Genome scale metabolic network reconstruction of the pathogen *Enterococcus faecalis*

Carla Portela*,**, Silas Villas-Bôas**, Isabel Rocha*, Eugénio C. Ferreira*

* IBB - Institute for Biotechnology and Bioengineering, Centre of Biological Engineering,
University of Minho, Campus de Gualtar, 4710-057 Braga-Portugal
** School of Biological Sciences, University of Auckland, 3A Symonds Street,
Auckland 1142, New Zealand

Abstract: Since the metabolic network reconstruction of *Haemophilus influenzae* was published in 1999, many other researchers have followed this trend. The possibilities and potential of having an *in silico* metabolic network for a specific organism is gaining emphasis among researchers that see the promise that the new era of genome-scale metabolic models could bring to the scientific scene.
Enterococcus faecalis is a Gram positive bacterium from the lactic acid group that inhabits commensally the gut of humans and warm-blooded animals. However, it acquired a relevant position in the medical scenery as the main cause of nosocomial infections in hospital environments and the reason behind multiple diseases, namely urinary tract infections, endocarditis, meningitis, to name a few.
In this work the first genome-scale model of *E. faecalis*, has been reconstructed to serve as a valuable tool to predict and better understand the phenotypic behaviour and its metabolism. The reconstruction process resulted in a network with 603 reactions and 576 metabolites. From the annotated genome it was possible to identify 366 genes with enzymatic activity. Performed simulations allowed us to validate the formation of end-products, as well as determine critical genes involved in the metabolism of this pathogen.

Keywords: *Enterococcus faecalis*, Pathogen, Systems biology, Genome scale reconstruction.

1. INTRODUCTION

Enterococcus faecalis is a Gram-positive cocci, non-spore forming, non-motile and facultative anaerobe that belongs to the lactic acid bacteria (LAB) group.

Although it is normally known for its influence in the medical panorama it is also gaining relevance in different fields namely in the food industry and more specifically in the process of ripening and aroma development in certain types of cheese. Nevertheless, it is still more prominent in the medical scenery as the cause of multiple diseases. Although initially it was regarded as harmless to humans and unimportant medically, *E. faecalis* became a problem in the 1970-1980's as a leading cause for hospital acquired infections and it is nowadays considered the main cause for hospital acquired infections of the bloodstream, urinary tract, surgical wounds and other sites (Gilmore *et al.* 2013). In addition, *E. faecalis* plays a problematic role in endodontic diseases. Although Enterococci makes up only a small proportion of the initial flora of untreated teeth with necrotic pulps, *E. faecalis* in particular, has been frequently found in obturated root canals exhibiting signs of chronic apical periodontitis (Kayaoglu & Orstavik 2004).

E. faecalis antibiotic resistance has been well documented over the past years. It is established that *E. faecalis* is inherently resistant to many antimicrobial drugs. In addition, it has also developed resistance to gentamycin and streptomycin (Murray 2000).

This problem acquires a new dimension when pathogens also transfer the antibiotic resistance genes to other bacteria. Population genetic studies revealed that recombination prevails as a driving force of genetic diversity in *E. faecalis, E. faecium, S. pneumonia* and *S. pyogenes* and, more importantly, horizontal acquisition of resistance and virulence genes play a key role in the emergence of new clinically relevant clones in these species (Willems *et al.* 2011).

E. faecalis is an opportunistic bacterium and normally carries out anaerobic metabolism. Likewise, the absence of a complete tricarboxilic cycle suggests that the energy production occurs *via* glycolysis or the pentose phosphate pathway (Paulsen *et al.* 2003).

Still, this organism has been shown to be able to carry aerobic respiration when in the presence of heme (Winstedt *et al.* 2000).

E. faecalis possesses an arsenal of strategic responses to fight oxidative stress which gives this organism an advantage towards its host and other bacteria.

In a world where scientific knowledge and data generation increases by the hour, it becomes more relevant to develop new tools to deal with all that information wisely.

