

Polihidroksibütirat Üretimi İçin Entegre Multi-Omik, Mekanistik Makine Öğrenmesi Ve
CRISPR Yaklaşımı

Blaise Manga Enuh

DOKTORA TEZİ

Biyoteknoloji ve Biyogüvenlik Anabilim Dalı

Aralık 2022

Integrated Multi-Omics, Mechanistic Machine Learning And CRISPR Approach For
Polyhydroxybutyrate Production

Blaise Manga ENUH

DOCTORAL DISSERTATION

Department of Biotechnology and Biosafety

December 2022

Integrated Multi-Omics, Mechanistic Machine Learning And CRISPR Approach For
Polyhydroxybutyrate Production

A thesis submitted to the Eskisehir Osmangazi University Graduate School of Natural
and Applied Sciences in partial fulfillment of the requirements for the degree of the Doctor of
Philosophy in Discipline of **Biotechnology and Biosafety** of the Department of
Biotechnology and Biosafety
by

Blaise Manga ENUH

Supervisor: Assoc. Prof. Dr. Pınar AYTAR ÇELİK

“This thesis was supported by Eskişehir Osmangazi University Scientific Research Projects
Commission with project number FDK-2022-2468”

December 2022



To my dear FATHER...

ETHICAL STATEMENT

I hereby declare that this thesis study titled “Integrated Multi-Omics, Mechanistic Machine Learning And CRISPR Approach For Polyhydroxybutyrate Production” has been prepared in accordance with the thesis writing rules of Eskişehir Osmangazi University Graduate School of Natural and Applied Sciences under academic consultancy of my supervisor Assoc. Prof. Dr. Pınar Aytar Çelik. I hereby declare that the work presented in this thesis is original. I also declare that, I have respected scientific ethical principles and rules in all stages of my thesis study, all information and data presented in this thesis have been obtained within the scope of scientific and academic ethical principles and rules, all materials used in this thesis which are not original to this work have been fully cited and referenced, and all knowledge, documents and results have been presented in accordance with scientific ethical principles and rules. 21/12/2022

Blaise Manga ENUH

Signature

ÖZET

Halofilik organizmalar, yüksek değerli biyotemelli ürünler üretebilme yetenekleri nedeniyle ilgi çekmektedirler. Ekstremofiller, endüstriyel biyoreaktörlerde aşırı koşullara dayanabilirler ve bu da kontaminasyona çok daha az maruz kalmasını sağlar. Bu yüksek katma değerli ürünler arasında enzimler, polihidroksialkanotlar (PHAs), ektoin, vanilik asit gibi ürünler bulunmaktadır. PHAs, besinin sınırlandırıldığı koşullar altında depo polimeri olarak sentezlenirken, ektoin gibi osmoprotektanlar ise yüksek tuzluluk koşullarına cevaben üretilirler ve tuz stresine karşı koruma sağlarlar.

Genom ölçekli metabolik modelleme (GEM) çalışmalarının uygulamaları çeşitlidir ve araştırmacılar, biyolojik sistemleri çalışmak için çok sayıda yenilikçi yöntem uygularlar. Özellikle, GEM'lerin kısıt tabanlı analizi, organizmaların genotip-fenotip ilişkilerini araştırmak ve istenilen bir işlevi değiştirmek için hipotez yönlendirilmiş metabolizma mühendisliği yapmak için kullanılabilir. Günümüzde, metabolik yolların ve gen silme etkilerinin tahmin edilmesine olanak tanıyan GEM içinde aynı zamanda gen amplifikasyonlarının kullanıldığı endüstriyel suş geliştirmede kullanılabilecek sayısal algoritmalar mevcuttur. Bu çalışmanın amacı, genomik yöntemler ve manuel derlenmiş GEM kullanılarak *Halomonas elongata*'da polihidroksibütirat (PHB) üretimini deneysel olarak analiz etmektir. Bu, PHB verimini artırmak için metabolik mühendislik stratejilerini araştırmayı mümkün kılacaktır.

Bu çalışmada, poli hidroksibütirat üretimi ve miktarını ölçmek için model organizma olarak *Halomonas elongata* suşu 153B kullanılmıştır. Bu bakterinin tüm genomu dizilenmiş, yapılandırılmış ve açıklanmıştır. İleri analizler, ortolog grupların kümeleri ve BLAST ile yorumlanan genlerin fonksiyonlarını keşfetmeye yardımcı olmuştur. Genomun içeriğinde; iyon taşıma, osmoregülasyon, oksidatif stres ile ilgili genler ile poli hidroksibütiratlar, sekonder metabolit kümeleri ve ksenobiotiklerin parçalanması için özelleşmiş genler bulunmaktaydı. Karşılaştırmalı genomik bir yaklaşım kullanılarak, *Halomonas elongata* suşu 153B'nin genomu ayrıca diğer 5 *Halomonas* suşu ve 7 türü ile de karşılaştırılmıştır. Karşılaştırma için kullanılan *Halomonas elongata* suşları HEK1, HAL363, BLLS135 ve MH25661 iken, referans genom

DSM 2581 idi. Karşılaştırmaya dahil edilen diğer türler ise *Halomonas caseinilytica*, *Halomonas. socia*, *Halomonas organivorans*, *Halomonas boliviensis*, *Halomonas zincidurans*, *Halomonas lutea* ve *Halomonas malpeensis* idi.

Annot edilmiş genom, *Halomonas elongata* ve bazı durumlarda yakın türe ait veriler kullanılarak literatür verileriyle sıkı bir şekilde manuel düzenlenmiş, standartlaştırılmış ve doğrulanmış bir GEM oluşturmak için kullanılmıştır. Model, büyümeyi, karbon kaynağı kullanımını ve osmoadaptasyon fenotipini, esansiyel genleri ve PHB, ektoin ve vanilik asit biyoproduktlarının üretim dinamiklerini taklit etmek için flux denge analiziyle analiz edilmiştir. CRISPR Cas benzeri gen silme çalışmalarının sonucu olarak ortaya çıkan fenotipik değişikliklerin nasıl PHB'nin biyosentezine etkilediği de araştırılmıştır.

H. elongata 153B'nin genel genom özellikleri ve birçok biyosistem için genleri, işlevleri ve yolları dahil olmak üzere ayrıntılı bir şekilde açıklandı. *H. elongata* 153B, diğer 4 rapor edilen türe ve referans türe göre değerlendirildiğinde *H. elongata* DSM 2581'e daha benzer özelliklere sahip olduğu görülmüştür. Ayrıca, *H. elongata* 153B suşunda yüksek osmolarite gösteren ortamlarda adaptasyon için önemli olan taşıma, osmoregülasyon, sinyalizasyon ve oksidatif stres ile ilgili birçok gen tespit edilmiştir. Bu genlerin bir koordine akış işbirliği ile adaptasyona yardımcı olduğu görülmüştür. *H. elongata* 153B'nin genomunun, ayrıca PHA, çeşitli vitaminler, ektoin, aromatik ve ilaç bileşiklerinin üretimine ilişkin genleri de içerdiği tespit edilmiştir.

GEM, büyük bir manuel kurgulama ve doğrulama işlemi sonrasında yeniden inşa edilmiştir. Modelin, iEB1239, sitosol, periplazm ve ekstrasellüler boşluk adında üç bölmeye sahip olduğu görülmüştür. İlk taslak modelinde 1534 metabolik maddenin, 2314 reaksiyonun ve 1239 genin olduğu görülmüştür. Modelin MEMOTE standartlaştırma skoru %71 olarak bulunmuştur. Bu model, kurulmuş minimal ortam ve biyomas reaksiyonu kullanılarak, *H. elongata* biyokimyasal ve osmoadaptasyon fenotip özelliklerini doğru bir şekilde tahmin edilmesini sağlamış ve ayrıca önceden test edilmiş ve test edilmemiş kaynaklar dahil 28 çeşit karbon kaynağında büyüme göstereceğini ortaya koymuştur. Bununla birlikte, GEM,

osmoregülasyonun inceliklerini doğru bir şekilde modelleme yeteneğine de sahip olduğundan PHB, ektoin ve vanillik asit üretiminin dinamiklerini de gösterebilmiştir.

Bu çalışmada elde edilen bilginin, halofilik genom, metabolizma ve fizyolojisini anlama, değerli enzimler elde etme ve biyoteknolojik araştırmalar için hızlandırma amacıyla güvenilir bir bilgi kaynağı olarak hizmet vereceği düşünülmektedir. Bunun yanı sıra, geniş ölçüde doğrulanmış GEM, CRISPR Cas mühendisliği ile PHB ve diğer biyoteknolojik ürünlerin verimini artırmak, ayrıca halofilik metabolik süreçlerde henüz keşfedilememiş mekanizmaları araştırmak amacıyla kullanılabilecek bir sistem olması açısından biyolojik bir platform sağlamaktadır.

Anahtar Kelimeler: *Halomonas elongata*, Tüm genom dizisi, Polihidroksiyalkanoatlar, Biyopolimerler, Halofil, Ekstremofil.

SUMMARY

Halomonas bacteria, which are halophilic gram-negative and aerobic, produce polyhydroxyalkanoates (PHAs), ectoine, and other valuable substances that can be utilized in various industries such as pharmaceuticals, food, chemicals, and cosmetics. These microorganisms have the ability to tolerate moderate levels of salinity and are classified as moderately halophilic.

The genome of *Halomonas elongata* strain 153B was analyzed and compared to the genomes of 5 other strains and 7 species. The annotated genome was used to create a genome-scale metabolic model (GEM), which was then validated using flux balance analysis and literature data. The analysis revealed that *H. elongata* 153B has more similar features to the reference strain *H. elongata* DSM 2581 compared to the other 4 reported strains. Additionally, several genes involved in transport, osmoregulation, signaling, and oxidative stress were identified as being important for adaptation to high osmolarity environments. These genes appear to work together in a coordinated network to allow adaptation. The genome of *H. elongata* 153B was also found to contain genes related to the synthesis of PHAs, vitamins, ectoine, and the degradation of aromatic and drugs compounds. The GEM was reconstructed following an extensive manual curation and validation process. The resulting model accurately predicted *H. elongata* phenotype features and showed growth on several carbon sources. Also, the GEM showed the ability to accurately model intricacies of osmoregulation as well as the dynamics of PHB, ectoine, and vanillic acid production. Halophile physiology can be studied with the help of the genome, which will also serve as a source of valuable enzymes and can expedite research for biotechnology applications. Furthermore, the extensively validated GEM provides a systems biology platform for use in further studies with CRISPR Cas engineering to increase the PHB yield and to investigate so far elusive mechanisms in halophilic metabolic processes.

Keywords: *Halomonas elongata*, Whole-genome sequence, Polyhydroxyalkanoates, Biopolymers, Halophile, Extremophile.

LIST OF CONTENTS

	<u>Page</u>
ÖZET	vi
SUMMARY	ix
LIST OF CONTENTS	x
LIST OF FIGURES.....	xiv
LIST OF TABLES.....	xvii
LIST OF ABBREVIATIONS AND SYMBOLS.....	xviii
1. INTRODUCTION AND PURPOSE.....	1
2. LITERATURE REVIEW	2
2.1. Halophilic organisms.....	2
2.2. Types of halophiles	2
2.3. Diversity of <i>Halomonas</i>	3
2.4. Adaptations to high salt concentration	3
2.5. Biotechnological importance.....	4
2.5.1. Antimicrobial compounds.....	4
2.5.2. Exopolysaccharide	5
2.5.3. Biopolymers (Polyhydroxybutyrates and Polyhydroxyvalerates)	5
2.5.4. Pigments.....	6
2.5.5. Biosurfactants	6
2.5.6. Compatible solutes.....	7
2.5.7. Other molecules	7

LIST OF CONTENTS (continued)

	<u>Page</u>
2.5.8. Enzymes.....	7
2.5.9. Bioremediation.....	8
2.6. Secondary metabolites.....	9
2.7. Next-generation genome sequencing	9
2.7.1. Definition	9
2.7.2. Whole-genome sequencing and comparative genomics in halophiles	11
2.8. Mechanistic modeling (Genome-Scale Metabolic Network Models).....	14
2.8.1. The Stoichiometric matrix	15
2.8.2. Analyzing the Stoichiometric matrix	16
2.8.3. Reconstruction process	18
2.8.4. Applications of GEMS.....	20
2.8.5. In silico systems metabolic engineering	21
2.9. The CRISPR Cas metabolic engineering tool	22
2.9.1. Approaches based on CRISPR tools.....	23
2.9.2. CRISPR Cas experimental considerations.....	24
2.9.3. CRISPR metabolic engineering in <i>Halomonas</i>	26
3. MATERIALS AND METHODS.....	28
3.1. Bacterial samples.....	28
3.2. Genomic DNA extraction and Sequencing	28
3.3. Genome Annotations.....	28
3.4. Comparative genomics	29

LIST OF CONTENTS (continued)

	<u>Page</u>
3.5. Phylogenetic classification.....	30
3.6. Bacterium strain and sequence.....	30
3.7. Genome-scale modeling.....	30
3.7.1. Draft metabolic model reconstruction	30
3.7.2. Biomass formulation.....	31
3.7.3. Model Benchmarking	31
3.7.4. Addition of annotations	32
3.7.5. Manual curation and gap analysis.....	32
3.7.6. Minimal medium.....	33
3.7.7. Model validation and analysis	34
3.7.8. Visualization	34
3.8. CRISPR Cas engineering	34
3.8.1. Plasmid design	34
4. RESULTS AND DISCUSSIONS.....	36
4.1. General genome features.....	36
4.2. Genomic features.....	37
4.3. Classification based on COG categories.....	40
4.4. Phylogenetic classification.....	41
4.5. Comparative genomics.....	42
4.6. Metabolism features.....	44
4.7. Transport systems.....	45
4.8. Osmoregulation	46

LIST OF CONTENTS (continued)

	<u>Page</u>
4.9. Oxidative stress	52
4.10. Polyhydroxybutyrates.....	52
4.11. Secondary metabolites biosynthesis clusters.....	53
4.12. Others	54
4.13. Genome-scale metabolic network modeling and analysis	55
4.13.1. General model features	55
4.13.2. Benchmarking.....	55
4.13.3. Added annotations	55
4.13.4. Model curation and gap analysis	56
4.13.5. Minimal medium.....	58
4.13.6. Biomass reaction.....	60
4.13.7. Validation on carbon source usage	62
4.13.8. Biochemical and osmoadaptation phenotypes.....	67
4.13.9. Gene essentiality	68
4.13.10. Production of bioproducts.....	69
4.14. CRISPR metabolic engineering	80
5. CONCLUSIONS AND RECOMMENDATIONS	81
REFERENCES	83
APPENDIX	83

LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
2.1. <i>Halomonas elongata</i> from TEM imaging	2
2.2. Sequencing generations and their properties	10
2.3. Description of working mechanisms of different sequencing generations.....	11
2.4. Reconstruction of GEMs. Development of the stoichiometric matrix, defining constraints, and the flux balance algorithm.	16
2.5. General process for the reconstruction of GEMs	19
2.6. The CRISPR/Cas system and the mechanism of DNA breaks.....	23
2.7. The CRISPR/Cas system mechanism by CRISPRi.....	25
3.1. The iterative GEM reconstruction process from the annotated genome to the final reconstruction.	33
3.2. pMB1-spect-PthrC3_25-eGFP plasmid map showing restriction sites for inclusion of other components.....	35
4.1. The allocation of specific subsystems within the genome of <i>H. elongata</i> 153B as determined by the RAST annotation process.	38
4.2. Circular genome visualization of <i>H. elongata</i> 153B whole genome showing open reading frames, contigs, GC skews, and GC content in that order.....	39
4.3. A circular representation of the genomic differences between <i>H. elongata</i> strains and other <i>Halomonas</i>	40

LIST OF FIGURES (continued)

<u>Figure</u>	<u>Page</u>
4.4. A phylogenetic tree showing organisms close to <i>H. elongata</i> strain 153B.....	42
4.5. Summary data on the comparative analysis of genomes in five <i>H. elongata</i> strains..	43
4.6. Metabolic pathway annotation of carbon metabolism in <i>H. elongata</i> 153B.....	45
4.7. Ectoine's biosynthesis and degradation pathway in <i>H. elongata</i> . Ectoine's biosynthesis and degradation pathway in <i>H. elongata</i>	48
4.8. Annotated PHB biosynthesis pathway in <i>H. elongata</i> 153B.....	53
4.9: Overall MEMOTE model benchmarking score for iEB1239.....	56
4.10. Escher map of the iEB1239 model showing main metabolic reactions.	57
4.11.KAAS metabolic pathway annotation from genome showing inability to synthesize Asparagine from precursors such as Oxaloacetate, or L-Aspartate.	59
4.12.Flux distribution in the iEB1239 model when glucose was used as the main carbon source in the minimal medium.....	64
4.13. Carbon sources showing growth with iEB1239	65
4.14. Flux distribution in iEB1239 when flux towards PHB biosynthesis is maximized	71
4.15. Phase plane analysis of PHB synthesis	72
4.16. Overexpression and Downregulation targets to increase the biosynthesis of PHB in <i>H. elongata</i> 153B	73

LIST OF FIGURES (continued)

<u>Figure</u>	<u>Page</u>
4.17. Flux distribution in iEB1239 when flux towards ectoine biosynthesis is maximized.....	75
4.18. Phenotypic phase plane of ectoine production.	76
4.19. Flux map of iEB1239 showing flux distribution towards ectoine biosynthesis in a oad mutant.....	77
4.20. Flux map of iEB1239 showing flux distribution towards ectoine biosynthesis in a Ppc mutant.....	78

LIST OF TABLES

<u>Table</u>	<u>Page</u>
2.1. Beneficial halophilic enzymes and their sources.....	8
4.1. General features of the <i>H. elongata</i> strain 153B genome.....	36
4.2. Statistics of genomes in other <i>Halomonas</i> sequencing projects.....	37
4.3. Value and percentage distribution of COGs in <i>H. elongata</i> 153B.	41
4.4. Genes in the subsystem for ectoine synthesis, hydroxy ectoine and ectoine uptake and catabolism.....	49
4.5. Genes predicted in the osmoregulation subsystem.....	50
4.6. Minimal media components used for simulations with iEB1239.....	59
4.7. Biomass equation representing high salinity conditions in the iEB1239 model.	61
4.8. Biomass equation representing high salinity conditions in the iEB1239 model (cont)	62
4.9. Growth on different carbon sources compared with other <i>Halomonas elongata</i> strains in the literature.....	66

LIST OF ABBREVIATIONS AND SYMBOLS

<u>Symbol</u>	<u>Meaning</u>
NaCl	Sodium Chloride
H ₂ O	Water
MgSO ₄	Magnesium sulphate
MgCl ₂	Magnesium chloride
NaBr	Sodium bromide
NaHCO ₃	Sodium bicarbonate
KCl	Potassium chloride
CaCl ₂	Calcium chloride
HCl	Hydrochloric acid
Na ₂ HPO ₄	Disodium phosphate
K ₂ HPO ₄	Dipotassium phosphate
pH	Power of hydrogen
°C	Celsius degree
L	Litre
mL	Millilitres
μL	Microlitres

1. INTRODUCTION AND PURPOSE

Halophilic organisms are attracting research interest because of their ability to synthesize many high-value bioproducts. As extremophiles, they can withstand extreme conditions in industrial bioreactors which enable their cultivation in media with relatively less contamination. Amongst these high-value products are enzymes, polyhydroxyalkanoates (PHAs), ectoine, vanillic acid, etc. PHAs are synthesized as nutrient storage polymers during nutrient-limiting conditions while osmoprotectants like ectoine are produced in response to high salinity and provide protection to intracellular proteins from salt stress.

Though interest in halophiles is growing in recent years, compared to common mesophilic organisms, there is limited research information to comprehensively understand their metabolism and their genotype-phenotype relationships. High-throughput whole-genome sequencing technologies facilitate the sequencing of genomes allowing for the characterization of an organism's entire gene pool. As a result, researchers can provide new insight into the organism's metabolic potential from its genome. Furthermore, the pool of omics data and information on molecular interactions have now enabled their consolidation in systems biology frameworks. Genome-scale metabolic network models provide a mechanistic approach to studying genotype-phenotype relationships in organisms and also investigate novel hypotheses that drive discoveries.

In this study, we cultured a newly isolated *Halomonas elongata* 153B and sequenced its genome to obtain insight into its metabolic potential to metabolize carbon sources, produce PHAs, and its adaptation to saline environments. We also compared its genome to other *Halomonas* species. Subsequently, we reconstructed and extensively curated a genome-scale metabolic network model to provide a reliable systems biology platform to guide CRISPR Cas metabolic engineering experiments to improve the yields of PHAs in *H. elongata* 153B.

2. LITERATURE REVIEW

2.1. Halophilic organisms

Halophilic bacteria can survive in environments of fluctuating salinity. In very low salt concentrations, their proteins can be denatured either through misfolding, aggregation, or precipitation. However, some halophiles can survive low salt concentrations and their mechanisms of reactivation have been associated with rapid retrieval of functional molecular dynamics in the proteome (Vauclare et al., 2020). *Halomonas elongata* is a moderately halophilic and Gram-negative rod-like bacteria (Figure 2.1) with many biochemical characteristics.

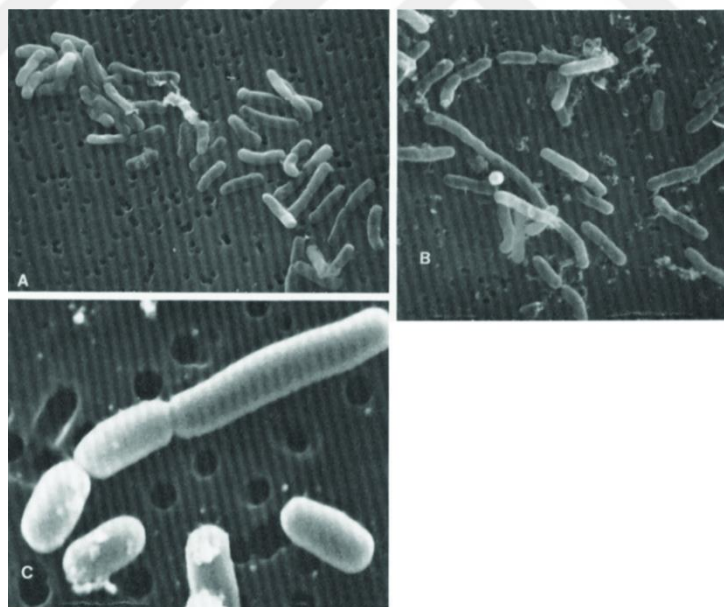


Figure 2.1. *Halomonas elongata* from TEM imaging (obtained from Vreeland et al., 1980)

2.2. Types of halophiles

All three domains, Archaea, Bacteria, and Eukarya have microbes that can survive high salt concentrations. Also, within these microbes, there is a rich metabolic diversity ranging from oxygenic and anoxygenic phototrophs, denitrifiers, methanogens, fermenters, aerobic

heterotrophs, and sulfate reducers. Generally, as salinity increases, the metabolic diversity also decreases due to the increase in energy cost for osmoadaptation (Oren, 2002).

2.3. Diversity of *Halomonas*

Halomonadaceae is a family of microorganisms that are found mostly in marine environments and can tolerate high levels of salt. They belong to the Gammaproteobacteria class and exhibit a diverse range of characteristics. Many of its present species were formerly classified in other genera because there were no main distinctive phenotypic features (Arahal and Ventosa, 2006). Gram-negative *Halomonas* bacteria are capable of aerobic respiration and thrive in moderately salty conditions. These microorganisms can survive in a range of environments, including hot and very hot, hypersaline, cold, alkaline, and acidic conditions. High concentrations of salts such as MgCl_2 , NaCl , CaSO_4 , MgSO_4 , and KCl are found in halophilic bacteria's salty and hypersaline habitats (Daoud and Ben Ali, 2020).

2.4. Adaptations to high salt concentration

To adapt to their high-salt and low-water-activity environments, halophilic bacteria employ a variety of strategies, including the accumulation of inorganic ions, such as Potassium ions, to balance osmotic differences with the environment (known as the "salt-in" strategy). They may also accumulate organic solutes, or "compatible solutes," including sugars, polyols, amino acids, N-derivatized carboxamides of glutamine, ectoine, betaines, and N-acetylated diamino acids to enable cellular processes to function properly (the "low salt-in" strategy). Additionally, these bacteria can store nutrients to survive in these harsh conditions (Gunde-Cimerman et al., 2018). Evolution has led to these species having metabolisms that are more efficient in high salt conditions and less efficient in low salt conditions (Pastor et al., 2013). Halophiles have also developed the ability to use a diverse range of carbon-based compounds as their sole source of carbon and to store high-energy polymers. One example of these compounds is polyhydroxybutyrate (PHB), a type of polyhydroxyalkanoate (PHA). PHAs are considered potential biodegradable alternatives to traditional plastics, which can contribute to pollution. These bacteria's unique properties make them a valuable resource for exploitation.

2.5. Biotechnological importance

One example of biotechnologically important compounds produced by halophiles is polyhydroxybutyrates (PHB), which have potential uses in various industries such as food, pharmaceuticals chemicals, and cosmetics. Other compounds produced by these organisms include extracellular enzymes adapted to high salt and stress environments, exopolysaccharides, and compatible solutes. These compounds are synthesized through unique metabolic pathways (Arahal and Ventosa, 2006; Çakmak et al., 2021). Exopolymers from halophilic microbes with varied chemical compositions and physico-chemical characteristics have been characterized, although with a poor production yield (Radchenkova et al., 2020). Despite these limitations in production, these chemicals are coveted by many industries that see their superior biodegradability, environmental friendliness, lesser toxicity, and versatility as a meaningful trade-off to make up for their cost of production. Below, we discuss further these advantages as they are fundamental to the widespread desire to adopt halophiles in biotechnology processes.

2.5.1. Antimicrobial compounds

Halophilic and halotolerant actinomycetes are known to be a source of bioactive metabolites such as antibiotics. So far, the small fraction of marine actinomycetes that have been studied, show potential for the production of beneficial bioactive compounds (Hamedi et al., 2013). It has also been observed that a higher proportion of industrially beneficial bioactive products are obtained from organisms in this group. A recent study with a sample of 15 isolates from different saline environments comprising *Streptomyces*, *Nocardiopsis*, and *Saccharopolyspora* showed different antibacterial, and antifungal activities against important pathogens (Tatar, 2021). Another study used the culture filtrate of *Streptomyces* sp. VPTSA18 isolated from a saltpan environment and tested against 11 and 2 bacterial and fungal pathogens respectively by well-diffusion. Antimicrobial activity was observed for all the pathogens tested for the ethyl acetate extract (Vijayakumar et al., 2012). The metabolites of halophilic bacteria have also shown antimicrobial activity against Multidrug-resistant bacteria that cause diabetic foot infections (Henciya et al., 2020) and also against pathogenic viruses (Suthindhiran et al., 2011).

