

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

Rebecca J. Hall^{1,2†}, Stephen Thorpe^{1,3†}, Gavin H. Thomas¹ and A. Jamie Wood^{1,4,*}

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 genus *Sodalis*, *Sodalis 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 internalized. 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.

DATA SUMMARY

The Python code for running the algorithm with an example data set is available from GitHub (<https://github.com/St659/SodalisFBAEvolution>). The data generated by the simulations are available in the York Research Database (<https://pure.york.ac.uk/portal>).

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 pseudogenization in symbiotic bacteria.

Computational biology is well established as a key tool of scientific discovery, now that vast amounts of data are generated quickly and cheaply from advancements in sequencing

Received 22 December 2019; Accepted 27 April 2020; Published 16 June 2020

Author affiliations: ¹Department of Biology, University of York, York, YO10 5NG, UK; ²School of Life Sciences, University of Nottingham, Nottingham, NG7 2TQ, UK; ³Department of Chemistry, University of Oxford, Oxford, OX1 3TA, UK; ⁴Department of Mathematics, University of York, York, YO10 5DD, UK.

*Correspondence: A. Jamie Wood, jamie.wood@york.ac.uk

Keywords: evolution; evolutionary algorithm; flux balance analysis; *Sodalis*; symbiosis.

Abbreviations: ASPTA, aspartate transaminase; dFBA, dynamic flux balance analysis; DW, dry weight; FBA, flux balance analysis; GalNAc, *N*-acetyl-D-galactosamine; MOEA, multi-objective evolutionary algorithm; PDH, pyruvate dehydrogenase; PPC, phosphoenolpyruvate carboxylase; TCA, tricarboxylic acid.

†These authors contributed equally to this work

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.

000378 © 2020 The Authors



This is an open-access article distributed under the terms of the Creative Commons Attribution NonCommercial License. This article was made open access via a Publish and Read agreement between the Microbiology Society and the corresponding author's institution.

technology [26, 27]. Genome-scale metabolic modelling of micro-organisms 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], 'Candidatus 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 pseudogenization [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 pseudogenizations 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 Pál *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, therefore, may not 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*

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 internalized 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.

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 (MOEAs) have been used in many disciplines for solving problems that have two or more conflicting objectives [62]. The use of MOEAs 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 optimization; symbiotic bacteria undergo genome reduction, whilst trying to maximize their individual growth.

A free-living organism within the genus *Sodalis* has been characterized 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*, iRH830. This model, and a previously presented model of *S. glossinidius* metabolism, iLF517 [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 *iRH830* 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 *iLF517*, 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 *iRH830* enabled us to ask in which order of the evolutionary sequence key pseudogenizations 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.

METHODS

Bacterial strains, growth conditions and reagents

S. praecaptivus was obtained from the DSMZ (German Collection of Microorganisms and Cell Cultures). Working stocks were established by incubating starter cultures on LB agar (Merck) plates overnight at 37 °C. A single colony was then sub-cultured onto 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 the protocol of Biolog, the manufacturer. Briefly, the colony was vortexed in IF-0 medium 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 medium was added per well. Optical density was measured at 590 and 730 nm in a microplate reader (Epoch; BioTek), and incubated with double orbital shaking at 37 °C for 24 h. 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, with supplementation 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 h.

Construction of the *S. praecaptivus* metabolic network

The annotated *S. praecaptivus* genome sequence, CP006569.1, was downloaded from the National Center for Biotechnology Information in GenBank format. Genes in *S. praecaptivus* with the same annotation as genes in the *E. coli* strain K-12 substrain MG1655 genome (ASM584v2) were highlighted, and the reactions encoded by these genes extracted from the BiGG Models database [70]. These processes were automated using custom scripts written in Python.

FBA models of *S. glossinidius* (*iLF517* [33]) and *E. coli* (*iJO1366* [54, 55], *iJR904* [53], *iAF1260* [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 [71, 72] and EcoCyc [73] databases were used to confirm the identity of the *E. coli* genes encoding each reaction. *S. glossinidius* gene assignments were taken from *iLF517* [33]. These orthologues in *E. coli* and *S. glossinidius*, with sequences taken from UniProt [74], were used as BLAST search queries.

KEGG, BiGG Models and MetaCyc [75] were used to assign reaction stoichiometry. Candidate pseudogenes were aligned with known functional orthologues using ClustalX 2.1 [76]. 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 [77]. The xylitol pathway components in *Morganelia morgani* subsp. *morgani* 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 *N*-acetyl-D-galactosamine (GalNAc) degradation pathways.

FBA

FBA solutions were generated using the GNU linear programming kit (GLPK) integrated with custom software in Java. Oxygen uptake was constrained to 20 mmol (g DW)⁻¹ h⁻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 × 10⁻⁵ mmol (g DW)⁻¹ h⁻¹] [7]. *iRH830* was supplied with either 6 mmol GlcNAc (g DW)⁻¹ h⁻¹ and 1 mmol thiamine (g DW)⁻¹ h⁻¹ ('famine'), a tsetse-specific medium ('blood'; Table S1, available with the online version of this article) or a sap-inspired medium ([77]; 'sap'). Full recipes are provided in Supplementary Data 1. The phenotype was considered viable if the biomass production rate was greater than 1 × 10⁻⁴ g DW (mmol glucose)⁻¹ h⁻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 reversibility with guidance from EcoCyc and BiGG Models.

To investigate the concordance between the *in vitro* screen and the *in silico* outputs, *iRH830* was, where possible, supplemented with the carbon sources analysed at an exogenous concentration of 6 mmol (g DW)⁻¹ h⁻¹. A qualitative presence/absence of a positive biomass output was noted. A full description of the model is provided in Supplementary Data 1.

Robustness analysis

Robustness analysis of the *iRH830* network was executed using COBRApy [78] to conduct single reaction deletions. *iRH830* 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 categorized as essential

if the resulting biomass output was less than $1 \times 10^{-3} \text{ g DW (mmol glucose)}^{-1} \text{ h}^{-1}$.

