Simulating the evolutionary trajectories of metabolic pathways for insect symbionts in the *Sodalis* genus

Rebecca J. Hall*[1,2], Stephen Thorpe*[1,3], Gavin H. Thomas[1] and A. Jamie Wood[1,4]

Author affiliations

*These authors contributed equally to this work.
[1]Department of Biology, University of York, York, YO10 5NG, UK
[2]School of Life Sciences, University of Nottingham, Nottingham, NG7 2TQ, UK
[3]Department of Chemistry, University of Oxford, Oxford, OX1 3TA, UK
[4]Department of Mathematics, University of York, York, YO10 5DD, UK
Corresponding author: A. Jamie Wood (jamie.wood@york.ac.uk)

Abstract

Insect-bacterial symbioses are ubiquitous, but there is still much to uncover about how these relationships establish, persist and evolve. The tsetse endosymbiont *Sodalis glossinidius* displays intriguing metabolic adaptations to its microenvironment, but the process by which this relationship evolved remains to be elucidated. The recent chance discovery of the free-living species of the *Sodalis* genus, *S. praecaptivus*, provides a serendipitous starting point from which to investigate the evolution of this symbiosis. Here, we present a flux balance model for *S. praecaptivus* and empirically verify its predictions. Metabolic modelling is used in combination with a multi-objective evolutionary algorithm to explore the trajectories that *S. glossinidius* may have undertaken from this starting point after becoming internalised. The order in which key genes are lost is shown to influence the evolved populations, providing possible targets for future *in vitro* genetic manipulation. This method provides a detailed perspective on

possible evolutionary trajectories for *S. glossinidius* in this fundamental process of evolutionary and ecological change.

Keywords

Symbiosis, evolution, *Sodalis*, flux balance analysis, evolutionary algorithm

Data Summary

The Python code for running the algorithm with an example data set is available on GitHub at https://github.com/St659/SodalisFBAEvolution. The data generated by the simulations are available on the York Research Database.

Data Statement

All supporting data, code and protocols have been provided within the article or through supplementary data files. Supplementary material is available with the online version of this article.

Impact Statement

Insect-microbe symbioses are challenging to study as the symbionts may not be amenable to *in vitro* culture or traditional genetic manipulation techniques. The establishment and tracking of symbiosis from initiation to infection also presents technical challenges. A metabolic model of a free-living plausible starting organism is presented and verified against empirical data. This work provides a computational method to examine the potential evolutionary trajectories that symbionts may have taken once becoming internalised by a host. It enables new questions to be asked about genome reduction and niche adaptation by symbiotic bacteria. This technique has wider implications beyond symbiosis, with potential applications in directed evolution for industrial biotechnology.

Introduction

Symbioses are both fundamental and ubiquitous in nature. Understanding their evolution poses an ongoing challenge, as well as an expanse of unresolved research questions. Bacterial symbionts of insects provide a range of benefits including stress tolerance [1, 2], protection from predation [1, 3, 4] and the provision of metabolites [5–10]. The latter forms arguably the strongest link within the symbioses. Host and symbiont frequently share metabolic substrates, as well as the products and components of individual biosynthetic pathways [7,11–15]. These relationships typically enable the host to survive on a nutritionally restricted diet, such as blood [8,16,17] or plant sap [18–20].

Deciphering the evolutionary pressures that affect the organisms within a symbiosis is an essential part of understanding the relationship. This includes establishing how the symbioses develop over time and the way in which the metabolism of the individuals is intertwined. It is, however, often hindered by biological difficulties. Symbiotic bacteria undergo genomic streamlining, may not be cultivatable *in vitro*, may no longer express stress response genes and might lack a sound outer membrane [4,5,21–25]. It is therefore impossible in many cases to test hypotheses about host-symbiont interactions in controlled experimental conditions. In these circumstances, computational techniques offer a viable, and currently the only, alternative to investigating metabolic potential and pseudogenisation in symbiotic bacteria.

Computational biology is now well established as a key tool of scientific discovery, now that vast amounts of data are generated quickly and cheaply from advancements in sequencing technology [26,27]. Genome scale metabolic modelling of microorganisms enables predictions to be made about metabolite preferences, transporter use and the functionality of biosynthetic pathways [26,28]. Microbial metabolism can be simulated using flux balance analysis (FBA), a constraint-based quantitative approach that reconstructs a metabolic network from a genome annotation [27,29]. FBA is a powerful tool when based on a well annotated genome and with the provision of *in vitro* experimental validation [29].

FBA is widely used for biotechnology applications, and this can be re-purposed to examine symbiosis. There are several published examples of using FBA to analyse the metabolism of symbiotic bacteria, including for *Buchnera aphidicola* [7,30,31], *Sodalis glossinidius* [32,33], *Portiera aleyrodidarum* [34], *Hamiltonella defensa* [34] and strains of *Blattabacterium* [35]. There are also models published for the *Synechocystis* species

used in the study of artificially induced symbiosis [36–40]. FBA is useful in this instance, as experiments that would not be possible empirically, due to culturability issues, can be performed *in silico*. Furthermore, the genomes of symbiotic bacteria are often unusual, with large pathway deletions or widespread pseudogenisation [6,25,41,42]. Analysis of the resulting metabolic network via FBA can suggest which pathways are being used when supplied with different media, and predict which external metabolites might be required to support growth *in vitro*.

FBA has been applied to several microbiological problems. Boolean logical operators have been incorporated into *Escherichia coli* metabolic models to investigate the impact of gene regulation on a system [43–46]. Dynamic FBA (dFBA), where a rate of change in flux constraints is included, has successfully modelled diauxic growth in *E. coli* [47]. FBA has been used to compare strains of *Blattabacterium* from separate cockroach lineages to assess their divergence [35], and to predict the evolution of metabolism from *E. coli* experimental data sets [48]. The evolution of metabolic networks in isolation has also been simulated with the aim of identifying key metabolites [49], and to investigate pseudogenisations in specific metabolic pathways [50]. FBA has not yet been harnessed to its full potential with regards to the investigation of symbiont evolution. This is perhaps surprising given that several models of *E. coli* metabolism are available as an evolutionary starting point [51–56]. The evolution of *B. aphidicola* and *Wigglesworthia glossinidia* from an *E. coli* ancestor has been simulated using FBA [56,57]. This work, whilst elegant, has a key limitation. Reactions that are lost at the start have no chance of being reintroduced. This limits the evolutionary space that can be explored, as the loss of a key reaction at the start will fundamentally affect which reactions can be lost subsequently. A similar approach to that used by Pal *et al*. [56] was used with dFBA to study the evolution of cooperation and cross-feeding in *E. coli* [58]. Using FBA in isolation to remove reactions successively may not therefore be the optimal way to simulate the evolution of symbiosis.

In silico evolution has been used increasingly in recent years to complement *in vivo* experimental evolution [59]. *In silico* evolution benefits from being able to test widely different ecological conditions whilst controlling key variables [60]. For example, it allows the investigation of groups of mutations that lead to a specific phenotype, or mutations that are difficult to induce *in vitro* [61]. This has enabled the study of many aspects of evolution, including simulating the reduction of genome size in an individual [60]. Multi-objective evolutionary algorithms (MOEA) have been used in many disciplines for solving problems that have two or more conflicting objectives [62]. The use of MOEA in combination with metabolic models has been implemented for the

design of minimal genomes [63] and for the production of industrially relevant molecules [64,65]. It has however seen only limited use for *in silico* evolution [66]. When viewed computationally, the evolution of symbiosis can be considered as a multi-objective optimisation; symbiotic bacteria undergo genome reduction whilst trying to maximise their individual growth.

A free-living organism within the *Sodalis* genus has been characterised and sequenced only recently [67,68]. *Sodalis praecaptivus* was isolated from a human wound, caused by an impalement on a crab apple tree branch, and it is assumed that the tree was the likely source of the *S. praecaptivus* infection. *S. praecaptivus* is a prototroph, capable of growth in minimal media and at 37°C [68]. The annotated genome sequence for *S. praecaptivus* is also available [67]. It is of particular interest given its close relation *S. glossinidius*, secondary symbiont of the tsetse [25]. The tsetse, genus *Glossina* is medically important as the vector for *Trypanosoma brucei*, causative agent of human African trypanosomiasis [69]. *S. praecaptivus* therefore provides a rich set of data from which to begin investigations into the origin of, and adaptations within, the tsetse-*S. glossinidius* symbiosis.