2.5.2. Exopolysaccharide

Exopolysaccharides (EPS) are biotechnologically beneficial in food, pharmaceutical, cosmetic, wastewater treatment, and bioremediation sectors. This is primarily due to their diverse structures, properties, and functions. Where they can function as viscosifying, or stabilizing agents (Joulak et al., 2020). Many studies have shown that *Halomonas* can produce exopolysaccharides. Among *Halomonas* species, *Halomonas* strain CRSS (Poli et al., 2004) produces up to 2.9gg-1 dry cells, *Halomonas smyrnensis* SVD III (Deshmukh et al., 2017), *Halomonas ventosae* and *Halomonas anticariensis* (Mata et al., 2006), *Halomonas smyrnensis* K2 (Joulak et al., 2020), *Halomonas aquamarina* (NB2) and *Halobacillus trueperi* (NB7, NB8, NB9, NB10, NB11) (Başer and Yuksekdağ, 2019) have all been shown to produce different kinds of EPS. The nature of polymers produced is affected by the media composition especially the source of carbon. For example, acetate was shown to be the most efficient source of carbon for EPS production in *Halomonas* strain CRSS (Poli et al., 2004). Most EPS are heteropolysaccharides and composed of different combinations of molecules. These can be composed of galactose, mannose, glucose, glucosamine, fructose, and galactosamine (Joulak et al., 2020; Poli et al., 2004). Studies also reveal that optimizing culture conditions can lead to fold increases in EPS secreted (Deshmukh et al., 2017; Mata et al., 2006).

2.5.3. Biopolymers (Polyhydroxybutyrates and Polyhydroxyvalerates)

Halophiles are well known for their production of high-energy reserve compounds in the form of polyhydroxyalkanoates. These are produced in different structures ranging from small, medium to long chain lengths polymers. The biosynthesis pathways can be different because different precursors can be used, however, common to most is towards the end of the pathways having the PhaA, PhaB, and PhaC enzymes. The structure and composition of the polymer produced are influenced primarily by the PhaC enzyme which is the last in the biosynthesis pathway. Also, the type of polymer produced is influenced by the carbon source fed in the culture (Akhlaq et al., 2022). Synthesis of PHBs has also been observed using waste materials and biomass feedstock (Sirohi et al., 2020). These polymers are biodegradable, and relatively environmentally friendly compared to their synthetic counterparts. Presently, their industrial

applications are numerous ranging from use as bioplastic packaging in the food industry to the pharmaceutical industry where their low cytotoxicity and structural versatility have aided in the design of novel drug carrier systems and adjuvants.

2.5.4. Pigments

Pigments from bacteria have beneficial properties such as anticancer, antibacterial, antioxidant, and cytotoxic properties, and are also valuable products in the food, textile, pharmaceutical, and cosmetic sectors. Pigmented compounds from halophiles are also favored due to their capacity to survive osmolytes, high salt environment, and alkaline pH (Nawaz et al., 2021). Carotenoids are the most common pigments produced by the *Halomonas* though others like bacterioruberin, aminophenoxazinones, and zeaxanthin have also been reported (Athmika et al., 2021).

Antimicrobial properties of *Halomonas* pigments have been shown against *Streptococcus pyogenes*, *Klebsiella* sp., *P. aeruginosa*, *S. aureus*, and *Escherichia coli* (Ravikumar et al., 2016) and other important human pathogens (Fariq et al., 2019). *Halomonas* have also served as chassis to further engineer biosynthetic pathways to produce higher amounts of β -carotene (Rodríguez-Sáiz et al., 2007).

2.5.5. Biosurfactants

Biosurfactants are generally one of the most sought-after bioproducts from microorganisms. They are amphiphilic molecules that interact with other molecules at surfaces. Halophilic bacteria of the genus *Bacillus* can produce biosurfactants (Donio et al., 2013). In addition, in the production process for biosurfactants, cell recovery and recycling can be eliminated reducing the downstream associated costs (Kebbouche-Gana et al., 2013).

Halophiles produce various types of biosurfactants with applications in the pharmaceutical, cosmetics, and food industries (Donio et al., 2013; Fariq and Yasmin, 2020). The advantage of using halophiles for biosurfactant production over more standard producers

like *Pseudomonas aeruginosa* include their resistance to harsh bioreactor conditions and less contamination during the production process (Liu et al., 2019).

2.5.6. Compatible solutes

Compatible solutes are small organic molecules accumulated by halophilic microorganisms that enable the cell to balance the osmotic potential between the cell and its immediate highly saline environment. This strategy is known as the “low salt-in” mechanism for halophilic adaptation often using ectoine, betaine, choline, etc. (Gunde-Cimerman et al., 2018). These compounds play a role in function preservation in industrial processes as such are useful in cosmetics, food preservation, and pharmaceuticals. Other compatible solutes such as betaine, choline, hydroxyectoine, carnitine, and glycine are also produced by halophilic bacteria (Czech et al., 2018). *Halomonas elongata* is the bacteria of choice used for the industrial production of ectoine (Czech et al., 2018).

2.5.7. Other molecules

Halophilic bacteria can synthesize salt-stable fluorescent nanoparticles also known as quantum dots. This has been reported in a polyextremophilic bacteria, that can survive high salt concentrations, acidic environments, and high concentrations of the heavy metal cadmium. This shows that halophilic bacteria can be used for the production of nanoparticles with improved or novel characteristics and properties useful in biotechnology (Bruna et al., 2019).

2.5.8. Enzymes

Enzymes from halophilic organisms also have several uses in the industry. For example, the biocatalytic features of halophilic α -amylases are unique to them (DasSarma and DasSarma, 2015). These enzymes are fully functional in one or more extremophilic conditions like high temperatures, salinity, and pH. Due to this novelty and beneficial properties, their demand is steadily on the rise (De Lourdes Moreno et al., 2013; Kumar et al., 2016). Table 2.1 shows a list

of halophilic enzymes that are beneficial in biotechnology and the sources from which they were obtained.

Table 2.1. Beneficial halophilic enzymes and their sources (DasSarma and DasSarma, 2015).

Enzyme	Source(s)	Enzyme	Source(s)
Malate dehydrogenase	<i>Haloarcula/Salinibacter</i>	β -glucosidase	<i>Halothermotrix</i>
Nucleoside diphosphate kinase	<i>Halomonas/Halobacterium</i>	Catalase/peroxidase	<i>Haloferax</i>
α -amylase	<i>Haloferax</i>	RNase H1	<i>Halobacterium</i>
Carbonic anhydrase	<i>Dunaliella</i>	D-mannonate dehydratase	<i>Chromohalobacter</i>
Glucose dehydrogenase	<i>Haloferax</i>	Alkaline phosphatase	<i>Halobacterium</i>
Dihydrofolate reductase	<i>Haloferax</i>		

2.5.9. Bioremediation

Halophilic bacteria can survive in environments with high salt concentrations; therefore, they can produce enzymes and other products that are functional in these environments. In salt-contaminated environments that inhibit the growth of plants, halophiles have the potential to promote plant growth through the release of growth-promoting enzymes. This approach can serve to improve crop yields in areas where fertility is affected by high salt concentrations (Arora et al., 2014).

Aromatic compounds are a key source of energy for organisms that live in the soil. The aromatic compounds protocatechuate and catechol are metabolized by the beta-ketoadipate route, which makes use of ring-cleaving enzymes (Moreno et al., 2011). Halophilic bacteria use protocatechuate to break down a variety of environmental pollutants (Wells and Ragauskas, 2012) while gentisate and salicylate have been recognized as intermediaries in the breakdown of naphthalene by *Rhodococcus* sp. strain B4 (Grund et al., 1992). The creation and improvement of bioremediation methods to handle environments with high salinity and aromatic chemical contamination are currently of growing interest. Studying the possible use of this

group of extremophiles in saline environment cleanup can be done by identifying catabolic genes in ring compound breakdown in halophiles (Moreno et al., 2011).

2.6. Secondary metabolites

A rich source of novel bioactive chemicals with potential pharmaceutical uses as antibiotics, anti-tumor medications, or cholesterol-lowering medications is found in the secondary metabolism of bacteria and fungi. Microbiologists are increasingly turning to the genome sequencing of bacteria to uncover new therapeutic possibilities. Common biosynthetic loci that encompass all known types of secondary metabolite compounds include polyketides, aminoglycosides, non-ribosomal peptides, terpenes, aminocoumarins, bacteriocins, nucleosides, indolocarbazoles, lantibiotics, beta-lactams, siderophores, melanins, butyrolactones, and others (Blin et al., 2019).

2.7. Next-generation genome sequencing

2.7.1. Definition

Next-generation sequencing is a term used to describe sequencing technologies that were developed after the foundational Maxam-Gilbert chemical cleavage and Sanger sequencing technologies. The latter relied on radioactive/fluorescence labels coupled with PCR amplification and detected only known sequences. Automated Sanger sequencing is also considered a next-generation sequencing technology as it improves both the equipment and speed of the initial sanger sequencing method. Meanwhile, NGS technologies do not require previous knowledge of the DNA sequence, and as such are hypothesis-free reducing PCR-related bias (Slatko et al., 2018). NGS is also high-throughput and employs parallel sequencing that allows for massive parallel analysis of large genomic sequences in hours or days as compared to years for sanger sequencing.

The fundamental steps in next-generation sequencing protocol are fragmenting DNA/RNA into separate bits, including adapters, sequencing the libraries, and putting everything back together to create a genomic sequence. The idea is comparable to capillary

electrophoresis in concept. The key distinction is that NGS improves speed and accuracy while lowering sequencing costs by massively parallel sequencing the millions of fragments created. The average lengths of fragments created in short-read and long-read sequencing platforms are now common, but short-read sequencing is cheaper and sample preparation has relatively shorter times.

Within the NGS, the lengths of the reads produced give rise to the second and third-generation technologies which are carried out on different platforms (Figure 2.2). The first and second generations are collectively known as short-read sequencing technologies while the third generation is known as long-read sequencing technologies. The second-generation technologies require the preparation of amplified sequencing libraries before the sequencing proper is done, while the third-generation technologies do not require this step, and hence have shorter processing times (Kulski, 2016). Aside from different sample processing, there are different output processing software and pipelines which as a result, affects the quality of the results produced. Nevertheless, these technologies are not inherently competitors per se but are individually suited for different research purposes. The individual mechanisms through which these technologies work is depicted in Figure 2.3 below.

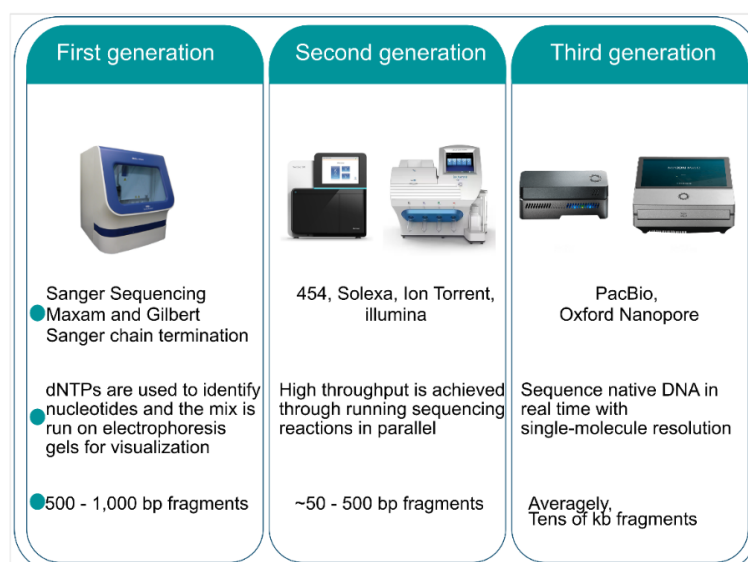


Figure 2.2. Sequencing generations and their properties

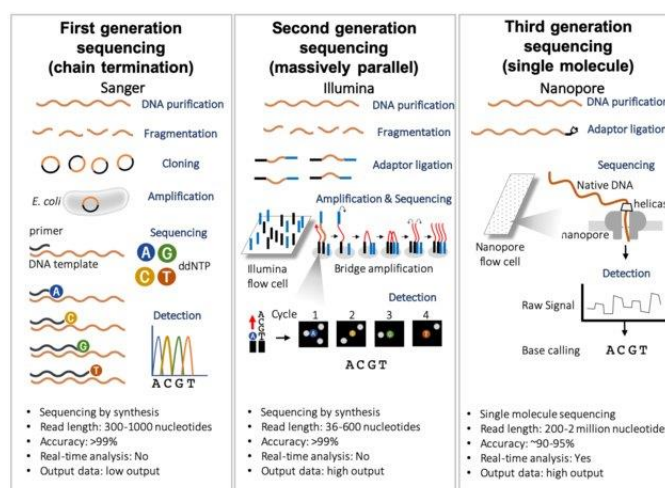


Figure 2.3. Description of working mechanisms of different sequencing generations

2.7.2. Whole-genome sequencing and comparative genomics in halophiles

NGS sequencing technologies have made possible the sequencing of whole bacterial genomes in a short time. Thanks to NGS whole-genome data, researchers have been able to gain insights into the structure and functions of individual cellular and organismal molecules, while understanding conservation and diversity between organisms. Despite these advancements, the pace of applications is different in each class of microbes. For example, in the *Halomonas*, at this time, a search in the gold database shows about 35 *Halomonas elongata* genomes, however only 4 were based on a study, and 5 were available publicly in the NCBI database. Furthermore, not many studies report an in-depth description of gene features. Comparing variations between different strains and species can be useful for contextualizing data, but at present, no study has analyzed the characteristics of the available strains. As more data is generated, it is important to compare this new information with existing data to identify emerging trends.

Whole genome sequencing of a bacteria genome enables the annotation of genes, and the prediction of their organization, characteristics, and functions. Whole genome sequencing can be used to study both cultivable and non-cultivable halophilic organisms in a habitat. Beyond identification, their genomes can be screened for beneficial enzymes and putative pathways important for industrial biotechnology (John et al., 2020). A comparison of the

predicted genes with those of related species also allows for contextualizing findings with respect to other genomic studies. Therefore, from a holistic point of view, pathways can be predicted as well as the potential of the organism to be exploited for biotechnological applications by describing the features specific to the organism (Thomas et al., 2019a).

Halophiles are a source of valuable genes that can be beneficial in genetic engineering to confer additional capabilities to other species. Genomic studies can help in the identification and isolation of these genes within the framework of comprehensive studies. Genes can be isolated and kept in vectors to transfer them to suitable hosts. For example, a genomic library of *Halobacillus trueperi* SS1 was prepared and constructed in plasmids. Recombinants were screened and some showed the ability to confer salt tolerance to *Escherichia coli*. This gene was identified to be a multidrug transport ATP binding/permease protein. Such methods can eventually be used as a medium for transfer to other organisms (Gupta et al., 2020).

By comparing the genomes of various organisms, researchers can gain insights into differences between them, track evolutionary patterns, and create effective strategies for industrial biotechnology in the future. Whole-genome sequencing and comparative genomics can enable the identification of enzyme orthologs, competing enzymes, etc that are valuable in designing optimal strategies for engineering (Ye et al., 2018). Both frameworks can be powerful tools to elucidate the functional genomics of bacteria and the evolutionary differences that exist between strains and species. Comparative genomics can help us understand the importance of genetic traits within species (by looking at the conserved functions), and the role they play in bringing forth the phenotypes of the species (de Castro et al., 2020). Generally, differences in genomes can be inferred from the differences in genome size, genetic features, and gene content in various functional categories (Chen et al., 2020).

While in the cases of organisms with secondary metabolism of interest, comparisons can reveal the presence of novel secondary metabolites that might be industrially beneficial (Chen et al., 2020). Tools such as Antibiotics and Secondary Metabolites Analysis Shell (antiSMASH) (Blin et al., 2019) are now used to accurately predict and analyze genes involved in secondary metabolite biosynthesis pathways. In a study, antiSMASH-predicted putative biosynthetic gene

clusters for secondary metabolites with high novelty and diversity in the *Paludifilum halophilum* genome, including biosynthesis of fusaricidin-E, mycosubtilin, streptomycin, zwittermycin A, colabomycine-A, and meilingmycin (Frikha-Dammak et al., 2021).

In the *Haloarchaea* compact genomes with high GC contents and large plasmids account for huge variation within them and support the possible existence of multiple operational taxonomic units in the halobacterium (DasSarma et al., 2019). Operational taxonomic units represent the lowest taxonomic level that can be assigned to a group of sequences based on a set similarity threshold. For example, a comparative genomics approach was used to investigate biosynthetic clusters as well as the clusters encoding the antibiotic Haloarcheocin (HalC8-like peptides) from a halophile. It was observed that the peptides could be classified into at least 4 different clades as the genetic conservation between the peptides was low. Amongst these peptides were uncharacterized major superfamily proteins, proteins with extracellular domains, DNA-binding transcriptional regulators, and transmembrane peptides. Further analyses showed that these proteins were associated with HalC8/HalC8-like biosynthesis (DasSarma et al., 2019).

In other cases, comparing genomes might reveal the homologs of enzymes within the same pathway which have underlying differences in amino acid composition but adopt a similar mechanism of action towards a substrate of interest; taking a common substrate and producing the same product (Tsallagov et al., 2019).

Whole genome comparisons enable us to identify the core genomic content of organisms in a species or clade, and this genomic content can be further analyzed for functions to see which of them are important within the clade. Specific differences need to be pointed out as well and possible reasons for these differences should be suggested (often, the differences are due to habitat) (Guan et al., 2019). Also, through comparative genomics, the comparisons of specific traits are possible across bacteria species and strains enabling the elucidation of mutation events such as insertion and deletion as divergence occurs within these organisms (Matsutani et al., 2021).

2.8. Mechanistic modeling (Genome-Scale Metabolic Network Models)

The field of sciences over the years has generated a lot of data to answer important isolated research questions and to investigate hypotheses. This contributed to the description of many metabolic pathways. Even more recently, the development of high throughput omics sequencing technologies has led to the generation of massive amounts of genomic, transcriptomics, proteomics, and metabolomics data. The currently massive repository of data remains largely unexploited but is a potential source of answers to many unanswered questions in research (Terzer et al., 2009).

The advent of the field of systems biology provides a foundation for the bottom-up exploitation of the massive amounts of data available. Specifically, genome-scale metabolic network models provide a framework based on mathematical formulations to represent the metabolism of an organism at the genome scale by making use of stoichiometric and constraint-based methodologies. The entire metabolic network is represented using stoichiometric matrices and clearly defined constraints on the fluxes of metabolic reactions. These are then resolved into a system of solvable differential equations that can be analyzed by advanced algorithms to predict optimal states from the information supplied in the model. The most common of these are flux balance analysis, pathway analysis, and some newer algorithms that are derivatives (Terzer et al., 2009).

Though there is a wealth of information on pathways and metabolism of organisms, even for the most well-studied organisms, the information is not near complete. This makes the beauty of computational methods which can use algorithmic approaches on data and make accurate and reproducible predictions with the limited information available. They have gained the interest of researchers who have applied them in many innovative ways over the years (Terzer et al., 2009).

2.8.1. The Stoichiometric matrix

As previously mentioned, the stoichiometric matrix is an abstraction of typically all the metabolic reactions present in a cell. The stoichiometric matrix is fundamental to genome-scale metabolic network approaches. The matrix is arranged systematically with the columns representing either a chemical or transport reaction, while values other than zero in the matrix represent the corresponding metabolites on the rows that participate in the reaction (Figure 2.4). These values represent the stoichiometric coefficient with negative values representing consumption and positive values representing production. The signs also provide an abstraction for reaction directionality. Hence the column provides a list of metabolites that participate in a metabolic reaction while the rows provide a list of reactions in which a metabolite participates. This comes in very handy when there is a need to check the mass balances of a metabolite in the model. In addition to this, constraints can be added to the matrix provided they can be mathematically defined. These constraints can be physicochemical, environmental, or self-imposed regulatory constraints in nature. (Terzer et al., 2009).

A system of differential equations is used to express mass balances in the system for metabolite concentrations. However, the rate of metabolism, regulatory events, or cell division is not the same hence to avoid the complexities associated with inputting the right rates, a pseudo-steady state is assumed where the concentrations of every metabolite do not change (Terzer et al., 2009). In other words, it is assumed that the quantity of a metabolite consumed is equal to the quantity of metabolite produced. This is critical for analyzing metabolic fluxes using the matrix.

To improve the predictive power of the model by limiting its solution space, it can be further annotated with constraints. Some of these constraints include reaction reversibility, compartmentalization of reactions, genetic and/or proteomic information, or kinetic parameters obtained from ^{13}C tracking experiments. These can serve to measure the changes in the concentration of metabolites that can be used to obtain the fluxes of the corresponding enzymatic reactions (Terzer et al., 2009). This is however more feasible for smaller pathways and difficult for entire metabolic networks. Linking genetic and proteomic information can be difficult and

complex because of the biological complexities. For example, enzymes can be made from different subunits located on different genes, or a reaction catalyzed by many different enzymes and vice versa. Literature and data from gene expression and knockout experiments are also important to understand these cellular processes (Terzer et al., 2009).

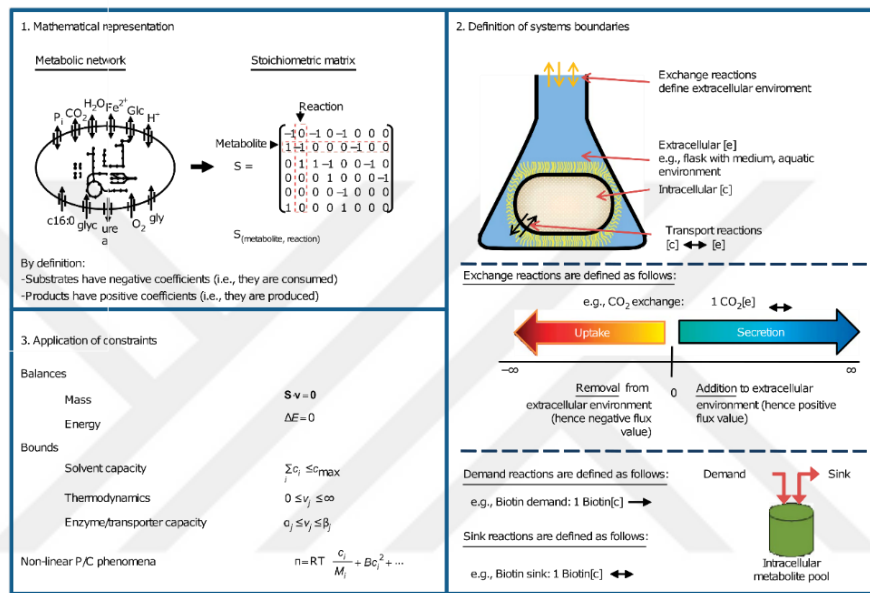


Figure 2.4. Reconstruction of GEMs. Development of the stoichiometric matrix, defining constraints, and the flux balance algorithm (Thiele and Palsson, 2010).

2.8.2. Analyzing the Stoichiometric matrix

Three approaches are commonly used to analyze stoichiometric networks. These are

1- To characterize the general structure of the network

Nullspace analysis to perform consistency validations on the metabolic network. These validations can apply to metabolites known as conserved moieties i.e their overall concentration should always be constant in the model e.g. NADH and NAD⁺, AMP, ADP, and ATP. This can also be used to analyze balanced metabolites. The implications of the results of a nullspace analysis have been explained (Terzer et al., 2009). Model consistency can also be inferred from minimizing and maximizing the fluxes of each reaction in the model.

2- Identify the distribution of fluxes

Linear programming (LP) is used to identify the distribution of fluxes in steady-state models. As the model is reduced to an LP problem, it is possible to identify points within the solution space that correspond to optima when using an objective function. The objective function usually is an artificially defined growth reaction or any other reaction within the model that we seek to optimize. To achieve this flux balance analysis (FBA) is used. The phrase "flux balance analysis" refers to the use of LP techniques to examine fluxes in the presence of balanced metabolite conditions. When using FBA, it's been noted that there are no unique optimal flux distributions but rather many different alternate optima can be a solution to the LP problem. This is one of the reasons why different solvers can give different FBA solutions. To reconcile this, other algorithms such as flux variability analysis (FVA) provide alternate optimal solutions that can be attained with the same objective function (Oberhardt et al., 2009; Terzer et al., 2009). Furthermore, if FBA is run in several iterations with changing constraints, it can generate a dynamic simulation in which phenotype changes can be predicted over longer periods and portrayed in a graph. In these types, kinetic information can be added to improve model accuracy.

3- Analyzing all the possible distributions of fluxes in the network

FBA can also be used iteratively to derive the essential genes in the network, other algorithms used for this purpose include MoMA and ROOM. Overall, the rationale behind these approaches is where MoMA and ROOM diverge from each other. MoMA has its roots in describing mathematically, the minimal responses to perturbations. To control gene expression, ROOM takes a more qualitative biological approach, presuming that a cell will ultimately try to reduce the number of large flux fluctuations (Terzer et al., 2009). Other analysis without the use of objective functions includes determining minimal cut sets and pathway analysis.

Challenges involved in using gene expression with FBA have been observed. For example, *E. coli* was found to express only about 50% of its genome in commonly used culture conditions. This implies that about half of the reactions in the stoichiometric matrix are supposed to be constrained. Their use is likely going to be more effective in the simplest of growth conditions. Nevertheless, novel algorithms such as regulatory FBA (rFBA) have been developed

that seem to provide more accurate predictions over FBA (Terzer et al., 2009). This brings us to a key assumption in constraint-based modeling which is the assumption of steady states.

In order to build GEMs, information is leveraged from several databases about enzyme kinetics, thermodynamics, genomics, proteins, metabolites, chemical reactions, etc. This information is curated and used to build up the metabolic reactions that make up the organism. Information from these databases helps promote the standardization of metabolic models through traceable annotations. Amongst the databases used BiGG (Schellenberger et al., 2010), BioCyc (Karp et al., 2019), MetaNetX (Moretti et al., 2021), CHEBI (Degtyarenko et al., 2008), KEGG (Kanehisa and Goto, 2000), HMDB (Wishart et al., 2007), RHEA (Alcántara et al., 2012), Inchikey (Heller et al., 2015), Reactome (Fabregat et al., 2018), Lipidmaps (Liebisch et al., 2020), and SEED EC-code (Seaver et al., 2020) are the most common.

Previously, the GEM building process used to be completely manual taking from months to years but has now been significantly reduced with the development of semiautomated network reconstruction pipelines that have reduced the process to a few weeks or months often requiring relatively minimal manual curation.

Genome-scale metabolic network models are increasingly used by researchers to achieve different goals; to investigate incomplete metabolic pathways, the synthesis of bioproducts, evolutionary research, and developing new hypotheses.