Implementation of a MOEA

A MOEA was used to explore possible evolutionary trajectories in the genus *Sodalis*. An overview of the process is provided in Fig. 1. The Non-dominated Sorted Genetic

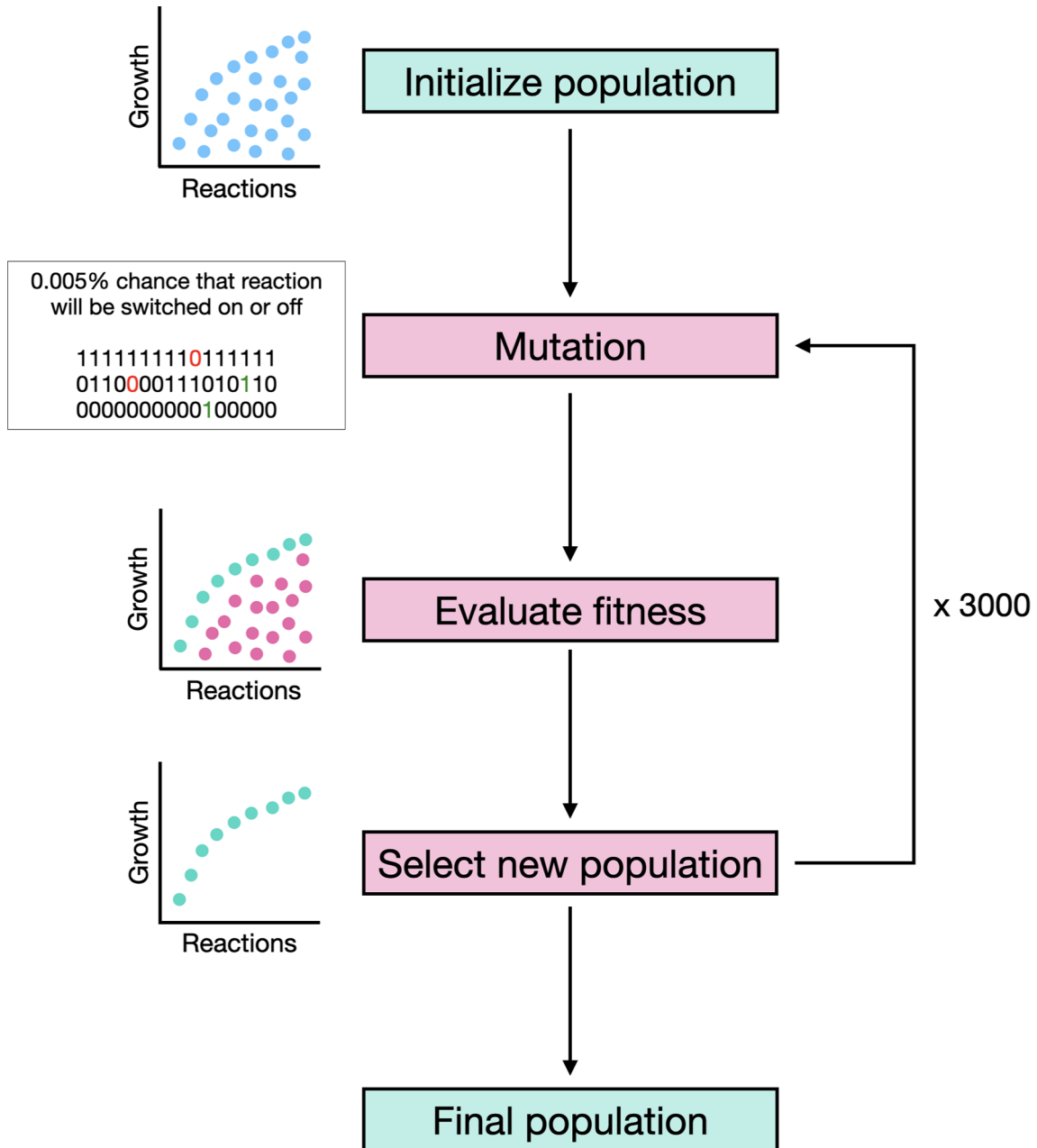


Fig. 1. Process of the MOEA. A starting population of individuals is initialized, and the fitness calculated by solving the FBA model to calculate biomass output and the number of active reactions. For each generation, the population is allowed to mutate and then the fitness of each individual is evaluated 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. Turquoise boxes represent the start and final populations; pink boxes represent the iterative process of mutation and selection.

Algorithm (NSGA-II) [79] from the Distributed Evolutionary Algorithms in Python (DEAP) [80] package was used in combination with the COBRAPy package [78] for FBA evaluation. Equal weight was placed on reducing the number of reactions used in the model, whilst maximizing the biomass output. The Python code for running the algorithm with an example data set is available from GitHub (<https://github.com/St659/SodalisFBAEvolution>). The full datasets generated by the simulations are available from the York Research Database (<https://pure.york.ac.uk/portal>).

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} g DW (mmol glucose) $^{-1}$ h $^{-1}$. 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. 1). Each genotype consisted of a binary number, where a one or zero 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. 1). 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 the current number of active reactions (Fig. 1). 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 maximized 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 minimized 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

Table 1. *In silico* evolution conditions

Conditions under which *i*RH830 and *i*LF517 were evolved, including WT or reaction knockouts and media type.

Scenario	Test	Model	Media
i	Effect of growth media	<i>i</i> RH830 (WT)	Blood, sap, famine
ii	Effect of gene loss	<i>i</i> RH830 (Δ ASPTA, Δ PDH, Δ PPC)	Blood
iii	Future of <i>S. glossinidius</i>	<i>i</i> RH830 (WT), <i>i</i> LF517 (WT)	Blood

genotypes that converged. This indicated that minimal new solutions were being found. The biomass output from the slim optimization 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 (aspartate transaminase), PDH (pyruvate dehydrogenase) and PPC (phosphoenolpyruvate carboxylase). 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.

RESULTS

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. 2). 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 (g DW) $^{-1}$ h $^{-1}$, reflecting the highly aerated