Here, we present a flux balance model for *S. praecaptivus*, *i*RH830. This model, and a previously presented model of *S. glossinidius* metabolism, *i*LF517 [33], both represent adaptations of the organisms to their contrasting environments. The *Sodalis* system is therefore an excellent candidate for assessing the ability of FBA to describe the evolution of symbioses. A MOEA has been used to evolve *i*RH830 under various biological conditions. The aim was to investigate computationally the route that *S. glossinidius* may have taken in its transition to symbiosis. It is not known whether the solutions found by *S. glossinidius*, described in *i*LF517, are the only possible outcomes given the metabolic constraints of the microenvironment, or whether the symbiont's unusual metabolic network evolved by chance. The application of the MOEA to *i*RH830 enabled us to ask in which order of the evolutionary sequence key pseudogenisations may have occurred. The effect of exposing the ancestral *Sodalis* to contrasting diets was modelled to mirror the different trajectories that this genus has taken within blood- and sap-feeding insects. The techniques used here could be applied to other symbiotic systems to drive forward the discovery of novel relationship criteria.

Results

A model of *S. praecaptivus* metabolism, *i*RH830

In order to investigate computationally the path that *S. glossinidius* has taken to symbiosis, a metabolic model describing its close, free-living relative *S. praecaptivus* was constructed (Fig. 1). Full details are given in Supplementary Data 1. *i*RH830 contains 830 genes, 891 metabolites and 1246 reactions (excluding pseudoreactions), and is a prototroph for all essential amino acids. An iterative process of gap filling was undertaken by comparing the draft *S. praecaptivus* model to *i*LF517 (*S. glossinidius*) and *i*JO1366, a model of *E. coli* metabolism [54]. *i*RH830 is supplied with an oxygen uptake value of 20 mmol gr DW^{-1} hr^{-1}, reflecting the highly aerated conditions the organism is grown in and to retain consistency with models of *E. coli* metabolism [53,54].

Fig. 1. The construction process for *i*RH830. The *S. praecaptivus* genome was mined for orthologues to metabolic genes in *E. coli* and *S. glossinidius*, before compiling into a draft model. An iterative process of testing and gap filling was then performed, using information provided in various databases (see key).

A series of biochemical screens were conducted using Biolog phenotypic microplates to strengthen the model. In total, 190 metabolites were tested for their ability to act as the sole carbon source for *S. praecaptivus*. Experiments were conducted in triplicate with full results detailed in Supplementary Data 2. Through this phenotypic screen, it was found that *S. praecaptivus* was able to use 19 of the metabolites tested as a sole source of carbon (Table S1). When these metabolites were tested *in silico* by the exogenous addition to *i*RH830, it was found that all but two mirrored the *in vitro* data; *N*-acetyl D-galactosamine (GalNAc) and xylitol. This was then confirmed quantitatively in a 96-well microplate with xylitol or GalNac supplemented into M9 minimal medium (Fig. S1a, Fig. S1b). Neither models of *S. glossinidius* (*i*LF517, [33]) or *E. coli* (*i*JO1366, [54]) were able to produce a positive biomass output with xylitol or GalNAc as sole sources of carbon (Table S1).

Comparison of the *S. praecaptivus* genome to other known D-xylitol consumers such as *Morganella morganii* subsp. *morganii* revealed a highly conserved catabolic operon containing the distinguishing D-xylitol dehydrogenase (79.7% identity between the *S. praecaptivus* orthologue, AFW03778/Sant_3108, and the *Morganella morganii* subsp.

morganii protein, Uniprot ID: Q59545). The cluster contains a xylulose reductase as the second enzyme required to convert D-xylitol to the central metabolite D-xylulose-5-phosphate (Fig. S1c) and a complete ABC transporter that is likely specific for D-xylitol. Interestingly, the pathway is also complete in the reduced *S. glossinidius* genome (Table S2), suggesting that this is a conserved metabolic trait of the *Sodalis* genus. The cluster is not present in *E. coli* with only weakly matching homologues are found fragmented over the genome (Table S2). The proposed pathway for GalNAc metabolism was also constructed using known pathways (Fig. S1d).

Robustness analysis of the *S. praecaptivus* metabolic network

Robustness analysis was used to examine reaction essentiality and therefore redundancy in the *i*RH830 network. *i*RH830 was run on a simple, tsetse-specific nutrient limited medium ("famine") and a blood medium simulating the internal tsetse environment and informed by *S. glossinidius* requirements [33] ("blood", Table S3). All media are detailed in Supplementary Data 1. Reactions were removed individually and the resulting effect on biomass output noted. The same analysis was also run on *i*LF517 in blood as a comparison.

There are 282 essential reactions in *i*RH830 when the medium (famine) is nutritionally limited, and 228 in the tsetse-specific blood medium (Fig. 2a). The overall pattern for the two conditions is very similar. The subsystem most represented in either condition is for cofactor and prosthetic group biosynthesis, with 88 and 87 essential reactions for the famine and blood media, respectively. The main difference at the subsystem level can be attributed to amino acid metabolism; 15.8% of the total number of essential reactions in blood and 29.8% in the famine medium that does not contain amino acids are involved in these pathways. Of these, the essential reactions involved in L-arginine, L-proline, L-threonine and L-lysine metabolism are highly prevalent in both media types (Fig. 2b).

Fig. 2. Robustness analysis of *i*RH830. (a) Essential reactions in famine (top) and blood (bottom) media. Essential reactions are categorised by subsystem. (b) Essential reactions involved in amino acid metabolism in *i*RH830 in famine (top) and blood (bottom) media.

Media provisioning affects evolutionary trajectories

NSGA-II is a heuristic multi-objective optimisation algorithm used to evaluate multi-objective problems without giving weight to any specific outcome. Evolution within a constrained environment, such as the tsetse microenvironment, can be considered a multi-objective optimisation problem of trying to reduce the genome size to increase replication speed, while still retaining sufficient capacity to grow [70]. The MOEA was used to explore the potential evolutionary trajectories of *S. praecaptivus* when exposed to similar environmental conditions to *S. glossinidius*. A graphical description of the MOEA is provided in Fig. 7. A key feature of this is the option of reactions that have been removed being re-introduced later in the simulation. This helps to prevent the model from consistently finding the same solutions and instead allows a greater evolutionary space to be explored. The conditions under which *i*RH830 and *i*LF517 were evolved are detailed in Table 1. In Scenario i, *i*RH830 was evolved in blood and famine growth media, as well as a medium that mimics plant sap (Supplementary Data 1), to examine the effect of metabolite availability. In Scenario ii, three key reactions, ASPTA, PDH and PPC, were removed from *i*RH830 prior to evolution to compare the trajectories that arise as a result of pseudogenisations, and to investigate if these were possible adaptations prior to symbiont establishment. The gene encoding PPC is thought to be pseudogenised in *S. glossinidius*, whereas the PDH and ASPTA reactions are predicted to be functional [33]. In Scenario iii, the MOEA is applied to *i*LF517 to investigate the possible future of *S. glossinidius* as a symbiont.

Table 1. *In silico* evolution conditions. Conditions under which *i*RH830 and *i*LF517 were evolved, including wild-type (WT) or reaction knockouts and media type.

Scenario	Test	Model	Media
i	Effect of growth media	*i*RH830 (WT)	Blood, sap, famine
ii	Effect of gene loss	*i*RH830 (ΔASPTA, ΔPDH, ΔPPC)	Blood
iii	Future of *S. glossinidius*	*i*RH830 (WT), *i*LF517 (WT)	Blood

Species of the *Sodalis* genus have been found in insects that feed on a variety of contrasting diets, including blood (e.g. tsetse [25] and ticks [71–73]) and plant tissue (e.g. weevils [74]). To replicate *Sodalis* evolution in different environments, the MOEA was applied to *i*RH830 that was supplied with the tsetse-specific blood medium, the famine medium, and a medium that mimics plant sap (Table 1, Scenario i). Sap was chosen as a comparison medium as *Sodalis*-allied symbionts have been identified in a range of phytophagic insects [75–80]. The algorithm underwent ten runs of 3000 generations and the resulting solutions were collated.