2.8.3. Reconstruction process

The reconstruction process of GEMs has seen significant improvements over the years moving from a completely manual process to now including semi-automated steps requiring less manual refinement. This has significantly reduced the labor-intensive nature of the process and the time it typically takes. There have also been many efforts to standardize the reconstruction process through conceptual reviews of the process. Moreover, a standard protocol has been published that outlines steps to reconstruct high-quality genome-scale metabolic networks using a bottom-up approach based on data from genomic and bibliomic databases

(Thiele and Palsson, 2010) (Figure 2.5). To generate a reconstruction, the genomic sequence of the organism is required, from which genes, enzymes, and metabolic functions can be inferred. Subsequently, a myriad of databases can be queried using the information from the genome to obtain further information about the functions, physiology, and biochemistry of the organism. Other well-researched organisms have organism-specific databases that make reconstruction easier.

The entire process consists in 4 main steps which include making a draft reconstruction, manual refinements, modeling the reconstruction, and evaluating the network. These main steps are further divided into 96 steps as outlined in Figure 2.5. However, as previously mentioned semi-automated reconstruction pipelines such as CarveMe (Machado et al., 2018), MODELSEED (Seaver et al., 2020), KBASE (Arkin et al., 2018), etc are reducing times and improving the accuracy of reconstructions.

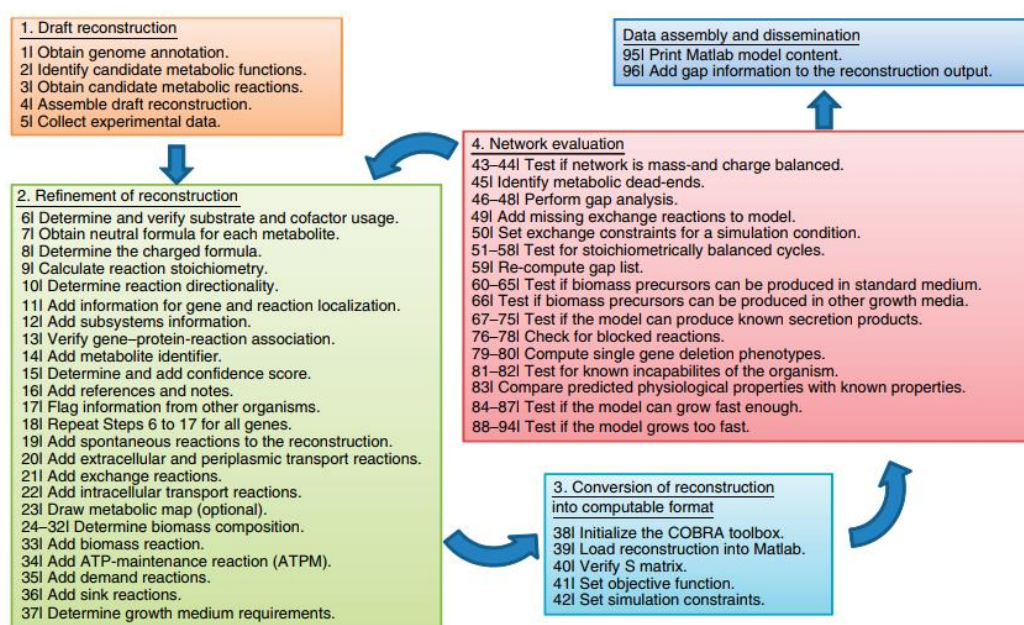


Figure 2.5. General process for the reconstruction of GEMs (Thiele and Palsson 2010)

Current challenges, however, lie in the discrepancies with genome annotations in different databases, and the lack of data for some organisms. Within the models, however, manual

curation is still required for substrate and cofactor usage by enzymes, reaction directionalities, intracellular pH, etc.

2.8.4. Applications of GEMS

The applications of GEMs are numerous as researchers are using them innovatively to study biological systems. Specifically, constraint-based analysis of GEMs allows for the investigation of genotype-phenotype relationships in organisms, as well as hypothesis-driven metabolic engineering to modify a desired function. As more high-throughput data becomes available, and more computational tools are developed, the applications of GEMs will continue to increase. Currently, computational algorithms allow for the prediction of pathways, gene deletion, and gene amplification effects that can be used for strain development (Kim et al., 2015). The integration of omics data and regulatory information also increases the scope of applications of GEMs (Occhipinti et al., 2018).

The emergence of GEMs has enabled a more systematic approach to genome engineering, unlike the traditional approaches that used several trials and errors before reaching an effective outcome. GEMs are systematic and organized knowledge bases with a variety of applications, such as transformation into computational models that evaluate and predict phenotypic states and the effects of genetic and environmental perturbations (Fang et al., 2020). Successes have been reported with genome-scale metabolic network models towards the improvement of yields of several bioproducts. In *Pseudomonas putida* KT244, a reconstructed GEM enabled investigation into its ability to synthesize PHAs and degrade aromatic compounds (Sohn et al., 2010). Another study used GEMs to investigate in *Caldicellulosiruptor bescii* the bottlenecks in the production of bioproducts using plant biomass. Mechanistic insights revealed optimal ethanol production was driven by the redox balance, the production of ATP, proton translocation mediated through the interplay of an ATP synthase, a bifurcating hydrogenase, a bifurcating NAD- and NADP-dependent oxidoreductase, and a membrane-bound hydrogenase. Validation of these predictions in the model showed close to a 2-fold increase in the yield of ethanol (Zhang et al., 2021). In Actinomycetes, GEMs are actively used as a mechanistic approach to gene manipulation for the production of antibiotics (Mohite et al., 2019). In

extremophilic *Chromohalobacter canadensis*, GEMs have shown the potential to model carbon usage phenotypes as well as dynamics involved in the biosynthesis of polyhydroxybutyrates (Enuh et al., 2022).

Other promising results involved improved yields of Lutein in *Auxenochlorella protothecoides* (Mekanik et al., 2022), using a threonine bypass in *Escherichia coli* to improve the PHB synthesis yields (Lin et al., 2015), Exopolysaccharide production in *Brevibacillus thermoruber* 423 (Yaşar Yildiz et al., 2019). Using these approaches, it is expected that the synthesis of bioproducts will become cost-effective as better engineering approaches will be implemented.

2.8.5. In silico systems metabolic engineering

For example, alongside engineering tools, knowing the genomics of organisms permits the elucidation of genes coding for important phenotypes that can be modified to improve the yields of bioproducts in the context of NGIB. This has been achieved through the manipulation of morphology (Jiang et al., 2017), the inclusion of de novo pathways and fine-tuning transcription (Ma et al., 2020), etc. Genomic information and genetic engineering tools, including CRISPR/Cas9, have allowed for the development of unconventional cell factories using species that were previously not considered for this purpose (Chen et al., 2018; Qin et al., 2018; Yin et al., 2014) showing massive future industrial potential.

Halomonas species can be cell factories for several beneficial products especially as their adaption to extreme conditions make them suitable for industrial use (Chen et al., 2018; Tohme et al., 2018). For example, they can produce high amounts of PHB even in unsterile and continuous fermentation processes (Tan et al., 2011) as well as use cheap and readily available wastes as feedstock (Kucera et al., 2018). Different pathways for PHB synthesis share a common precursor either acyl-CoA or acetyl-CoA. One way to determine the direction of the pathway is by considering the type of carbon source being utilized and the characteristics of the PhaC enzyme present in the organism (Thomas et al., 2019a). Through in silico systems metabolic engineering, it has been shown that these characteristics have the potential to improve the cell

factory capabilities of *Halomonas* species. Moreover, metabolic engineering can further enhance cost competitiveness (Ye et al., 2018).

2.9. The CRISPR Cas metabolic engineering tool

The CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) system is a defense mechanism used by bacteria to combat bacteriophages. *Streptococcus pyogenes*' CRISPR system has recently been made programmable for precise genetic material editing. A Cas9 (spCas9) nuclease, a scaffold RNA (tracrRNA), and a CRISPR RNA (crRNA), which together make up the gRNA, are the components of this system (Jinek et al., 2012).

The crRNA is created using exogenous DNA that has already been incorporated into the bacteria's CRISPR arrays. Additionally, it possesses a protospacer region that, upon a subsequent encounter with the cell, can identify its counterpart sequence in the same source. The formation of the gRNA/Cas9 complex results in double-strand breaks (DSBs), which can be repaired through either a homology-directed repair pathway (HDR) or non-homologous end joining pathways (NHEJ). Since most prokaryotes lack or have a limited NHEJ pathway, executing clean editing and mending double-strand breaks can be difficult. But using the HDR route, precise changes can be made by mutating the DNA or inserting a donor with either a single or double strand (Sander and Joung, 2014) (Figure 2.6).

Nowadays, CRISPR technologies play a significant role in the development of tools for next-generation metabolic engineering. These tools are used to improve strains to increase the competitiveness of industrial bioprocesses that generate biobased compounds. In this context, to boost commercial competitiveness against synthetic products, strategies aimed at improving the utilization of inexpensive substrates and energy-efficient methods are beneficial (Yu et al., 2019). Gene and genome editing, as well as gene expression modulation, are general approaches in metabolic engineering that have made use of CRISPR technologies. These have been used to generate novel strains, break down xenobiotics, enhance microbial physiology, boost microbial product outputs, etc. (Yu et al., 2019).

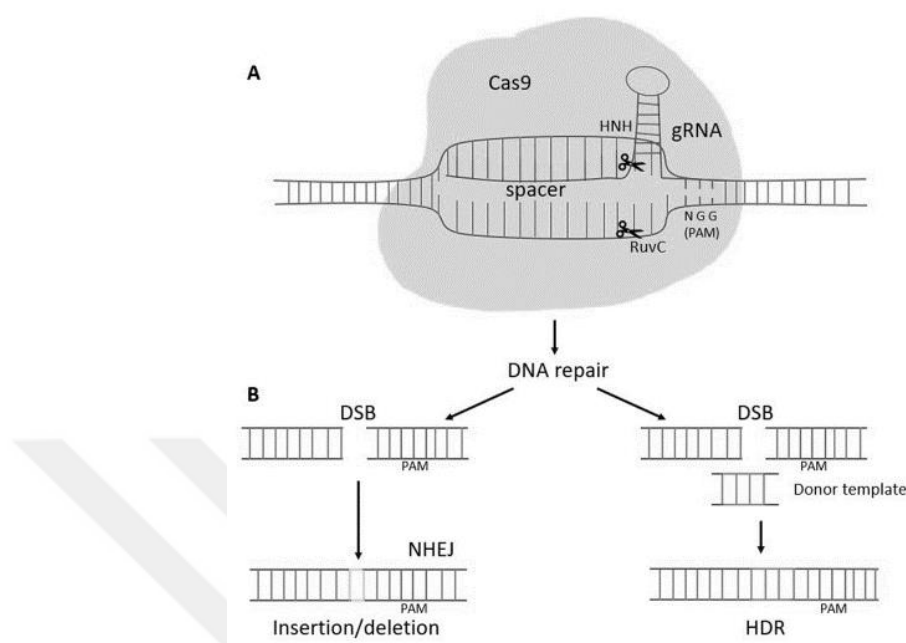


Figure 2.6. The CRISPR/Cas system and the mechanism of DNA breaks. A) The Cas9 protein showing the HNH and RuvC domains and the cutting site near the NGG PAM recognition sequence of the foreign DNA. B) The HDR and NHEJ DNA repair mechanisms (Lee et al. 2020).

2.9.1. Approaches based on CRISPR tools

i) Gene editing

The nuclease domain of the Cas protein is responsible for strand breaks in the genome. For gene editing studies, the widely used nuclease spCas9 is employed since it is well defined, only needs a gRNA that can be synthesized to create a targeted cleavage of genetic material, and functions in a variety of hosts (Freed et al., 2018; Jinek et al., 2012). Other Cpf1-based systems, CasX, and CasY (with highly compact systems uncovered from uncultivated microbes through metagenomics), and native Cas proteins (used to avoid heterologous Cas protein toxicity) are systems with potential for metabolic engineering but have not yet been fully characterized (Freed et al., 2018). The discovery and characterization of new nucleases will reveal new structural and functional information that could lead to the development of more efficient methods for gene editing with further study.

ii) Gene regulation

The ability of RNA polymerase to transcribe DNA is affected by the activity of activator and repressor proteins in cells, which are principally responsible for gene regulation. When a repressor prevents RNA polymerase from attaching to a promoter or from moving along a DNA strand that has to be transcribed, gene repression can occur. The suppression of a natural repressor or the introduction of a stronger promoter that increases the recruitment of RNA polymerase, on the other hand, causes gene activation.

The Cas9 protein's catalytic activity is carried out by its HNH and RuvC domains. Each of them breaks a strand, producing DNA double-strand breaks close to the target's PAM sequence Figure 2.7. The sequences encoding these domains can undergo mutations to change their catalytic activity (Adli, 2018). Complete cleavage activity is lost when the Cas9 nuclease protein is altered to catalytically inactive dCas9 (nuclease-dead Cas9). It only affects the beginning and end of DNA transcription when paired with gRNA, but it can still be directed to a target location. They have sequence-specific action and can either inhibit or activate DNA transcription (Schultenkämper et al., 2020). In CRISPR interference (CRISPRi), transcription is halted through the blockage of RNA polymerase while for CRISPR activation (CRISPRa) dCas9 binds upstream on promoters recruiting further binding activation components thereby activating transcription (Mougiakos et al., 2016; Yao et al., 2018).

2.9.2. CRISPR Cas experimental considerations

The CRISPR components are designed in-silico (Mougiakos et al., 2016). This can be carried out on openly accessible platforms or with the help of other specialists. However, it is crucial to keep in mind that prediction of the tracrRNA, which is essential for the type II CRISPR system to work, is frequently difficult due to structural variations and non-canonical base pairs that result from its interactions with the CRISPR RNA, Cas9, or itself. It is crucial to develop a successful tool because of the fundamental function that its structure plays. Considering this, computational models to better correctly predict tracrRNA are constantly being developed

(Chyou and Brown, 2019). It could also take a series of iterative wet lab experiments to try different combinations in order to obtain what works best.

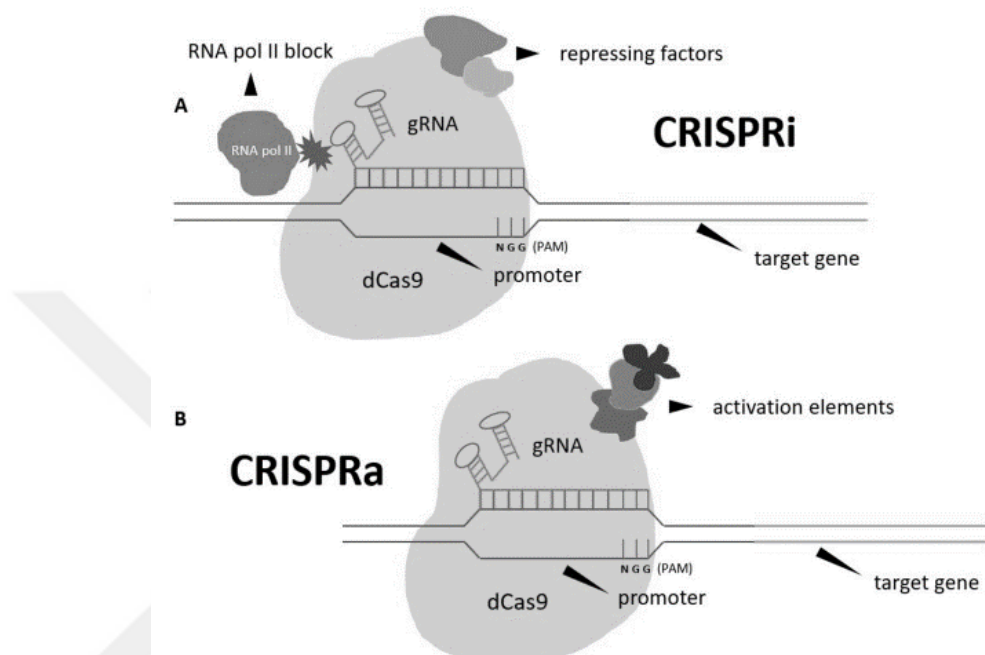


Figure 2.7. The CRISPR/Cas system mechanism by CRISPRi. A) dCas9 acquires repression factors and blocks RNA polymerase. B) CRISPRa showing activation by recruiting activation elements. The cleaving domains of Cas9 beside the PAM recognition site are inactive (Lee et al., 2020).

Vehicles like replicating plasmids (Jinek et al., 2012; Ungerer and Pakrasi, 2016) and suicide/integrating plasmids (Wang et al., 2019) can be used to introduce and sustain the CRISPR system in the host cell after the parts have been designed, synthesized, and assembled. Exploring the inherent mechanisms of microorganisms, such as competence and conjugation, as well as physical manipulation by electroporation or chemical induction, could be used to deliver vehicles. Host cell components, such as restriction enzymes, native CRISPR Cas systems, and integrons (via homologous recombination and employing transposons), which represent its defense mechanism against foreign DNA, must be taken into account in order to allow CRISPR component absorption (Freed et al., 2018).

2.9.3. CRISPR metabolic engineering in *Halomonas*

Genetic engineering has been more common in model microorganisms like *Saccharomyces cerevisiae* and *E. coli*. For non-model organisms like *Halomonas* sp. very few studies have been reported due to the scarcity of molecular tools. Recent studies have explored different CRISPR approaches to carry out genetic engineering in *Halomonas* to study their metabolism as well as improve the synthesis of high-value products.

Using CRISPR, osmoregulation mechanisms in *H. elongata* have been studied. A "Push, Pull, Block" approach for metabolic engineering was applied using CRISPRi to study targets in the biosynthesis of ectoine. Effects of mechanosensitive channels on cell growth and ectoine excretion, the effect of CRISPRi on lysine biosynthesis, and the effect of CRISPRi on Entner-Duodoroff pathway gene expression were examined. The PUSH results point to a metabolic overflow mechanism for glycolysis, where fluxes are consistently high even at lower salinities when no resources are needed for ectoine production. BLOCK with the aid of CRISPRi, lysine biosynthesis was suppressed, which significantly increased the production of ectoine. Ectoine excretion was partially mediated by the PULL mechanosensitive channels (Vandrich, 2019).

Genetic engineering approaches have been attempted as well to increase the yield of PHAs in *Halomonas*. *Halomonas* sp. TD01 was subjected to the CRISPRi system in order to reduce the expression of the *ftsZ* gene. The *prpC* and *gltA* genes were targeted by the CRISPRi system to channel more substrates into the synthesis of PHBV and PHB, and the *ftsZ* gene to produce longer cell sizes. The result was an increase in polymer synthesis (Tao et al., 2017).

More efforts at metabolic engineering were applied to increase the output of PHA biopolymers in *Halomonas*. The conversion efficiency of propionic acid to 3-hydroxyvalerate monomer fraction in random PHBV copolymers of 3-hydroxybutyrate and 3-hydroxyvalerate rose from about 10% to almost 100% when the gene encoding 2-methylcitrate synthase in *Halomonas* TD01 was removed (Fu et al., 2014).

To expedite *Halomonas* engineering, another CRISPR/Cas9 genome editing technique was developed. In *Halomonas*, a CRISPR/Cas9 system was designed for genome editing that is quick, effective, and scar-free. 100% was the method's greatest efficiency level. Shake flask experiments revealed a roughly 16-fold increase in the 3HV fraction of the copolymers, indicating the improved ability of a *prpC* mutant to manufacture PHBV. By cutting engineering times from roughly 6 months to three weeks, the genome engineering technique considerably quickens the studies on *Halomonas* engineering (Qin et al., 2018b). These studies have shown that CRISPR Cas can be used to carry out markerless metabolic engineering in non-model *Halomonas* species in an efficient, fast, and cost-effective way.

3. MATERIALS AND METHODS

3.1. Bacterial samples

Bacteria samples were revived on nutrient agar cultures. A loop inoculum was obtained and transferred to a minimal salt medium made of NaCl (96 g), $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$ (12 g), $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ (14 g), KCl (2.8 g), NaHCO_3 (0.008 g), $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$ (2 g), yeast extract (1 g), NaBr (0.32 g), peptone (5 g), and glucose (20 g). The culture was incubated for 3 days at 35 °C and 150 rpm in 250 mL Erlenmeyer flasks for polymer production (Gedikli et al., 2019).

3.2. Genomic DNA extraction and Sequencing

To obtain genomic DNA, 2 mL of bacterium suspension was taken from the bacterial cultures. The PureLink Microbiome DNA purification kit (Invitrogen) was used to extract the DNA following the manufacturer's guide. The extracted DNA was run on a one percent gel of agarose for 70 minutes at 100v. The resulting gel was viewed on a BIO-RAD gel documentation system (USA).

3.3. Genome Annotations

Genome sequencing was done at BM laboratories with the Illumina Next-generation sequencing platform. The sequence read quality was analyzed using FASTQC v0.11.9 and was trimmed using default settings. The genomic reads were then assembled in PATRIC (Wattam et al., 2018, 2017) using the default parameters of the auto assembly strategy and annotation was done with RASTtk v3.6.9 (Brettin et al., 2015) in the annotation web services provided by RAST (<https://rast.nmpdr.org>) and PATRIC (<https://www.patricbrc.org/>) (Brettin et al., 2015; Wattam et al., 2018).

EggNOG (Evolutionary Genealogy of Genes: Non-supervised Orthologous Groups) was used for orthology annotation. This was to support homology annotation with functional metadata for the interpretation of gene functions (Huerta-Cepas et al., 2019). KAAS (KEGG

Automatic Annotation Server) annotation server (Moriya et al., 2007) with BLAST and BBH (bi-directional best hit) was used to reconstruct metabolic pathways. Details in metabolic pathways, genes, and visualization were done in the KEGG database (<http://www.genome.jp/kegg/>) (Kanehisa and Goto, 2000) and BioCyc (Karp et al., 2019).

To confirm the features and functions of genes that were to be further researched, we manually used BLASTp (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) on the DNA sequences from the annotations. These sequences were compared to both the NCBI non-redundant database and the Swiss-Prot and UniProt databases (Boutet et al., 2007), and the information was combined to obtain the characteristics of proteins. To create visualizations of genomic features and characteristics, the circular genome viewer tool server (CGView) was utilized (Stothard et al., 2019). This tool allows for the creation of genomic maps by using annotated genome data from annotation pipelines.

3.4. Comparative genomics

We examined the genome of *H. elongata* 153B and compared it to the genomes of other strains of *H. elongata* that have been previously reported in the literature and have genome datasets available in NCBI. The accession numbers for all the strains and species that were used can be found in the appendix. The *H. elongata* strains that were included in the comparison were HEK1, HAL363, BLLS135, and MH25661 as well as the reference genome strain DSM 2581. Other species that were included in the comparison were *H. caseinilytica*, *H. socia*, *H. organivorans*, *H. boliviensis*, *H. zincidurans*, *H. lutea*, and *H. malpeensis*.

To compare whole genomes, we used the BLAST Ring Image Generator to align the reference genome DSM 2581 with both *H. elongata* 153B and the previously mentioned genomes (Alikhan et al., 2011). Then, we compared the whole genome orthologous clusters for 5 of the *H. elongata* strains using OrthoVenn2 <https://orthovenn2.bioinfotoolkits.net> (Xu et al., 2019). Investigations of secondary metabolite clusters were done with antiSMASH v6.1.1 (Blin et al., 2019).

3.5. Phylogenetic classification

The phylogenetic tree for genome-level comparison was obtained from PATRIC as part of its comprehensive genome report. Details of the tree building have been previously described. The tree was edited in FigTree v1.4.4 and Inkscape for better aesthetics.

To generate a phylogenetic tree in Molecular Evolutionary Genetics Analysis MEGA X, the 16S ribosomal subunit sequence from the annotated genome was taken and a sequence blast in the NCBI database was run. From the results, the top 35 were selected and were used for the tree (Kumar et al., 2018).

3.6. Bacterium strain and sequence

This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession **JAJQTQ000000000**. The version described in this paper is version **JAJQTQ010000000**.

3.7. Genome-scale modeling

3.7.1. Draft metabolic model reconstruction

CarveMe v1.4.1 (Machado et al., 2018) was used with default pipeline arguments to curate a draft reconstruction from the genome of *H. elongata* 153B (Figure 3.1). CarveMe is a fast and scalable tool for generating single-species and community models using a top-down approach. The pipeline relies on the BIGG database for metabolite and reaction information and has been shown to produce models that closely match experimental data on factors like gene essentiality and substrate use. To reconstruct the GEM, a genome file with annotations in the FASTA format was retrieved from the RAST server and input into the CarveMe pipeline using the command "\$carve --dna genome.fna" in the command line.

3.7.2. Biomass formulation

A biomass equation for cellular constituents is necessary to model cell growth. There hasn't yet been any other thorough investigation of *Halomonas elongata*'s biomass composition. As a result, data on biomass from additional Halophile *Chromohalobacter salexigens* species was employed. Biomass shared the same complete macromolecular composition as *Chromohalobacter salexigens*, including proteins, DNA, RNA, lipids, lipoteichoic acid, cell wall, iron, and metabolites. The GEM of *Chromohalobacter salexigens* was used to determine the amino acid content of proteins. Based on information from the *Chromohalobacter salexigens* genome, DNA and RNA components were determined. Other biomass macromolecule compositions were treated similarly to *Chromohalobacter salexigens* composition. Values of non-GAM and growth-associated ATP maintenance were obtained from *Chromohalobacter salexigens*.

A distinct biosynthetic reaction was defined for each macromolecule. Calculations of the molar fractions of the constituent parts of the macromolecules led to the stoichiometric coefficients for these reactions. The genome sequence of *H. elongata* 153B was used to estimate the molar fractions of amino acids for proteins and nucleotides for DNA and RNA, as previously proposed (Thiele and Palsson, 2010).

3.7.3. Model Benchmarking

To evaluate the model's adherence to standardized principles of model descriptions and identify areas for further improvement, we used the command-line version of MEMOTE v0.11.1 (Lieven et al., 2020). The results of the tests and annotations provided by MEMOTE guided our curation efforts, including addressing consistency issues, identifying metabolic gaps, assigning charges to metabolites, and adjusting reaction bounds (Figure 3.1). We repeated the MEMOTE analysis after each round of manual curation to ensure that the model achieved the highest possible score (Lieven et al., 2020).

3.7.4. Addition of annotations

In order to add more annotations to the model, ModelPolisher v2.0.1 was employed (Römer et al., 2016). This tool compares the entity IDs of the model with the BiGG model database and retrieves metadata that conforms to SBO terms (Schellenberger et al., 2010). All the relevant information and data about the corresponding instance are then incorporated into the initial draft reconstruction as annotations for each related entry in the BiGG database.

3.7.5. Manual curation and gap analysis

The first draft of the model was refined through manual steps using COBRApy v0.22.1 (Ebrahim et al., 2013). This process involved verifying the reactions in the model with literature evidence on *H. elongata* and adding missing reactions, genes, or metabolites that were mistakenly left out during annotation. To trace the presence of genes, we used annotation data from RAST, and for gene ontologies, we relied on information from EggNOG.

