communities**

In perspective, understanding the evolutionary and mechanistic properties of natural microbial communities might allow the design of synthetic ones of potential industrial interest (Großkopf and Soyer, 2014; Hosoda and Yomo, 2011). For instance, the common laboratory bacteria *E. coli* have been used to study artificial communities (Hosoda et al., 2011; Pande et al., 2014). Interestingly,

synthetic communities of cross-feeders based on *E. coli* mutant strains (with both over-producers and auxotrophs) were found to grow efficiently, even faster than the growth of the wild-type strain (Germerodt et al., 2016; Pande et al., 2014). Our model can be extended to these microbial communities by including the cost of protein synthesis in a genome-scale metabolic network, which has been recently been done in the context of constrained-based modeling for *E. coli* (Mori et al., 2016). However, the inclusion of allosteric regulation and diffusion still present some challenges (Machado et al., 2015; Tepper et al., 2013). This approach may eventually lead to automated design of microbial communities, which would be of extreme interest for industrial applications.

## Material & Methods

### Optimization procedure

Optimizations were solved using the *fmincon* function in Matlab 2015a for Linux. Due to the non-linear formulation of the optimization problem, the cost function exhibits multiple local minimums. Therefore, local optimization strategies such that gradient descent are not always able to find the correct minimum. However, if the parameters of the model are varied smoothly, such as in the case of the permeabilities  $D_X$  and  $D_P$  in Fig. 2, one may rely on the fact that the optimal solution should also vary smoothly (excluding the frontiers between the symmetric and the non-symmetric solutions). Therefore, for each combination of parameters, we optimized the cost function starting from a random seed until we got the correct minimum in all cases.

### Sensitivity analysis procedure

In the sensitivity analysis, we chose a point in the parameter space, roughly corresponding to the border between the symmetric and asymmetric region. Then, we randomized by various amounts the kinetic parameters, and for each set of parameters we computed the optimal concentrations,  $\{x^*\}$ , and the corresponding optimal  $E_S$  protein asymmetry,  $A_S = ([E_{1,S}] - [E_{2,S}]) / ([E_{1,S}] + [E_{2,S}])$ , where the asterisk indicates the optimal solution. Then, for each parameters  $k_i$ , we compute the following correlation:

$$C_i = \langle A_S(x^*) \cdot k_i \rangle - \langle A_S(x^*) \rangle \cdot \langle k_i \rangle ,$$

where the angled brackets represent the average over the  $N$  samples. In order to ensure that the optimal concentration corresponded to real absolute minima of the cost function, an iterative approach was developed, which is described in detail in Supplementary Text S2. The results of the sensitivity analysis are instead in Supplementary Table S1.



## Multiple Sequence Alignment

Amino acid sequences of the anthranilate synthase component I (*trpE*) from twelve different *Buchnera* strains available, as well as *Escherichia coli* (strain K12) and *Serratia marcescens*, were downloaded from the UniProtKB database (UniProt-Consortium, 2014). The sequences in fasta format can be found in the Supplementary file Data S1. The multiple sequences alignment was performed using T-coffee package using default parameters (Notredame et al., 2000). The alignment representation was done using Jalview (Waterhouse et al., 2009).

## Estimation of metabolites permeability

A precise quantification of the diffusion constants and membrane permeability for any intermediate metabolite is not a trivial task. Therefore, we used different parameters as estimators of the permeability or diffusion capacity of the intermediate metabolites in different amino acid biosynthetic pathways we analyzed. For each compound we considered: (i) the molecular weight; (ii) number of hydrogen bond acceptors and donors; (iii) the octanol-water partition coefficient; and (iv) the topological polar surface area (TPSA). The structures of the amino acids biosynthetic pathways were retrieved from MetaCyc (Caspi et al., 2014). The list of pathways analyzed includes the super-pathway of aromatic amino acid biosynthesis (L-Trp, L-Tyr, L-Phe) and the super-pathway of branched chain amino acid biosynthesis (L-Leu, L-Ile, L-Val). Furthermore, all the physicochemical parameters listed above were computed using OpenBabel (O'Boyle et al., 2011), (accessed through the Python wrapper PyBel (O'Boyle et al., 2008)) with the exception of the the logarithm of the octanol-water partition coefficient (LogP) for which two different estimators were chosen, named ALogP (Tetko and Poda, 2004) and XLOP3 (Cheng et al., 2007). Results can be found in Supplementary Table S2.