In all conditions, the models evolved to completion, demonstrated by the convergence of solutions to the left of the plots (Fig. 3a). The number of reactions decreases over

evolutionary time, with the majority of solutions clustering at the maximum biomass output. This is an indication that sub-optimal solutions are being removed successfully. After 3000 generations, there are a range of solution sizes at the maximum biomass output found in sap, whereas in blood and famine all solutions at this time point cluster at the minimum number of reactions. The two complex media, blood and sap (Fig. 3a), produce a lot of metabolic flexibility, with a complete range of possible biomass outputs produced by the smallest models. When grown in the simple famine medium, there is significantly less flexibility in terms of possible solutions found (Fig. 3a). Here, the majority of the solutions cluster at the minimum reactions/maximum biomass output. This is as expected, given the fitness function of the MOEA. In blood and sap, the biomass outputs reach near zero, made possible by the variety of available substrates. In famine, the options for streamlining are limited, resulting in few solutions that are able to deviate away from what is selected by the fitness function.

Fig. 3. *i*RH830 evolved under different starting conditions. (a) Evolution of *i*RH830 in a tsetse-specific blood medium (left), a nutritionally-limited famine medium (centre), and a medium mimicking plant sap (right). (b) *i*RH830 evolution in a blood medium with the reactions ASPTA (left), PDH (centre) and PPC (right) removed at the start. The MOEA was run for 3000 generations, with the plot depicting new populations every 50 generations (blue to green). Black boxes indicate individual solutions selected for further analysis.

A number of individual solutions that were representative of the biomass output of *i*LF517 [33] were then selected from each of these simulations (Fig. 3, black boxes). The raw, binary data were translated back into reaction names and this was subsequently processed to produce a list of "core non-essential reactions". These reactions are found in all individuals selected, and do not produce a lethal phenotype when removed. A full list of all core non-essential reactions described here can be found in Table S4. There are 14 core non-essentials reactions found in all 1194 of the individuals examined when *i*RH830 was supplied with blood; AGDC, ARGabc, ASNt2r, G6PDA, H2Ot, HISt2r, ILEt2r, NH4t, RPE, TKT1, TKT2, TMK, TRPt2r and TYRt2r. In sap, only one non-essential reaction is found in all 1888 individuals; the L-arginine ABC transporter reaction ARGabc. As anticipated, when grown in the limited famine medium there are a higher number of core nonessential reactions (22 found in each of the 2989 individuals tested); ATPS4r, CO2t,
ENO, FORt, GAPD, GHMT2r, GLUDy, ORNDC, PAPSR, PGCD, PGK, PGM, PPPGO3, PSERT, PSP_L, RPE, TALA, THRAi, TKT1, TPI, TRDR and TRPS1.

The rare core non-essential reactions were then calculated. In famine, there are 73 unique reactions that occur in less than 0.1% of the 2989 evolved models. This is significantly more than for sap (13 in less than 0.1% of 1888 models) and blood (five in less than 0.1% of 1194 models).

These core non-essential reactions were then analysed by subsystem to assess themes across the different conditions. In blood, over half (eight of 14) of these are secondary transporter reactions (Fig. 4a). This reflects what is observed in *S. glossinidius*, which has retained, for example, secondary amino acid transporters, as well as losing metabolic pathways whilst maintaining functional transporters in order to scavenge free metabolites [33, 81]. As mentioned previously, the only core nonessential reaction in sap is a transport reaction. In contrast, the set of core nonessentials are more varied when metabolites are limited (famine), with a particular emphasis towards central metabolism and amino acid metabolism.

Fig. 4. Core non-essential reactions in evolved *i*RH830 populations. (a) The proportion of core non-essential reactions per conditions by subsystem when the ancestral *i*RH830 is exposed to blood (left), famine (centre), or sap (right) media. (b) Core non-essential reactions in ΔASPTA (left), ΔPDH (centre), and ΔPPC (right) *i*RH830 models in a blood medium, grouped by subsystem.

The order of gene loss can be estimated

A characteristic of *S. glossinidius* and other symbiotic bacteria is their propensity to accumulate pseudogenes [41]. It is not known whether certain genes are lost early in the tsetse-*Sodalis* symbiosis in order to facilitate the establishment of the relationship, or whether their loss is an inevitable consequence of genomic streamlining. To investigate the effect that pseudogenising key genes early in evolutionary time has on the trajectory of a symbiont, the MOEA was run on *i*RH830 with one of three reactions involved in the TCA cycle removed at the start, with the resulting solutions compared to wild-type (WT) (Table 1, Scenario ii). An assumption is made in these simulations that pseudogenes are non-functional. The first reaction selected was PPC (phosphoenolpyruvate carboxylase) (Fig. 5). The gene encoding this reaction, *ppc*, is pseudogenised in *S. glossinidius* [33] and it was thought that the loss of this gene would have had a significant impact on the resulting evolution of the symbiont. The two other reactions selected were PDH (pyruvate dehydrogenase) and ASPTA (aspartate transaminase). These reactions are both encoded by genes predicted to be functional in *S. glossinidius* (Fig. 5). It was hypothesised that the loss of PPC may result in a different

evolutionary trajectory compared to the loss of PDH or ASPTA, with the former potentially producing solutions that were more similar to *S. glossinidius* metabolism.

Fig. 5. TCA cycle reactions examined by the MOEA. Three reactions were removed from *i*RH830 to investigate the resulting trajectories following application of the MOEA; ASPTA, PDH and PPC. Blue arrows show reactions functional in *S. praecaptivus* and *S. glossinidius*, white arrows show reaction not functional in *S. glossinidius*. Gene associations in *S. praecaptivus* and *S. glossinidius* are given in blue and black text, respectively. Adapted from Hall *et al*. [33].

When considering the population plots, there is minimal qualitative difference between ΔPDH and ΔPPC (Fig. 3b). ΔASPTA, in contrast, produces solutions with a much lower biomass output and with fewer individuals that deviate away from the optimum as defined by the fitness function. A selection of individuals were then examined and the number of core non-essential reactions in the evolved models analysed as described previously (Fig. 3b, black boxes). There are one, eleven and nine core non-essential reactions in the WT, ΔPDH and ΔPPC solutions, respectively, whereas there are 61 in ΔASPTA. These 61 reactions function in a variety of subsystems, particularly transport, central metabolism, amino acid metabolism and nucleotide salvage pathways. There are minimal differences between the core non-essential reactions at the subsystem level between ΔPDH and ΔPPC (Fig. 4b). The main difference of note is the presence of reactions involved in amino acid metabolism in the ΔASPTA, but not the ΔPDH or ΔPPC, solutions. This could be of relevance given the amino acid-rich hematophagic tsetse diet. The order in which key pseudogenisations occurred can therefore be estimated, using the resulting evolutionary trajectories as a guide. The gene encoding PPC could have been lost early by *S. glossinidius* in the sequence of pseudogenisations with minimal impact on its fitness.

A prediction of the evolutionary future of *S. glossinidius* as a symbiont

S. glossinidius is a secondary symbiont. Both bacterium and insect can survive independently of one another, and the former is likely a more recent acquisition [25]. It is however unclear how recently *S. glossinidius* was captured, or, given the pseudogenisations already present, how much more streamlined its genome can become. The algorithm was therefore applied to *i*LF517 in a blood medium with the aim

of evaluating potential future evolutionary trajectories within the bounds of its relationship with host and primary symbiont (Table 1, Scenario iii). There are a spread of biomass outputs found at the end of the simulation (Fig. 6a), as observed when *i*RH830 was evolved in blood. The smallest solutions contain approximately 300 reactions. After 3000 generations the *i*LF517 model retained 51 ± 1.1% of the starting 606 reactions compared to *i*RH830 which was reduced to 27 ± 1.8% of the 1247 starting reactions. The *S. glossinidius* genome can therefore reduce its potential coding capacity for metabolic genes to approximately half the size that it is currently.