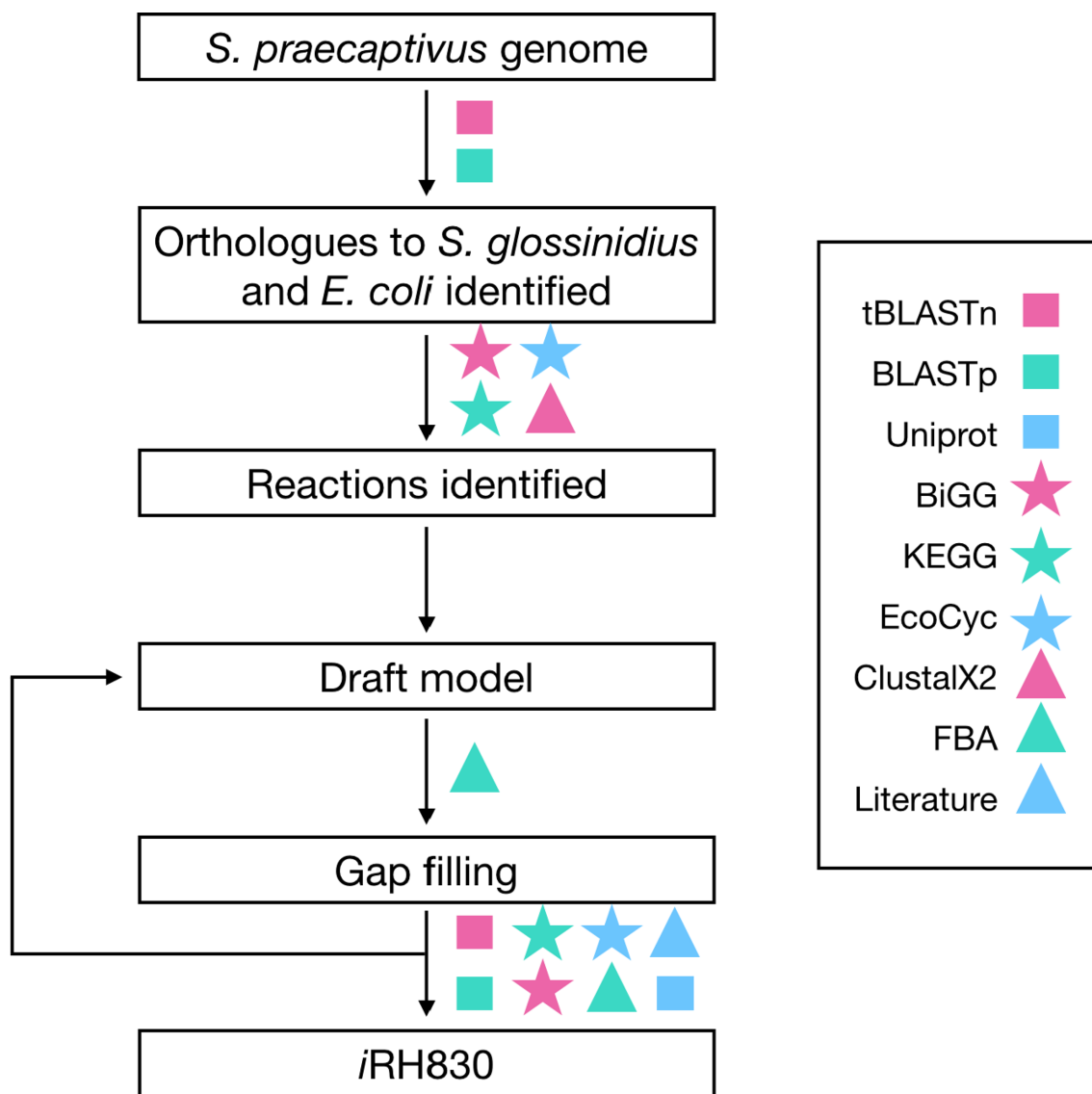


Fig. 2. The construction process for *iRH830*. The *S. praecaptivus* genome was mined for orthologues to metabolic genes in *E. coli* and *S. glossinidius* before compiling into a draft model. An iterative process of testing and gap filling was then performed, using information provided in various databases (see the key).

conditions the organism is grown in and to retain consistency with models of *E. coli* metabolism [53, 54].

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 S2). When these metabolites were tested *in silico* by the exogenous addition to *iRH830*, it was found that all but two mirrored the *in vitro* data: GalNAc and xylitol. This was then confirmed quantitatively in a 96-well microplate with xylitol or GalNAc supplemented

into M9 minimal medium (Fig. S1a, b). Neither models of *S. glossinidius* (*iLF517* [33]) or *E. coli* (*iJO1366* [54]) were able to produce a positive biomass output with xylitol or GalNAc as sole sources of carbon (Table S2).

Comparison of the *S. praecaptivus* genome to other known D-xylitol consumers, such as *M. morgani* subsp. *morgani*, 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 *M. morgani* subsp. *morgani* 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.

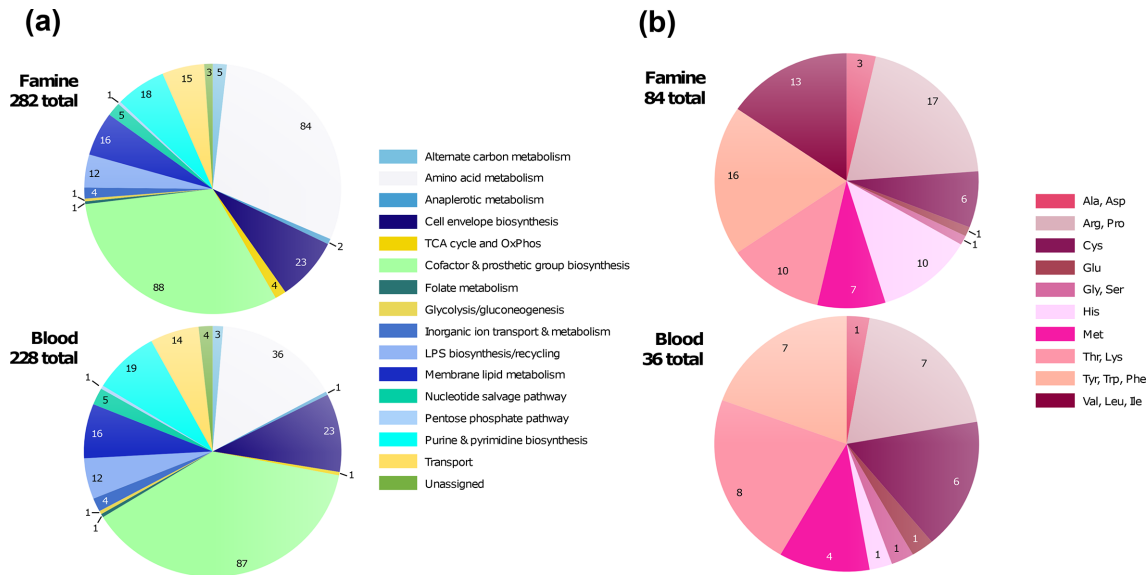


Fig. 3. Robustness analysis of *iRH830*. (a) Essential reactions in famine (top) and blood (bottom) media. Essential reactions are categorized by subsystem. (b) Essential reactions involved in amino acid metabolism in *iRH830* in famine (top) and blood (bottom) media.

Interestingly, the pathway is also complete in the reduced *S. glossinidius* genome (Table S3), suggesting that this is a conserved metabolic trait of the genus *Sodalis*. The cluster is not present in *E. coli*, with only weakly matching homologues found fragmented over the genome (Table S3). 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 *iRH830* network. *iRH830* 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 S1). 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 *iLF517* in blood as a comparison.

There are 282 essential reactions in *iRH830* when the medium (famine) is nutritionally limited, and 228 in the tsetse-specific blood medium (Fig. 3a). 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. 3b).

Media provisioning affects evolutionary trajectories

NSGA-II is a heuristic multi-objective optimization 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 optimization problem of trying to reduce the genome size to increase replication speed, while still retaining sufficient capacity to grow [81]. 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. 1. 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 *iRH830* and *iLF517* were evolved are detailed in Table 1. In scenario i, *iRH830* 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 *iRH830* prior to evolution to compare the trajectories that arise as a result of pseudogenizations, and to investigate if these were possible adaptations prior to symbiont establishment. The gene encoding PPC is thought to be pseudogenized in *S. glossinidius*, whereas the PDH and ASPTA reactions are predicted to be functional [33]. In scenario iii, the MOEA was applied to *iLF517* to investigate the possible future of *S. glossinidius* as a symbiont.