## Acknowledgments

MM acknowledges financial support by Università La Sapienza of Rome, as well as Universidad Complutense de Madrid and University of California, San Diego for technical support. MPL would like to thank Fundación Española para la Cooperación Internacional, Salud y Política Social for allowing him to carry out the research activities corresponding to this paper. Financial support from Spanish Government (grant reference: BFU2012-39816-C02-01 co-financed by FEDER funds and Ministerio de Economía y Competitividad) and Generalitat Valenciana (grant reference: PROMETEOII/2014/ 065) is grateful acknowledged. MP would like to thank Martín Graña for fruitful discussions and comments.

# References

- Baumann, L., Baumann, P., Moran, N. A., Sandström, J., and Thao, M. L. (1999). Genetic Characterization of Plasmids Containing Genes Encoding Enzymes of Leucine Biosynthesis in Endosymbionts (Buchnera) of Aphids. *J. Mol. Evol.* 48, 77–85. doi:10.1007/PL00006447.
- Baumann, P., Moran, N. A., and Baumann, L. (2006). *The Prokaryotes*. , eds. M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer, and E. Stackebrandt New York, NY: Springer New York doi:10.1007/0-387-30741-9.
- Burke, G. R., and Moran, N. A. (2011). Massive genomic decay in *Serratia symbiotica*, a recently evolved symbiont of aphids. *Genome Biol. Evol.* 3, 195–208. doi:10.1093/gbe/evr002.
- Caligiuri, M. G., and Bauerle, R. (1991). Identification of amino acid residues involved in feedback regulation of the anthranilate synthase complex from *Salmonella typhimurium*. Evidence for an amino-terminal regulatory site. *J. Biol. Chem.* 266, 8328–35. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/2022650> [Accessed March 29, 2016].
- Caspi, R., Altman, T., Billington, R., Dreher, K., Foerster, H., Fulcher, C. A., et al. (2014). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res.* 42, D459–71. doi:10.1093/nar/gkt1103.
- Charles, H., Balmand, S., Lamelas, A., Cottret, L., Pérez-Brocal, V., Burdin, B., et al. (2011). A genomic reappraisal of symbiotic function in the aphid/Buchnera symbiosis: reduced transporter sets and variable membrane organisations. *PLoS One* 6, e29096. doi:10.1371/journal.pone.0029096.
- Cheng, T., Zhao, Y., Li, X., Lin, F., Xu, Y., Zhang, X., et al. (2007). Computation of octanol-water partition coefficients by guiding an additive model with knowledge. *J. Chem. Inf. Model.* 47, 2140–2148. doi:10.1021/ci700257y.
- Cornish-Bowden, A. (2013). *Fundamentals of Enzyme Kinetics*. John Wiley & Sons Available at: [https://books.google.com/books?id=\\_3fqaYnrdoC&pgis=1](https://books.google.com/books?id=_3fqaYnrdoC&pgis=1) [Accessed April 8, 2016].
- Crawford, I. P. (1989). Evolution of a biosynthetic pathway: the tryptophan paradigm. *Annu. Rev. Microbiol.* 43, 567–600. doi:10.1146/annurev.mi.43.100189.003031.
- Dekel, E., and Alon, U. (2005). Optimality and evolutionary tuning of the expression level of a protein. *Nature* 436, 588–92. doi:10.1038/nature03842.
- Douglas, A. E. (1992). Requirement of pea aphids (*Acyrtosiphon pisum*) for their symbiotic bacteria. *Entomol. Exp. Appl.* 65, 195–198. doi:10.1111/j.1570-7458.1992.tb01643.x.
- Douglas, A. E. (1998). Host benefit and the evolution of specialization in symbiosis. *Heredity (Edinb.)*. 81, 599–603. doi:10.1046/j.1365-2540.1998.00455.x.
- Dunn, M. F. (2012). Allosteric regulation of substrate channeling and catalysis in the tryptophan synthase bienzyme complex. *Arch. Biochem. Biophys.* 519, 154–66. doi:10.1016/j.abb.2012.01.016.