Scores for reactions that were added to the model were also recorded based on information obtained from literature sources. Metabolites that were blocked were identified using COBRApy (Ebrahim et al., 2013), and the identifiers were used to search the KEGG (Kanehisa and Goto, 2000) and BioCyc (Karp et al., 2019) databases to curate missing reactions and fill in metabolic gaps. If present, the reactions were checked for mass and charge balance and corrected if needed before being added to the model. The final model was tested for SBML compliance using the COBRApy library in Python 3.8. (Figure 3.1).

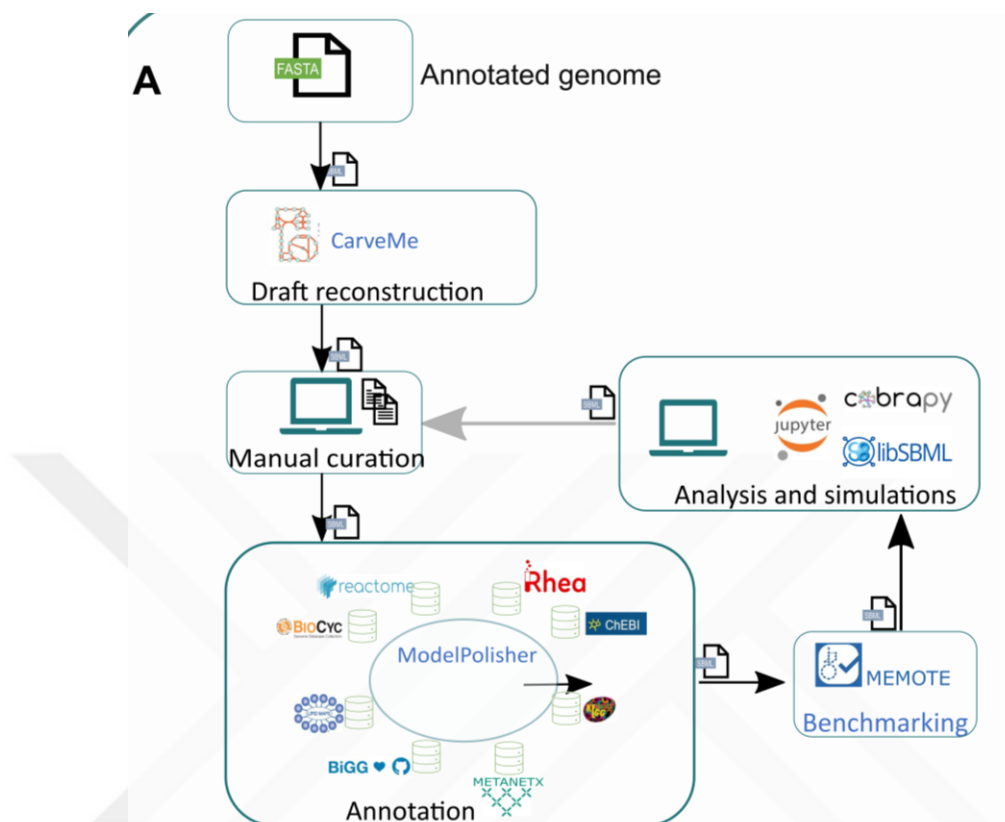


Figure 3.1. The iterative GEM reconstruction process from the annotated genome to the final reconstruction.

3.7.6. Minimal medium

To ensure the accuracy of metabolite essentialities in the medium, we conducted a thorough analysis using in-silico simulations. Specifically, we limited the availability of each metabolite and evaluated the model's performance. If the model did not grow when the availability of a particular metabolite was limited, we determined that the metabolite was essential. Based on these results, we compiled a list of essential media components and used them to create a minimal medium for the model.

3.7.7. Model validation and analysis

Using the minimal medium obtained from simulations, the in-silico growth capabilities of *Halomonas elongata* 153B on different carbon sources were examined. The lower bound for the carbon source's exchange reactions being studied was set at $-5 \text{ mmol gDW}^{-1} \text{ h}^{-1}$. Each carbon source was evaluated individually by allowing only the tested carbon source's exchange reaction and optimizing the model for growth through FBA (Orth et al., 2010). If the simulation had a flux value of zero, it was interpreted as the model's inability to grow on the carbon source being utilized. Additional analyses of reaction fluxes in optimal states were conducted using FVA, and the ability to produce bioproducts was assessed through phenotypic phase plane analysis with CAMEO (Cardoso et al., 2018) in python 3.8.

3.7.8. Visualization

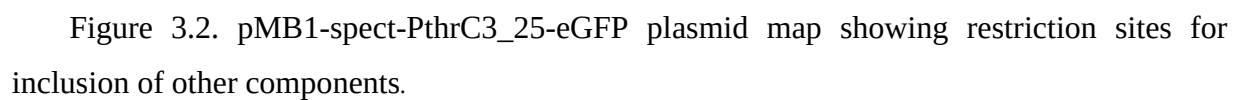
The software Escher was utilized to assist in the process of organizing and analyzing pathways in a metabolic model. This tool allows for the creation of visualizations of metabolic pathways by incorporating reactions, metabolites, and genes within the context of the organism's metabolism. The Escher Python package v1.7.1 (King et al., 2015) was specifically utilized to create customized metabolic maps of *H. elongata* 153B within Jupyter notebooks, as it is compatible with COBRApy. Additionally, ggplot2 (Wickham, 2009) in R studio version 4.1.1 (RStudio Team, 2015) was utilized to plot graphs for carbon source predictions.

3.8. CRISPR Cas engineering

To further implement the results of this research, overexpression with a CRISPR engineering tool was experimented with.

3.8.1. Plasmid design

The constructed plasmid pMB1-spect-PthrC3_25-eGFP used in this study is shown below. The plasmid information has been previously described (Anilionyte et al., 2018).



4. RESULTS AND DISCUSSIONS

4.1. General genome features

The whole genome was a total of 4,191,662 base pairs long, with an average Guanine-Cytosine content of 63.7%. There were 4080 coding sequences present. The general features of the strain ((Genomic features

Of all the approximately one-third genomic features annotated, the categories of stress response and membrane transport had a quite higher number of genes in their subsystems, with 86 and 119 genes respectively (Figure 4.1). This can be related to the importance of these systems as halophilic organisms are able to survive in extreme conditions due to their ability to effectively transport ions and solutes across the cell membrane and properly respond to stress. A genome map illustrating the circular organization of genes on the chromosome can be seen in Figure 4.2.

Table 4.2) are provided below.

Table 4.1) and context with other strains (Genomic features

Of all the approximately one-third genomic features annotated, the categories of stress response and membrane transport had a quite higher number of genes in their subsystems, with 86 and 119 genes respectively (Figure 4.1). This can be related to the importance of these systems as halophilic organisms are able to survive in extreme conditions due to their ability to effectively transport ions and solutes across the cell membrane and properly respond to stress. A genome map illustrating the circular organization of genes on the chromosome can be seen in Figure 4.2.

Table 4.2) are provided below.

Table 4.1. General features of the *H. elongata* strain 153B genome.

Characteristic	Value
Size	4,191,662
GC Content	63.7
N50	135344
L50	11
Contigs with protein-encoding genes (PEGs)	239
Subsystems	348
Coding Sequences	4080
RNAs	77

4.2. Genomic features

Of all the approximately one-third genomic features annotated, the categories of stress response and membrane transport had a quite higher number of genes in their subsystems, with 86 and 119 genes respectively (Figure 4.1). This can be related to the importance of these systems as halophilic organisms are able to survive in extreme conditions due to their ability to effectively transport ions and solutes across the cell membrane and properly respond to stress. A genome map illustrating the circular organization of genes on the chromosome can be seen in Figure 4.2.

Table 4.2. Statistics of genomes in other *Halomonas* sequencing projects.

Specie	Genome length	CDS	G+C content	Reference
<i>Halomonas shambharensis</i>	1,533,947	3847	63.60%	(Jadhav et al., 2021)
<i>Halomonas</i> sp. strain SF2003	4,358,421	4013	62.5%	(Thomas et al., 2019a)
<i>Halomonas malpeensis</i> strain YU-PRIM-29 ^T	3,607,821	3337	63.8%	(Athmika et al., 2021)
<i>Halomonas alkalicola</i> CICC 11012s	3,511,738	3295	67.67%	(Zhai et al., 2018)
<i>Halomonas</i> sp. Soap Lake #7	4,317,101	4,279	52.71 %	(Mormile et al., 2019)

<i>Halomonas boliviensis</i> strain kknpp38	4,960,786	4476	55%	(Sushmitha et al., 2021)
<i>Halomonas socia</i> CKY01	4,753,299	4627	64 %	(Park et al., 2020)
<i>Halomonas elongata</i> strain 153B	4,191,662	4080	63.7	This study
<i>Halomonas elongata</i> DSM 2581T	4,061,296	3,473	63.6%	(Schwibbert et al., 2011)
<i>Halomonas</i> sp. strain KM-1	4,992,811	4,220	63.9%	(Kawata et al., 2012)
<i>Halomonas</i> sp. Strain HAL1	4,347,024	3,439	54.1%	(Lin et al., 2012)
<i>Halomonas desertis</i> G11	3,963,288	3622	57.82%	(Neifar et al., 2019)

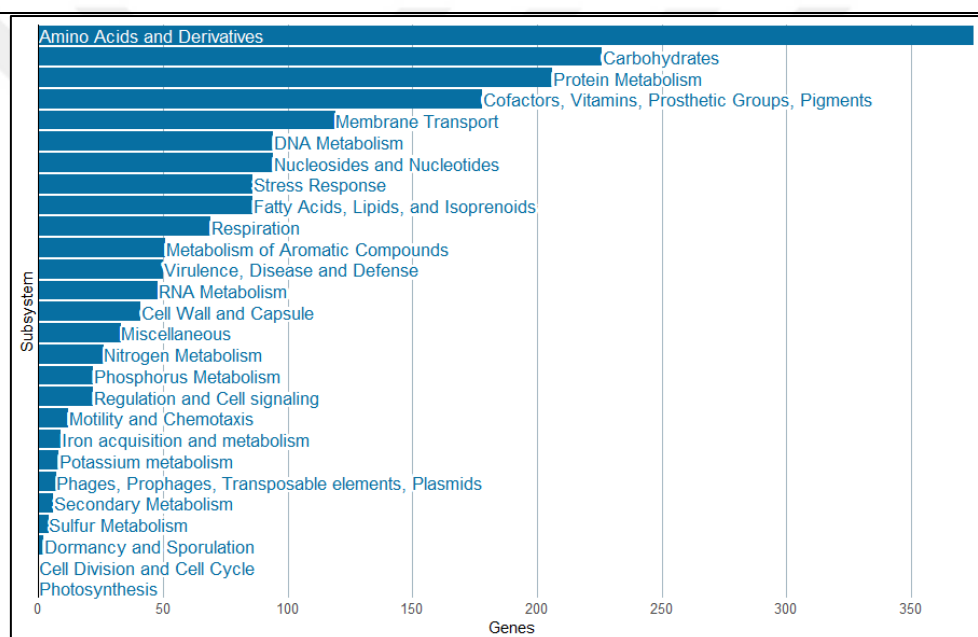


Figure 4.1. The allocation of specific subsystems within the genome of *H. elongata* 153B as determined by the RAST annotation process.

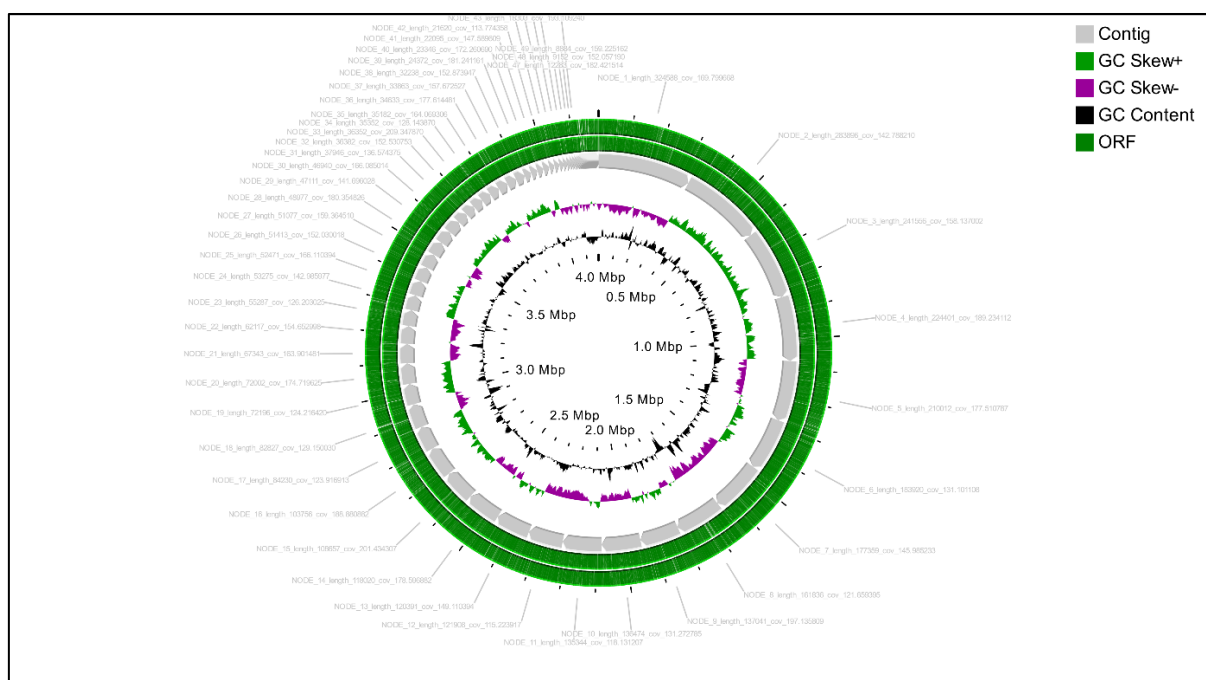


Figure 4.2. Circular genome visualization of *H. elongata* 153B whole genome showing open reading frames, contigs, GC skews, and GC content in that order.

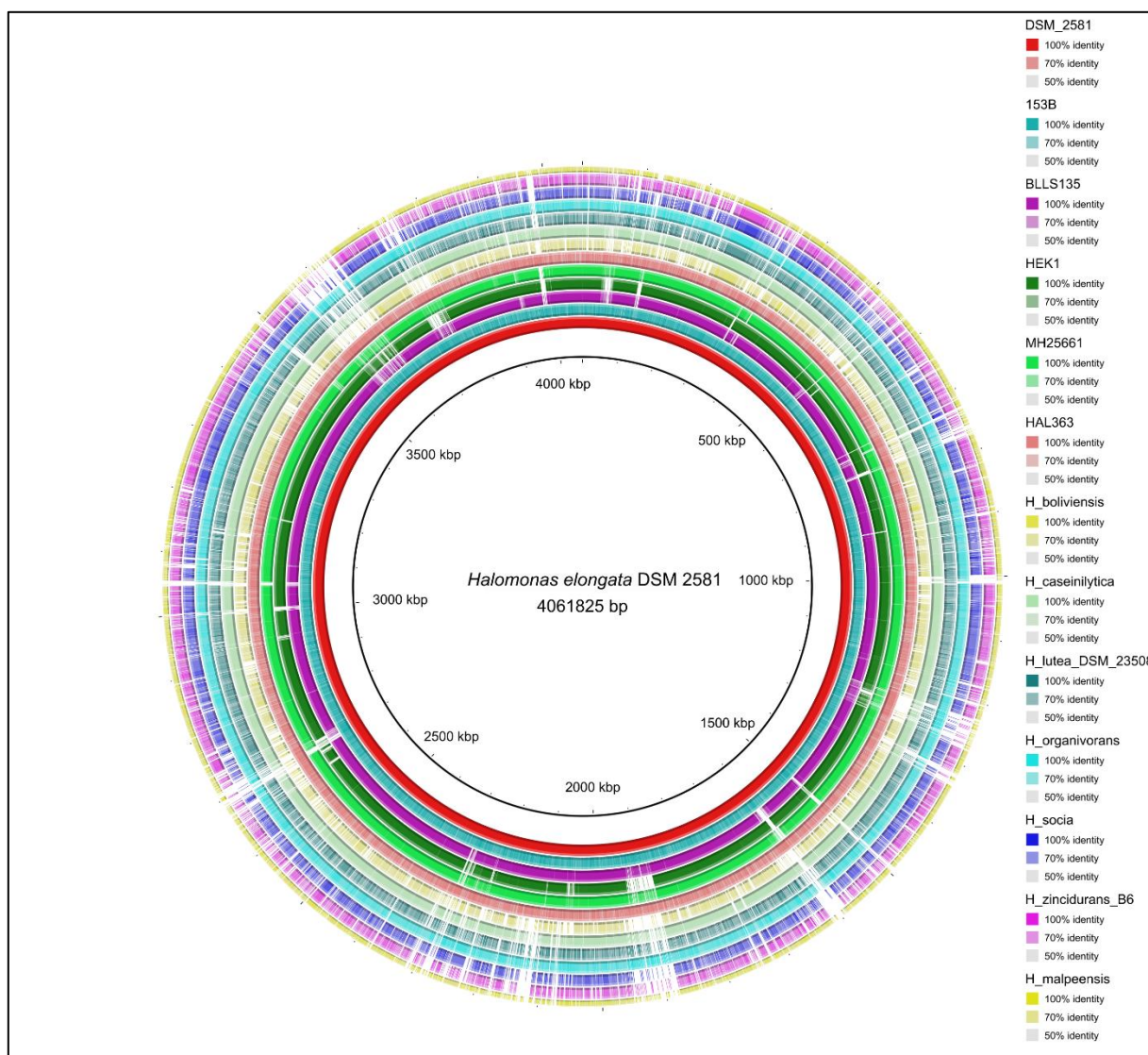


Figure 4.3: A circular representation of the genomic differences between *H. elongata* strains and other *Halomonas* is depicted. The reference *H. elongata* DSM 2581, is depicted in the center, while the subsequent genomes are arranged according to the legend.

4.3. Classification based on COG categories

There were similar patterns seen in the annotation based on subsystems and the annotation based on orthology. Of the genes annotated, 7.48% were involved in the transport and metabolism of inorganic ions. Further information about other orthologous clusters can be found in Table 4.3.

Table 4.3. Value and percentage distribution of COGs in *H. elongata* 153B.

Code	Value	Percentage	Description
E	371	10.13%	Amino acid metabolism and transport
K	327	8.93%	Transcription
P	274	7.48%	Inorganic ion transport and metabolism
C	255	6.96%	Energy production and conversion
G	254	6.94%	Carbohydrate metabolism and transport
M	215	5.87%	Cell wall structure and biogenesis and outer membrane
J	202	5.52%	Translation, including ribosome structure and biogenesis
H	170	4.64%	Coenzyme metabolism
L	167	4.56%	Replication, recombination, and repair
I	142	3.88%	Lipid metabolism
O	136	3.71%	Molecular chaperones and related functions
T	135	3.69%	Signal transduction
Q	127	3.47%	Secondary metabolites biosynthesis, transport, and catabolism
F	96	2.62%	Nucleotide metabolism and transport
U	85	2.32%	Intracellular trafficking, secretion, and vesicular transport
N	69	1.88%	Secretion, motility, and chemotaxis
D	46	1.26%	Cell division and chromosome partitioning
V	37	1.01%	Defense mechanisms
S	707	19.31%	No functional prediction
-	127	3.47%	Null

4.4. Phylogenetic classification

The classification of the *H. elongata* 153B genome in relation to other genomes is displayed in Figure 4.4. The *H. elongata* strains DSM 2581 and AAD6 were the most similar to the *H. elongata* 153B described in this work. A homology search of the 16S rRNA sequences of *H. elongata* 153B via the Blast program yielded a similarity of 98.90% with the DSM 2581

reference strain. The Blast ring image visualization also depicts a pattern that supports the blast outcomes (Figure 4.3).

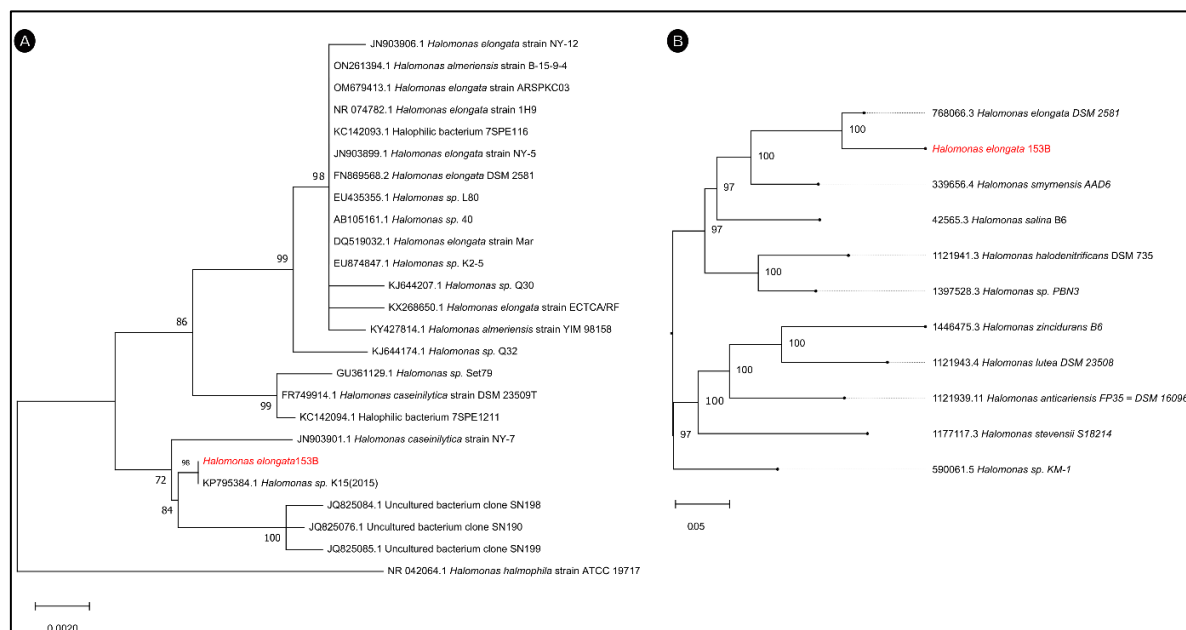


Figure 4.4. A phylogenetic tree showing organisms close to *H. elongata* strain 153B.

4.5. Comparative genomics

The genomic sequences of the following *H. elongata* strains MH25661, HEK1, HAL363, and BLLS135, the reference DSM 2581, and the species *Halomonas organivorans*, *H. zincidurans*, *H. malpeensis*, *H. boliviensis*, *H. caseinilytica*, *H. socia*, and *H. lutea* were used to generate a comparative visualization for the Blast Image Generator (Figure 4.3).

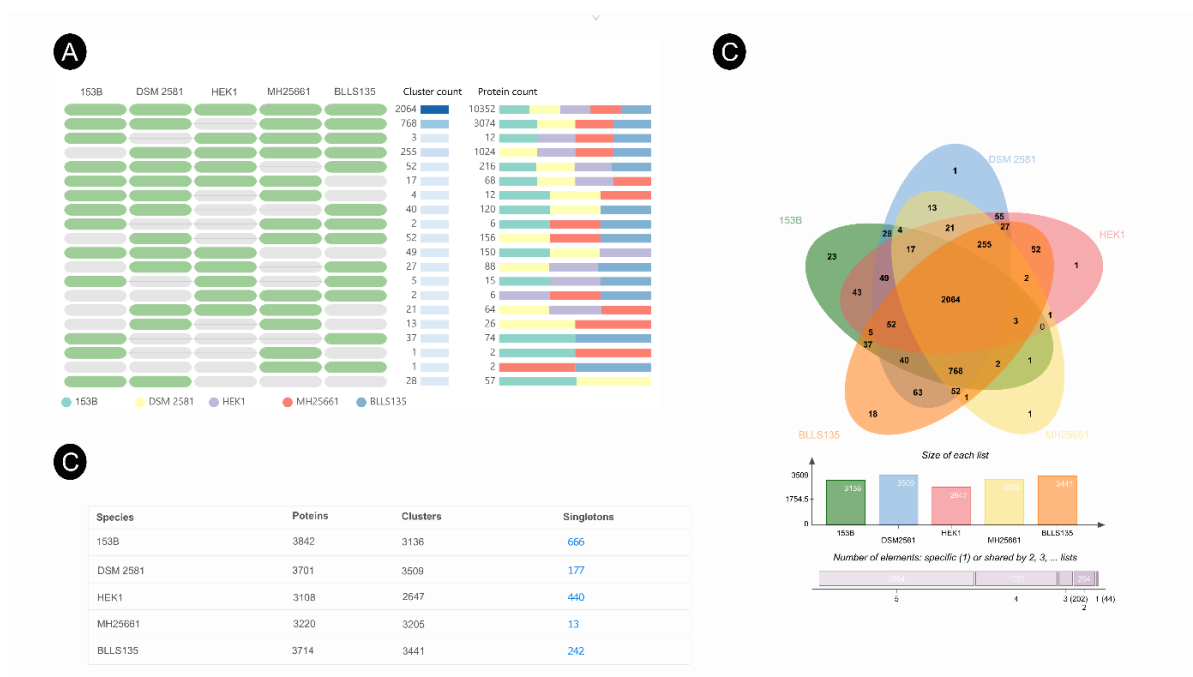


Figure 4.5. Summary data on the comparative analysis of genomes in five *H. elongata* strains. A) The distribution of the different clusters and protein counts in all 5 strains is shown. Clusters that are shared are shown in green, while absent clusters are shown in grey. The specific number of clusters shared between the strains is shown in the cluster count column while the distribution of proteins in the various clusters is shown in the stacked column. B) A visual representation of how the orthologs are distributed is depicted. C) More specific information on individual cluster counts in the various strains. Importantly, the singletons are shown as well.

In comparing the genomes to the reference genome *H. elongata* DSM 2581, that of *H. elongata* 153B was found to be the most similar. However, other strains of *H. elongata* exhibited areas of missing sequence, with the gaps becoming more pronounced when comparing different *Halomonas* species. Using OrthoVenn with default parameters, five *H. elongata* strains were clustered, with strain HAL363 excluded due to insufficient data. The analysis resulted in the identification of 3699 orthologous clusters, including 2064 core genome orthologs and 1661 clusters containing at least two species. There were also 2038 single-copy gene clusters identified, with strain 153B having the highest number of singletons (Figure 4.5).