The prospect of understanding the massive amount of biological data using mathematical models to describe the structure of a system and its response to external perturbations

is particularly appealing and could allow the researchers to answer a new level of questions.

The field of genome scale reconstructions has expanded rapidly over the years and currently accounts for more than 60 genome-scale network reconstructions (GENRE) for bacterial species alone. When considering *archea* and eukaryotes the total number of GENREs raises up to 90. The available information about published models can be found elsewhere (http://gcrg.ucsd.edu/InSilicoOrganisms/OtherOrganisms) (Feist *et al.* 2009).

To reconstruct a genome scale metabolic model, it is fundamental to collect all the biochemical reactions and corresponding stoichiometry for the organism to be reconstructed. Hence, metabolic models tend to be organism-specific.

The reconstruction of a GENRE is very laborious and time consuming and requires a fully annotated genome to start with. Efforts have been made in order to standardize the reconstruction pipeline (Thiele & Palsson 2010) but as each reconstruction tends to be organism-specific it is therefore unique.

Overall, GENREs are powerful tools that allow the simulation of the phenotypic behavior under different sets of conditions and can also be a valuable resource to predict and improve microbial cell factories to produce relevant industrial products or to help the physiological characterization and identification of drug targets in pathogens.

Our aim is to provide the first genome scale metabolic model of *E. faecalis* to serve as a tool for better understanding the phenotypic behavior of this pathogen as well as its innumerous capabilities to survive harsh and diverse environments.

2. RECONSTRUCTION PROCESS PIPELINE

There are several reviews and detailed papers published elsewhere which describe with different degree of detail, the process of genome scale metabolic network reconstruction (Pinney *et al.* 2007; Durot *et al.* 2009; Feist *et al.* 2009; Notebaart *et al.* 2006; Francke *et al.* 2005; Rocha and Förster 2008). A succinct description is pictured in figure 1 and is described next.

2.1 Compilation of available information

The reconstruction process of any organism starts with its annotated genome. Genome annotation can be found in different online databases that store collections of genome annotations and are access free. Some databases can be organism specific such as EcoCyc for *Escherichia coli* or SGD (*Saccharomyces* Genome Database) for *Saccharomyces cerevisae,* while others comprise the annotated genomes of multiple organisms such as, EntrezGene, CMR (Comprehensive Microbial Resource) or IMG (Integrated Microbial Genomes).
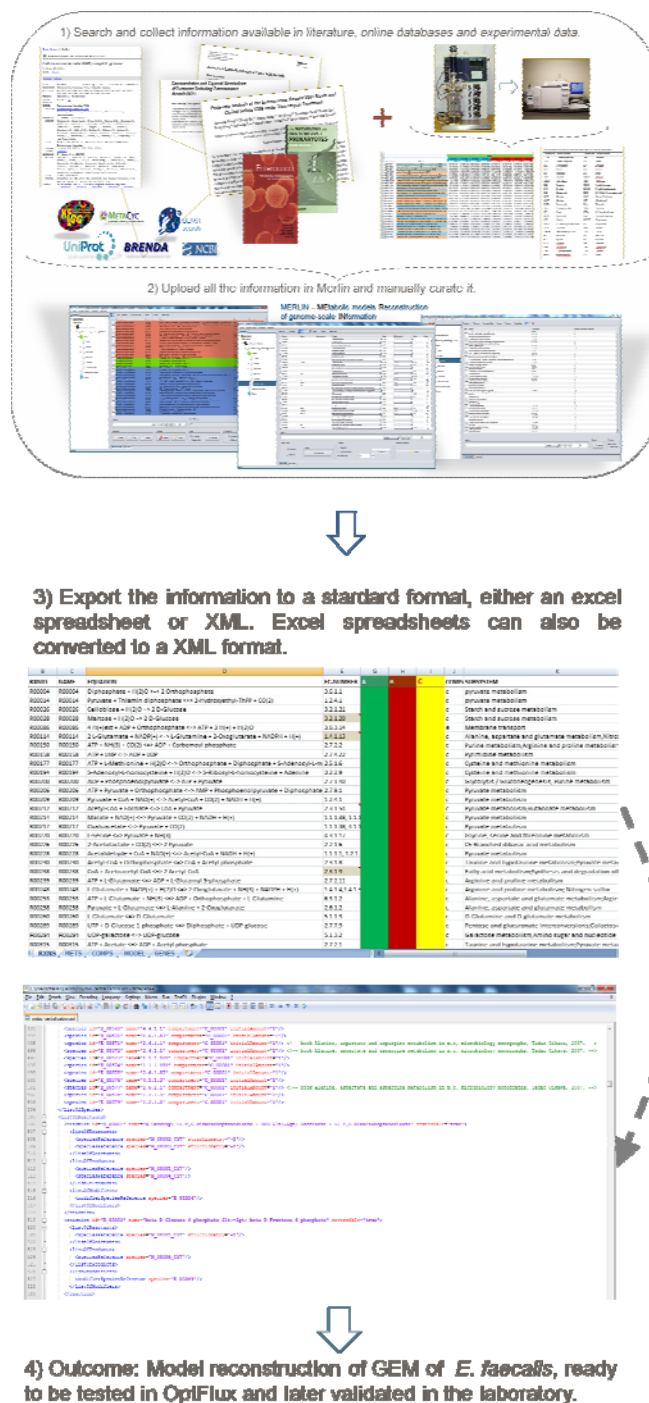
Figure 1. Data integration of different levels of information in the process of reconstruction of the genome scale metabolic model of *E. faecalis.*