Species of the genus *Sodalis* have been found in insects that feed on a variety of contrasting diets, including blood

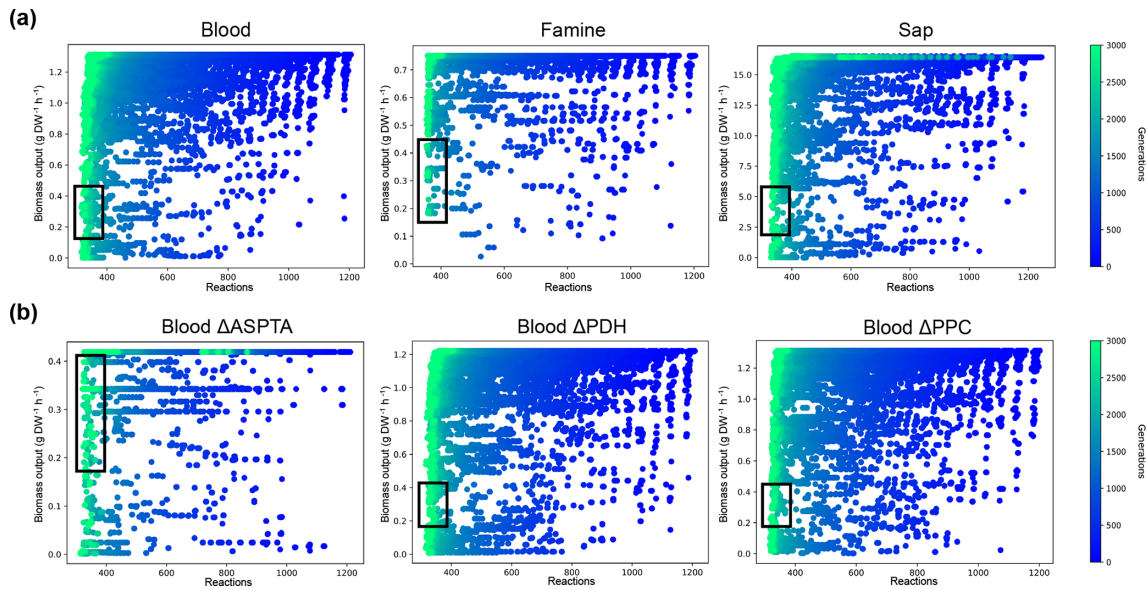


Fig. 4. *iRH830* evolved under different starting conditions. (a) Evolution of *iRH830* in a tsetse-specific blood medium (left), a nutritionally limited famine medium (centre) and a medium mimicking plant sap (right). (b) *iRH830* evolution in a blood medium with the reactions ASPTA (left), PDH (centre) 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.

(e.g. tsetse [25] and ticks [82–84]) and plant tissue (e.g. weevils [85]). To replicate *Sodalis* evolution in different environments, the MOEA was applied to *iRH830* that was supplied with a tsetse-specific blood medium, a 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 phytophagous insects [86–91]. 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. 4a). 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. 4a), 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. 4a). 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.

A number of individual solutions that were representative of the biomass output of *iLF517* [33] were then selected from each of these simulations (Fig. 4, black boxes). The raw, binary data were translated back into reaction names and these were 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-essential reactions found in all 1194 of the individuals examined when *iRH830* was supplied with blood: AGDC, ARGabc, ASNt2r, G6PDA, H2Ot, HlSt2r, 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 (5 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 (8 of 14) of these are secondary transporter reactions (Fig. 5a). This reflects what is observed in

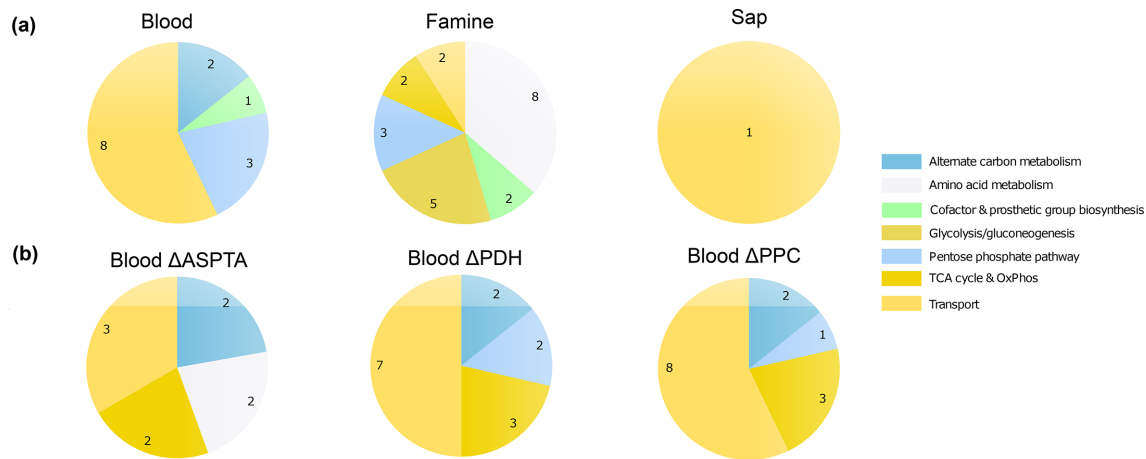


Fig. 5. Core non-essential reactions in evolved *iRH830* populations. (a) The proportion of core non-essential reactions per condition by subsystem when the ancestral *iRH830* is exposed to blood (left), famine (centre) or sap (right) media. (b) Core non-essential reactions in Δ ASPTA (left), Δ PDH (centre) and Δ PPC (right) *iRH830* models in a blood medium, grouped by subsystem.

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, 92]. 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.

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 pseudogenizing key genes early in evolutionary time has on the trajectory of a symbiont, the MOEA was run on *iRH830* with one of three reactions involved in the TCA (tricarboxylic acid) cycle removed at the start, with the resulting solutions compared to WT (Table 1, scenario ii). An assumption is made in these simulations that pseudogenes are non-functional. The first reaction selected was PPC (Fig. 6). The gene encoding this reaction, *ppc*, is pseudogenized 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 and ASPTA. These reactions are both encoded by genes predicted to be functional in *S. glossinidius* (Fig. 6). It was hypothesized 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.

When considering the population plots, there is minimal qualitative difference between Δ PDH and Δ PPC (Fig. 4b). Δ ASPTA, 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 was then examined and the number of core non-essential reactions in the evolved models analysed as described previously (Fig. 4b, black boxes). There are 1, 11 and 9 core non-essential reactions in the WT, Δ PDH and Δ PPC solutions, respectively, whereas there are 61 in Δ ASPTA. 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 Δ PDH and Δ PPC (Fig. 5b). The main difference of note is the presence of reactions involved in amino acid metabolism in the Δ ASPTA, but not the Δ PDH or Δ PPC solutions. This could be of relevance given the amino acid-rich haematophagous tsetse diet. The order in which key pseudogenizations 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 pseudogenizations with minimal impact on its fitness.