4.6. Metabolism features

An analysis of the genome of *H. elongata* strain 153B revealed that it contains pathways for a variety of metabolic processes, including glycolysis, the Embden-Meyerhof-Parnas pathway, gluconeogenesis, pyruvate oxidation, the pentose phosphate pathway (with the exception of the reaction catalyzed by phosphogluconate dehydrogenase EC. 1.1.1.44), complete tricarboxylic acid cycle, the Entner-Doudoroff pathway, and the glyoxalate shunt. The genome also includes pathways for the metabolism of essential amino acids and the biosynthesis of various vitamins, including vitamin B6, riboflavin, thiamine, pantothenate, biotin, CoA, and folate. These results are consistent with previous studies that have identified similar pathways in *H. elongata* AAD6T (Thomas et al., 2019b) (Figure 4.6).

According to annotations, the *H. elongata* strain 153B can utilize a wide range of carbohydrates for energy production, such as sucrose, galactose, glucose, and fructose. Additionally, 35 genes involved in oxidative phosphorylation were identified as being necessary for this process. The carbon fixation pathways in photosynthesis, a complete Calvin cycle as well as a CAM (Crassulacean acid metabolism), and dark reactions for the conversion of CO₂ to malate were annotated. No light reaction genes were predicted. GABA (gamma-Aminobutyrate) shunt for GABA synthesis from succinate was predicted. Applications in the production of GABA for health purposes could be considered.

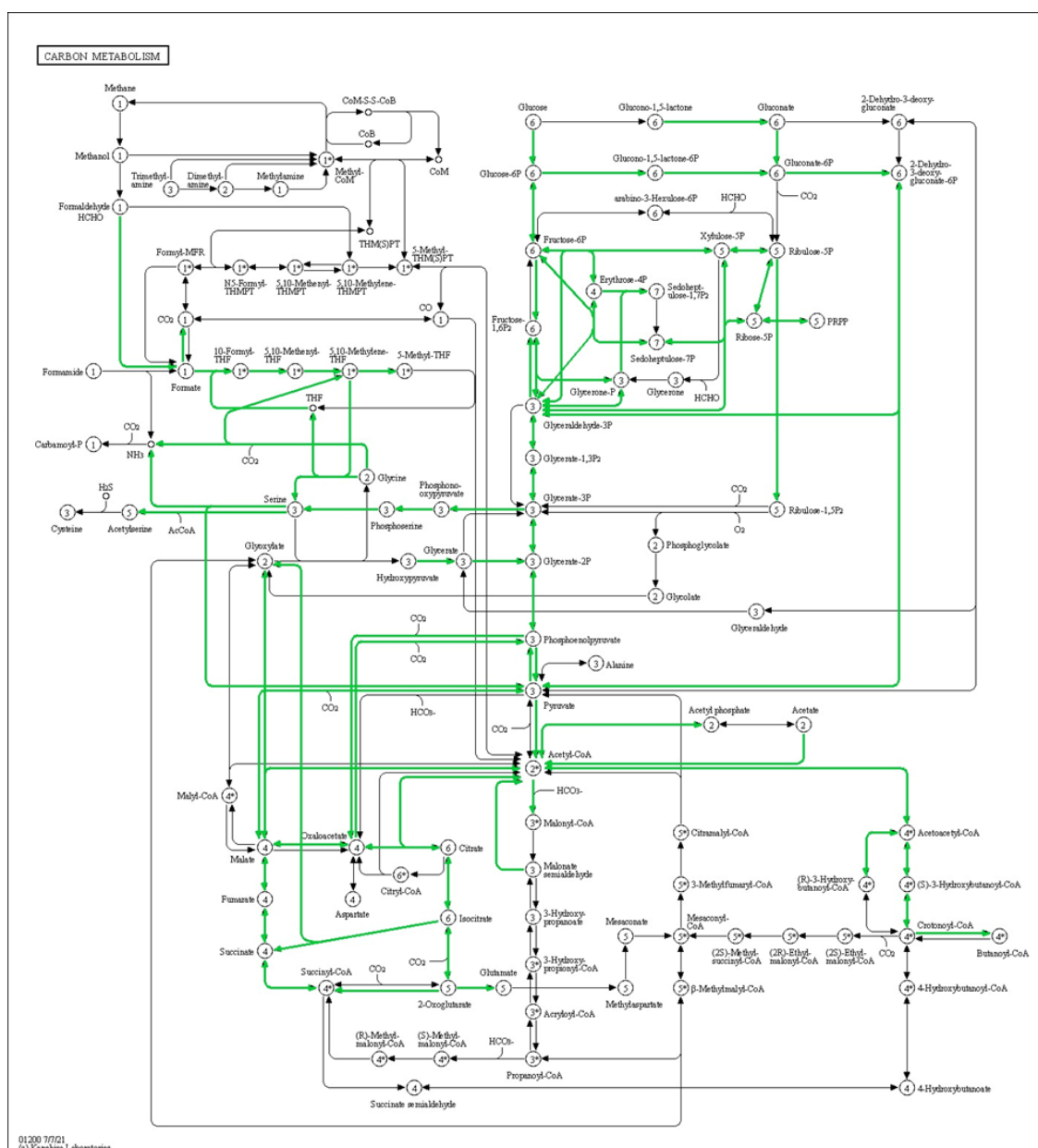


Figure 4.6. Metabolic pathway annotation of carbon metabolism in *H. elongata* 153B. The green path indicates the hypothetical route of metabolite flow along the path of genes that were annotated.

4.7. Transport systems

Transport systems are important for microorganisms to maintain homeostasis. Halophiles, such as *H. elongata* strain 153B, rely on various transport mechanisms to acquire

the necessary nutrients and expel by-products in order to maintain optimal osmotic pressure through the balance of protons and salt concentrations. Approximately 7.48% of the genome of *H. elongata* strain 153B has been identified as playing a role in the transport of inorganic ions, which is crucial for the adaptation to high salt environments.

Up to 130 ABC transporters were revealed in the genome amongst which ProX, ProW, and ProV were the glycine betaine/Proline transporter, and for osmoprotectant OpuBC, OpuBB, and OpuBA, 72 two-component systems, quorum sensing (42) and biofilm formation *Vibrio cholerae* (15), *Pseudomonas aeruginosa* (7), and *Escherichia coli* (20). Eukaryotes lack Tripartite ATP-independent periplasmic (TRAP) transporters, but these transporters are present in bacteria and archaea. TRAP transporters serve as a means of transporting solutes and are essential for the proper functioning of microorganisms. Halophilic transport systems, however, tend to differ from those found in other microorganisms, likely due to the need to maintain optimal homeostasis. The *H. smyrnensis* AAD6T strain, for example, requires the presence of sodium and chloride ions in order to develop properly (Poli et al., 2013). Halophiles, or organisms that thrive in high salt environments, often have very active Na^+/H^+ antiporters which help to maintain low intracellular sodium levels and pH balance.

It is worth mentioning that intracellular cations are more commonly potassium ions rather than sodium ions, and the most common anions are chloride ions (Daoud and Ben Ali, 2020). For potassium homeostasis subsystems annotation revealed several features: Glutathione-regulated potassium-efflux system protein KefB, FKBP-type peptidyl-prolyl cis-trans isomerase SlyD (EC 5.2.1.8), cAMP-dependent Kef-type K^+ transport system, Potassium efflux system KefA protein, Potassium voltage-gated channel subfamily KQT, and Glutathione-regulated potassium-efflux system and associated functions. Annotated figures are provided in the appendix.

4.8. Osmoregulation

Adaptation to extreme environments by halophiles often involves the use of compatible solutes, which are substances that are easily dissolved in water, have a low molecular weight,

do not alter pH or disrupt cell metabolism, and cannot pass through the cytoplasmic membrane without the assistance of specialized transporters. Examples of compatible solutes include ectoine, glycine, trehalose, and betaine. The concentrations of these solutes within cells can change in response to variations in external salinity (Daoud and Ben Ali, 2020). In order to survive and thrive, bacteria must be able to quickly adjust to changes in the osmotic pressure of their environment by ensuring that their internal osmotic pressure is equal to that of the external environment. This helps maintain the necessary turgor pressure for their normal growth. To do this, they can utilize these compounds which they can either produce themselves or take in from the outside through specialized transport proteins (Ongagna-Yhombi et al., 2015).

Ion-based osmoregulation, which involves the regulation of water and electrolyte balance in cells, relies on proteins such as *trkh* and *trka*. These proteins are part of the Trk potassium uptake system and are involved in the uptake of potassium under hyperosmotic conditions. Genomic analysis revealed that *TrkH* and a locus of *TrkA* are physically close to each other, while another gene for *TrkH* is located elsewhere in the genome. In a study involving *H. elongata*, the absence of both *trkh* and *trki* transporters did not result in a buildup of potassium (Kraegeloh et al., 2005). Both *TrkH* and *TrkI* are dependent on one type of *trka* protein, as K transport is absent in a *TrkA* deletion mutant. During reverse transcriptase PCR and northern hybridization analysis of the *TrkAH* locus, it was found that *TrkAH* was transcribed together as well as a monocistronic transcript containing only *trkA* (Kraegeloh et al., 2005). Consequently, these genetic elements may serve as a crucial factor in preserving ionic equilibrium within *H. elongata* strain 153B due to their direct connection to halophilic osmotic adaptation as potassium transporters (Kraegeloh et al., 2005). The uptake of potassium and hydrogen ions along with the release of sodium could potentially act as a means of maintaining osmotic balance within *H. elongata* strain 153B.

Genome analysis showed that the ectoine biosynthesis pathway had the *EctA*, *EctB*, *EctC*, and *EctD* genes in solute-based osmoregulation, as demonstrated in Table 4.4. Ectoine is hydrolyzed to become N-acetyl-l-2,4-diaminobutyric acid (*DoeA*), which is then deacetylated to form diaminobutyric acid, which is how ectoine (*doe*) gets degraded (*DoeB*). Aspartate or the ectoine synthesis route can be re-entered by diaminobutyric acid in *H. elongata*, creating a cycle

of ectoine degradation and synthesis (Figure 4.7). Genomic analysis showed that non-salt-loving bacteria lacking the ability to synthesize ectoine serve as the primary host organisms for the ectoine degradation process (Schwibbert et al., 2011). The osmoregulated transporter TeaABC allows *H. elongata* to get ectoine from its surroundings in addition to relying on synthesis. Ectoine is constantly being lost to the medium in strains where the TeaABC-mediated ectoine absorption is disrupted (Vandrich et al., 2020).

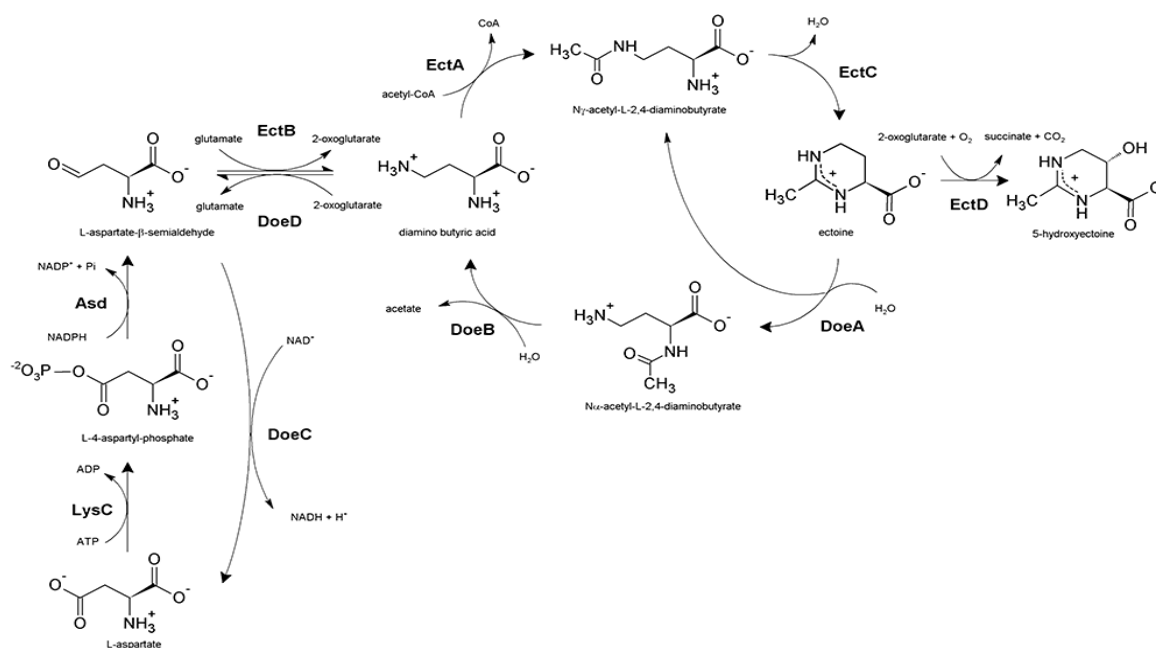


Figure 4.7. Ectoine's biosynthesis and degradation pathway in *H. elongata*. Ectoine's biosynthesis and degradation pathway in *H. elongata*. The ectoine biosynthetic pathway from KEGG is illustrated. Asd: -aspartate-semialdehyde-dehydrogenase; LysC: aspartate kinase; L-2,4-Diaminobutyric Acid Transaminase is known as EctB. DoeA: ectoine hydrolase; EctA: l-2,4-diaminobutyric acid N-acetyltransferase; EctC: ectoine synthase; EctD: ectoine hydroxylase; DoeD stands for l-2,4-diaminobutyric acid transaminase, DoeC for aspartate-semialdehyde dehydrogenase, and DoeB for N-acetyl-l-2,4-diaminobutyric acid deacetylase-adapted from (Schwibbert et al., 2011).

Table 4.4. Genes in the subsystem for ectoine synthesis, hydroxy ectoine and ectoine uptake and catabolism.

Gene ID	Gene Product
fig 2745.490.peg.964	Aspartate-semialdehyde dehydrogenase (EC 1.2.1.11)
fig 2745.490.peg.320	Aspartate-semialdehyde dehydrogenase DoeC in ectoine degradation (EC 1.2.1.-)
fig 2745.490.peg.2432	Diaminobutyrate--2-oxoglutarate transaminase (EC 2.6.1.76)
fig 2745.490.peg.321	DNA-binding protein DoeX, ectoine utilization regulator
fig 2745.490.peg.323	Ectoine hydrolase
fig 2745.490.peg.323	
fig 2745.490.peg.2574	Ectoine hydroxylase
fig 2745.490.peg.318	Ectoine utilization protein EutB, threonine dehydratase-like
fig 2745.490.peg.317	Ectoine utilization protein EutC, similar to ornithine cyclodeaminase
fig 2745.490.peg.2123	Ectoine/hydroxyectoine TRAP transporter large permease protein TeaC
fig 2745.490.peg.2124	Ectoine/hydroxyectoine TRAP transporter small permease protein TeaB
fig 2745.490.peg.2125	Ectoine/hydroxyectoine TRAP transporter substrate-binding periplasmic protein TeaA
fig 2745.490.peg.2433	L-2,4-diaminobutyric acid acetyltransferase (EC 2.3.1.178)
fig 2745.490.peg.319	L-2,4-diaminobutyric acid transaminase DoeD
fig 2745.490.peg.2431	L-ectoine synthase (EC 4.2.1.108)
fig 2745.490.peg.322	N-alpha-acetyl-L-2,4-diaminobutyrate deacetylase
fig 2745.490.peg.322	
fig 2745.490.peg.2122	Universal stress protein UspA in Ectoine TRAP cluster

Table 4.5. Genes predicted in the osmoregulation subsystem (choline uptake and conversion to betaine clusters)

Gene ID	Gene Product
fig 2745.490.peg.1448	Betaine ABC transporter, substrate-binding protein BetX
fig 2745.490.peg.1894	Betaine aldehyde dehydrogenase (EC 1.2.1.8)
fig 2745.490.peg.3518	Betaine/carnitine/choline transporter (BCCT) family
fig 2745.490.peg.1892	Choline binding ABC transport system substrate-binding protein ChoX
fig 2745.490.peg.1895	Choline dehydrogenase (EC 1.1.99.1)
fig 2745.490.peg.1403	Choline-sulfatase (EC 3.1.6.6)
fig 2745.490.peg.1450	Glycine betaine ABC transport system, ATP-binding protein OpuAA (EC 3.6.3.32)
fig 2745.490.peg.3192	Glycine betaine ABC transport system, glycine betaine-binding protein OpuAC
fig 2745.490.peg.1449	Glycine betaine ABC transport system, permease protein OpuAB
fig 2745.490.peg.1393	Glycine betaine/L-proline transport substrate-binding protein ProX (TC 3.A.1.12.1)
fig 2745.490.peg.3528	
fig 2745.490.peg.3521	
fig 2745.490.peg.956	
fig 2745.490.peg.1167	High-affinity choline uptake protein BetT
fig 2745.490.peg.1402	
fig 2745.490.peg.1446	
fig 2745.490.peg.3087	
fig 2745.490.peg.352	
fig 2745.490.peg.3222	
fig 2745.490.peg.3810	
fig 2745.490.peg.589	Osmoprotectant ABC transporter ATP-binding subunit YehX
fig 2745.490.peg.587	Osmoprotectant ABC transporter binding protein YehZ
fig 2745.490.peg.590	Osmoprotectant ABC transporter inner membrane protein YehW
fig 2745.490.peg.588	Osmoprotectant ABC transporter permease protein YehY
fig 2745.490.peg.1893	Transcriptional regulator BetI, TetR family

The osmoregulation subsystem has been found to contain Aquaporin Z and OmpA. Research on Aquaporin Z in *Escherichia coli* has helped scientists understand its role in facilitating fast water movement through membrane pathways, which enables bacteria to quickly adapt to changes in osmolality in their surroundings. This protein is necessary for cells that are growing quickly and plays a role in both immediate and long-term osmoregulatory responses (Calamita, 2000). A protein known as YehZYXW, which is an osmoprotectant ABC transporter found in Enterobacteriales, was identified as a component of this subsystem. This transporter is activated under high osmotic conditions and during the stationary phase of growth. The osmotic and growth-phase controls are controlled by a single promoter, *yehZp*, and are completely reliant on the stress-related sigma factor sigma (Lang et al., 2015). However, low-affinity glycine betaine absorption, mediated by YehZYXW, was found to be hindered by high salinity in a study using transport experiments. The betaine transporter YehZYXW was similar to the *E. coli* ProU in terms of substrate specificity. However, the osmoregulatory function of YehZYXW differs from that of ProU (Lang et al., 2015).

Approximately 19 genes were identified to be part of the subsystem that represents the biosynthesis and uptake of betaine and choline, with *BetT* present at seven different locations in the genome. *BetT* is a low-affinity choline transporter found in *Pseudomonas syringae* that is responsible for choline's improved osmoprotection over glycine betaine (Chen and Beattie, 2008). An operon-like arrangement of three ORFs (*betIBA*) was discovered by nucleotide sequence analysis of a 4–6 kb stretch from the cloned DNA. The gene product of the deduced *betI* gene had characteristics of DNA-binding regulatory proteins. The inferred *BetB* and *BetA* proteins had striking similarities to membrane-bound choline dehydrogenases and soluble glycine betaine aldehyde dehydrogenases from a number of species, respectively. All of these demonstrated that *BetA* can facilitate both processes in the synthesis of glycine betaine by being able to oxidize both choline and glycine betaine aldehyde (Cánovas et al., 2000). Strain 153B of *H. elongata* has the potential to produce glycine betaine from choline due to the presence of the necessary enzymes. (Table 4.5).

4.9. Oxidative stress

In halophilic species, oxidative stress is required, and researchers have theorized a relationship between a cell's capabilities for antioxidation and halotolerance (Gostinčar and Gunde-Cimerman, 2018). The response to an excess of reactive oxygen species can be described as oxygen stress. This term encompasses a range of different stressors, including superoxide, hydrogen peroxide, peroxynitrite, and singlet oxygen. Oxygen stress can also be triggered by certain inducers. Some important proteins associated with oxygen stress include paraquat-inducible protein, ferric uptake regulation A, and superoxide dismutase. The paraquat-inducible protein A is a transmembrane protein similar to those in some *Pseudomonas* species.

Here, 39 genes that relate to oxidative stress were identified. Several genes were identified that are involved in the protection against reactive oxygen and nitrogen species, as well as the redox cycle and non-redox reactions of glutathione and the gamma-glutamyl cycle. Another such gene is rubrerythrin (*Rbr*), which is a nonheme iron protein that helps bacteria resist oxidative stress. According to studies, rubrerythrin may have the ability to shield cells from oxidative stress by catalyzing the removal of hydrogen peroxide within cells in situations where other peroxidases are not present (Sztukowska et al., 2002).

4.10. Polyhydroxybutyrates

Under conditions of limited nutrients and an excess of carbon, *Halomonas* bacteria are capable of synthesizing polyhydroxybutyrate (PHB) as an energy reserve within their cells. The rate of PHB buildup can also be affected by sodium chloride concentration (Quillaguamán et al., 2006). PhbA (acetyl-CoA C-acetyltransferase [EC:2.3.1.9]), PhbB (acetoacetyl-CoA reductase [EC:1.1.1.36]), and PhbC (PHA synthetase [EC:2.3.1.304]) genes involved in polyhydroxybutyrate biosynthesis from Acetyl-CoA were predicted (Figure 4.8). The genome also contained a prediction for the Polyhydroxyalkanoate synthesis inhibitor called PhaR.

The *PhaC* gene displayed strong similarity (>80%) with numerous other class 1 PHA synthases that are common within the *Halomonas* during a blast search. A Uniprot database

blast revealed a high degree of similarity to the reference PhaC1 enzyme (97%), which also supports the findings. The PHA synthase structures of *Halomonas elongata* (strain ATCC 33173/ DSM 2581/ NBRC 15536/ NCIMB 2198/ 1H9) were the closest matches. The *PhaC* gene of *H. elongata* 153B was analyzed in relation to that of various organisms from different classes to establish evolutionary connections. The alignment comprised 23 nucleotide sequences in total. All positions were free of holes and missing information (complete deletion option). The figure is provided in the appendix.

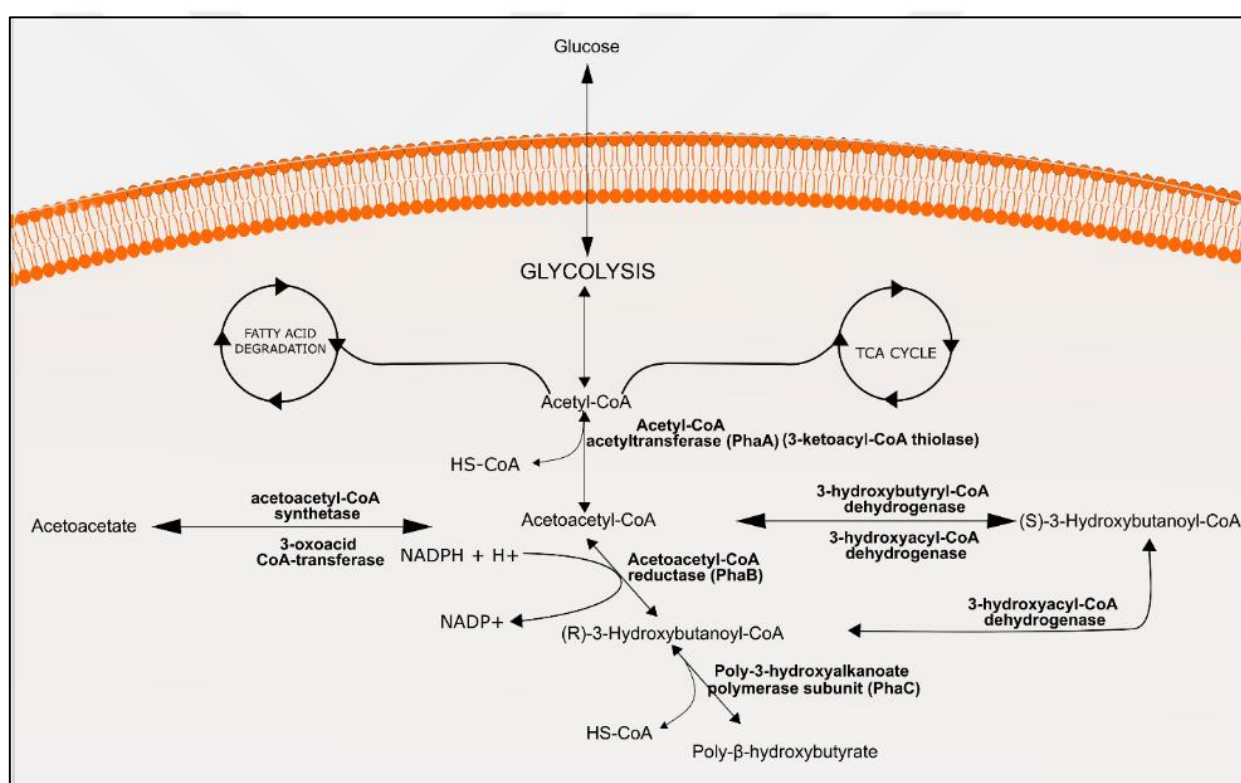


Figure 4.8. Annotated PHB biosynthesis pathway in *H. elongata* 153B

4.11. Secondary metabolites biosynthesis clusters

Bacteria secondary metabolite biosynthesis clusters are groups of genes that are responsible for the synthesis of small molecules that are not essential for bacterial growth and survival, but rather play a role in the bacteria's interactions with its environment and other organisms. These clusters contain the genes needed to synthesize compounds, as well as

regulatory genes that control the expression of biosynthetic genes. Ectoine biosynthesis cluster with a similarity of 85% was annotated in the genome. No other cluster was annotated.

4.12. Others

In saline habitats, industrial waste is often found and can have high levels of aromatic substances. Bacteria in these areas have been known to adapt and use these compounds as an energy source through the process of degradation. *Halomonas*, in particular, have been observed to be able to break down petroleum hydrocarbons and several aromatic compounds (García et al., 2004; Moreno et al., 2011; Neifar et al., 2019). The *Halomonas elongata* strain 153B can break down 6-Mercaptopurine and Azathioprine, which are commonly used chemotherapy drugs to treat acute lymphoblastic leukemia. In the genome, Fluorouracil, isoniazid degradation genes, as well as beta-lactam resistance through the presence of penicillin-binding proteins, were annotated. An analysis of subsystems showed that 51 genes are involved in the breakdown of aromatic compounds, including the catabolism pathway genes for quinate, benzoate, biphenyl, and p-Hydroxybenzoate. Quinate degradation enzymes are an uncommon family of dehydratases found in *Acinetobacter calcoaceticus* and are linked to quinate catabolism (Peek et al., 2017). Biphenyl metabolism genes in *Rhodococcus* sp. strain RHA1 are engaged in limiting the formation of unanticipated toxic metabolites that interfere with RHA1 growth as well as the primary metabolism of HPD during growth on both biphenyl and ethylbenzene (Sakai et al., 2003).