The evolved solutions were then compared to the evolved *i*RH830 models to assess their similarity. *i*LF517 converges on a minimum after approximately 1000 generations, compared to the 2500 generations taken by the evolved *i*RH830 model to find the minimum number of reactions (Fig. S2). The greater standard deviation of the *i*RH830 solutions compared to *i*LF517 is likely due to the larger starting point of the free-living model. The evolved *i*RH830 and *i*LF517 solutions ultimately converge at a similar point. To investigate this further, ten evolved models for both *i*RH830 and *i*LF517 were analysed. All exchange reactions and those that carried zero flux were removed from further analysis. Full evolved models with fluxes can be found in Supplementary Data 3. Of the 383 unique reactions that carry flux in the evolved *i*RH830 models, 289 (75.5%) are found in all ten. For the *i*LF517 solutions, 301 of the 316 (95.3%) unique reactions that carry flux are found in all ten. This suggests that the smaller *S. glossinidius* model has fewer viable trajectories compared to the larger *S. praecaptivus* model. Of the 441 unique reactions across the 20 evolved models, 225 (51%) were found in all of the *i*RH830 and *i*LF517 evolved solutions; they are core across the two species. The biomass outputs for the evolved *i*RH830 and *i*LF517 solutions range from 0.064 to 0.281 (gr DW (mmol glucose)$^{-1}$ hr^{-1}), and 0.075 to 0.331 (gr DW (mmol glucose)$^{-1}$ hr^{-1}), respectively (Fig. 6b). Given the differences between the solutions from the two simulations, and the lower proportion of core conserved reactions, it is possible that *S. praecaptivus* may not be the free-living species of *Sodalis* most closely related to *S. glossinidius*, and that there may be others yet to be discovered, or may now be extinct or unrecognisable from the *S. glossinidius* progenitor.

Fig. 6. Evolution of *i*LF517. (a) Evolution of *i*LF517 in a blood medium. MOEA was run for 3000 generations (blue to green). (b) Biomass output (gr DW (mmol glucose)$^{-1}$ hr^{-1}) and the number of reactions carrying flux in evolved *i*RH830 (blue triangle) and evolved *i*RH830 (yellow circle) models. The evolved solutions produce comparable biomass outputs. Ten evolved solutions are given for each, some duplicates are present.

Discussion

Classical studies of microbial evolution, whilst useful, are ultimately limited by their inherent inability to replicate adaptations over large evolutionary time scales. Here, we present a computational approach by combining a MOEA with FBA, with the *Sodalis* system as a model for this. The *Sodalis* genus is ideal for the study of the evolution of symbiosis in this way, as within the genus are a free-living and a host restricted species, both well defined with complete genome sequences and existing protocols for culture.

Here, we present a model for *S. praecaptivus* metabolism, *i*RH830, accompanied by experimental verification, that has been used in subsequent *in silico* evolution experiments. Supplying the ancestral *i*RH830 with contrasting growth media demonstrates the effect that nutrient provisioning may have on evolutionary trajectories of symbiotic bacteria. Exposure to the famine medium reflects what might be expected in a nutrient-limited environment *in vivo*, in which evolutionary pressures result in the retention of pathways to synthesise key, essential metabolites. Here, this has shown to be particularly evident in the pathways retained for glycolysis/gluconeogenesis, the pentose phosphate pathway, and amino acid metabolism. This indicates that key pathways in central metabolism are being retained when the external environment is nutrient-limited. The evolved solutions therefore reflect what is observed in symbiotic bacteria; the retention or loss of pathways can be used to inform about the microenvironment it resides within. Some reactions in these pathways are also likely to be retained as they produce essential components of the biomass reaction. It is expected that the biomass reaction for symbiotic bacteria will change over evolutionary time, and therefore this work is limited by maintaining a consistent biomass reaction throughout the simulation. Incorporating a biologically accurate, variable biomass reaction, for example one which will at certain time points lose components that the host can synthesise, would be an interesting progression to this work.

It has been shown in the simulations presented here that the evolved famine solutions contain a much greater number of core non-essential reactions that are present in a small percentage of the solutions. This suggests a lack of flexibility in the evolved network; either the reaction is found repeatedly, or not at all. This is not observed in the solutions provided with complex media (blood or sap), where a greater degree of flexibility is demonstrated by more reactions being included repeatedly across the evolutionary space. This implies that, *in vivo*, there are many possible trajectories for an early symbiont if there are sufficient nutrients in the microenvironment.

The work here demonstrates the power of evolutionary algorithms in the study of symbiont evolution. A strength of this system is that removal of a reaction from the model is not irreversible; it is possible for a reaction to be added back into an individual at any point. This reduces the likelihood of repeatedly encountering the same evolutionary solutions as a result of losing the same key reaction, or reactions, early in the simulation. Although according to Muller's ratchet [82] the reduction of symbiont genomes should be irreversible [70], *S. praecaptivus* is as yet not obligately intracellular. A simple model to allow the (re-)acquisition of reactions is therefore appropriate for this system. Whilst there is no evidence currently for horizontal gene transfer (HGT) within species of the *Sodalis* genus, the NSGA-II algorithm is only intended to be used as a tool to explore the possible evolutionary space rather than as a biologically accurate model of genome reduction. Previous examples of evolving minimal metabolic networks do not allow for full exploration of the possible evolutionary space [50,56,57]. Decisions that are made at the start of process persist, which, whilst biologically relevant, does not allow the full complement of evolutionary routes to be examined. Expanding this algorithm to include the possibility of the model acquiring reactions that are not currently encoded by *S. praecaptivus*, therefore more closely reflecting HGT, may be an interesting avenue of future research.

This tool can produce biologically relevant simulations that accurately reflect the metabolic pressures that symbionts are exposed to. An example of this was demonstrated by the investigation of key knockouts in *S. glossinidius*. The symbiont has a pseudogenisation in *ppc*, a key gene in central metabolism [33]. It is not possible to deduce when this loss occurred from the genome annotation alone. Reactions are related to genes in multiple ways through the gene-protein-reaction relations which may or may not be 1:1; here we evolve reactions to focus on the phenotypic effects. As the MOEA enables a flexible search methodology this will cause minimal difference in outcomes to a gene-centred approach. By removing the PPC reaction from *S. praecaptivus* at the start of the simulation, the resulting trajectories can be analysed and compared to WT. The loss of PPC appeared to have minimal effect on the evolved populations compared to WT, in contrast to what was observed when the ASPTA reaction was removed at the start (Fig. 3). This would indicate that, *in vivo*, the loss of the gene encoding the ASPTA reaction would have a greater impact on a bacterial symbiont if it was lost early in the relationship in comparison to the lower burden that the loss of the genes encoding PDH or PPC would have. This result can then be used to infer the possible sequence of gene loss in the tsetse-*S. glossinidius* symbiosis. *S. glossinidius* has lost the *ppc* gene, whereas it has retained the genes encoding the PDH

(*SG0467-9*) and ASPTA (*SG1006*) reactions [25,33]. As the profile of ΔPPC evolution is similar to that of WT, it could be suggested that the *ppc* gene could have been lost early in evolutionary time without heavily bottlenecking *S. glossinidius* evolution subsequently. The gene encoding the ASPTA reaction may have been retained by *S. glossinidius* because of the detrimental impact that its loss may have caused. This is therefore a useful tool for making general predictions about when key pseudogenisations in insect-bacterial symbioses may have occurred.

It has been shown here that is it possible for *S. glossinidius* to reduce its metabolic network to approximately half of the size that it is currently. This provides support for the published hypothesis that *S. glossinidius* is a recent acquisition by the tsetse [25]. The number of reactions remains slightly higher in evolved *i*RH830 models compared to evolved *i*LF517 solutions, possibly due to difficulties in finding the minima from a larger starting point. *i*RH830 can however be reduced down to look phenotypically similar to *i*LF517 at the level of biomass output, but with differences at the individual reaction level (Supplementary Data 3). These results suggest therefore that the route that *S. glossinidius* has taken within the tsetse is perhaps just one of several possible routes. The differences also indicate that *S. praecaptivus* may not be the ancestor that initiated the tsetse-*S. glossinidius* symbiosis. The unusual ability of *S. praecaptivus* to metabolise xylitol may be related to the frequent presence of the *Sodalis* genus as a symbiont amongst sap-feeding insects, by hinting that it may naturally subsist on this important plant-derived sugar.