- Faust, K., and Raes, J. (2012). Microbial interactions: from networks to models. *Nat. Rev. Microbiol.* 10, 538–50. doi:10.1038/nrmicro2832.
- Flamholz, A., Noor, E., Bar-Even, A., Liebermeister, W., and Milo, R. (2013). Glycolytic strategy as a tradeoff between energy yield and protein cost. *Proc. Natl. Acad. Sci. U. S. A.* 110, 10039–44. doi:10.1073/pnas.1215283110.
- Germerodt, S., Bohl, K., Lück, A., Pande, S., Schröter, A., Kaleta, C., et al. (2016). Pervasive Selection for Cooperative Cross-Feeding in Bacterial Communities. *PLOS Comput. Biol.* 12, e1004986. doi:10.1371/journal.pcbi.1004986.
- Goelzer, a, and Fromion, V. (2011). Bacterial growth rate reflects a bottleneck in resource allocation. *Biochim. Biophys. Acta* 1810, 978–88. doi:10.1016/j.bbagen.2011.05.014.
- Gosalbes, M. J., Lamelas, A., Moya, A., and Latorre, A. (2008). The striking case of tryptophan provision in the cedar aphid *Cinara cedri*. *J. Bacteriol.* 190, 6026–9. doi:10.1128/JB.00525-08.
- Großkopf, T., and Soyer, O. S. (2014). Synthetic microbial communities. *Curr. Opin. Microbiol.* 18, 72–77. doi:10.1016/j.mib.2014.02.002.
- Großkopf, T., and Soyer, O. S. (2016). Microbial diversity arising from thermodynamic constraints. *ISME J.* doi:10.1038/ismej.2016.49.
- Hansen, A. K., and Moran, N. A. (2011). Aphid genome expression reveals host-symbiont cooperation in the production of amino acids. *Proc. Natl. Acad. Sci. U. S. A.* 108, 2849–54. doi:10.1073/pnas.1013465108.
- Hansen, S. K., Rainey, P. B., Haagensen, J. A. J., and Molin, S. (2007). Evolution of species interactions in a biofilm community. *Nature* 445, 533–6. doi:10.1038/nature05514.
- Hardin, G. (1960). The competitive exclusion principle. *Science* 131, 1292–7. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/14399717> [Accessed December 9, 2015].
- Heinrich, R., and Schuster, S. (1996). *The Regulation of Cellular Systems*. Springer Available at: <http://www.amazon.com/Regulation-Cellular-Systems-Reinhart-Heinrich/dp/0412032619> [Accessed February 1, 2013].
- Hosoda, K., Suzuki, S., Yamauchi, Y., Shiroguchi, Y., Kashiwagi, A., Ono, N., et al. (2011). Cooperative adaptation to establishment of a synthetic bacterial mutualism. *PLoS One* 6. doi:10.1371/journal.pone.0017105.
- Hosoda, K., and Yomo, T. (2011). Designing symbiosis. *Bioeng. Bugs* 2, 338–341. doi:10.4161/bbug.2.6.16801.
- Jiang, Z.-F. F., Xia, F., Johnson, K. W., Bartom, E., Tuteja, J. H., Stevens, R., et al. (2012). Genome sequences of the primary endosymbiont “*Candidatus portiera aleyrodidarum*” in the whitefly *Bemisia tabaci* B and Q biotypes. *J. Bacteriol.* 194, 6678–6679. doi:10.1128/JB.01841-12.
- Jiménez, N., González-Candelas, F., and Silva, F. J. (2000). Prephenate dehydratase from the aphid endosymbiont (*Buchnera*) displays changes in the regulatory domain that suggest its