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 pseudogenizations already present, how much more streamlined its genome can become. The algorithm was, therefore, applied to *iLF517* 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 is a spread of biomass outputs found at the end of the simulation (Fig. 7a), as observed when *iRH830* was evolved in blood. The smallest solutions

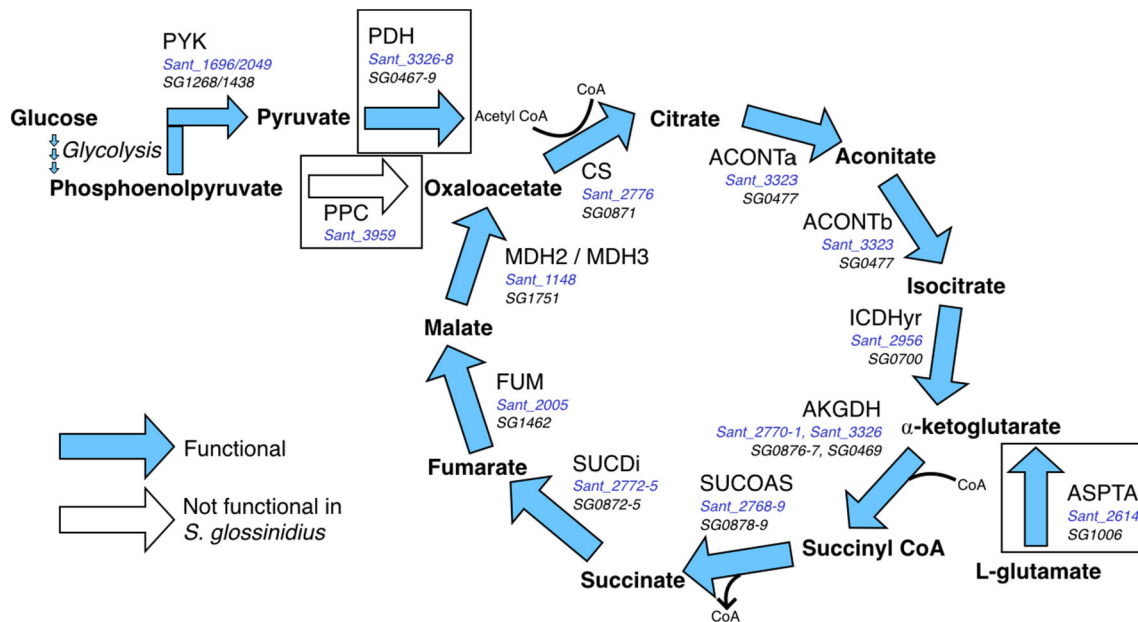


Fig. 6. TCA (tricarboxylic acid) cycle reactions examined by the MOEA. Three reactions were removed from *iRH830* to investigate the resulting trajectories following application of the MOEA: ASPTA, PDH and PPC. Blue arrows show reactions functional in *S. praecaptivus* and *S. glossinidius*; the white arrow shows a reaction not functional in *S. glossinidius*. Gene associations in *S. praecaptivus* and *S. glossinidius* are given in blue and black text, respectively. Adapted from previous work (Hall et al.) [33].

contain approximately 300 reactions. After 3000 generations, the *iLF517* model retained $51 \pm 1.1\%$ of the starting 606 reactions compared to *iRH830*, which was reduced to $27 \pm 1.8\%$ 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 *iRH830* models to assess their similarity. *iLF517* converges on a minimum after approximately 1000 generations, compared to the 2500 generations taken by the evolved *iRH830* model to find the minimum number of reactions (Fig. S2). The greater standard deviation of the *iRH830* solutions compared to *iLF517* is likely due to the larger starting point of the free-living model. The evolved *iRH830* and *iLF517* solutions ultimately converge at a similar point. To investigate this further, ten evolved models for both *iRH830* and *iLF517* 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 *iRH830* models, 289 (75.5%) are found in all ten. For the *iLF517* 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 *iRH830* and *iLF517* evolved solutions; they are core across the two species. The biomass outputs for the evolved *iRH830* and *iLF517* solutions range from 0.064 to 0.281 g DW $(\text{mmol glucose})^{-1} \text{h}^{-1}$, and 0.075 to 0.331 g DW $(\text{mmol$

glucose) $^{-1} \text{h}^{-1}$, respectively (Fig. 7b). 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 unrecognizable from the *S. glossinidius* progenitor.

DISCUSSION

Classical studies of microbial evolution, whilst useful, are ultimately limited by their inherent inability to replicate adaptations over large evolutionary timescales. Here, we present a computational approach by combining a MOEA with FBA, with the *Sodalis* system as a model for this. The genus *Sodalis* 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, *iRH830*, accompanied by experimental verification, that has been used in subsequent *in silico* evolution experiments. Supplying the ancestral *iRH830* 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 synthesize key, essential metabolites. Here, this has shown to be particularly evident in the pathways retained for glycolysis/