Metabolic databases, such as KEGG (Kyoto Encyclopaedia of Genes and Genomes) (Kanehisa *et al.* 2004), BRENDA (Schomburg *et al.* 2004), BioCyc/Metacyc (Caspi *et al.* 2006) and the Transport DB (Ren *et al.* 2007) are also an important source of information as they provide collections of biochemical reactions that occur in different organisms. Through these databases, it is possible to collect all the information that relates enzymes and the reactions catalyzed by them.

The procedure for the reconstruction of the genome-scale model of *E. faecalis* involved an exhaustive search of the available information about the organism. Using the annotated genome as a starting point, a reaction database was compiled from KEGG. Although there is no manual curation of the information in this database, it still contains the most extent repertoire of reactions known from all available databases.

2.2 Manual curation of the proposed model

The automated or semi-automated extraction of the metabolic reactions from the different databases provides a set of biochemical reactions that constitute the initial draft model. However, this information is frequently inaccurate or incomplete and therefore, a manual curation is required. This step is undoubtedly the most time consuming and tedious of the reconstruction process and requires specific information on the organism, laboratory data and literature for validation of the information.

In this work, each pathway associated with *E. faecalis* metabolism was manually curated to add additional information that could be missing. The methodology involved the analysis of the enzymes marked as not present according to the annotated genome. Each enzyme was blasted (BLASTP) against NCBI database in order to find homologous sequences in other genomes. Merlin software was used to perform this task (Dias *et al.* 2010). This software aims to semi-automate the process of genome scale reconstructions and, among other tasks, it allows performing a BLAST search of the sequenced genome against the entire NCBI database. It further allows viewing the information of the possible homologues for each gene and the associated *e-value* score, in a user friendly view. As a result, 25 reactions involved in the metabolism of *E. faecalis*, were added to the database.

All the information added to the model was subjected to deep analysis of the literature as well as complementary databases such as BioCyc, Brenda and Uniprot (Apweiler *et al.* 2004). A particular emphasis was given to reactions reversibility. The reversibility was manually curated with focus on the tools, eQuilibrator (Flamholz *et al.* 2012), BioCyc and Brenda databases. Reactions whose reversibility was not defined were considered to be reversible.