- desensitization to inhibition by phenylalanine. *J. Bacteriol.* 182, 2967–9. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=102009&tool=pmcentrez&rendertype=abstract> [Accessed May 14, 2016].
- Kacser, H., and Burns, J. A. (1973). The control of flux. *Symp. Soc. Exp. Biol.* 27, 65–104. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/4148886> [Accessed May 16, 2016].
- Kafri, M., Metzl-Raz, E., Jona, G., and Barkai, N. (2015). The Cost of Protein Production. *Cell Rep.* 14, 22–31. doi:10.1016/j.celrep.2015.12.015.
- Kirschner, K., Lane, A. N., and Strasser, A. W. (1991). Reciprocal communication between the lyase and synthase active sites of the tryptophan synthase holoenzyme complex. *Biochemistry* 30, 472–8. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/1899027> [Accessed March 24, 2016].
- Kwak, J.-H., Hong, K.-W., Lee, S.-H., Hong, J.-H., and Lee, S.-Y. (1999). Identification of Amino Acid Residues Involved in Feedback Inhibition of the Anthranilate Synthase in *Escherichia coli*. *BMB Rep.* 32, 20–24. Available at: [http://www.koreascience.or.kr/article/ArticleFullRecord.jsp?cn=E1MBB7\\_1999\\_v32n1\\_20](http://www.koreascience.or.kr/article/ArticleFullRecord.jsp?cn=E1MBB7_1999_v32n1_20) [Accessed March 23, 2016].
- Lai, C. Y., Baumann, L., and Baumann, P. (1994). Amplification of trpEG: adaptation of *Buchnera aphidicola* to an endosymbiotic association with aphids. *Proc. Natl. Acad. Sci. U. S. A.* 91, 3819–23. doi:10.1073/pnas.91.9.3819.
- Laidler, K. J., and Shuler, K. E. (1949). The Kinetics of Membrane Processes. I. The Mechanism and the Kinetic Laws for Diffusion through Membranes. *J. Chem. Phys.* 17, 851. doi:10.1063/1.1747076.
- Lamelas, A., Gosalbes, M. J., Manzano-Marín, A., Peretó, J., Moya, A., and Latorre, A. (2011). *Serratia symbiotica* from the aphid *Cinara cedri*: a missing link from facultative to obligate insect endosymbiont. *PLoS Genet.* 7, e1002357. doi:10.1371/journal.pgen.1002357.
- Van Leuven, J. T., Meister, R. C., Simon, C., and McCutcheon, J. P. (2014). Sympatric speciation in a bacterial endosymbiont results in two genomes with the functionality of one. *Cell* 158, 1270–80. doi:10.1016/j.cell.2014.07.047.
- López-Sánchez, M. J., Neef, A., Peretó, J., Patiño-Navarrete, R., Pignatelli, M., Latorre, A., et al. (2009). Evolutionary convergence and nitrogen metabolism in *Blattabacterium* strain Bge, primary endosymbiont of the cockroach *Blattella germanica*. *PLoS Genet.* 5, e1000721. doi:10.1371/journal.pgen.1000721.
- Macdonald, S. J., Lin, G. G., Russell, C. W., Thomas, G. H., and Douglas, A. E. (2012). The central role of the host cell in symbiotic nitrogen metabolism. *Proc. Biol. Sci.* 279, 2965–73. doi:10.1098/rspb.2012.0414.
- Machado, D., Herrgård, M. J., and Rocha, I. (2015). Modeling the Contribution of Allosteric Regulation for Flux Control in the Central Carbon Metabolism of *E. coli*. *Front. Bioeng. Biotechnol.* 3, 154. doi:10.3389/fbioe.2015.00154.