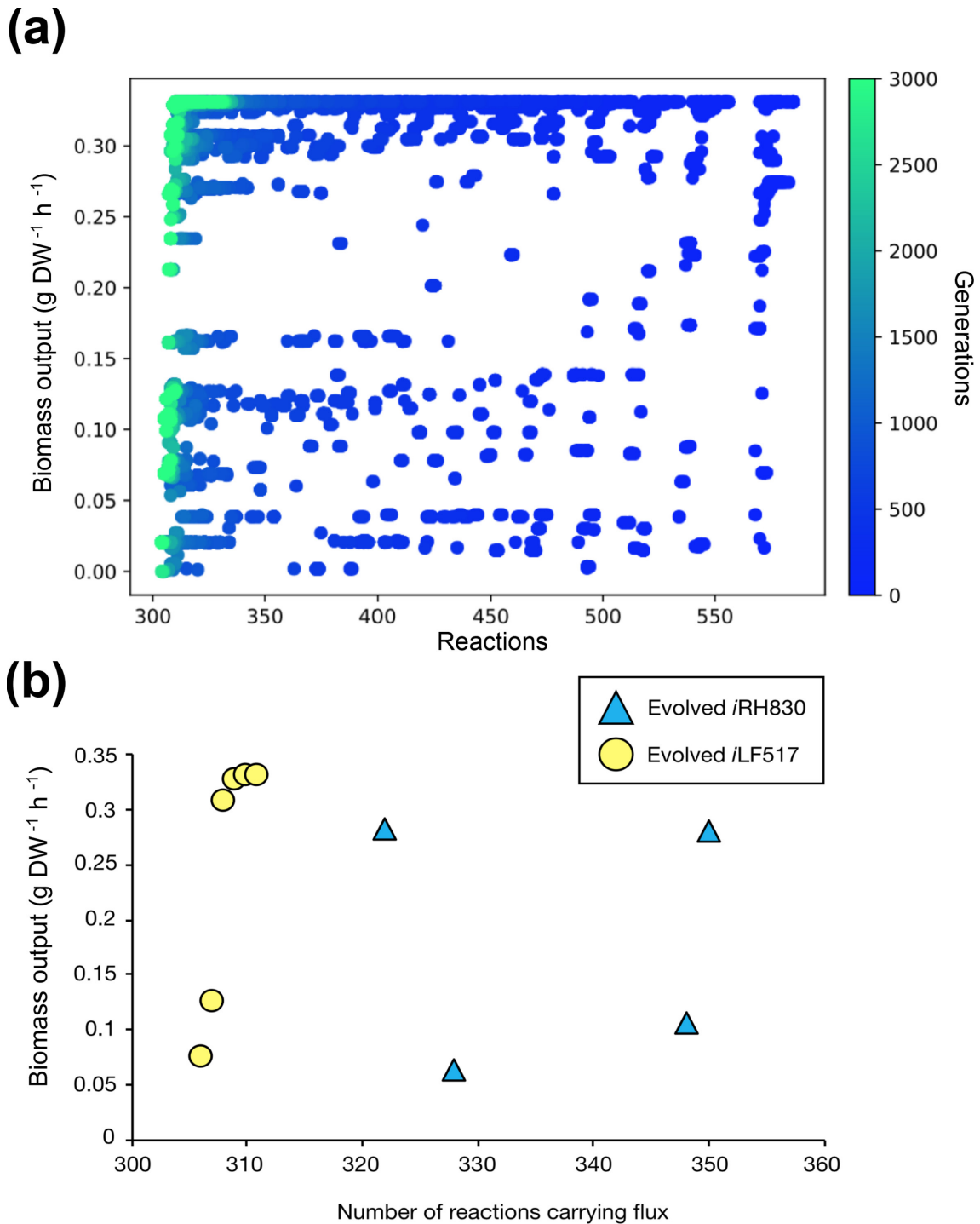


Fig. 7. Evolution of *iLF517*. (a) Evolution of *iLF517* in a blood medium. MOEA was run for 3000 generations (blue to green). (b) Biomass output [$\text{g DW} (\text{mmol glucose})^{-1} \text{h}^{-1}$] and the number of reactions carrying flux in evolved *iRH830* (blue triangles) and evolved *iLF517* (yellow circles) models. The evolved solutions produce comparable biomass outputs. Ten evolved solutions are given for each, some duplicates are present.

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 synthesize, 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 [93] the reduction of symbiont genomes should be irreversible [81], *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 within species of the genus *Sodalis*, 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 horizontal gene transfer, 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 pseudogenization 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. 4). 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 Δ PPC 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 pseudogenizations in insect-bacterial symbioses may have occurred.

It has been shown here that it is 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 *iRH830* models compared to evolved *iLF517* solutions, possibly due to difficulties in finding the minima from a larger starting point. *iRH830* can, however, be reduced down to look phenotypically similar to *iLF517* 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 metabolize xylitol may be related to the frequent presence of the genus *Sodalis* 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*, which produces antimicrobial peptides to constrict the symbiont population size [94]. Alternatively, the host may benefit from reduced costs of symbiont maintenance [95], 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 micro-organisms or to the study of rapid genome evolution in pathogenic bacteria.

Funding information

R.J.H. was funded by the BBSRC White Rose DTP (BB/M011151/1) and S.T. by the Wellcome Trust CIDCATS programme (WT095024MA).

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 that there are no conflicts of interest.

Data Bibliography

York Research Database 10.15124/2ff387bd-246f-49f6-9d3b-3da730afb9c8 (2020)

References

1. Wilcox JL, Dunbar HE, Wolfinger RD, Moran NA. Consequences of reductive evolution for gene expression in an obligate endosymbiont. *Mol Microbiol* 2003;48:1491–1500.
2. Dunbar HE, Wilson ACC, Ferguson NR, Moran NA. Aphid thermal tolerance is governed by a point mutation in bacterial symbionts. *PLoS Biol* 2007;5:e96.
3. Oliver KM, Russell JA, Moran NA, Hunter MS. Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. *Proc Natl Acad Sci USA* 2003;100:1803–1807.
4. Nakabachi A, Ueoka R, Oshima K, Teta R, Mangoni A et al. Defensive bacteriome symbiont with a drastically reduced genome. *Curr Biol* 2013;23:1478–1484.
5. Aksoy S. *Wigglesworthia* gen. nov. and *Wigglesworthia glossinidia* sp. nov., taxa consisting of the mycetocyte-associated, primary endosymbionts of tsetse flies. *Int J Syst Bacteriol* 1995;45:848–851.
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 et al. A fragile metabolic network adapted for cooperation in the symbiotic bacterium *Buchnera aphidicola*. *BMC Syst Biol* 2009;3:24.
8. Snyder AK, Rio RVM. "Wigglesworthia morsitans" folate (vitamin B9) biosynthesis contributes to tsetse host fitness. *Appl Environ Microbiol* 2015;81:5375–5386.
9. Hrusa G, Farmer W, Weiss BL, Applebaum T, Roma JS et al. TonB-dependent heme iron acquisition in the tsetse fly symbiont *Sodalis glossinidius*. *Appl Environ Microbiol* 2015;81:2900–2909.
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 Biol Evol* 2015;7:2871–2884.
11. McCutcheon JP, Moran NA. Parallel genomic evolution and metabolic interdependence in an ancient symbiosis. *Proc Natl Acad Sci USA* 2007;104:19392–19397.
12. McCutcheon JP, McDonald BR, Moran NA. Convergent evolution of metabolic roles in bacterial co-symbionts of insects. *Proc Natl Acad Sci USA* 2009;106:15394–15399.
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 Genet* 2009;5:e1000565.
14. Wilson ACC, Ashton PD, Calevro F, Charles H, Colella S et al. Genomic insight into the amino acid relations of the pea aphid, *Acyrtosiphon pisum*, with its symbiotic bacterium *Buchnera aphidicola*. *Insect Mol Biol* 2010;19:249–258.
15. McCutcheon JP, von Dohlen CD. An interdependent metabolic patchwork in the nested symbiosis of mealybugs. *Curr Biol* 2011;21:1366–1372.
16. Rio RVM, Lefevre C, Heddi A, Aksoy S. Comparative genomics of insect-symbiotic bacteria: influence of host environment on microbial genome composition. *Appl Environ Microbiol* 2003;69:6825–6832.
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. *Appl Environ Microbiol* 2014;80:5844–5853.
18. Baumann P, Baumann L, Lai C-Y, Rouhbakhsh D, Moran NA et al. Genetics, physiology, and evolutionary relationships of the genus *Buchnera*: intracellular symbionts of aphids. *Annu Rev Microbiol* 1995;49:55–94.
19. Akman Gündüz E, Douglas AE. Symbiotic bacteria enable insect to use a nutritionally inadequate diet. *Proc. R. Soc. B.* 2009;276:987–991.
20. International Aphid Genomics Consortium. Genome sequence of the pea aphid *Acyrtosiphon pisum*. *PLoS Biol* 2010;8:e1000313.
21. Wu M, Sun LV, Vamathevan J, Riegler M, Deboy R et al. Phylogenomics of the reproductive parasite *Wolbachia pipientis* wMel: a streamlined genome overrun by mobile genetic elements. *PLoS Biol* 2004;2:e69.
22. Pérez-Brocail V, Gil R, Ramos S, Lamelas A, Postigo M et al. A small microbial genome: the end of a long symbiotic relationship? *Science* 2006;314:312–313.
23. Moran NA, McCutcheon JP, Nakabachi A. Genomics and evolution of heritable bacterial symbionts. *Annu Rev Genet* 2008;42:165–190.
24. Akman L, Yamashita A, Watanabe H, Oshima K, Shiba T et al. Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*. *Nat Genet* 2002;32:402–407.
25. Dale C, Maudlin I. *Sodalis* gen. nov. and *Sodalis glossinidius* sp. nov., a microaerophilic secondary endosymbiont of the tsetse fly *Glossina morsitans morsitans*. *Int J Syst Bacteriol* 1999;49:267–275.
26. Edwards JS, Covert M, Palsson B. Metabolic modelling of microbes: the flux-balance approach. *Environ Microbiol* 2002;4:133–140.
27. Kauffman KJ, Prakash P, Edwards JS. Advances in flux balance analysis. *Curr Opin Biotechnol* 2003;14:491–496.
28. Lewis NE, Nagarajan H, Palsson BO. Constraining the metabolic genotype-phenotype relationship using a phylogeny of *in silico* methods. *Nat Rev Microbiol* 2012;10:291–305.
29. Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? *Nat Biotechnol* 2010;28:245–248.
30. Macdonald SJ, Lin GG, Russell CW, Thomas GH, Douglas AE. The central role of the host cell in symbiotic nitrogen metabolism. *Proc. R. Soc. B* 2012;279:2965–2973.