Finally, it is also necessary to define the biomass equation. As it will be explained in the further section, to be able to solve a metabolic model, a biomass equation is requires.

The biomass equation encompasses a set of reactions that translate the necessary building blocks (e.g., aminoacids, nucleotides, DNA) required for biomass formation. Each component can either be determined experimentally or, in alternative, adopted from the literature or from a close related organism.

For the present model, biomass equation from *Lactococcus lactis* genome scale model (Oliveira *et al.* 2005), was adopted as the corresponding experimental determination is under way.

2.3 Mathematical Representation and Simulation methods

Once the model is debugged, it is crucial to convert the biochemical information into a computational language to further explore the capabilities of the reconstructed model.

The stoichiometry of the model is contained in a stoichiometric matrix S that, together with additional constraints over the flux values originating from reversibility information, for example, represents the hypercone of feasible solutions, assuming a steady state approximation.

To find a biologically meaningful flux distribution, flux balance analysis (FBA) is the most used methodology and requires the definition of an objective function (Raman & Chandra 2009). Usually, this consists on the biomass assembly, since it is assumed that the main purpose of a microorganism would be to continuously grow.

In this work, the set of reactions were initially collected converted to an SBML model using the merlin software. Additional determination of the stoichiometric matrix was accomplished within the software OptFlux (Rocha *et al.* 2010). Also, model simulations and examination of the network capabilities were performed using the same software. This user friendly tool allows the user to perform different analysis from wild type simulations, strain optimization, definition of minimal nutrient requirements to name a few.

2.4 Model validation and evaluation

When the GEM model is ready to simulate, it is fundamental to analyse if the predictions are in accordance with the *in vivo* data.

Generally, continuous culture is a good strategy for model validation as it implies a steady state condition and the entire variables are controlled by the operator. Additionally, it is possible to test different stress conditions or nutrient availability to then compare with the *in silico* simulations.

High throughput *omic* techniques play nowadays a major role in the model validation and evaluation. *Omic* techniques such as metabolomics, proteomics and fluxomics provide important information that assists the refinement of the model and open new channels to discover unknown pathways, genes, reactions and metabolite utilization.

All these available data, are in fact one of the main reasons that lead to the development of new and more powerful tools in terms of computational biology. The ability of understanding such amount of data in a global and integrated perspective is a good challenge that GEM is able to answer.

Similarly, with the model, it is possible to simulate different growth media to define minimal medium requirements, perform gene deletion or optimization analysis to target a specific metabolite production.

Hence, it is clear that the reconstruction process of a GEM is an iterative cycle that uses the model to simulate a certain activity, but also relies on the laboratory data to feed the current model with more information to make it more sensitive, robust and accurate.

3. RESULTS

The present work focuses on the reconstruction of the first genome scale metabolic model of *E. faecalis* strain V583.

E. faecalis is metabolically diverse and possesses on its genome four DNA molecules: one chromosome and 3 plasmids. The main molecule is the chromosome with around 3000 kilo base pairs (bp) while the plasmids are smaller structures of 20,000 to 60,000 bp. These molecules harbour most of the proteins that encode multiple drug resistance and regulatory systems. Table 1 depicts the overview of the identity card of *E. faecalis* V583.

Table 1. *E. faecalis* identity card

	Base pairs
Chromosome	3,218,030
Plasmid-1	66,320
Plasmid-2	17,963
Plasmid-3	57,660

	ORF
Total	3,500
Assigned function	2,333
No assigned function	1,167

	GC content
Average	37.4%

3.1 General statistics

The reconstruction process resulted in a network with 603 reactions and 576 metabolites. From the annotated genome it is possible to identify 366 genes with enzymatic activity. There are 592 reactions with gene association which corresponds to 87% of the total reactions. The remaining 13% (reactions without gene association) are mainly transport reactions or reactions not catalyzed by enzymes or spontaneous. This information is compiled in Table 2.