A possible, if challenging, extension to this work could be to incorporate the influence of the host and other members of the microbiome on the evolution of *Sodalis*. We acknowledge that this modelling method does not account for changes in host fitness that may arise from the evolution of the symbiont. The host could, for example, constrain the population of symbiont when it increases beyond a certain density, as demonstrated by the weevil *Sitophilus oryzae* that produce antimicrobial peptides to constrict the symbiont population size [83]. Alternatively, the host may benefit from reduced costs of symbiont maintenance [84], or an increased fitness or efficiency of the symbiont via the provision of metabolites. It may also suffer if the bacterial population becomes less fit. The latter is less likely to be an issue here, given that it is not yet known for certain whether *S. glossinidius* provides a benefit to the tsetse. This level of nuance is not captured by FBA as it focuses entirely on the fitness of an individual, with the only reference here to a population being the selection of the next generation. This tool is therefore most useful as a technique to examine broad changes that may occur during microbial evolution.

Previous uses of metabolic models to simulate evolution have focused on *E. coli* and *B. aphidicola* as a proof of concept [56,57]. The availability of both a genome sequence and a culturable organism for a free-living and symbiont of the same genus makes the *Sodalis* system a candidate model system to investigate the evolution of symbiosis. The work described here has augmented knowledge about the loss of key genes in *S. glossinidius* central metabolism. Combining FBA with a MOEA in this way could be used for any organism for which a well-annotated genome is available. It could be applied not only to the evolution of symbiosis but to the directed evolution of, for example, industrially relevant microorganisms or to the study of rapid genome evolution in pathogenic bacteria.

Materials and Methods

Bacterial strains, growth conditions and reagents

S. praecaptivus was obtained from DMSZ (Brunswick, Germany). Working stocks were established by incubating starter cultures on LB (Merck, Darmstadt, Germany) agar plates overnight at 37°C. A single colony was then sub-cultured on to a fresh LB plate and incubated overnight at 37°C. A single colony was selected with a sterile pipette tip and used for downstream experimentation as per Biolog, Inc (Hayward, CA, USA) manufacturer protocol. Briefly, the colony was vortexed in IF-0 media before a redox dye was added (Biolog). Phenotypic microplates were used to screen for the ability of *S. praecaptivus* to grow on a range of carbon sources, using PM1 and PM2A microplates (Biolog). A 100 μL bacterial suspension in the relevant media was added per well. Optical density was measured at 590 nm and 730 nm in a microplate reader (Epoch, BioTek, Winooski, VT, USA), and incubated with double orbital shaking at 37°C for 24 hours.

Discrepancies between *in silico* and *in vitro* Biolog results were re-examined by establishing individual cultures of *S. praecaptivus* in M9 salts in 96-well microplates, and supplemented with the metabolite of interest at a range of concentrations from 25 mM to 50 μM. Cultures were incubated in a microplate reader with double orbital shaking at 37°C for 36 hours.

Construction of the *S. praecaptivus* metabolic network

The annotated *S. praecaptivus* genome sequence, CP006569.1, was downloaded from NCBI in GenBank format. Genes in *S. praecaptivus* with the same annotation as genes in the *E. coli* str. K-12 substr. MG1655 genome (ASM584v2) were highlighted, and the reactions encoded by these genes extracted from the BiGG Models database [85]. These processes were automated using custom scripts written in Python.

FBA models of *S. glossinidius* (*i*LF517 [33]) and *E. coli* (*i*JO1366 [54,55], *i*JR904 [53], *i*AF1260 [52]) were then used to aid the identification of missing reactions. The reactions and corresponding gene assignments in these published models were compared to the draft *S. praecaptivus* model. These gene assignments were then used to guide translated nucleotide and protein BLAST searches of the *S. praecaptivus* genome. KEGG [86,87] and EcoCyc [88] databases were used to confirm the identity of the *E. coli* genes encoding each reaction. *S. glossinidius* gene assignments were taken from *i*LF517 [33]. These orthologues in *E. coli* and *S. glossinidius*, with sequences taken from UniProt [89], were used as BLAST search queries.

KEGG, BiGG Models, and MetaCyc [90] were used to assign reaction stoichiometry. Candidate pseudogenes were aligned with known functional orthologues using ClustalX 2.1 [91]. Those with sequences missing or mutations in key residues were not included in the model. FBA and literature searches were used to identify and fill gaps in metabolic pathways appropriately [92]. The xylitol pathway components in *Morganella morganii* subsp. *morganii* were identified using KEGG, with candidate protein sequences extracted from UniProt and used in a protein BLAST search against *S. praecaptivus*. KEGG was also used to identify known GalNAc degradation pathways.

Flux balance analysis

FBA solutions were generated using the GNU linear programming kit (GLPK) integrated with custom software in Java. Oxygen uptake was constrained to 20 mmol gr DW^{-1} hr^{-1}, comparable to other models of free-living Gram negative bacteria. The uptake of ammonia, water, phosphate, sulphate, potassium, sodium, calcium, carbon dioxide, protons and essential transition metals was unconstrained for all media conditions. Cofactor constraints were implemented by introducing these metabolites to the biomass functions at small fluxes (1 x 10^{-5} mmol gr DW^{-1} hr^{-1}) [7]. *i*RH830 was supplied with either 6 mmol gr DW^{-1} hr^{-1} GlcNAc and 1 mmol gr DW^{-1} hr^{-1} thiamine ("famine"), a

tsetse-specific medium ("blood", Table S3), or a sap-inspired medium (from [92], "sap"). Full recipes are provided in Supplementary Data 1. The phenotype was considered viable if the biomass production rate was greater than 1 x 10^{-4} gr DW (mmol glucose)$^{-1}$ hr^{-1}. Futile cycles, closed loops of a number of reactions, were detected by the presence of unsustainably large fluxes. Futile cycles often occur when several reversible reactions are present in which the product of one becomes the substrate of another. These reactions were examined individually, and solved by adjusting the reversability with guidance from EcoCyc and BiGG Models.

To investigate the concordance between the *in vitro* screen and the *in silico* outputs, *i*RH830 was, where possible, supplemented with the carbon sources analysed at an exogenous concentration of 6 mmol gr DW^{-1} hr^{-1}. A qualitative presence/absence of a positive biomass output was noted. Full description of the model is provided in Supplementary Data 1.

Robustness analysis

Robustness analysis of the *i*RH830 network was executed using COBRApy [94] to conduct single reaction deletions. *i*RH830 was supplied with either famine or blood media under aerated conditions. The flux through reactions was set to zero individually and the resulting effect on biomass output measured. Reactions were categorised as essential if the resulting biomass output was less than 1 x 10^{-3} gr DW (mmol glucose)$^{-1}$ hr^{-1}.

Implementation of multi-objective evolutionary algorithm

A MOEA was used to explore possible evolutionary trajectories in the *Sodalis* genus. An overview of the process is provided in Fig. 7. The non-dominated sorted genetic algorithm (NSGA-II) [95] from the Distributed Evolutionary Algorithms in Python (DEAP) [96] package was used in combination with the COBRApy package [94] for FBA evaluation. Equal weight was placed on reducing the number of reactions used in the model whilst maximising the biomass output. The Python code for running the algorithm with an example data set is available on GitHub (https://github.com/St659/SodalisFBAEvolution). The full datasets generated by the simulations are available on the York Research Database.

Fig. 7. Process of the MOEA. A starting population of individuals is initialised, and the fitness calculated by solving the FBA model to calculate biomass output and the number of active reactions. For each generation the population is allowed to mutate and then the fitness of each individual is evaluated from the biomass output and the sum of the active reactions. A new population is then selected using nondominated sorting, generating a Pareto front of biomass output to active reactions. The process of mutation and selection is repeated for 3000 generations resulting in a final population. Green boxes represent the start and final populations, pink boxes represent the iterative process of mutation and selection.

Population initiation

Prior to starting an evolutionary run, reactions essential to growth were identified using a single reaction knockout. Essential reactions were defined as those producing a biomass output of less than 1×10^{-3} gr DW $(\text{mmol glucose})^{-1}\ \text{hr}^{-1}$. Reactions that were identified as essential were not included in the subsequent mutation strategy, therefore reducing the solution space and computational time taken to run the MOEA. The essential reactions were added back to the evolved populations for downstream analysis.