The degradation of ringed molecules includes catechol's breakdown (part of the beta-ketoadipate pathway), gentisate, salicylate, protocatechuate, and homogentisate. These molecules are an important source of energy for microorganisms that inhabit the soil. The beta-ketoadipate pathway involves the use of ring-cleaving enzymes in the metabolism of protocatechuate and catechol (Moreno et al., 2011). There has been a surge in interest in finding ways to improve bioremediation techniques to address contamination problems involving high levels of salt and aromatic substances. It has been discovered that halophilic bacteria, such as those of the genus *Protocatechuate*, can break down several environmental contaminants (Wells and Ragauskas, 2012). In *Rhodococcus* sp. strain B4, salicylate, and gentisate were found to be

intermediates in the breakdown of naphthalene (Grund et al., 1992). Exploring the catabolic genes that allow halophiles to break down phenol can serve as a model for understanding the usefulness of these extremophiles in remediation efforts in saline environments (Moreno et al., 2011).

4.13. Genome-scale metabolic network modeling and analysis

4.13.1. General model features

Following the reconstruction of the draft, an iterative process was used to develop the model (Figure 3.1). The entire model represented the 3 bacterial compartments which are the extracellular space, periplasm, and cytosol. There were 1534 metabolites, 2314 reactions, and 1239 genes in the model. According to established conventions for naming GEMs, the model was given the name *iEB1239*, with *i* standing for in silico, EB standing for the first letters of the researcher who curated the model, and 1239 standing for how many genes are represented in the model.

4.13.2. Benchmarking

MEMOTE's initial model output scored 27%, with the lowest scores coming from subpar annotations. The draft GEM of *H. elongata* 153B was further curated and given annotations, resulting in a MEMOTE score of 71%. As this is one of the earliest genome-scale models for this species and there is a scarcity of experimental information to supplement it, this score can be seen as a promising start for future research on the model (Figure 4.9).

4.13.3. Added annotations

CarveMe models generate annotations that appear in the model's Notes section. MEMOTE does not, however, notice this when benchmarking. A high MEMOTE score was made possible by annotations for the model's included genes, SBO terms, and metabolites. Annotations that MEMOTE can identify were allowed to be included in the appropriate fields thanks to ModelPolisher. Annotation databases that were queried include BiGG (Schellenberger

et al., 2010), BioCyc (Karp et al., 2019), MetaNetX (Moretti et al., 2021), CHEBI (Degtyarenko et al., 2008), KEGG (Kanehisa and Goto, 2000), HMDB (Wishart et al., 2007), RHEA (Alcántara et al., 2012)Inchikey (Heller et al., 2015), Reactome (Fabregat et al., 2018), Lipidmaps (Liebisch et al., 2020), and SEED EC-code (Seaver et al., 2020).

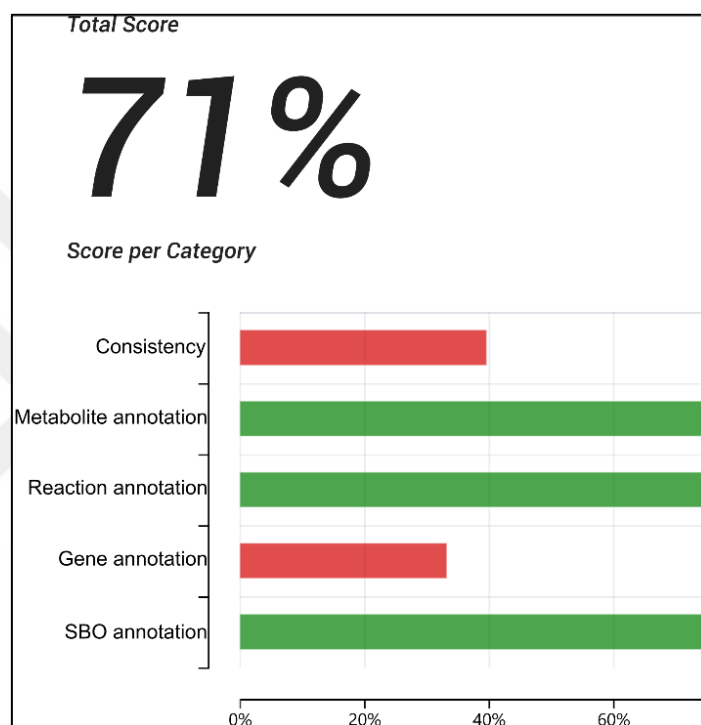


Figure 4.9: Overall MEMOTE model benchmarking score for iEB1239

4.13.4. Model curation and gap analysis

A total of 40 blocked reactions were identified in the model, which were primarily categorized as exchange reactions, prosthetic groups, and cofactors according to an analysis of metadata provided in the BIGG database. Escher maps allowed for the depiction of metabolic pathways, allowing for the identification of gaps in such pathways and their filling (Figure 4.10).

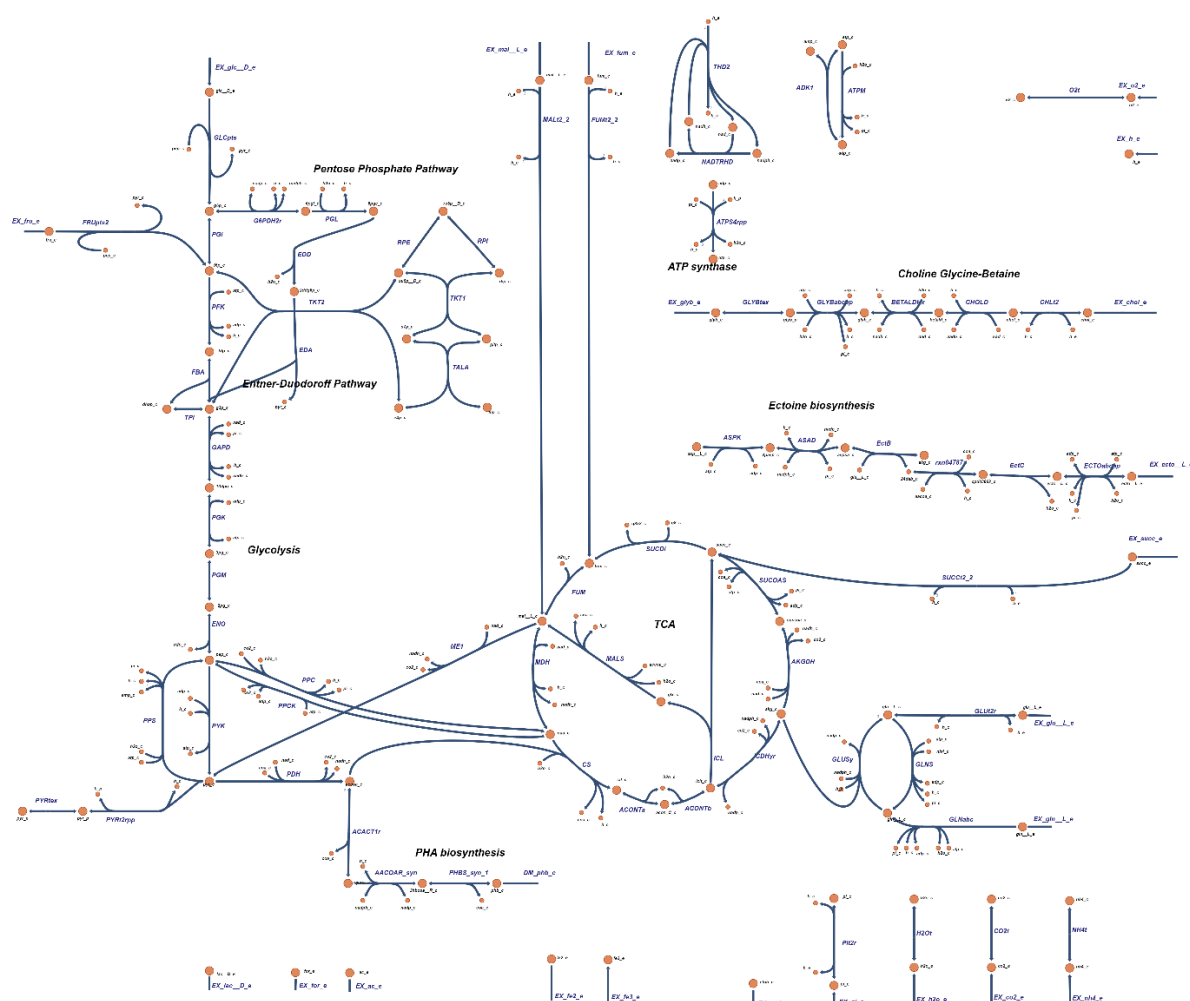


Figure 4.10. Escher map of the iEB1239 model showing main metabolic reactions (including Glycolysis, Pentose phosphate and Entner-Duodoroff pathway, PHB and Ectoine biosynthesis, Choline Glycine-Betaine, ATP synthase, and metabolite flow in and out of the cell).

During model curation, reaction reversibilities were checked. For example, the draft model showed that 95% ATP was being produced from the aspartate kinase reaction. This is obviously not correct as *H. elongata* can completely oxidize glucose to CO₂ and H₂O by oxidative phosphorylation. So, to curate and balance energy cycles, we checked the reaction reversibility of Aspartate Kinase (ASPK). After making the reaction reversible, the model produced a more realistic biomass yield with energy generated from ATP synthase,

phosphoglucokinase, pyruvate kinase, and succinyl_CoA. The directionality of proton transport processes was also changed to ensure that no proton flux towards the outside of the cell is permitted if it is not mediated by respiratory chain components or associated with an energy cost. This was to ensure that proton flux does not occur against a concentration gradient. The periplasmic compartment was designed in general to better represent the proton gradient profile.

The majority of the pathway gaps could not be further studied due to the dearth of information on *H. elongata* in the major databases. These were left because they were thought to be knowledge gaps that would be filled by expanding research. However, MEMOTE found a high metabolite connection with a score of 100%. All issues were fixed when the output model underwent additional SBML compliance testing using the Python COBRApy (Ebrahim et al., 2013) package. The completed model includes all necessary SBML fields.

4.13.5. Minimal medium

By repeatedly testing the model's growth under various limiting nutrients, we were able to determine the minimum requirements for its survival. The simulations used glucose as the sole source of carbon and examined the effects of various salts and ions on the model. To make sure that just CO₂ was expelled, the secretion of other carbon-containing substances was closely monitored. Table 4.6 lists the complete list of necessary metabolites known as the minimal media.

We noticed that the model had an auxotrophy for Asparagine, and low amounts of asparagine caused the model to be ineffective in glucose metabolism producing glycerate and acetate besides carbon dioxide as excretion products Figure 4.11. L-asparaginase is a very important enzyme as a study has outlined the anticancer properties of recombinant L-asparaginase obtained from *Halomonas elongata* (Ghasemi et al., 2017).

Table 4.6. Minimal media components used for simulations with iEB1239.

Component	Name	Flux (mmol / gDW h)
EX_glc__D_e	D-Glucose	5.0
EX_cl_e	Chloride	0.004448022
EX_pi_e	Phosphate	0.842682732
EX_asn__L_e	Asparagine	10.0
EX_fe3_e	Ferric iron	12.09894427
EX_k_e	Potassium	0.166805535
EX_ca2_e	Calcium	0.004448022
EX_mg2_e	Manganese	0.00741337
EX_mn2_e	Magnesium	0.000590506
EX_cobalt2_e	Cobalt	8.55E-05
EX_zn2_e	Zinc	0.000291407
EX_cu2_e	Copper	0.000605888
EX_o2_e	Oxygen	14.95368827
EX_so4_e	Sulfate	0.214176758

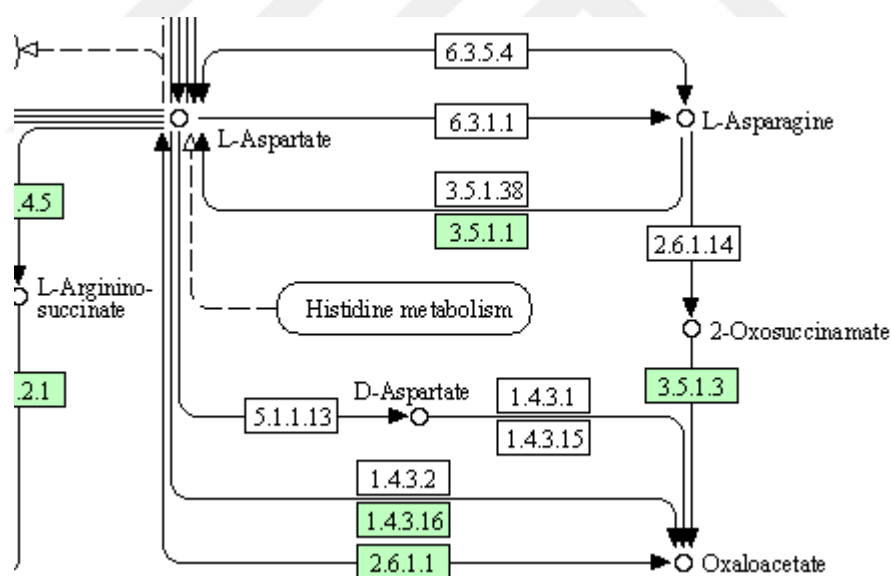


Figure 4.11. KAAS metabolic pathway annotation from genome showing inability to synthesize Asparagine from precursors such as Oxaloacetate, or L-Aspartate.

4.13.6. Biomass reaction

The process of manual curation involves several steps including the addition of reactions necessary for the production of biomass precursors (gap-filling), correcting the balance of elements and the directionality of reactions, identifying cyclical reactions that do not result in a product, and eliminating blocked reactions and dead-end metabolites. Failure to address these factors can result in incorrect predictions of phenotypes such as incorrect ATP production, biomass yields, incorrect nutritional requirements, and excess incorrect identification of essential genes (Machado et al., 2018; Thiele and Palsson, 2010).

Mass fractions of DNA, RNA, lipid, protein, carbohydrate, fatty acids, and peptidoglycan which make up the biomass equation was composed as described below. *Halomonas elongata* responds to increases in the salt concentration of its growth medium by undergoing several physiological changes. The internal coherence and compactness of the cell wall have been observed to increase in a high-salinity environment. A greater proportion of negatively charged lipids are present overall in the lipid profiles (Vreeland et al., 1984). Other studies support that *H. elongata* changes the affinity of its outermost cell surface for water to adapt to various NaCl concentrations (Hart and Vreeland, 1988). With these in mind, we curated the biomass reaction to reflect the biomass of *H. elongata* at high salinity. The biomass components and their proportions in these conditions have not been reported yet. Every new reconstruction requires a human curation process that takes time and involves repetitive work.

In the CarveMe pipeline, the draft universal model of bacterial metabolism was created using the entire BiGG database and extensively curated to produce accurate results. First, all eukaryotic compartment processes and metabolites were eliminated. The model was subsequently supplemented with a general biomass equation. This equation was modified from the biomass composition of *Escherichia coli* (Orth et al., 2011) and also according to research on universal biomass components in prokaryotes (Xavier et al., 2017). The biomass components of closely related halophilic strains previously experimentally determined were used to improve the biomass equation (Piubeli et al., 2018). This is a useful approach to solving the challenges of obtaining strain-specific data (Thiele and Palsson, 2010). The GC content of *H. elongata*

153B was used to calculate the amounts of DNA nucleotides required for the biomass while other components were maintained. The resulting biomass equation for high salinity was composed as shown in Table 4.7 below.

Table 4.7. Biomass equation representing high salinity conditions in the iEB1239 model.

Macromolecule	overall wt%	composition (molar fraction)	mmol/gDW (Calc.)	metabolite
Protein	0.205	0.115	0.215287	ala-L
		0.075	0.139700	arg-L
		0.024	0.045647	asn-L
		0.061	0.114285	asp-L
		0.010	0.017859	cys-L
		0.037	0.070118	gln-L
		0.062	0.116021	glu-L
		0.081	0.151978	gly
		0.026	0.047708	his-L
		0.046	0.085461	ile-L
		0.113	0.210541	leu-L
		0.025	0.046844	lys-L
		0.025	0.046992	met-L
		0.034	0.064392	phe-L
		0.049	0.091691	pro-L
		0.054	0.100388	ser-L
		0.052	0.097638	thr-L
		0.015	0.027963	trp-L
		0.024	0.044776	tyr-L
		0.072	0.134842	val-L
DNA	0.031	0.180	0.018178	datp
		0.318	0.032029	dctp
		0.321	0.032414	dgtg
		0.181	0.018210	dtg
RNA	0.21	0.242	0.152315	ctp
		0.375	0.236514	gtp
		0.179	0.113148	utp
		0.231	0.145440	atp**
glycogen	0.025	1	0.374000	glycogen
murein	0.025	0.4	0.028000	murein5p5p
LPS	0.034	1	0.019456	kdo2lipid4
lipid	0.093	0.1800	0.027172	pe160
		0.0408	0.006159	pe161
		0.1498	0.022607	pe181
		0.0833	0.012567	pg160
		0.0189	0.002848	pg161
		0.0693	0.010456	pg181

Table 4.8. Biomass equation representing high salinity conditions in the iEB1239 model (cont)

Macromolecule	overall wt%	composition (molar fraction)	mmol/gDW (Calc.)	metabolite
inorganic ions	0.01	0.7143	0.169185	k
		0.0476	0.011279	nh4
		0.0317	0.007519	mg2
		0.0190	0.004512	ca2
		0.0286	0.006767	fe2
		0.0286	0.006767	fe3
		0.0127	0.003008	cu2
		0.0127	0.003008	mn2
		0.0127	0.003008	mobd
		0.0127	0.003008	cobalt2
		0.0127	0.003008	zn2
		0.0190	0.004512	cl
		0.0159	0.003760	so4
		0.0159	0.003760	pi
			0.000279	accoa
			0.000168	coa
			0.000098	succoa
			0.001787	nad
			0.000045	nadh
			0.000112	nadp
			0.665011	ect-L
			0.377131	hdect
			0.000335	nadph
			0.000223	fad
			0.000223	5mthf
			0.000223	hemeO
			0.000223	sheme
		1	59.810000	atp
		1	59.810000	h2o
			59.955440	atp
			57.939868	h2o

4.13.7. Validation on carbon source usage

Microbes belonging to the Halomonadaceae family have diverse metabolic capabilities. Within a single species, the ability to use a particular carbon source for growth may vary (Vreeland et al., 1980). GEMs can be used to understand how metabolic pathways, carbon sources, and the synthesis of important metabolites interact within a holistic system (Ates,

2015). Predictions are crucial for directing experiments that need to be labeled or for the creation of bioproducts.

In silico predictions were conducted on the previously obtained minimal medium, with the biomass equation serving as the objective function and glucose being the sole carbon source. The metabolism of glucose in the glycolysis, pentose phosphate, TCA pathways, and electron transport chain were accurately predicted by the model (Figure 4.12). 10.83% ATP was obtained from Phosphoglucokinase, 0.27% from pyruvate kinase, 7.74% from succinyl-CoA synthetase, and 81.6% of ATP was synthesized by ATP synthase using 4 protons for each ATP molecule. The proportions here shown are represented by the complete oxidation of glucose. Flux distributions are shown in Figure 4.12 below.

To evaluate the ability of *iEB1239* to accurately predict carbon consumption characteristics and replicate laboratory outcomes, we utilized FBA to simulate growth on various carbon sources with each carbon source serving as the main carbon provider in separate simulations, using an uptake rate of -10 mmol/gDW/h. The model was used to depict growth on 28 different carbon sources with varying flux rates (Figure 4.13). It is worth noting that the uptake rate of -10 mmol/gDW/h was derived from data obtained from *E. coli*, and with the absence of specific experimental data on flux values, biomass yields produced through simulation that exceeded 2 g/mmol require further experimentation to establish these states.

The in silico predictions with experimental results reported in the literature from different studies were compared in Table 4.9. From the table, it is observed that contrary to experimental data, *iEB1239* was not able to produce growth on succinate, mannose, glycerol, gluconate, cellobiose, and alanine. Despite the presence of transport reactions, these were not metabolized by the model. The reason for this is the absence of genome annotation for the required enzymes which meant the corresponding reaction was not reflected in the model. The table also shows that growth on carbon sources can be variable within *H. elongata* strains. As a result, we can hypothesize that this is a strain-specific characteristic of *H. elongata* 153B. As seen in Table 4.9, the model predicted growth on other carbon sources that are yet to be reported

in the literature. Also of interest, is the ability to use xylose which indicates that the strain can use cheap agricultural biomass as substrate in industrial processes.

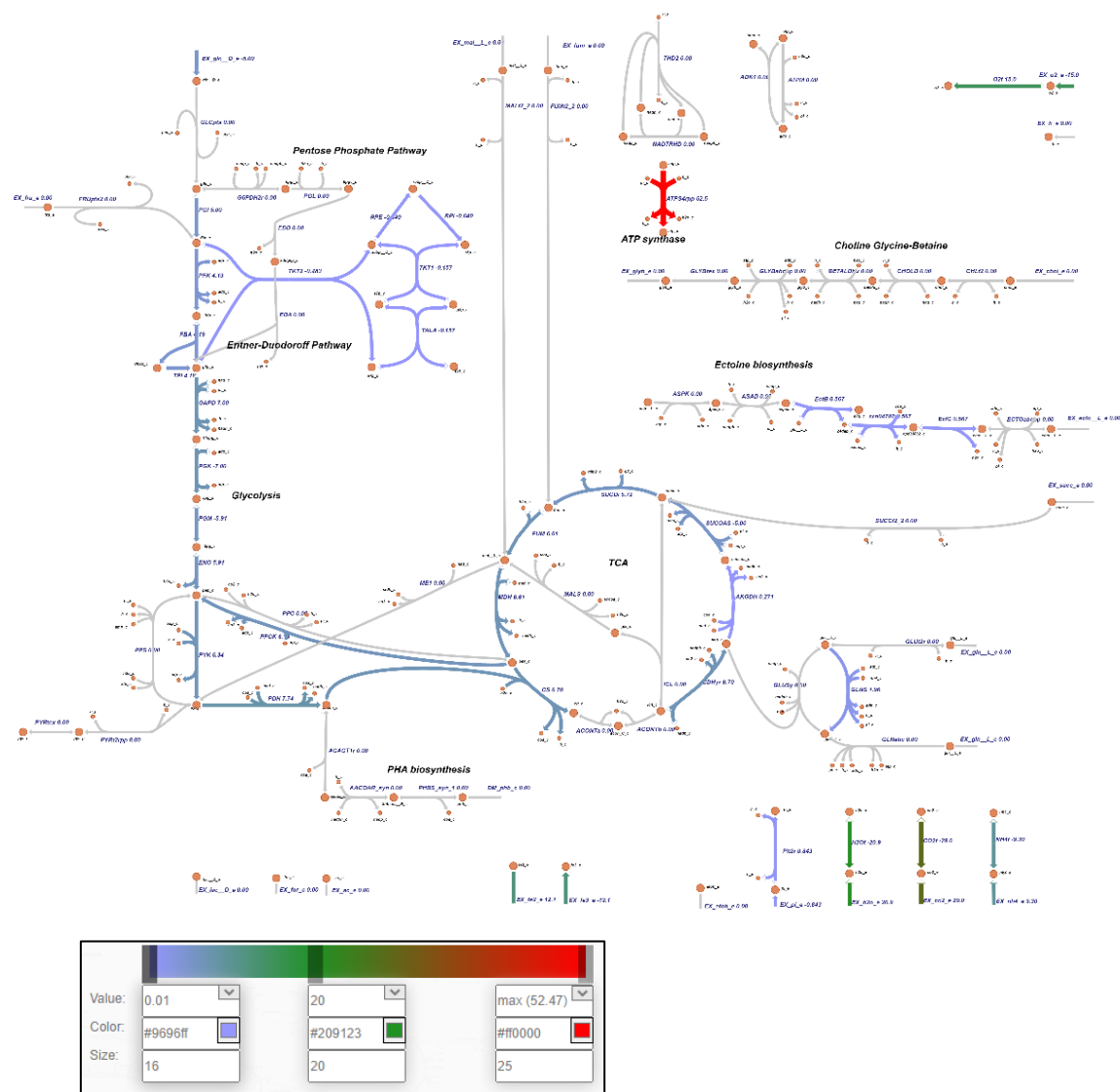


Figure 4.12. Flux distribution in the iEB1239 model when glucose was used as the main carbon source in the minimal medium. Flux map shows the use of the glycolysis pathway, TCA cycle, oxidative phosphorylation, and the movement of metabolites with the external environment.

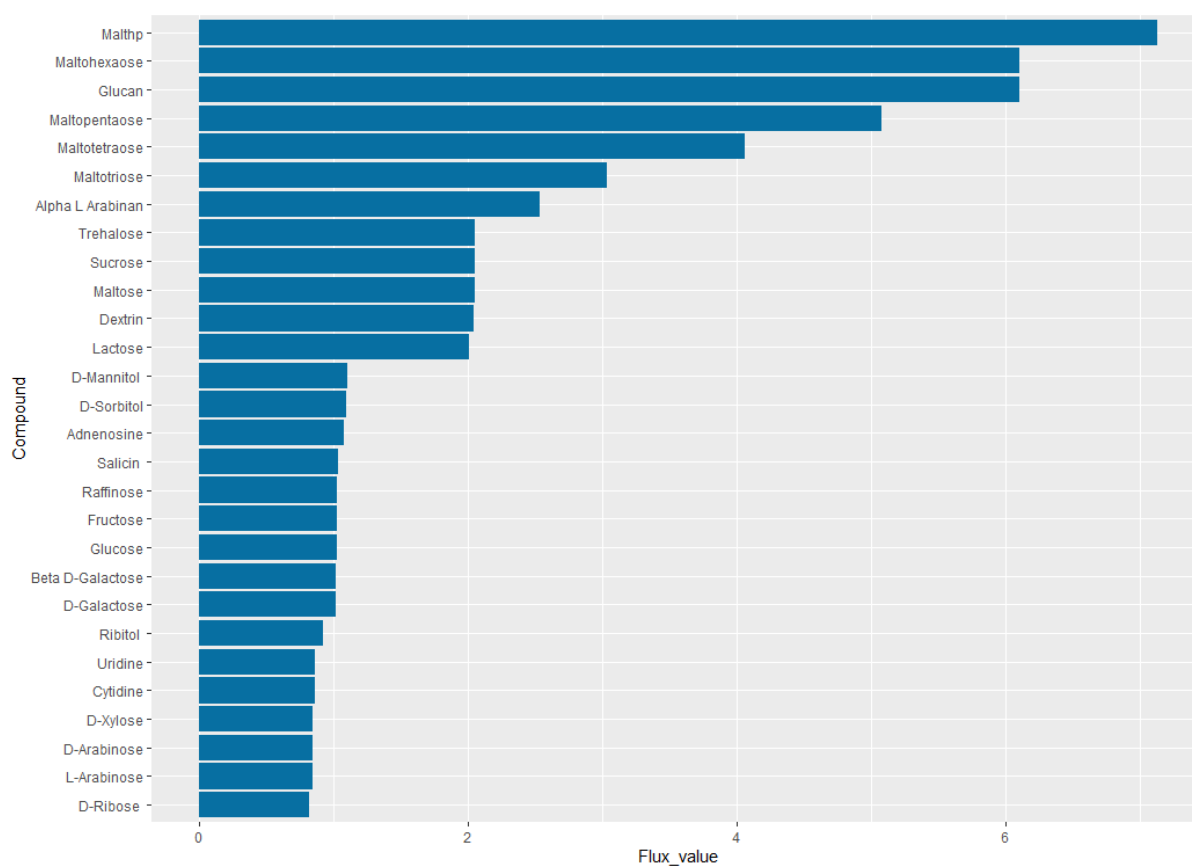


Figure 4.13. Carbon sources showing growth with iEB1239

Table 4.9. Growth on different carbon sources compared with other *Halomonas elongata* strains in the literature. + growth observed, - no growth observed, $\pm$ very slow growth.