**Table 2. Basic network properties of GEM**

Basic Network Properties	Compartment	Number	Total
Reactions	cytosol	603	682
	external	79	
Metabolites	cytosol	576	653
	external	77	
Genes		366	366
Reactions with gene association		592	592

3.2 Simulations

Once the model was debugged and the gaps filled, it was possible to perform simulations and to generate biomass.

Simulating a wild type growth with biomass maximization allowed us to determine the specific growth rate of *E. faecalis* for a defined environmental condition. The value observed was 0.17 h-1 which did not differ much from what is observed *in vivo* (0.15 h-1).

Critical genes of a model are genes that, if deleted, lead to cell death. This information becomes extremely relevant when designing metabolic engineering experiments. The critical genes/reaction determined for *E. faecalis* indicated a total of 59 gene/reactions mainly involved in the aminoacids pathways as depicted in Figure 2.
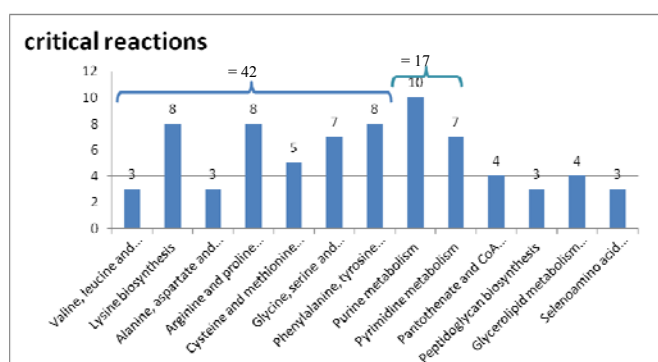
**Figure 2.** Critical reactions calculated for *E. faecalis*.

3.3 Validation

To validate the model we used data from Mehmeti *et al.* (2012), that tested E. faecalis V583 growth at different dilution rates and glucose limited chemostat. Simulations using the same environmental conditions were performed to test the outcome of the in silico simulation and compare to what is described in the literature. Although the work from Mehmeti *et al.* group was independent from the work proposed in this paper, it represents an interesting approach to validate the capability of prediction of the first genome-scale metabolic model of *Enterococcus faecalis* V583.

It was observed that the results obtained in silico were accurate and that the model predicted very similar flux distribution of the end-products in vivo.

As evidenced in Figure 3, a comparison between the simulation results and the experimental data (Mehmeti *et al.* 2006) shows that the model could predict all the end-products detected in a chemostast experiment. Additionally the simulation predictions were in accordance with what is observable in vivo with an error margin always lower than 13%. Therefore, high level of accuracy of the simulation results was achieved.

Further analysis are being undertaken to evaluate the capability of the model to detect the shift from homolactic to heterolactic fermentation and also to test the sensitivity of the model to oxygen perturbations.
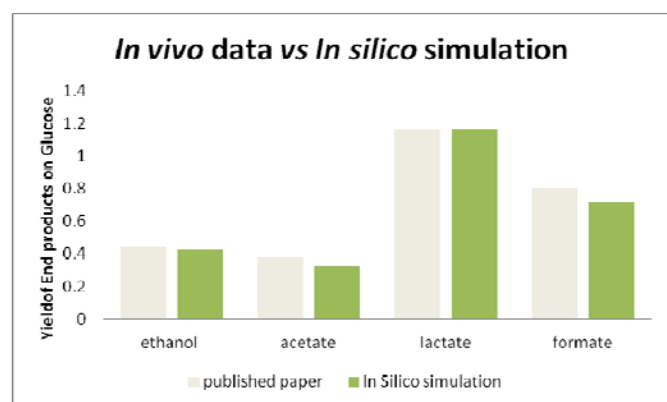
Figure 3. Simulation performed using Optflux (Rocha *et al.*, 2010) to determine the in silico flux distribuition of end-products and comparison with in vivo data.