At the start of the algorithm an initial population of 100 genotype copies was created, with all non-essential reactions being active (Fig. 7). Each genotype consisted of a binary number, where a 1 or 0 corresponded to the reaction being active or inactive, respectively. This is a proxy for gene loss, where a one-to-one gene-protein reaction mapping is assumed. Reactions, rather than genes, were used to reduce the potential search space whilst maintaining the key phenotypic effect. All post-evolution analysis focused on the reactions lost or retained.

Mutation

Mutation was performed on each genotype by flipping the value of each reaction with a probability of 0.005 (Fig. 7). The fitness of each individual is evaluated by solving the FBA model to calculate both its biomass output and the sum of number of active reactions.

Fitness evaluation and selection

The population was first evaluated for non-dominated individuals. This gave a population of individuals that has the highest biomass output for their current number

of active reactions (Fig. 7). From the non-dominated population, the Euclidean distance between each individual was calculated. A greater priority was given to selecting individuals with a larger Euclidean distance. This prevented the clustering of similar potential solutions, thereby reducing the likelihood of becoming trapped in sub-optimal local minima within the search space. The resulting population maximised the convergence on the highest biomass output, lowest number of reactions, and the distribution of those solutions. There will be a set of solutions whereby the number of reactions cannot be minimised further without also reducing the corresponding biomass output. This set of solutions is known as a Pareto front. The algorithm was repeated for 3000 generations to produce genotypes that converged. This indicated that minimal new solutions were being found. The biomass output from the slim optimisation COBRApy function and the summation of the number of active reactions was used to evaluate the fitness.

MOEA variations

The MOEA was run under several conditions in order to investigate aspects of symbiont evolution. Full details are provided in Table 1. Scenario i investigated the trajectories taken when the *S. praecaptivus* model was provided with blood, sap, and famine growth media. In Scenario ii, gene knockouts were simulated by removing individual reactions from the *S. praecaptivus* model prior to commencing the evolution. The reactions chosen were ASPTA, PDH, and PPC. In Scenario iii, the MOEA was applied to a model of *S. glossinidius* metabolism, *i*LF517 [33]. Here, *i*LF517 was supplied with the blood medium for 3000 generations.

Analysis of evolved populations

For all conditions the algorithm was independently run ten times, giving a total of 1000 final solutions. All of the solutions were pooled together for analysis. To identify key reactions in the evolved populations, individuals were selected from each condition and the remaining non-essential reactions extracted. The subset of reactions that were present in every individual selected were designated as "core nonessentials", and are referred to hereafter as such. When examining the similarity between evolved models, exchange reactions and reactions carrying zero flux were discounted.

Acknowledgements

The authors would like to thank Greg Hurst for insightful comments on this work, and the two anonymous reviewers for their constructive and thorough feedback.

Conflicts of interest

The authors declare no conflicts of interest.

Funding information

RJH was funded by the BBSRC White Rose DTP (BB/M011151/1) and ST by the Wellcome Trust CIDCATS programme (WT095024MA).

References

[1] Wilcox JL, Dunbar HE, Wolfinger RD, Moran NA. Consequences of reductive evolution for gene expression in an obligate endosymbiont. Molecular Microbiology. 2003 5;48(6):1491 – 1500.

[2] Dunbar HE, Wilson ACC, Ferguson NR, Moran NA. Aphid thermal tolerance is governed by a point mutation in bacterial symbionts. PLoS Biology. 2007 4;5(5):e96.

[3] Oliver KM, Russell JA, Moran NA, Hunter MS. Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. Proceedings of the National Academy of Sciences. 2003;100(4):1803–1807.

[4] Nakabachi A, Ueoka R, Oshima K, Teta R, Mangoni A, Gurgui M, et al. Defensive bacteriome symbiont with a drastically reduced genome. Current Biology. 2013 8;23(15):1478–1484.

[5] Aksoy S. Wigglesworthia gen. nov. and Wigglesworthia glossinidia sp. nov., taxa consisting of the mycetocyte-associated, primary endosymbionts of tsetse flies. International Journal of Systematic and Evolutionary Microbiology. 1995;45(4):848–851.

[6] Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H. Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. APS. Nature. 2000;407:81–86.

[7] Thomas GH, Zucker J, Macdonald SJ, Sorokin A, Goryanin I, Douglas AE. A fragile metabolic network adapted for cooperation in the symbiotic bacterium Buchnera aphidicola. BMC Systems Biology. 2009;3:24.

[8] Snyder AK, Rio RVM. Wigglesworthia morsitans folate (Vitamin B9) biosynthesis contributes to tsetse host fitness. Applied and Environmental Microbiology. 2015;81(16):5375 – 5386.

[9] Hrusa G, Farmer W, Weiss BL, Applebaum T, Roma JS, Szeto L, et al. TonBdependent heme iron acquisition in the tsetse fly symbiont Sodalis glossinidius. Applied and Environmental Microbiology. 2015;81(February):04166–14.

[10] Manzano-Marín A, Oceguera-Figueroa A, Latorre A, Jiménez-García LF, Moya A. Solving a bloody mess: B-vitamin independent metabolic convergence among gammaproteobacterial obligate endosymbionts from blood-feeding arthropods and the leech Haementeria officinalis. Genome Biology and Evolution. 2015;7(10):2871–2884.

[11] McCutcheon JP, Moran NA. Parallel genomic evolution and metabolic interdependence in an ancient symbiosis. Proceedings of the National Academy of Sciences. 2007 12;104(49):19392–19397.

[12] McCutcheon JP, McDonald BR, Moran NA. Convergent evolution of metabolic roles in bacterial co-symbionts of insects. Proceedings of the National Academy of Sciences. 2009 9;106(36):15394–9.

[13] McCutcheon JP, McDonald BR, Moran NA. Origin of an alternative genetic code in the extremely small and GC–rich genome of a bacterial symbiont. PLoS Genetics. 2009 7;5(7):e1000565.

[14] Wilson ACC, Ashton PD, Calevro F, Charles H, Colella S, Febvay G, et al. Genomic insight into the amino acid relations of the pea aphid, Acyrthosiphon pisum, with its symbiotic bacterium Buchnera aphidicola. Insect Molecular Biology. 2010 2;19:249–258.

[15] McCutcheon JP, von Dohlen CD. An interdependent metabolic patchwork in the nested symbiosis of mealybugs. Current Biology. 2011 8;21(16):1366– 1372.

[16] Rio RVM, Lefevre C, Heddi A, Aksoy S. Comparative genomics of insectsymbiotic bacteria: influence of host environment on microbial genome composition. Applied and Environmental Microbiology. 2003;69(11):6825–6832.

[17] Michalkova V, Benoit JB, Weiss BL, Attardo GM, Aksoy S. Vitamin B6 generated by obligate symbionts is critical for maintaining proline homeostasis and fecundity in tsetse flies. Applied and Environmental Microbiology. 2014;80(18):5844–5853.

[18] Baumann P, Baumann L, Lai CY, Rouhbakhsh D, Moran NA, Clark MA. Genetics, physiology, and evolutionary relationships of the genus Buchnera: Intracellular symbionts of aphids. Annual Review of Microbiology. 1995;49:55– 94.

[19] Akman Gündüz E, Douglas AE. Symbiotic bacteria enable insect to use a nutritionally inadequate diet. Proceedings of the Royal Society B. 2009;276:987– 991.

[20] Richards S, Gibbs RA, Gerardo NM, Moran N, Nakabachi A, Richards S, et al. Genome Sequence of the Pea Aphid Acyrthosiphon pisum. PLoS Biology. 2010 2;8(2):e1000313.

[21] Wu M, Sun LV, Vamathevan J, Riegler M, Deboy R, Brownlie JC, et al. Phylogenomics of the reproductive parasite Wolbachia pipientis wMel: A streamlined genome overrun by mobile genetic elements. PLoS Biology. 2004;2(3):e69.

[22] Pérez-Brocal V, Gil R, Ramos S, Lamelas A, Postigo M, Michelena JM, et al. A small microbial genome: The end of a long symbiotic relationship? Science. 2006 10;314(5797):312–313.