- Martínez-Cano, D. J., Reyes-Prieto, M., Martínez-Romero, E., Partida-Martínez, L. P., Latorre, A., Moya, A., et al. (2015). Evolution of small prokaryotic genomes. *Front. Microbiol.* 5, 742. doi:10.3389/fmicb.2014.00742.
- Martino, P. Di, Fursy, R., Bret, L., Sundararaju, B., and Phillips, R. S. (2003). Indole can act as an extracellular signal to regulate biofilm formation of *Escherichia coli* and other indole-producing bacteria. *Can. J. Microbiol.* 49, 443–9. doi:10.1139/w03-056.
- McCutcheon, J. P., and von Dohlen, C. D. (2011). An interdependent metabolic patchwork in the nested symbiosis of mealybugs. *Curr. Biol.* 21, 1366–72. doi:10.1016/j.cub.2011.06.051.
- Miles, E. W. (2013). The tryptophan synthase  $\alpha\beta_2$  complex: a model for substrate channeling, allosteric communication, and pyridoxal phosphate catalysis. *J. Biol. Chem.* 288, 10084–91. doi:10.1074/jbc.X113.463331.
- Moran, N. A. (1996). Accelerated Evolution and Muller's Ratchet in Endosymbiotic Bacteria. *Proc. Natl. Acad. Sci.* 93, 2873–2878. Available at: <http://www.pnas.org/content/93/7/2873> [Accessed May 29, 2012].
- Moran, N. A., and Bennett, G. M. (2014). The Tiniest Tiny Genomes. *Annu. Rev. Microbiol.* 68, 195–215. doi:10.1146/annurev-micro-091213-112901.
- Mori, M., Hwa, T., Martin, O. C., De Martino, A., and Marinari, E. (2016). Constrained Allocation Flux Balance Analysis. *PLOS Comput. Biol.* 12, e1004913. doi:10.1371/journal.pcbi.1004913.
- Morris, J., Lenski, R. E., Zinser, E. R., Black, T., Hypothesis, Q., Loss, A. G., et al. (2012). The Black Queen Hypothesis: Evolution of Dependencies through Adaptive Gene Loss. doi:10.1128/mBio.00036-12.Updated.
- Noor, E., Flamholz, A., Liebermeister, W., Bar-Even, A., and Milo, R. (2013). A note on the kinetics of enzyme action: a decomposition that highlights thermodynamic effects. *FEBS Lett.* 587, 2772–7. doi:10.1016/j.febslet.2013.07.028.
- Notredame, C., Higgins, D. G., and Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* 302, 205–17. doi:10.1006/jmbi.2000.4042.
- O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., and Hutchison, G. R. (2011). Open Babel: An open chemical toolbox. *J. Cheminform.* 3, 33. doi:10.1186/1758-2946-3-33.
- O'Boyle, N. M., Morley, C., and Hutchison, G. R. (2008). Pybel: a Python wrapper for the OpenBabel cheminformatics toolkit. *Chem. Cent. J.* 2, 5. doi:10.1186/1752-153X-2-5.
- Pande, S., Merker, H., Bohl, K., Reichelt, M., Schuster, S., de Figueiredo, L. F., et al. (2014). Fitness and stability of obligate cross-feeding interactions that emerge upon gene loss in bacteria. *ISME J.* 8, 953–62. doi:10.1038/ismej.2013.211.
- Pérez-Brocá, V., Gil, R., Ramos, S., Lamelas, A., Postigo, M., Michelena, J. M., et al. (2006). A small microbial genome: the end of a long symbiotic relationship? *Science (80-. )*. 314, 312–

313. doi:10.1126/science.1130441.