3.4 Minimal growth requirements

To determine the minimal growth requirements of *E. faecalis*, model constraints were imposed and biomass production was simulated. Inorganic compounds such as phosphate, minerals, ammonium and water are freely available, which means that lower and upper bounds were left unconstrained. For qualitative purposes, a glucose flux was set to 12 mmol/gDW.h. Next, each amino acid, was individually tested for its essentiality. Individual amino acid were supplied at a rate of 10 mmol/gDW.h. Single amino acid exclusion test was performed and essentiality analysed.

From literature it is known that lactic acid bacteria have complex nutritional requirements as they are usually associated with nutrient rich environments (Fernández & Zúñiga 2006). *E. faecalis* is a fastidious organism that requires a number of amino acids for its survival. Although no studies have been focused on the definition of a minimal growth medium of *E. faecalis*, some authors have developed a chemically defined medium to serve a specific purpose (Khan *et al.* 2013; Zhang *et al.* 2009). The authors have determined that the amino acid requirements were different in each study, most likely due to the fact that they were working with different bacterial strains. Hence it is expected that *E. faecalis* V583 might also exhibit different amino acid essentiality.

Table 3 resumes the amino acid essentiality for each strain (OG1RF, KA177 and V583) and the *in silico* prediction of amino acid essentiality for V583. Histidine (H), leucine (L), methionine (M) and valine (V) were common among the different studies. Both studies indicate that *E. faecalis* OG1RF, KA177 and B9510 are auxotrophic for arginine (R) while V583 appears to be able to synthesise this amino acid. On the other hand, glutamine (Q) (or glutamate as they interconvert on one another) was defined as essential for V583 but not for the other strains.

4. CONCLUSIONS

The first genome scale metabolic model of the pathogen *E. faecalis* is presented here. The model was reconstructed based on genomic, biochemical and physiological information, comprising a total of 603 reaction and 576 metabolites.

Metabolic network analysis was performed using FBA methodology and the genome scale model has shown to be able to predict with accuracy some of the experimental observations given the specific environmental constraints.

Table 3. Essential amino acids for each strain tested in different published studies and in silico simulation (present study). Yes means the amino acid is essential.

Amino acid	Zhang *et al.* OG1RF and KA177	Khan *et al.* B9510	*In silico* V583
R	yes	yes	no
Q	no	no	yes
G	yes	no	no
H	yes	yes	yes
I	no	yes	yes
L	yes	yes	yes
M	yes	yes	yes
W	yes	no	yes
V	yes	yes	yes

Further analysis need to be done to better improve the model capabilities and also to test for its robustness. Nevertheless, it is already possible to analyse and predict important phenotypic outcomes and also understand how the flux distribution is influenced.

In the near future the GEM will be used as a powerful tool to test and develop novel strategies useful for the industry or medical area.

REFERENCES

Apweiler, R. *et al.*, 2004. UniProt: the Universal Protein knowledgebase. *Nucleic acids research*, 32(Database issue), pp.D115–9.

Caspi, R. *et al.*, 2006. MetaCyc: a multiorganism database of metabolic pathways and enzymes. *Nucleic acids research*, 34(Database issue), pp.511–6.

Dias, O. *et al.*, 2010. Merlin : Metabolic Models Reconstruction using Genome-Scale Information. , (International Symposium on Computer Applications in Biotechnology), pp.120–125.

Durot, M., Bourguignon, P.-Y. & Schachter, V., 2009. Genome-scale models of bacterial metabolism: reconstruction and applications. *FEMS microbiology reviews*, 33(1), pp.164–90.

Feist, A.M. *et al.*, 2009. Reconstruction of biochemical networks in microorganisms. *Nature reviews. Microbiology*, 7(2), pp.129–43.

Fernández, M. & Zúñiga, M., 2006. Amino acid catabolic pathways of lactic acid bacteria. *Critical reviews in microbiology*, 32(3), pp.155–83.