[23] Moran NA, Mccutcheon JP, Nakabachi A. Genomics and evolution of heritable bacterial symbionts. Annual Review of Genetics. 2008;42:165–190.

[24] Akman L, Yamashita A, Watanabe H, Oshima K, Shiba T, Hattori M, et al. Genome sequence of the endocellular obligate symbiont of tsetse flies, Wigglesworthia glossinidia. Nature Genetics. 2002;32(3):402–407.

[25] Dale C, Maudlin I. Sodalis gen. nov. and Sodalis glossindius sp. nov., a microaerophilic secondary endosymbiont of the tsetse fly Glossina morsitans morsitans. International Journal of Systematic Bacteriology. 1999;49:267–275.

[26] Edwards JS, Covert M, Palsson B. Metabolic modelling of microbes: the fluxbalance approach. Environmental Microbiology. 2002;4(3):133–140.

[27] Kauffman KJ, Prakash P, Edwards JS. Advances in flux balance analysis. Current Opinion in Biotechnology. 2003 10;14(5):491–496.

[28] Lewis NE, Nagarajan H, Palsson BO. Constraining the metabolic genotype–phenotype relationship using a phylogeny of in silico methods. Nature Reviews Microbiology. 2012 4;10(4):291–305.

[29] Orth JD, Thiele I, Palsson B. What is flux balance analysis? Nature Biotechnology. 2010;28(3):245–248.

[30] Macdonald SJ, Lin GG, Russell CW, Thomas GH, Douglas AE. The central role of the host cell in symbiotic nitrogen metabolism. Proceedings of the Royal Society B. 2012 8;279:2965–2973.

[31] Macdonald SJ, Thomas GH, Douglas AE. Genetic and metabolic determinants of nutritional phenotype in an insect-bacterial symbiosis. Molecular Ecology. 2011 5;20(10):2073–2084.

[32] Belda E, Silva FJ, Peretó J, Moya A. Metabolic networks of Sodalis glossinidius: A systems biology approach to reductive evolution. PLoS ONE. 2012;7(1):e30652.

[33] Hall RJ, Flanagan LA, Bottery MJ, Springthorpe V, Thorpe S, Darby AC, et al. A tale of three species: Adaptation of Sodalis glossinidius to tsetse biology, Wigglesworthia metabolism, and host diet. mBio. 2019 1;10(1):02106– 18.

[34] Ankrah NYD, Luan J, Douglas AE. Cooperative metabolism in a three-partner insect-bacterial symbiosis revealed by metabolic modeling. Journal of Bacteriology. 2017 8;199(15):00872–16.

[35] González-Domenech C, Belda E, Patiño-Navarrete R, Moya A, Peretó J, Latorre A. Metabolic stasis in an ancient symbiosis: Genome-scale metabolic networks from two Blattabacterium cuenoti strains, primary endosymbionts of cockroaches. BMC Microbiology. 2012 1;12(Suppl 1):S5.

[36] Sørensen MES, Cameron DD, Brockhurst MA, Wood AJ. Metabolic constraints for a novel symbiosis. Royal Society Open Science. 2016;3(3):150708.

[37] Shastri AA, Morgan JA. Flux balance analysis of photoautotrophic metabolism. Biotechnology Progress. 2005 12;21(6):1617–1626.

[38] Nogales J, Gudmundsson S, Knight EM, Palsson BO, Thiele I. Detailing the optimality of photosynthesis in cyanobacteria through systems biology analysis. Proceedings of the National Academy of Sciences. 2012 2;109(7):2678– 83.

[39] Knoop H, Zilliges Y, Lockau W, Steuer R. The metabolic network of Synechocystis sp. PCC 6803: Systemic properties of autotrophic growth. Plant Physiology. 2010 9;154(1):410–422.

763 [40] Knoop H, Gründel M, Zilliges Y, Lehmann R, Hoffmann S, Lockau W, et al. Flux
764 balance analysis of cyanobacterial metabolism: The netabolic network of
765 Synechocystis sp. PCC 6803. PLoS Computational Biology. 2013 6;9(6):e1003081.
766 pcbi.1003081.

767 [41] Toh H, Weiss BL, Perkin SAH, Yamashita A, Oshima K, Hattori M, et al. Massive
768 genome erosion and functional adaptations provide insights into the symbiotic
769 lifestyle of Sodalis glossinidius in the tsetse host. Genome Research.
770 2006;16(2):149–156.

771 [42] Goodhead I, Blow F, Brownridge P, Hughes M, Kenny J, Krishna R, et al.
772 Large scale and significant expression from pseudogenes in Sodalis glossinidius - a
773 facultative bacterial endosymbiont. bioRxiv. 2018 4;p. 124388.

774 [43] Covert MW, Palsson BO. Transcriptional regulation in constraints-based metabolic
775 models of Escherichia coli. The Journal of Biological Chemistry. 2002
776 8;277(31):28058–28064.

777 [44] Lee JM, Gianchandani EP, Papin JA. Flux balance analysis in the era of
778 metabolomics. Briefings in Bioinformatics. 2006;7(2):140–150.

779 [45] Covert MW, Palsson BO. Constraints-based models: regulation of gene expression
780 reduces the steady-state solution space. Journal of Theoretical Biology.
781 2003;221:309–325.

782 [46] Covert MW, Schilling CH, Palsson B. Regulation of gene expression in flux balance
783 models of metabolism. Journal of Theoretical Biology. 2001 11;213(1):73– 88.

784 [47] Mahadevan R, Edwards JS, Doyle FJ. Dynamic flux balance analysis of diauxic growth
785 in Escherichia coli. Biophysical Journal. 2002 9;83(3):1331– 1340.

786 [48] Harcombe WR, Delaney NF, Leiby N, Klitgord N, Marx CJ. The ability of flux balance
787 analysis to predict evolution of central metabolism scales with the initial distance
788 to the optimum. PLoS Computational Biology. 2013 6;9(6):e1003091.

789 [49] Pfeiffer T, Soyer OS, Bonhoeffer S. The evolution of connectivity in metabolic
790 networks. PLoS Biology. 2005;3(7):e228.

791 [50] Ponce-de Leon M, Tamarit D, Calle-Espinosa J, Mori M, Latorre A, Montero F, et al.
792 Determinism and contingency shape metabolic complementation in an
793 endosymbiotic consortium. Frontiers in Microbiology. 2017;8:2290.

794 [51] Edwards JS, Palsson BO. Metabolic flux balance analysis and the in silico analysis of
795 Escherichia coli K-12 gene deletions. BMC Bioinformatics. 2000 7;1(1):1.

796 [52] Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, et al. A genome-
797 scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for
798 1260 ORFs and thermodynamic information. Molecular Systems Biology. 2007
799 1;3(1):121.

800 [53] Reed JL, Vo TD, Schilling CH, Palsson BO. An expanded genome-scale model of
801 Escherichia coli K-12 (iJR904 GSM/GPR). Genome Biology. 2003;4(9):R54.

802 [54] Orth JD, Conrad TM, Na J, Lerman JA, Nam H, Feist AM, et al. A comprehensive
803 genome-scale reconstruction of Escherichia coli metabolism–2011. Molecular
804 Systems Biology. 2011 10;7(1):535.

805 [55] Orth JD, Palsson BO. Gap-filling analysis of the iJO1366 Escherichia coli metabolic
806 network reconstruction for discovery of metabolic functions. BMC Systems Biology.
807 2012 5;6(1):30.

808 [56] Pál C, Papp B, Lercher MJ, Csermely P, Oliver SG, Hurst LD. Chance and necessity in
809 the evolution of minimal metabolic networks. Nature. 2006 3;440(7084):667–670.

810 [57] Yizhak K, Tuller T, Papp B, Ruppin E. Metabolic modeling of endosymbiont genome
811 reduction on a temporal scale. Molecular Systems Biology. 2011 1;7(1):479.
812 McNally CP, Borenstein E. Metabolic model-based analysis of the emergence of
813 bacterial cross-feeding through extensive gene loss. bioRxiv. 2017;p.
814 10.1101/180208.

815 [58] McNally CP, Borenstein E. Metabolic model-based analysis of the emergence of
816 bacterial cross-feeding through extensive gene loss. bioRxiv. 2017.
817 p.10.1101/18020.