- Pfeiffer, T., and Bonhoeffer, S. (2004). Evolution of cross-feeding in microbial populations. *Am. Nat.* 163, E126–35. doi:10.1086/383593.
- Piñero-Fernandez, S., Chimere, C., Keyser, U. F., and Summers, D. K. (2011). Indole transport across *Escherichia coli* membranes. *J. Bacteriol.* 193, 1793–8. doi:10.1128/JB.01477-10.
- Poliakov, A., Russell, C. W., Ponnala, L., Hoops, H. J., Sun, Q., Douglas, A. E., et al. (2011). Large-scale label-free quantitative proteomics of the pea aphid-*Buchnera* symbiosis. *Mol. Cell. Proteomics* 10, M110.007039. doi:10.1074/mcp.M110.007039.
- Ponce de León, M., Cancela, H., and Acerenza, L. (2008). A strategy to calculate the patterns of nutrient consumption by microorganisms applying a two-level optimisation principle to reconstructed metabolic networks. *J. Biol. Phys.* 34, 73–90. doi:10.1007/s10867-008-9067-2.
- Ponce-de-Leon, M., Montero, F., Peretó, J., Ponce-de-Leon, M., Montero, F., Peretó, J., et al. (2013). Solving gap metabolites and blocked reactions in genome-scale models: application to the metabolic network of *Blattabacterium cuenoti*. *BMC Syst. Biol.* 7, 114. doi:10.1186/1752-0509-7-114.
- Price, D. R. G., Feng, H., Baker, J. D., Bavan, S., Luetje, C. W., and Wilson, A. C. C. (2014). Aphid amino acid transporter regulates glutamine supply to intracellular bacterial symbionts. *Proc. Natl. Acad. Sci. U. S. A.* 111, 320–5. doi:10.1073/pnas.1306068111.
- Rao, Q., Rollat-Farnier, P.-A., Zhu, D.-T., Santos-Garcia, D., Silva, F. J., Moya, A., et al. (2015). Genome reduction and potential metabolic complementation of the dual endosymbionts in the whitefly *Bemisia tabaci*. *BMC Genomics* 16, 226. doi:10.1186/s12864-015-1379-6.
- Ringemann, C., Ebenho, O., Heinrich, R., and Ginsburg, H. (2006). Can biochemical properties serve as selective pressure for gene selection during inter-species and endosymbiotic lateral gene transfer ? 153. doi:10.1049/ip-syb.
- Rouhbakhsh, D., Clark, M. A., Baumann, L., Moran, N. A., and Baumann, P. (1997). Evolution of the tryptophan biosynthetic pathway in *Buchnera* (aphid endosymbionts): studies of plasmid-associated trpEG within the genus *Uroleucon*. *Mol. Phylogenet. Evol.* 8, 167–76. doi:10.1006/mpev.1997.0419.
- Rouhbakhsh, D., Lai, C. Y., von Dohlen, C. D., Clark, M. A., Baumann, L., Baumann, P., et al. (1996). The tryptophan biosynthetic pathway of aphid endosymbionts (*Buchnera*): genetics and evolution of plasmid-associated anthranilate synthase (trpEG) within the aphididae. *J. Mol. Evol.* 42, 414–21. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/8642610> [Accessed March 23, 2016].
- Russell, C. W., Bouvaine, S., Newell, P. D., and Douglas, A. E. (2013). Shared metabolic pathways in a coevolved insect-bacterial symbiosis. *Appl. Environ. Microbiol.* 79, 6117–23. doi:10.1128/AEM.01543-13.
- Russell, C. W., Poliakov, A., Haribal, M., Jander, G., van Wijk, K. J., and Douglas, A. E. (2014).