Flamholz, A. *et al.*, 2012. eQuilibrator-the biochemical thermodynamics calculator. *Nucleic acids research*, 40(Database issue), pp.D770–5.

Francke, C., Siezen, R.J. & Teusink, B., 2005. Reconstructing the metabolic network of a bacterium from its genome. *Trends in microbiology*, 13(11), pp.550–8.

Gilmore, M.S., Lebreton, F. & van Schaik, W., 2013. Genomic transition of enterococci from gut commensals to leading causes of multidrug-resistant hospital infection in the antibiotic era. *Current Opinion in Microbiology*, pp.1–7.

Kanehisa, M. *et al.*, 2004. The KEGG resource for deciphering the genome. *Nucleic acids research*, 32(Database issue), pp.D277–80.

Kayaoglu, G. & Orstavik, D., 2004. Virulence Factors of *Enterococcus faecalis*: Relationship To Endodontic Disease. *Critical Reviews in Oral Biology & Medicine*, 15(5), pp.308–320. Available at: http://cro.sagepub.com/cgi/doi/10.1177/1544111304015 00506 [Accessed February 12, 2013].

Khan, H., Flint, S.H. & Yu, P.-L., 2013. Development of a chemically defined medium for the production of enterolysin A from *Enterococcus faecalis* B9510. *Journal of applied microbiology*, 114(4), pp.1092–102.

Murray, B.E., 2000. Vancomycin-resistant enterococcal infections. *The New England Journal of Medicine*, 342(10), pp.710–21.

Notebaart, R. a *et al.*, 2006. Accelerating the reconstruction of genome-scale metabolic networks. *BMC bioinformatics*, 7, p.296.

Oliveira, A.P., Nielsen, J. & Förster, J., 2005. Modeling Lactococcus lactis using a genome-scale flux model. *BMC microbiology*, 5, p.39.

Paulsen, I.T. *et al.*, 2003. Role of mobile DNA in the evolution of vancomycin-resistant *Enterococcus faecalis*. *Science*, 299(5615), pp.2071–4.

Pinney, J.W. *et al.*, 2007. Metabolic reconstruction and analysis for parasite genomes. *Trends in parasitology*, 23(11), pp.548–54.

Raman, K. & Chandra, N., 2009. Flux balance analysis of biological systems: applications and challenges. *Briefings in bioinformatics*, 10(4), pp.435–49.

Ren, Q., Chen, K. & Paulsen, I.T., 2007. TransportDB: a comprehensive database resource for cytoplasmic membrane transport systems and outer membrane channels. *Nucleic acids research*, 35, pp.274–9.

Rocha, I. & Förster, J.N.J., 2008. Microbial gene essentiality - protocols and bioinformatics. In *Microbial gene essentiality - protocols and bioinformatics*. pp. 409–431.

Rocha, I. *et al.*, 2010. OptFlux: an open-source software platform for in silico metabolic engineering. *BMC systems biology*, 4(1), p.45.

Schomburg, I. *et al.*, 2004. BRENDA, the enzyme database: updates and major new developments. *Nucleic acids research*, 32, pp.431–3.

Thiele, I. & Palsson, B.Ø., 2010. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature protocols*, 5(1), pp.93–121.

Willems, R.J.L. *et al.*, 2011. Population biology of Gram-positive pathogens: high-risk clones for dissemination of antibiotic resistance. *FEMS microbiology reviews*, 35(5), pp.872–900.

Winstedt, L. *et al.*, 2000. *Enterococcus faecalis* V583 contains a cytochrome bd-type respiratory oxidase. *Journal of bacteriology*, 182(13), pp.3863–6.

Zhang, G., Mills, D.A. & Block, D.E., 2009. Development of chemically defined media supporting high-cell-density growth of lactococci, enterococci, and streptococci. *Applied and environmental microbiology*, 75(4), pp.1080–7.