Compound	153B	1H9	1H11	1H15
Adenosine	+			
Alanine	-	+	NT	NT
Alpha L Arabinan	+			
Beta D-Galactose	+			
Cellobiose	-	+	$\pm$	+
Cytidine	+			
D-Arabinose	+			
Dextrin	+			
D-Galactose	+			
D-Mannitol	+	+	NT	NT
D-Ribose	+			
D-Sorbitol	+			
D-Xylose	+	+		
Fructose	+			
Glucan	+			
Gluconate	$\pm$	+	+	+
Glucose	+	+	+	+
Glutamine	-	+	NT	NT
Glycerol	$\pm$	+	+	+
Lactose	+	+	$\pm$	+
L-Arabinose	+			
Maltose	+			
Maltotetraose	+			
Maltotriose	+			
Mannose	$\pm$	+	+	+
Raffinose	+			
Ribitol	+			
Salicin	+			
Succinate	$\pm$	+	NT	NT
Sucrose	+	+	+	+
Trehalose	+			
Uridine	+			

4.13.8. Biochemical and osmoadaptation phenotypes

An important question GEMs help to answer is if we have enough information to fully explain observable phenotypes in organisms. A model to investigate osmotolerance in *H. elongata* had previously been reported, but it was made up of only a few reactions and pathways of interest (Ceylan et al., 2012). More complex phenotypes such as dependence on NaCl have been modeled with substrate inhibition kinetics (Dötsch et al., 2008). Here we are presenting the most extensive GEM for *H. elongata* that includes a genome-wide representation of its metabolism rather than specific reactions or pathways. Furthermore, the model can be easily used by other researchers to investigate other hypotheses.

Biochemical experiments on *H. elongata* show that they are able to reduce NO_3 to NO_2 , i.e. it could grow anaerobically in the presence of NO_3 (Vreeland et al., 1980). *iEB1239* simulations did not agree with these observations. When oxygen was limited in the minimal medium but supplemented with NO_3 , growth was not supported, indicating that in place of oxygen, it cannot be used as a terminal electron acceptor. In this case, it is very likely that *H. elongata* 153B doesn't grow anaerobically on nitrate. It could also be a case of gap in the model that requires filling. Specific biochemical experiments need to be carried out for confirmation. In the *Halomonas campisalis* species, researchers discovered an increase in trans fatty acids when salinity was not present, which may be due to a disruption in the membrane. In addition, cyclic fatty acids also increased when no salinity was present and at very high salinities, which indicates that the cells were under stress in these conditions. The production of compatible solutes varied based on the level of salinity, the rate of growth, and the final electron acceptor (either oxygen or nitrate). When nitrate was the final electron acceptor, there was a decrease in ectoine production compared to when oxygen was used (Aston and Peyton, 2007). As our model did not produce growth on NO_3^- , we could not verify its impact on the production of ectoine.

H. elongata has the ability to produce ectoine and hydroxyectoine to adapt to changes in salt concentration in its surroundings. It can also absorb various organic substances that can act as osmoprotectants when externally present. There have been studies on the transport of choline and its transformation into betaine an osmoprotectant compound in *H. elongata*. In the genetic

organization of the *betIBA* region, the *betI* gene encodes a putative regulatory protein, *BetB* codes for a glycine betaine aldehyde dehydrogenase (BADH) homologue, *BetA* codes for a choline dehydrogenase (CDH) homologue (Cánovas et al., 2000). The reactions corresponding to this pathway were present in the model and the corresponding flux along this pathway is shown in Figure 4.12. *iEB1239* therefore is useful for simulating osmoregulatory processes as well as the dynamics in the conversion of choline to glycine betaine osmoprotectant.

When the *iEB1239* minimal medium was supplemented with exogenous choline at uptake values of -10mmol/g/DW/h there was an increase in the growth rate depicted by an increase in the biomass. Furthermore, glycine betaine was produced as one of the secreted products showing that choline was converted to glycine betaine when supplemented in the medium. Adding choline to the growth medium resulted in simulations predicting the occurrence of states that directed metabolic flux to the production of ectoine and the synthesis of little quantities of glycine betaine, according to our FBA analysis.

Studies also show that *H. elongata* grew more quickly on 3 M NaCl-glucose-M63 plates when betaine, choline, or choline-O-sulfate were added externally (but not ectoine, proline, or proline betaine), showing that these substances can act as osmoprotectants. *H. elongata* absorbed exogenous choline and converted it to betaine, proving that this moderately halophilic bacterium has a choline-glycine betaine route. The results of the experiment showed that the movement and conversion of choline in the presence of betaine or not were quicker in a 2 M NaCl solution compared to a 0.5 M NaCl solution. This suggests that the salt content of the medium plays a significant role in regulating the choline-betaine pathway of *Halophila elongata* under the tested conditions (Cánovas et al., 1996). After choline was added to the medium, transport and conversion of choline to betaine started right away (Cánovas et al., 1998).

4.13.9. Gene essentiality

The model *iEB1239* was analyzed for predictions of genes that can be considered essential to survival by iteratively knocking out single-genes in the model and then doing an optimization for growth. This function is comparable to transposon mutagenesis and tn-seq in vivo research. Each time growth was not produced by the simulation, the gene that was knocked

out including the reaction it is associated with were considered essential for survival in the organism. In total, 12 lethal essential genes were predicted which was unusually small. While 51 genes produced negative, and extremely low growth. The list of essential genes predicted by *iEB1239* is provided as appendix. Gene essentiality data showed that the most essential genes are those used for amino acid metabolism, metabolism of osmoprotectants and the transportation of ions. For halophilic bacteria, these are indeed very necessary for survival in high salinity environments. A systems biology approach could be used to examine the mechanisms by which essential genes contribute to the adaptation and survival of *Halomonas elongata*. Integrating transcriptomics and other omics data could further improve the results (Occhipinti et al., 2021).

4.13.10. Production of bioproducts

Halophilic organisms like *H. elongata* can produce high-value bioproducts such as ectoine, PHAs, vanillic acid as well as beneficial enzymes like L-Asparaginase that have been shown to have anticancer effects. The industrial production of these high-value products however remains challenging as in most cases, reliance is on trial-and-error optimization experiments. GEMs provide a systematic systems biology-based approach that enable more efficient investigation of feasible optimal product biosynthesis and simulate possible transposon mutagenesis gene knock-out outcomes for metabolic engineering. In this section we attempt to reproduce experimentally determined phenotypes related to the biosynthesis of PHB, ectoine, and vanillic acid. We also provide hypothesis to explain the dynamics of these phenotypes as our model employs a novel and more comprehensive systems biology approach. We show through these analyses the robustness of *iEB1239* as a reliable platform for further biotechnology research.

iEB1239 was tested for its capacity to synthesize PHB, ectoine, and vanillic acid by first simulating the model using Flux balance analysis using the biomass equation as objective for optimum growth and then monitoring the flux of the reactions yielding the metabolites. Second, Flux variability analysis was used to look into the possibility of more optimal states. Thirdly, a simulation was run with the objective altered to the demand reaction of the final product in the corresponding pathways to simulate the maximum product synthesis. It is important to note that

because FBA simulations only anticipate optimal solutions, they do not account for the effect of gene regulation when predicting product biosynthesis rates. As a result, to match in silico FBA predictions, culture conditions that generate the best responses must be identified while verifying simulations in vivo. In these situations, FBA predictions in cases when regulatory genes are known, offer ideal product biosynthesis rates following gene knockout (O'Brien et al., 2015). Also, GEMs are predicated on the notion of spatial homogeneity and a well-mixed environment (such as a chemostat). Fourthly, a phenotypic phase plane study was carried out and plotted to examine the model's suitability to synthesize these metabolites under experimentally ideal circumstances. Finally, using the same objective function, we mimicked transposon mutagenesis gene knock-out simulations to reproduce experimentally determined phenotypes related to metabolic engineering.

i) Polyhydroxalkanoates

PHB plays a fundamental role in the adaptation of halophiles as they are used as an energy store in nutrient limiting conditions. Studies have reported the production of PHBs from *H. elongata*. In one such study, according to the authors, in the presence of 100 g/L NaCl, *H. elongata* accumulated up to 50 % w/w PHA within 2–3 days under non-optimized conditions (Mothes et al., 2008). At high salt concentrations, *H. elongata* strain 2FF was able to synthesize sizable amounts of PHB from D-glucose and PHAs from a variety of carbon substrates (Cristea et al., 2018).

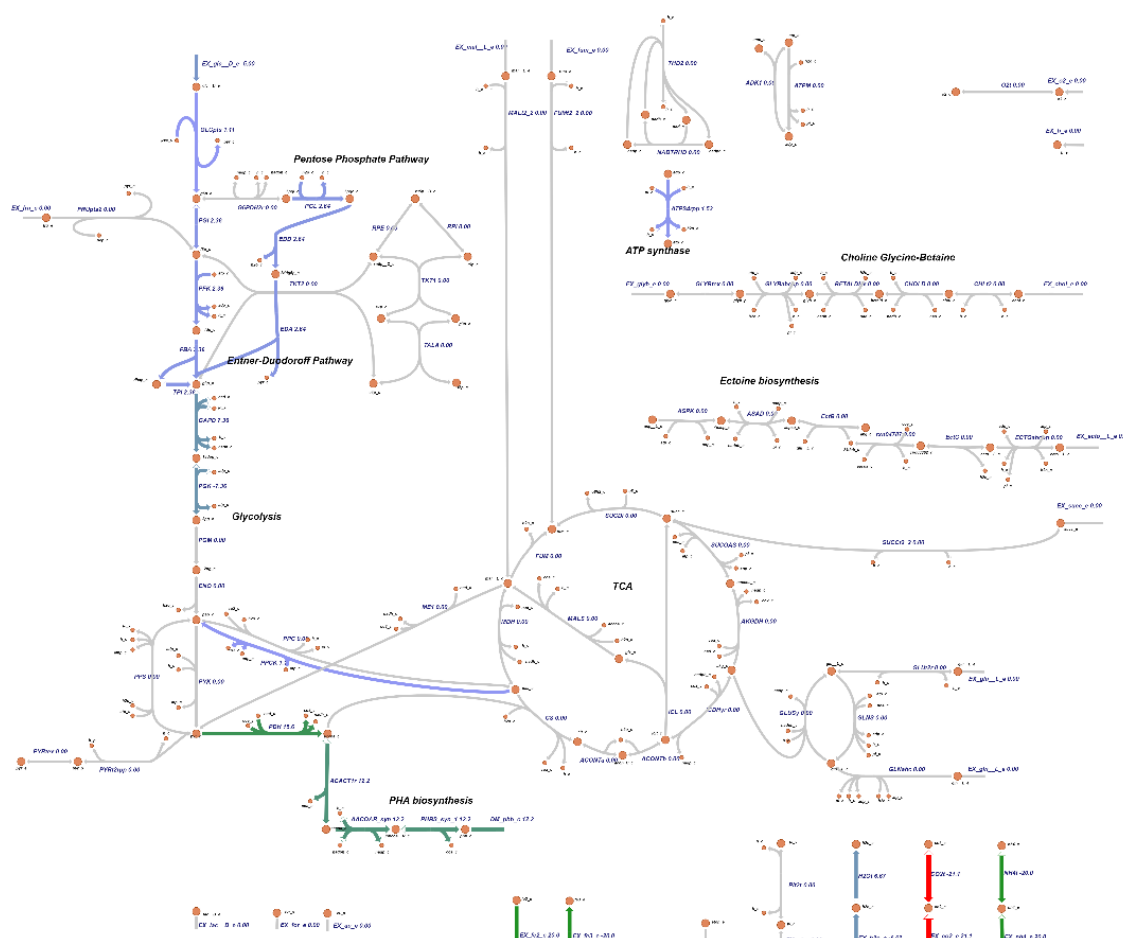


Figure 4.14. Flux distribution in iEB1239 when flux towards PHB biosynthesis is maximized

iEB1239 was extensively curated to ensure that all reactions and precursor reactions in the PHB biosynthesis pathway were included in the model. For reactions that were not in the BiGG database, considering gene evidence, we obtained these and their corresponding identifiers from the MODELSEED database. The pathway curation and simulations were also guided by Escher maps. First, using the biomass as objective, according to FVA simulations, the minimum and maximum fluxes for PHB synthesis were 0.0 mmol/gDW/h and -7.6×10^{-12} mmol/gDW/h, respectively. When the objective was changed to the demand reaction for PHB the maximum flux rate for PHB biosynthesis was 12.22 mmol/gDW/h. The flux map shows a switch from pentose phosphate pathway to the Entner-Duodoroff pathway as PHB biosynthesis is maximized (Figure 4.14).

iEB1239 is able to produce PHBs, and the rate at which it does so appears to be inversely proportional to its growth rate. The highest PHB production, reaching 12.22 mmol/gDW/h, was seen at the lowest growth rate. However, as the nitrogen uptake rate (NH_4^+) increases beyond approximately 17 mmol/gDW/h, the phase plane analysis revealed a decrease in PHB production (Figure 4.15). This means that if *H. elongata* reaches an optimal growth rate, reducing the rate at which NH_4^+ is taken up to initiate secondary metabolism would result in an increase in PHB production that is roughly proportional. These results are in line with PHB production dynamics in industrial and laboratory settings (Koller, 2018; McAdam et al., 2020).

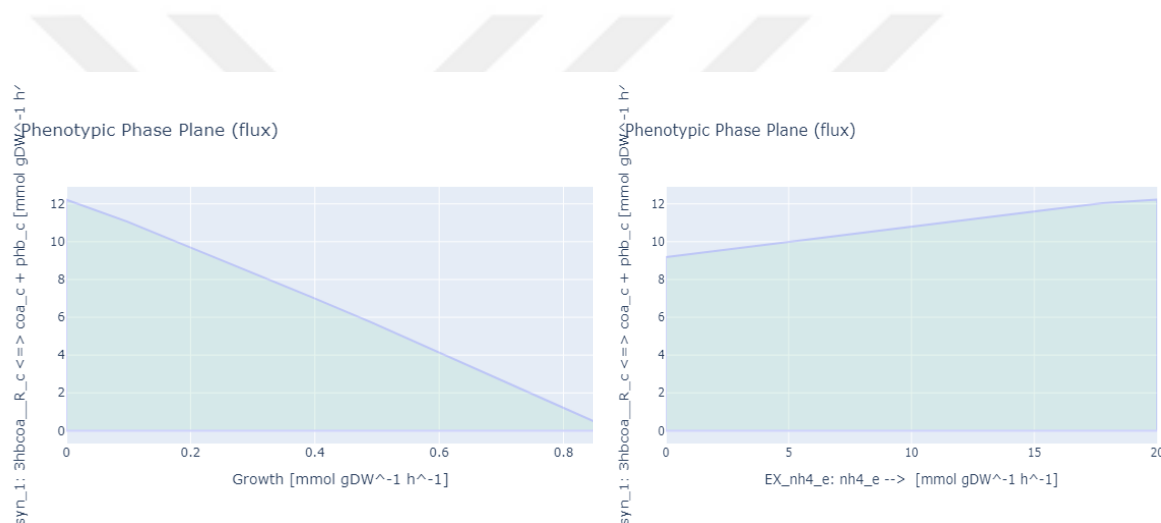


Figure 4.15: Phase plane analysis of PHB synthesis A) with changes in biomass B) with changes in nitrogen source uptake

Prediction of overexpression and downregulation targets to improve the biosynthesis of PHB in *H. elongata* 153B was done as well with the CAMEO package. The suggested targets are provided Figure 4.16. These were further investigated insilico, but there is still a need to experimentally carryout CRISPR engineering studies to validate these findings.

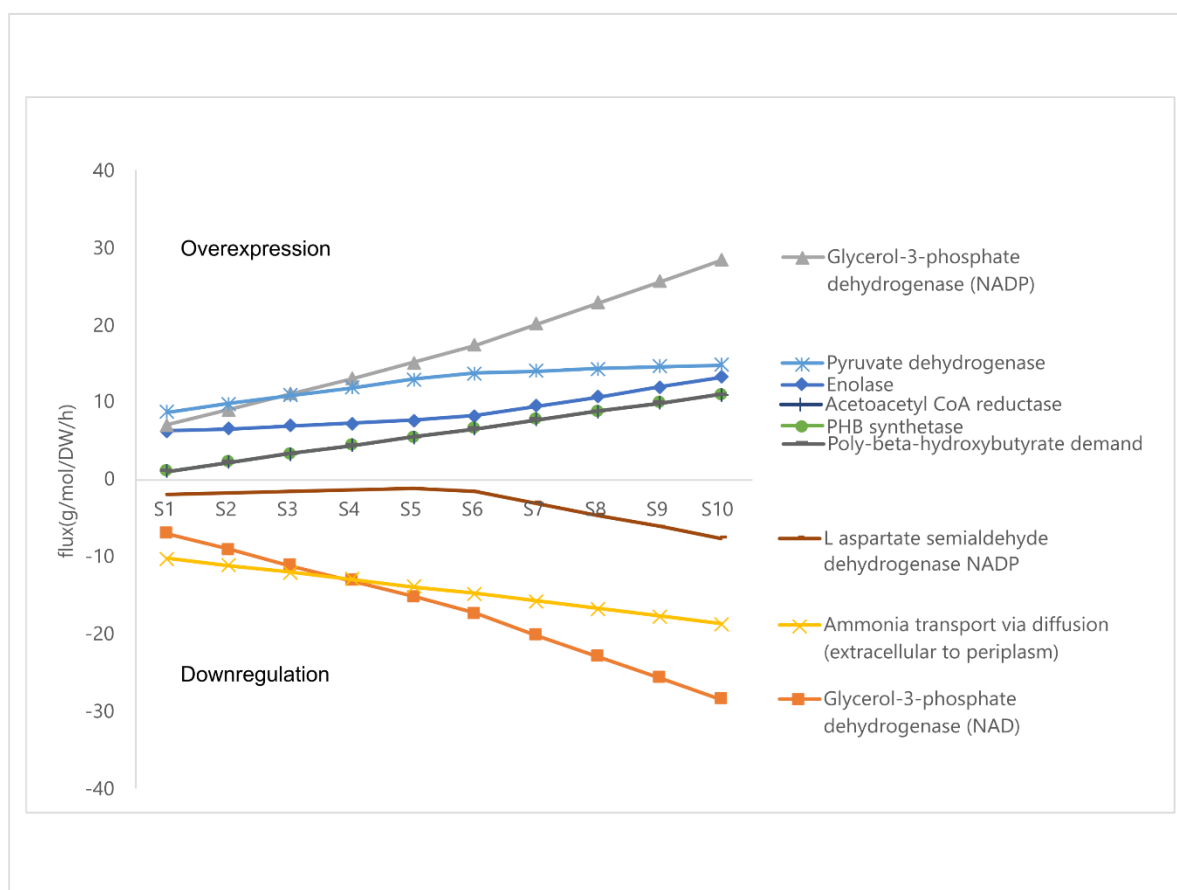


Figure 4.16: Overexpression and Downregulation targets to increase the biosynthesis of PHB in *H. elongata* 153B

ii) Ectoine

The metabolism of ectoine production is of great interest since it can illuminate the metabolic underpinnings of halotolerance and open the door to more efficient biotechnological synthesis of this compatible solute. The biosynthetic family of amino acids generated from aspartate includes ectoine. Oxaloacetate is converted to aspartate, which links the synthesis of ectoine with the anaplerotic processes that add carbon to the tricarboxylic acid cycle (TCA cycle). This puts a lot of strain on these processes, necessitating their regulation not only in response to growth but also in response to the level of extracellular salt (Hobmeier et al., 2020).

The GEM included an extensively curated pathway for ectoine biosynthesis and degradation. We also ensured that the pathway was not a standalone but connected to other pathways by means of precursors. The reactions curated for ectoine biosynthesis included Asd: -aspartate-semialdehyde-dehydrogenase; LysC: aspartate kinase; L-2,4-Diaminobutyric Acid Transaminase -EctB; DoeA: ectoine hydrolase; EctA: l-2,4-diaminobutyric acid N-acetyltransferase; EctC: ectoine synthase; EctD: ectoine hydroxylase; l-2,4-diaminobutyric acid transaminase -DoeD; aspartate-semialdehyde dehydrogenase – DoeC; N-acetyl-l-2,4-diaminobutyric acid deacetylase – DoeB as previously displayed in Figure 4.7. Pathway curation was particularly difficult when using genome annotations from one pipeline only. Genome annotations from the RAST server were especially important in this step as it provided more genetic evidence which facilitated confidence scoring from genome annotation.

The biomass reaction included 5-hydroxyectoine and ectoine and the amount of each was determined using the values and fractions previously calculated and implemented in the *C. salexigens* GEM established by Piubeli et al., 2018. These calculations were done taking into account the different molarities of NaCl. As both species are closely related and have similar abilities to adapt to high levels of salt, this was a useful method for estimating the amounts.

In order to accurately simulate intracellular ectoine production within our model, we included demand reactions. Thiele and Palsson (2010) explained that reactions that represent demand of a metabolite can be included for substances for which there is literature or experimental evidence that the organism in question is able to produce it. This is also valid for substances that are produced as a result of environmental influence (Thiele and Palsson, 2010). This information is useful when trying to predict osmoadaptation behavior. According to our simulations, the synthesis of ectoine is inversely related to growth and requires specific conditions in order to occur. Additionally, ectoine synthesis has been found to be energy-intensive and utilize several intermediates within the tricarboxylic acid cycle. (Piubeli et al., 2018). Also in a more recent study, the transcriptome analysis of *H. elongata* at various NaCl concentrations revealed that both 8% and 13% NaCl conditions led to an increase in the expression of genes related to the flagellar assembly pathway, the Entner-Doudoroff pathway, the pentose phosphate pathway, and the ectoine metabolism, but had a negative impact on the

tricarboxylic acid cycle and fatty acid metabolism (Yu et al., 2022). These results emphasize further the strain on anaplerotic pathways supplying precursors for ectoine synthesis. The model was initially simulated with FBA and FVA with the biomass as objective.

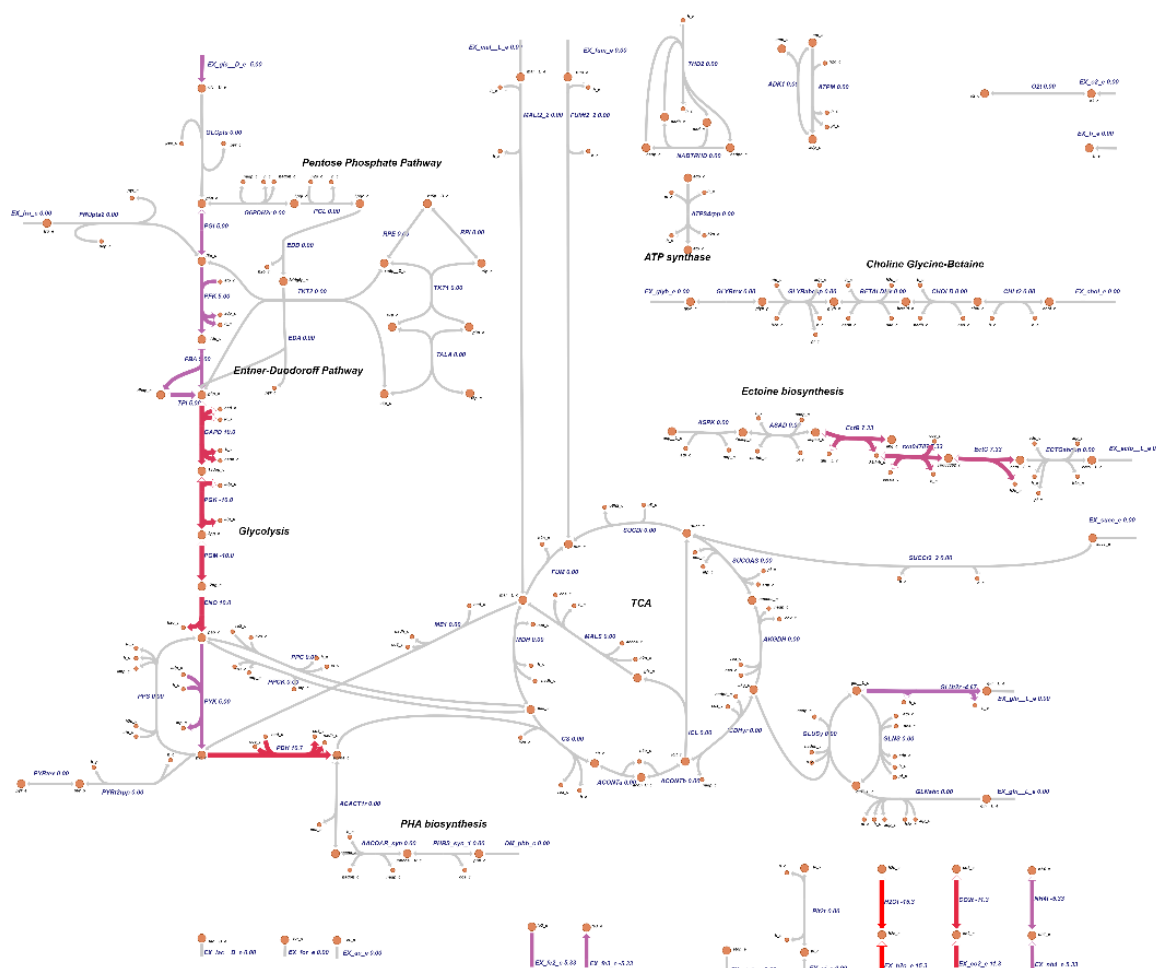


Figure 4.17. Flux distribution in iEB1239 when flux towards ectoine biosynthesis is maximized.

The minimum and maximum flux for ectoine synthesis were 0.0 and 3.6e-12 mmol/gDW/h respectively. When the biomass was changed to the ectoine demand, Ectoine production is negatively correlated with growth rate, according to the fitness of iEB1159 to make it, and when the growth rate is slowed, ectoine production can reach up to 7.33

mmol/gDW/h (Figure 4.18). A flux map showing the flow of metabolites is shown in Figure 4.17.