- Matching the supply of bacterial nutrients to the nutritional demand of the animal host. *Proc. Biol. Sci.* 281, 20141163. doi:10.1098/rspb.2014.1163.
- Seth, E. C., and Taga, M. E. (2014). Nutrient cross-feeding in the microbial world. *Front. Microbiol.* 5, 350. doi:10.3389/fmicb.2014.00350.
- Shachrai, I., Zaslaver, A., Alon, U., and Dekel, E. (2010). Cost of unneeded proteins in *E. coli* is reduced after several generations in exponential growth. *Mol. Cell* 38, 758–67. doi:10.1016/j.molcel.2010.04.015.
- Sloan, D. B., and Moran, N. A. (2012). Genome reduction and co-evolution between the primary and secondary bacterial symbionts of psyllids. *Mol. Biol. Evol.* 29, 3781–92. doi:10.1093/molbev/mss180.
- Spraggon, G., Kim, C., Nguyen-Huu, X., Yee, M. C., Yanofsky, C., and Mills, S. E. (2001). The structures of anthranilate synthase of *Serratia marcescens* crystallized in the presence of (i) its substrates, chorismate and glutamine, and a product, glutamate, and (ii) its end-product inhibitor, L-tryptophan. *Proc. Natl. Acad. Sci. U. S. A.* 98, 6021–6. doi:10.1073/pnas.111150298.
- Tan, J., Zuniga, C., and Zengler, K. (2015). Unraveling interactions in microbial communities - from co-cultures to microbiomes. *J. Microbiol.* 53, 295–305. doi:10.1007/s12275-015-5060-1.
- Tang, X. F., Ezaki, S., Atomi, H., and Imanaka, T. (2001). Anthranilate synthase without an LLES motif from a hyperthermophilic archaeon is inhibited by tryptophan. *Biochem. Biophys. Res. Commun.* 281, 858–65. doi:10.1006/bbrc.2001.4428.
- Tepper, N., Noor, E., Amador-Noguez, D., Haraldsdóttir, H. S., Milo, R., Rabinowitz, J., et al. (2013). Steady-state metabolite concentrations reflect a balance between maximizing enzyme efficiency and minimizing total metabolite load. *PLoS One* 8, e75370. doi:10.1371/journal.pone.0075370.
- Tetko, I. V., and Poda, G. I. (2004). Application of ALOGPS 2.1 to predict log D distribution coefficient for Pfizer proprietary compounds. *J. Med. Chem.* 47, 5601–4. doi:10.1021/jm049509l.
- UniProt-Consortium (2014). UniProt: a hub for protein information. *Nucleic Acids Res.* 43, D204–12. doi:10.1093/nar/gku989.
- Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–91. doi:10.1093/bioinformatics/btp033.
- White, D., Drummond, J. T., and Fuqua, C. (2012). *The Physiology and Biochemistry of Prokaryotes*. Oxford University Press Available at: <https://books.google.com/books?id=ToF6pwAACAAJ&pgis=1> [Accessed March 31, 2016].
- Wilcox, J. L., Dunbar, H. E., Wolfinger, R. D., and Moran, N. A. (2003). Consequences of reductive evolution for gene expression in an obligate endosymbiont. *Mol. Microbiol.* 48, 1491–500.



Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12791133> [Accessed February 25, 2016].

- Wu, D., Daugherty, S. C., Van Aken, S. E., Pai, G. H., Watkins, K. L., Khouri, H., et al. (2006). Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. *PLoS Biol.* 4, e188. doi:10.1371/journal.pbio.0040188.
- Zelezniak, A., Andrejev, S., Ponomarova, O., Mende, D. R., Bork, P., and Patil, K. R. (2015). Metabolic dependencies drive species co-occurrence in diverse microbial communities. *Proc. Natl. Acad. Sci.* 112, 201421834. doi:10.1073/pnas.1421834112.
- Zientz, E., Dandekar, T., and Gross, R. (2004). Metabolic interdependence of obligate intracellular bacteria and their insect hosts. *Microbiol. Mol. Biol. Rev.* 68, 745–70. doi:10.1128/MMBR.68.4.745-770.2004.