31. Macdonald SJ, Thomas GH, Douglas AE. Genetic and metabolic determinants of nutritional phenotype in an insect-bacterial symbiosis. *Mol Ecol* 2011;20:2073–2084.
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:e30652.
33. Hall RJ, Flanagan LA, Bottery MJ, Springthorpe V, Thorpe S et al. A tale of three species: adaptation of *Sodalis glossinidius* to tsetse biology, *Wigglesworthia* metabolism, and host diet. *mBio* 2019;10:e02106–18.
34. Ankrah NYD, Luan J, Douglas AE. Cooperative metabolism in a three-partner insect-bacterial symbiosis revealed by metabolic modeling. *J Bacteriol* 2017;199:e00872–16.
35. González-Domenech C, Belda E, Patiño-Navarrete R, Moya A, Peretó J et al. Metabolic stasis in an ancient symbiosis: genome-scale metabolic networks from two *Blattabacterium cuenoti* strains, primary endosymbionts of cockroaches. *BMC Microbiol* 2012;12:S5.
36. Sørensen MES, Cameron DD, Brockhurst MA, Wood AJ. Metabolic constraints for a novel symbiosis. *R Soc Open Sci* 2016;3:150708.
37. Shastri AA, Morgan JA. Flux balance analysis of photoautotrophic metabolism. *Biotechnol Prog* 2005;21:1617–1626.
38. Nogales J, Gudmundsson S, Knight EM, Palsson BO, Thiele I. Detailing the optimality of photosynthesis in cyanobacteria through systems biology analysis. *Proc Natl Acad Sci USA* 2012;109:2678–2683.
39. Knoop H, Zilliges Y, Lockau W, Steuer R. The metabolic network of *Synechocystis* sp. PCC 6803: systemic properties of autotrophic growth. *Plant Physiol* 2010;154:410–422.
40. Knoop H, Gründel M, Zilliges Y, Lehmann R, Hoffmann S et al. Flux balance analysis of cyanobacterial metabolism: the metabolic network of *Synechocystis* sp. PCC 6803. *PLoS Comput Biol* 2013;9:e1003081.
41. Toh H et al. Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of *Sodalis glossinidius* in the tsetse host. *Genome Res* 2006;16:149–156.
42. Goodhead I, Blow F, Brownridge P, Hughes M, Kenny J et al. Large-scale and significant expression from pseudogenes in *Sodalis glossinidius* – a facultative bacterial endosymbiont. *Microb Genom* 2020;6:e000285.
43. Covert MW, Palsson BØ. Transcriptional regulation in constraints-based metabolic models of *Escherichia coli*. *J Biol Chem* 2002;277:28058–28064.
44. Lee JM, Gianchandani EP, Papin JA. Flux balance analysis in the era of metabolomics. *Brief Bioinform* 2006;7:140–150.
45. Covert MW, Palsson BO. Constraints-based models: regulation of gene expression reduces the steady-state solution space. *J Theor Biol* 2003;221:309–325.
46. Covert MW, Schilling CH, Palsson B. Regulation of gene expression in flux balance models of metabolism. *J Theor Biol* 2001;213:73–88.
47. Mahadevan R, Edwards JS, Doyle FJ. Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophys J* 2002;83:1331–1340.
48. Harcombe WR, Delaney NF, Leiby N, Klitgord N, Marx CJ. The ability of flux balance analysis to predict evolution of central metabolism scales with the initial distance to the optimum. *PLoS Comput Biol* 2013;9:e1003091.
49. Pfeiffer T, Soyer OS, Bonhoeffer S. The evolution of connectivity in metabolic networks. *PLoS Biol* 2005;3:e228.
50. Ponce-de-Leon M, Tamarit D, Calle-Espinosa J, Mori M, Latorre A et al. Determinism and contingency shape metabolic complementation in an endosymbiotic consortium. *Front Microbiol* 2017;8:2290.
51. Edwards JS, Palsson BØ. Metabolic flux balance analysis and the *in silico* analysis of *Escherichia coli* K-12 gene deletions. *BMC Bioinformatics* 2000;1:1.
52. Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR et al. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol Syst Biol* 2007;3:121.
53. Reed JL, Vo TD, Schilling CH, Palsson BO. An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biol* 2003;4:R54.
54. Orth JD, Conrad TM, Na J, Lerman JA, Nam H et al. A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism—2011. *Mol Syst Biol* 2011;7:535.
55. Orth JD, Palsson BØ. Gap-filling analysis of the iJO1366 *Escherichia coli* metabolic network reconstruction for discovery of metabolic functions. *BMC Syst Biol* 2012;6:30.
56. Pál C, Papp B, Lercher MJ, Csermely P, Oliver SG et al. Chance and necessity in the evolution of minimal metabolic networks. *Nature* 2006;440:667–670.
57. Yizhak K, Tuller T, Papp B, Ruppin E. Metabolic modeling of endosymbiont genome reduction on a temporal scale. *Mol Syst Biol* 2011;7:479.
58. McNally CP, Borenstein E. Metabolic model-based analysis of the emergence of bacterial cross-feeding via extensive gene loss. *BMC Syst Biol* 2018;12:69.
59. Hindré T, Knibbe C, Beslon G, Schneider D. New insights into bacterial adaptation through *in vivo* and *in silico* experimental evolution. *Nat Rev Microbiol* 2012;10:352–365.
60. Batut B, Parsons DP, Fischer S, Beslon G, Knibbe C. *In silico* experimental evolution: a tool to test evolutionary scenarios. *BMC Bioinformatics* 2013;14:S11.
61. François P, Hakim V. Design of genetic networks with specified functions by evolution *in silico*. *Proc Natl Acad Sci USA* 2004;101:580–585.
62. Budinich M, Bourdon J, Larhlimi A, Eveillard D. A multi-objective constraint-based approach for modeling genome-scale microbial ecosystems. *PLoS One* 2017;12:e0171744.
63. Wang L, Maranas CD. MinGenome: an *in silico* top-down approach for the synthesis of minimized genomes. *ACS Synth Biol* 2018;7:462–473.
64. Fong SS, Burgard AP, Herring CD, Knight EM, Blattner FR et al. *In silico* design and adaptive evolution of *Escherichia coli* for production of lactic acid. *Biotechnol Bioeng* 2005;91:643–648.
65. Garcia S, Trinh CT. Comparison of multi-objective evolutionary algorithms to solve the modular cell design problem for novel biocatalysis. *Processes* 2019;7:361.
66. Machado D, Herrgård MJ. Co-evolution of strain design methods based on flux balance and elementary mode analysis. *Metab Eng Commun* 2015;2:85–92.
67. Clayton AL, Oakeson KF, Gutin M, Pontes A, Dunn DM et al. A novel human-infection-derived bacterium provides insights into the evolutionary origins of mutualistic insect-bacterial symbioses. *PLoS Genet* 2012;8:e1002990.
68. Chari A, Oakeson KF, Enomoto S, Jackson DG, Fisher MA et al. Phenotypic characterization of *Sodalis praecaptivus* sp. nov., a close non-insect-associated member of the *Sodalis*-allied lineage of insect endosymbionts. *Int J Syst Evol Microbiol* 2015;65:1400–1405.
69. WHO. *Control and Surveillance of African Trypanosomiasis: Report of a WHO Expert Committee (WHO Technical Report Series 881)*. Geneva: World Health Organization; 1998.
70. King ZA, Lu J, Dräger A, Miller P, Federowicz S et al. BiGG Models: a platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res* 2016;44:D515–D522.
71. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* 2000;28:27–30.
72. Kanehisa M, Sato Y, Furumichi M, Morishima K, Tanabe M. New approach for understanding genome variations in KEGG. *Nucleic Acids Res* 2019;47:D590–D595.
73. Keseler IM, Mackie A, Santos-Zavaleta A, Billington R, Bonavides-Martínez C et al. The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12. *Nucleic Acids Res* 2017;45:D543–D550.
74. Bateman A, Martin MJ, Orchard S, Magrane M, Alpi E et al. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* 2019;47:D506–D515.