818 [59] Hindré T, Knibbe C, Beslon G, Schneider D. New insights into bacterial adaptation
819 through in vivo and in silico experimental evolution. Nature Reviews Microbiology.
820 2012 5;10(5):352–365.

821 [60] Batut B, Parsons DP, Fischer S, Beslon G, Knibbe C. In silico experimental evolution:
822 a tool to test evolutionary scenarios. BMC Bioinformatics. 2013 10;14(Suppl
823 15):S11.

824 [61] François P, Hakim V. Design of genetic networks with specified functions by
825 evolution in silico. Proceedings of the National Academy of Sciences.
826 2004;101(2):580–585.

827 [62] Budinich M, Bourdon J, Larhlimi A, Eveillard D. A multi-objective constraintbased
828 approach for modeling genome-scale microbial ecosystems. PLoS ONE. 2017
829 2;12(2).

830 [63] Wang L, Maranas CD. MinGenome: An n silico top-down approach for the synthesis
831 of minimized genomes. ACS Synthetic Biology. 2018;7:462–473.

832 [64] Fong SS, Burgard AP, Herring CD, Knight EM, Blattner FR, Maranas CD, et al. In silico
833 design and adaptive evolution of Escherichia coli for production of lactic acid.
834 Biotechnology and Bioengineering. 2005 9;91(5):643–648.

835 [65] Garcia S, Trinh C. Comparison of multi-objective evolutionary algorithms to solve
836 the modular cell design problem for novel biocatalysis. bioRxiv. 2019;p.
837 10.1101/616078.

838 [66] Machado D, Herrgård MJ. Co-evolution of strain design methods based on flux
839 balance and elementary mode analysis. Metabolic Engineering Communications.
840 2015 12;2:85–92.

841 [67] Clayton AL, Oakeson KF, Gutin M, Pontes A, Dunn DM, von Niederhausern AC, et al.
842 A novel human-infection-derived bacterium provides insights into the evolutionary
843 origins of mutualistic insect–bacterial symbioses. PLoS Genetics.
844 2012;8(11):e1002990.

845 [68] Chari A, Oakeson KF, Enomoto S, Jackson DG, Fisher MA, Dale C. Phenotypic
846 characterisation of Sodalis praecaptivus sp. nov., a close non-insect associated
847 member of the Sodalis-allied lineage of insect endosymbionts. International Journal
848 of Systematic and Evolutionary Microbiology. 2015;65:1400–1405.

849 [69] WHO. Control and surveillance of African trypanosomiasis. WHO Technical Report
850 Series. 1998;881:1-113.

851 [70] Moran NA. Accelerated evolution and Muller's rachet in endosymbiotic bacteria.
852 Proceedings of the National Academy of Sciences. 1996;93:2873–2878.

853 [71] Novakova E, Hypsa V. A new Sodalis lineage from bloodsucking fly Craterina melbae
854 (Diptera, Hippoboscoidea) originated independently of the tsetse flies symbiont
855 Sodalis glossinidius. FEMS Microbiology Letters. 2007 4;269(1):131– 135.

[72] Chrudimský T, Husník F, Nováková E, Hypša V. Candidatus Sodalis melophagi sp. nov.: Phylogenetically independent comparative model to the tsetse fly symbiont Sodalis glossinidius. PLoS ONE. 2012 7;7(7):e40354.

[73] Boyd BM, Allen JM, Koga R, Fukatsu T, Sweet AD, Johnson KP, et al. Two bacterial genera, Sodalis and Rickettsia, associated with the seal louse Proechinophthirus fluctus (Phthiraptera: Anoplura).

[74] Oakeson KF, Gil R, Clayton AL, Dunn DM, von Niederhausern AC, Hamil C, et al. Genome Degeneration and Adaptation in a Nascent Stage of Symbiosis. Genome Biology and Evolution. 2014;6(1):76–93.

[75] Szklarzewicz T, Kalandyk-Kołodziejczyk M, Michalik K, Jankowska W, Michalik A. Symbiotic microorganisms in Puto superbus (Leonardi, 1907) (Insecta, Hemiptera, Coccomorpha: Putoidae). Protoplasma. 2018 1;255(1):129–138.

[76] Koga R, Bennett GM, Cryan JR, Moran NA. Evolutionary replacement of obligate symbionts in an ancient and diverse insect lineage. Environmental Microbiology. 2013 7;15(7):2073–2081.

[77] Koga R, Moran NA. Swapping symbionts in spittlebugs: evolutionary replacement of a reduced genome symbiont. The ISME Journal. 2014 1;8:1237.

[78] Michalik A, Jankowska W, Kot M, Gołas A, Szklarzewicz T. Symbiosis in the green leafhopper, Cicadella viridis (Hemiptera, Cicadellidae). Association in statu nascendi? Arthropod Structure & Development. 2014 11;43(6):579– 587.

[79] Kaiwa N, Hosokawa T, Kikuchi Y, Nikoh N, Meng XY, Kimura N, et al. Primary gut symbiont and secondary, Sodalis-allied symbiont of the Scutellerid stinkbug Cantao ocellatus. Applied and Environmental Microbiology. 2010 6;76(11):3486–94.

[80] Hosokawa T, Kaiwa N, Matsuura Y, Kikuchi Y, Fukatsu T. Infection prevalence of Sodalis symbionts among stinkbugs. Zoological Letters. 2015 12;1(1):5.

[81] Snyder AK, Deberry JW, Runyen-Janecky L, Rio RVM. Nutrient provisioning facilitates homeostasis between tsetse fly (Diptera: Glossinidae) symbionts. Proceedings of the Royal Society B: Biological Sciences. 2010;277(1692):2389–2397.

[82] Muller HJ. The relation of recombination to mutational advance. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis. 1964 5;1(1):2–9.

[83] Login FH, Balmand S, Vallier A, Vincent-Monégat C, Vigneron A, WeissGayet M, et al. Antimicrobial peptides keep insect endosymbionts under control. Science. 2011 10;334(6054):362 – 365.

[84] Ankrah NYD, Chouaia B, Douglas AE. The cost of metabolic interactions in symbioses between insects and bacteria with reduced genomes. mBio. 2018 9:e01433-18.

[85] King ZA, Lu J, Dräger A, Miller P, Federowicz S, Lerman JA, et al. BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. Nucleic Acids Research. 2016 1;44(D1):D515–D522.

[86] Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research. 2000 1;28(1):27–30.

[87] Kanehisa M, Sato Y, Furumichi M, Morishima K, Tanabe M. New approach for understanding genome variations in KEGG. Nucleic Acids Research. 2019 1;47(D1):D590–D595.

[88] Keseler IM, Mackie A, Santos-Zavaleta A, Billington R, Bonavides-Martínez C, Caspi R, et al. The EcoCyc database: reflecting new knowledge about Escherichia coli K-12. Nucleic Acids Research. 2017 1;45(D1):D543–D550.

[89] Bateman A, Martin MJ, Orchard S, Magrane M, Alpi E, Bely B, et al. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Research. 2019 1;47(D1):D506–D515.

[90] Caspi R, Foerster H, Fulcher CA, Kaipa P, Krummenacker M, Latendresse M, et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Research. 2007 12;36(Database):D623–D631.

[91] Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007 11;23(21):2947–2948.

[92] Thiele I, Palsson BO. A protocol for generating a high-quality genome-scale metabolic reconstruction. Nature Protocols. 2010;5(1):93–121.

[93] Krishnan HB, Natarajan SS, Bennett JO, Sicher RC. Protein and metabolite composition of xylem sap from field-grown soybeans (Glycine max). Planta. 2011;233:921–931.

[94] Ebrahim A, Lerman JA, Palsson BO, Hyduke DR. COBRApy: COnstraintsBased Reconstruction and Analysis for Python. BMC Systems Biology. 2013 8;7(1):74.

[95] Deb K, Pratap A, Agarwal S, Meyarivan T. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation. 2002;6(2):182–197.

[96] De Rainville FM, Fortin FA, Gardner MA, Parizeau M, Gagné C. DEAP: A Python framework for evolutionary algorithms. In: Proceedings of the 14th International Conference on Genetic and Evolutionary Computation; 2012. p. 85–92.