75. Caspi R, Foerster H, Fulcher CA, Kaipa P, Krummenacker M *et al.* The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res* 2008;36:D623–D631.
76. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* 2007;23:2947–2948.
77. Thiele I, Palsson BØ. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc* 2010;5:93–121.
78. Ebrahim A, Lerman JA, Palsson BO, Hyduke DR. COBRApy: constraints-based reconstruction and analysis for python. *BMC Syst Biol* 2013;7:74.
79. Deb K, Pratap A, Agarwal S, Meyarivan T. A fast and elitist multi-objective genetic algorithm: NSGA-II. *IEEE Trans Evol Comput* 2002;6:182–197.
80. De Rainville F-M, Fortin F-A, Gardner M-A, Parizeau M, Gagné C. DEAP: a Python framework for evolutionary algorithms. In: *Proceedings of the 14th International Conference on Genetic and Evolutionary Computation*. New York: Association for Computing Machinery; 2012. pp. 85–92.
81. Moran NA. Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. *Proc Natl Acad Sci USA* 1996;93:2873–2878.
82. Novakova E, Hypsa V. A new *Sodalis* lineage from bloodsucking fly *Craterina melbae* (Diptera, Hippoboscoidea) originated independently of the tsetse flies symbiont *Sodalis glossinidius*. *FEMS Microbiol Lett* 2007;269:131–135.
83. 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:e40354.
84. Boyd BM, Allen JM, Koga R, Fukatsu T, Sweet AD *et al.* Two bacterial genera, *Sodalis* and *Rickettsia*, associated with the seal louse *Proechinophthirus fluctus* (Phthiraptera: Anoplura). *Appl Environ Microbiol* 2016;82:3185–3197.
85. Oakeson KF, Gil R, Clayton AL, Dunn DM, von Niederhausern AC *et al.* Genome degeneration and adaptation in a nascent stage of symbiosis. *Genome Biol Evol* 2014;6:76–93.
86. Szklarzewicz T, Kalandyk-Kołodziejczyk M, Michalik K, Jankowska W, Michalik A. Symbiotic microorganisms in *Puto superbus* (Leonardi, 1907) (Insecta, Hemiptera, Cocco-morpha: Putoidae). *Protoplasma* 2018;255:129–138.
87. Koga R, Bennett GM, Cryan JR, Moran NA. Evolutionary replacement of obligate symbionts in an ancient and diverse insect lineage. *Environ Microbiol* 2013;15:2073–2081.
88. Koga R, Moran NA. Swapping symbionts in spittlebugs: evolutionary replacement of a reduced genome symbiont. *Isme J* 2014;8:1237–.
89. Michalik A, Jankowska W, Kot M, Gołas A, Szklarzewicz T. Symbiosis in the green leafhopper, *Cicadella viridis* (Hemiptera, Cicadellidae). Association in statu nascendi? *Arthropod Struct Dev* 2014;43:579–587.
90. Kaiwa N, Hosokawa T, Kikuchi Y, Nikoh N, Meng XY *et al.* Primary gut symbiont and secondary, *Sodalis*-allied symbiont of the Scutellerid stinkbug *Cantao ocellatus*. *Appl Environ Microbiol* 2010;76:3486–3494.
91. Hosokawa T, Kaiwa N, Matsuura Y, Kikuchi Y, Fukatsu T. Infection prevalence of *Sodalis* symbionts among stinkbugs. *Zoological Lett* 2015;1:5.
92. Snyder AK, Deberry JW, Runyen-Janecky L, Rio RVM. Nutrient provisioning facilitates homeostasis between tsetse fly (Diptera: Glossinidae) symbionts. *Proc Biol Sci* 2010;277:2389–2397.
93. Muller HJ. The relation of recombination to mutational advance. *Mutat Res* 1964;106:2–9.
94. Login FH, Balmand S, Vallier A, Vincent-Monégat C, Vigneron A *et al.* Antimicrobial peptides keep insect endosymbionts under control. *Science* 2011;334:362–365.
95. 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.

Five reasons to publish your next article with a Microbiology Society journal

1. The Microbiology Society is a not-for-profit organization.
2. We offer fast and rigorous peer review – average time to first decision is 4–6 weeks.
3. Our journals have a global readership with subscriptions held in research institutions around the world.
4. 80% of our authors rate our submission process as 'excellent' or 'very good'.
5. Your article will be published on an interactive journal platform with advanced metrics.

Find out more and submit your article at microbiologyresearch.org.