Accordingly, the flux in the Embden-Meyerhof pathway increased supplying minimal energy for growth as well as an increase in flux towards the synthesis of glutamate which is a precursor for ectoine synthesis. For 5-hydroxyectoine, a comparable pattern was also seen.

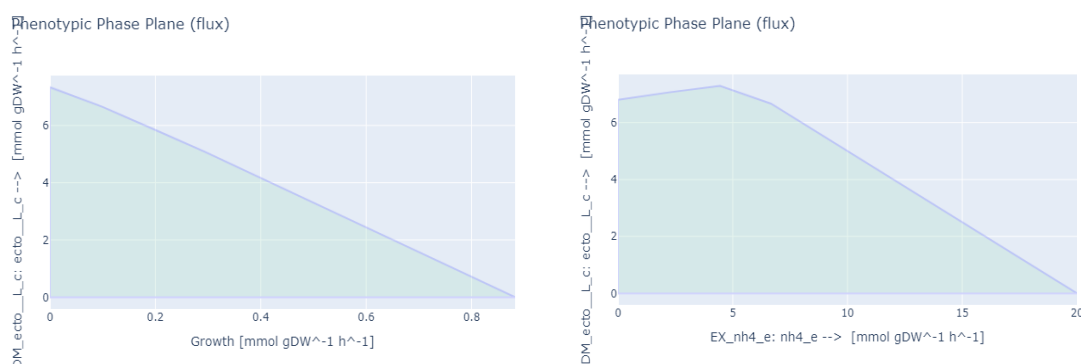


Figure 4.18. Phenotypic phase plane of ectoine production. A) with changes in growth rate B) with changes in the uptake rates of the nitrogen source

We also carried out single gene deletion simulations to investigate the role of precursors from anaplerotic pathways. FBA was carried out and L-ectoine synthesis was considered as the objective function. Gene deletion on each of oxaloacetate decarboxylase (Oad) (Figure 4.19) and phosphoenolpyruvate carboxylase (Ppc) (Figure 4.20) showed no reduction in the flux of ectoine synthesis.

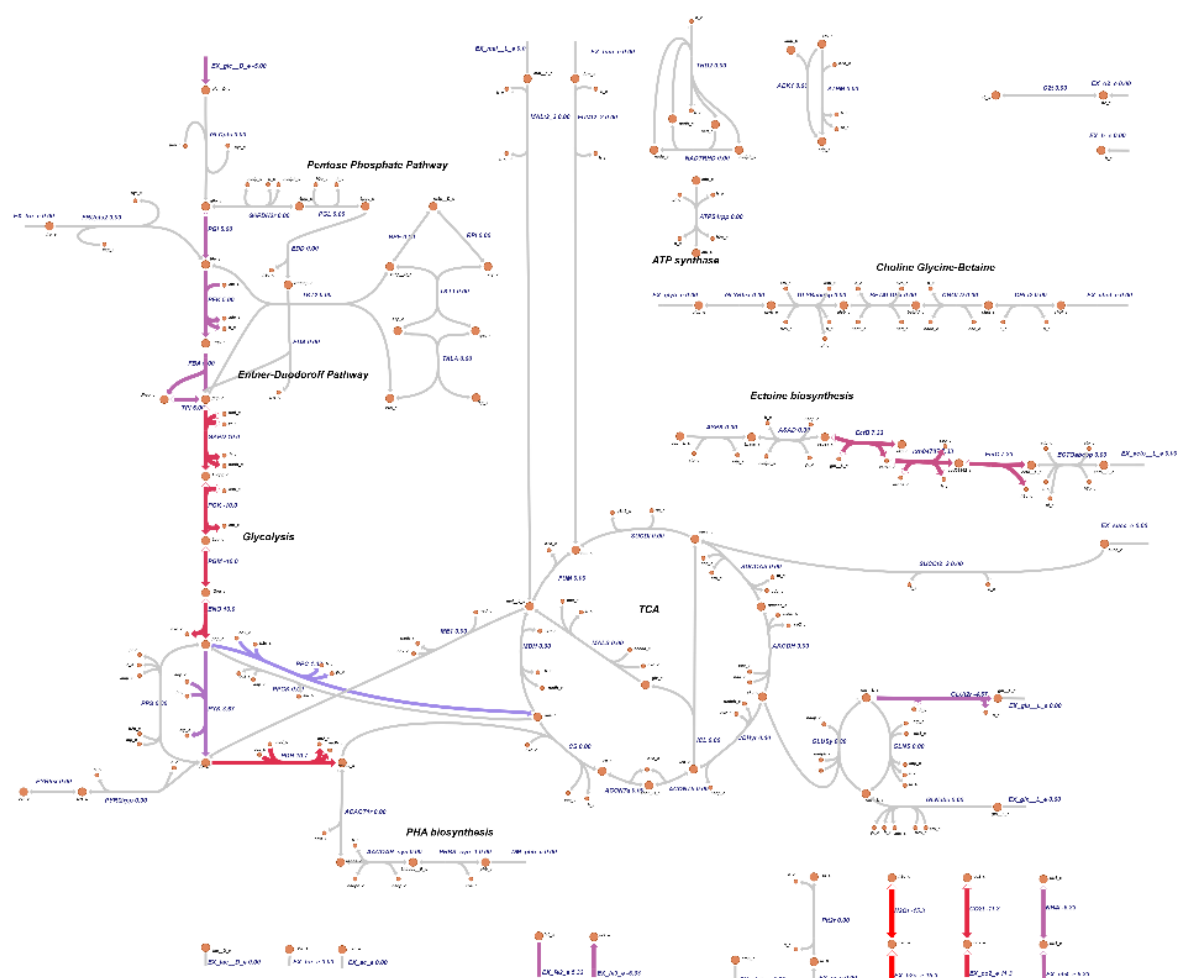


Figure 4.19. Flux map of iEB1239 showing flux distribution towards ectoine biosynthesis in a oad mutant

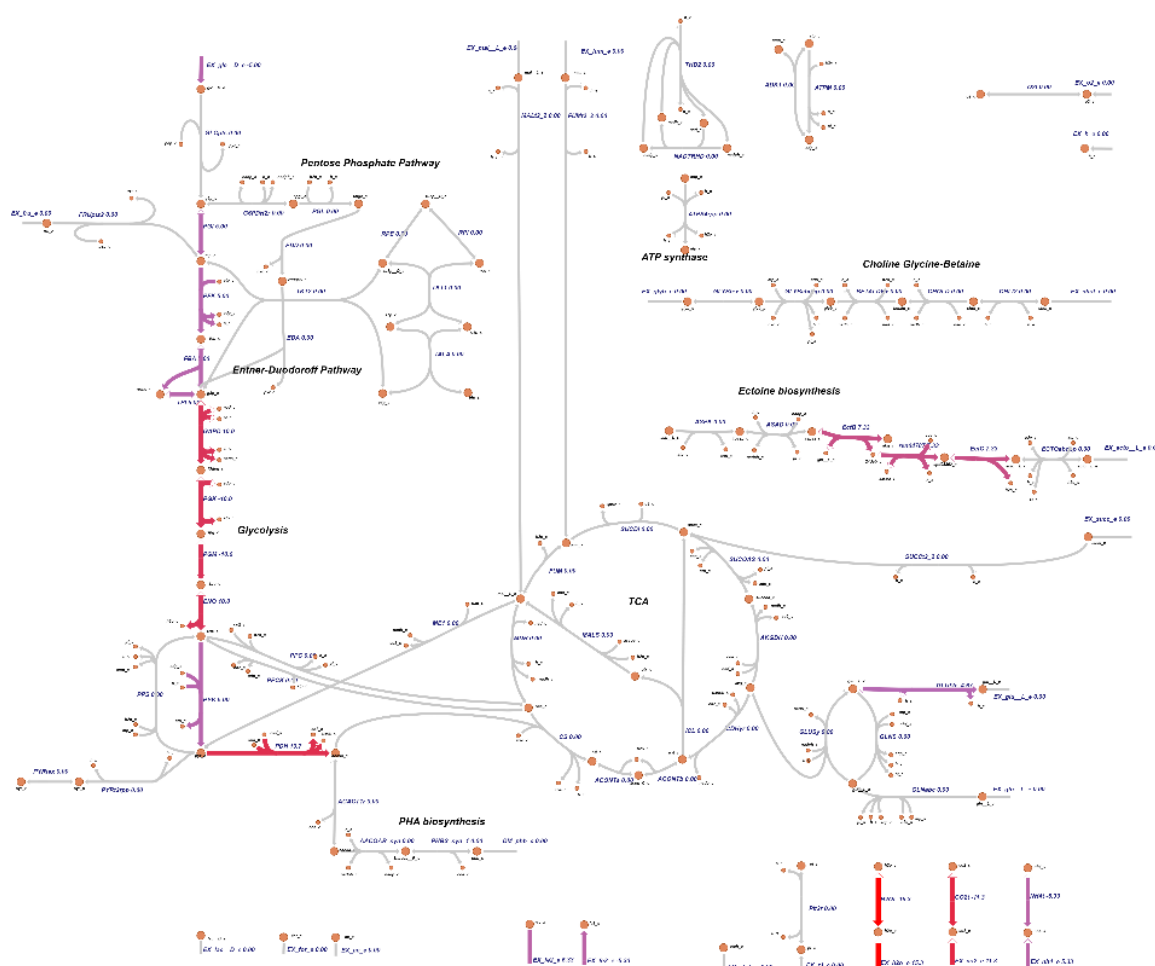


Figure 4.20. Flux map of iEB1239 showing flux distribution towards ectoine biosynthesis in a *Ppc* mutant

These results show that each of these reactions are not on their own indispensable for the synthesis of ectoine. This ensures that ectoine can be produced irrespective of conditions and osmoregulation is ensured.

The flux diagram for these changes is shown in Figure 4.19 and Figure 4.20 while the flux values are also shown in the figures as well. A similar experimental characterization was done with the same anaplerotic reactions phosphoenolpyruvate carboxylase and oxaloacetate decarboxylase. Their findings suggested that these two enzymes work together to ensure the availability of oxaloacetate for biosynthetic processes (Hobmeier et al., 2020). Our findings

agree with these reports and are significant in understanding the dynamics of ectoine biosynthesis in *H. elongata*.

The results show that *H. elongata* is distinct from other gram-negative bacteria in that its anaplerotic responses are undoubtedly influenced by selection pressures. Ectoine-producing halophiles have greater oxaloacetate metabolic requirements, and Ppc's anaplerotic efficiency is reduced by the majority of marine bacteria's reliance on the Entner-Doudoroff pathway. One of the main enzymes carrying out this job is generally ppc. In addition to Ppc, which has not yet been shown to affect anaplerotic flux in vivo, Oad is another enzyme that does so in *H. elongata*.

Findings from other systems biology studies suggest that important osmoregulatory processes occur outside of the ectoine production route. There is a systemic reaction that calls for coordinated activity from various subsystems, including salt transport and the PEP-pyruvate-oxaloacetate node (Kindzierski et al., 2017). Simulations about the adaptive capabilities in *Halomonas* sp. AAD12, the total flux via the energy-producing pathways, specifically the pentose phosphate route and gluconeogenesis, was much reduced, and the carbon source that entered the system as citrate was primarily diverted to osmolyte synthesis at high salinity (Ceylan et al., 2012).

Our model was also able to synthesize ectoine when xylose and a glucose/xylose combination was used as the carbon source. These were used to represent a hypothetical lignocellulosic biomass substrate. Ectoine production flux on xylose alone was 6.64mmol/gDW/h while that on glucose alone was 7.33 mmol/gDW/h. While a combination of xylose and glucose resulted in a flux of 10.00 mmol/gDW/h. Other studies have reported a slight difference in dynamic. Ectoine production by *H. elongata* when grown on xylose as the only carbon source was 333 mmol/kg fresh cell weight (FW), which was 1.4 times more than that of growth on glucose (Tanimura et al., 2013). Using both carbon sources resulted in an overall increase in ectoine synthesis also. Overall, iEB1239 shows a reliability in predicting the dynamics of ectoine biosynthesis by accurately predicting the dynamics for optimal ectoine biosynthesis. It also provides a reliable platform to investigate novel hypothesis on the role of different important reactions in bioproduct synthesis pathways.

iii) Vanillic acid

Vanillic acid has various bioactivity against diseases such as obesity, neurodegenerative, cancer, diabetes, cardiovascular, and hepatic diseases through blocking pathways. Additionally, its derivatives have the ability to treat bacterial and fungal infections, as well as autoimmune illnesses (Kaur et al., 2022). The first halophilic bacteria to be demonstrated to convert ferulic acid to vanillic acid under hypersaline conditions is *Halomonas elongata* strain Mar (=CCUG 52759), which was isolated from table-olive fermentation. This strain was able to facilitate the synthesis of large amounts of vanillic acid and trace amounts of vanillin during growth on ferulic acid (Abdelkafi et al., 2006). To research its production with our model, we curated the related pathways in *iEB1239*. Reactions included were the conversion of vanillin to vanillate by vanillin dehydrogenase [EC:1.2.1.67] (*vdh*). Simulations were run with glucose, xylose, and glucose/xylose as the carbon source. As the reaction was a one-step reaction, we did not do any gene knock-out simulations.

The model produced vanillic acid when vanillin was included in the medium at an uptake rate of -10mmol/gDW/h/. There was also an overall increase in the flux towards vanillate production when both glucose and xylose were included in the medium. This shows that vanillic acid can be produced from biomass containing xylose using *H. elongata* 153B.

4.14. CRISPR metabolic engineering

CRISPR metabolic engineering results are still pending. Research towards designing plasmids and overexpressing the *PhaC* gene will enable the validation of results produced in this work and engineering of a strain producing higher yields of PHBs.

5. CONCLUSIONS AND RECOMMENDATIONS

The *Halomonas* bacteria are gaining attention due to their versatility and the ability to synthesize various useful bioproducts. While there have been numerous genome projects involving *Halomonas* species, there is a lack of information on their genetic characteristics with regard to biotechnology and no comparisons between strains. The sequenced genome herein described shared many features with the other strains of the same species. However, *H. elongata* 153B was closest to *H. elongata* DSM 2581 which is the reference genome. Differences in gene content and the locus was even more pronounced when comparisons were made with seven other *Halomonas* species. As such the strains were much closer to each other than they are to other species. Specifically, within the *Halomonas elongata* strains 2064 core genome orthologs were common but *H. elongata* 153B alone had 666 singletons. Extensive analysis of the genome led to the description of genes involved in the transport of metabolites, osmotic regulation in hypersaline environments, and survival in the presence of oxidative stress which all form the basis for tolerance to hypersaline environments. In addition, genomic features related to the production of PHB and ectoine, which are valuable bioproducts, were also identified.

A genome-scale metabolic network model of *H. elongata* 153B's metabolism was also developed. We started by creating a draft reconstruction, which was then iteratively manually checked, curated, then annotated with database metadata, and then subsequently used for simulations when ready. Finally, we used data from the literature to validate the model. Our concept offers a framework for combining machine learning and deep learning methods with multi-omic data integration. There is a lack of detailed experimental data available for *H. elongata*, compared to other organisms such as *S. cerevisiae* or *E. coli* which suggests that there are still many unknowns and research opportunities, particularly when it comes to using *H. elongata* in modeling and optimization for specific conditions (Czajka et al., 2021; Vijayakumar and Angione, 2021; Zhang et al., 2020).

In this research, we developed a draft metabolic network model that has been validated and can be modified to align with other Genome-Scale Metabolic Models (GEMs).

Additionally, this model can be further enhanced by incorporating context-specific modeling techniques, particularly when omics data related to a particular condition is available.

However, we observe that GEMs continue to be effective tools even when the knowledge base is incomplete. The model developed here, for example, accurately predicts growth on various carbon sources in minimum media, as well as the generation of polyhydroxybutyrates, ectoines, and vanillic acid.

The findings provided here will contribute to our understanding of the metabolic processes and adaption mechanisms in halophilic species, enabling biotechnological exploitation to produce valuable products as well as the extraction of useful enzymes. The model will allow researchers from a variety of fields to better understand the metabolism of the organism and generate new hypotheses about how it may be used in industrial biotechnology. Further research will involve the design of a CRISPR Cas engineering tool to carry out overexpression of the PHA genes in *Halomonas elongata* 153B to improve the yield of PHA.

REFERENCES

- Abdelkafi, S., Sayadi, S., Ali Gam, Z.B., Casalot, L., Labat, M. (2006). *Bioconversion of ferulic acid to vanillic acid by Halomonas elongata isolated from table-olive fermentation*. FEMS Microbiology Letters, 262, 115–120. <https://doi.org/10.1111/j.1574-6968.2006.00381.x>
- Adli, M. (2018). *The CRISPR tool kit for genome editing and beyond*. Nat. Commun. 9, 1911. <https://doi.org/10.1038/s41467-018-04252-2>
- Akhlaq, S., Singh, D., Mittal, N., Srivastava, G., Siddiqui, S., Faridi, S.A., Siddiqui, M.H. (2022). *Polyhydroxybutyrate biosynthesis from different waste materials, degradation, and analytic methods: a short review*. Polym. Bull. <https://doi.org/10.1007/s00289-022-04406-9>
- Alcántara, R., Axelsen, K.B., Morgat, A., Belda, E., Coudert, E., Bridge, A., Cao, H., de Matos, P., Ennis, M., Turner, S., Owen, G., Bougueleret, L., Xenarios, I., Steinbeck, C. (2012). *Rhea—a manually curated resource of biochemical reactions*. Nucleic Acids Res, 40, D754–D760. <https://doi.org/10.1093/nar/gkr1126>
- Alikhan, N.-F., Petty, N.K., Ben Zakour, N.L., Beatson, S.A. (2011). *BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons*. BMC Genomics, 12, 402. <https://doi.org/10.1186/1471-2164-12-402>
- Anilionyte, O., Liang, H., Ma, X., Yang, L., Zhou, K., 2018. Short, auto-inducible promoters for well-controlled protein expression in Escherichia coli. Appl Microbiol Biotechnol 102, 7007–7015. <https://doi.org/10.1007/s00253-018-9141-z>
- Arahal, D.R., Ventosa, A. (2006). *The Family Halomonadaceae*, in: The Family Halomonadaceae. Springer.
- Arkin, A.P., Cottingham, R.W., Henry, C.S., Harris, N.L., Stevens, R.L., Maslov, S., Dehal, P., Ware, D., Perez, F., Canon, S., Sneddon, M.W., Henderson, M.L., Riehl, W.J., Murphy-Olson, D., Chan, S.Y., Kamimura, R.T., Kumari, S., Drake, M.M., Brettin, T.S., Glass, E.M., Chivian, D., Gunter, D., Weston, D.J., Allen, B.H., Baumohl, J., Best, A.A., Bowen, B., Brenner, S.E., Bun, C.C., Chandonia, J.-M., Chia, J.-M., Colasanti, R., Conrad, N., Davis, J.J., Davison, B.H., DeJongh, M., Devoid, S., Dietrich, E., Dubchak, I., Edirisinghe, J.N., Fang, G., Faria, J.P., Frybarger, P.M., Gerlach, W., Gerstein, M., Greiner, A., Gurtowski, J., Haun, H.L., He, F., Jain, R., Joachimiak, M.P., Keegan, K.P., Kondo, S., Kumar, V., Land, M.L., Meyer, F., Mills, M., Novichkov, P.S., Oh, T., Olsen, G.J., Olson, R., Parrello, B., Pasternak, S., Pearson, E., Poon, S.S., Price, G.A., Ramakrishnan, S., Ranjan, P., Ronald, P.C., Schatz, M.C., Seaver, S.M.D., Shukla, M., Sutormin, R.A., Syed, M.H., Thomason, J., Tintle, N.L., et al (2018). *KBase: The United States Department of Energy Systems Biology Knowledgebase*. Nature Biotechnology 36, 566–569. <https://doi.org/10.1038/nbt.4163>

REFERENCES (continued)

- Arora, S., Vanza*, M.J., Mehta, R., Bhuva, C., Patel, P.N. (2014). *Halophilic microbes for bio-remediation of salt affected soils*. AJMR, 8, 3070–3078. <https://doi.org/10.5897/AJMR2014.6960>
- Aston, J.E., Peyton, B.M. (2007). *Response of Halomonas campisalis to saline stress: changes in growth kinetics, compatible solute production and membrane phospholipid fatty acid composition*. FEMS Microbiology Letters, 274, 196–203. <https://doi.org/10.1111/j.1574-6968.2007.00851.x>
- Athmika, Ghate, S.D., Arun, A.B., Rao, S.S., Kumar, S.T.A., Kandiyil, M.K., Saptami, K., Rekha, P.D. (2021). *Genome analysis of a halophilic bacterium Halomonas malpeensis YU-PRIM-29T reveals its exopolysaccharide and pigment producing capabilities*. Sci Rep, 11. <https://doi.org/10.1038/s41598-021-81395-1>
- Başer, H., Yuksekdağ, Z. (2019). *Exopolysaccharide (EPS) production of halophilic bacteria*. Eurasian Journal of Biological and Chemical Sciences, 2, 84–89.
- Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S.Y., Medema, M.H., Weber, T., (2019). *AntiSMASH 5.0: updates to the secondary metabolite genome mining pipeline*. Nucleic Acids Research, 47, W81–W87. <https://doi.org/10.1093/nar/gkz310>
- Boutet, E., Lieberherr, D., Tognolli, M., Schneider, M., Bairoch, A. (2007). *UniProtKB/Swiss-Prot*. Methods Mol Biol 406, 89–112. https://doi.org/10.1007/978-1-59745-535-0_4
- Brettin, T., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Olsen, G.J., Olson, R., Overbeek, R., Parrello, B., Pusch, G.D., Shukla, M., Thomason, J.A., Stevens, R., Vonstein, V., Wattam, A.R., Xia, F. (2015). *RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes*. Sci Rep 5, 8365. <https://doi.org/10.1038/srep08365>
- Bruna, N., Collao, B., Tello, A., Caravantes, P., Díaz-Silva, N., Monrás, J.P., Órdenes-Aenishanslins, N., Flores, M., Espinoza-Gonzalez, R., Bravo, D., Pérez-Donoso, J.M., (2019). *Synthesis of salt-stable fluorescent nanoparticles (quantum dots) by polyextremophile halophilic bacteria*. Sci Rep 9, 1953. <https://doi.org/10.1038/s41598-018-38330-8>
- Çakmak, H., Çelik, P.A., Çınar, S., Hoşgün, E.Z., Mutlu, M.B., Çabuk, A. (2021). *Levan production potentials from different hypersaline environments in Turkey*. Journal of Microbiology, Biotechnology and Food Sciences 2021, 61–64. <https://doi.org/10.15414/jmbfs.2020.10.1.61-64>
- Calamita, G. (2000). *The Escherichia coli aquaporin-Z water channel*. Mol Microbiol 37, 254–262. <https://doi.org/10.1046/j.1365-2958.2000.02016.x>

REFERENCES (continued)

- Cánovas, D., Vargas, C., Csonka, L.N., Ventosa, A., Nieto, J.J. (1998). *Synthesis of Glycine Betaine from Exogenous Choline in the Moderately Halophilic Bacterium Halomonas elongata*. *Applied and Environmental Microbiology*, 64, 4095–4097. <https://doi.org/10.1128/AEM.64.10.4095-4097.1998>
- Cánovas, D., Vargas, C., Csonka, L.N., Ventosa, A., Nieto, J.J. (1996). *Osmoprotectants in Halomonas elongata: high-affinity betaine transport system and choline-betaine pathway*. *Journal of Bacteriology*, 178, 7221–7226. <https://doi.org/10.1128/jb.178.24.7221-7226.1996>
- Cánovas, D., Vargas, C., Kneip, S., Morón, M.-J., Ventosa, A., Bremer, E., Nieto, J.J.Y. (2000). *Genes for the synthesis of the osmoprotectant glycine betaine from choline in the moderately halophilic bacterium Halomonas elongata DSM 3043*. *Microbiology*, 146, 455–463. <https://doi.org/10.1099/00221287-146-2-455>
- Cardoso, J.G.R., Jensen, K., Lieven, C., Lærke Hansen, A.S., Galkina, S., Beber, M., Özdemir, E., Herrgård, M.J., Redestig, H., Sonnenschein, N. (2018). *Cameo: A Python Library for Computer Aided Metabolic Engineering and Optimization of Cell Factories*. *ACS Synth. Biol.* 7, 1163–1166. <https://doi.org/10.1021/acssynbio.7b00423>
- Ceylan, S., Yilan, G., Akbulut, B.S., Poli, A., Kazan, D. (2012). *Interplay of adaptive capabilities of Halomonas sp AAD12 under salt stress*. *J. Biosci. Bioeng.* 114, 45–52. <https://doi.org/10.1016/j.jbiosc.2012.02.030>
- Chen, C., Beattie, G.A. (2008). *Pseudomonas syringae BetT Is a Low-Affinity Choline Transporter That Is Responsible for Superior Osmoprotection by Choline over Glycine Betaine*. *J Bacteriol* 190, 2717–2725. <https://doi.org/10.1128/JB.01585-07>
- Chen, D.-D., Tian, Y., Jiao, J.-Y., Zhang, X.-T., Zhang, Y.-G., Dong, Z.-Y., Xiong, M.-J., Xiao, M., Shu, W.-S., Li, W.-J. (2020). *Comparative genomics analysis of Nitriliruptoria reveals the genomic differences and salt adaptation strategies*. *Extremophiles* 24, 249–264. <https://doi.org/10.1007/s00792-019-01150-3>
- Chen, X., Yu, L., Qiao, G., Chen, G.-Q. (2018). *Reprogramming Halomonas for industrial production of chemicals*. *Journal of Industrial Microbiology and Biotechnology* 45, 545–554. <https://doi.org/10.1007/s10295-018-2055-z>
- Chyou, T., Brown, C.M. (2019). *Prediction and diversity of tracrRNAs from type II CRISPR-Cas systems*. *RNA Biol.* 16, 423–434. <https://doi.org/10.1080/15476286.2018.1498281>

REFERENCES (continued)

- Cristea, A., Baricz, A., Leopold, N., Floare, C. g., Borodi, G., Kacso, I., Tripon, S., Bulzu, P. a., Andrei, A.-Ș., Cadar, O., Levei, E. a., Banciu, H. l. (2018). *Polyhydroxybutyrate production by an extremely halotolerant Halomonas elongata strain isolated from the hypersaline meromictic Fără Fund Lake (Transylvanian Basin, Romania)*. Journal of Applied Microbiology 125, 1343–1357. <https://doi.org/10.1111/jam.14029>
- Czajka, J.J., Oyetunde, T., Tang, Y.J. (2021). *Integrated knowledge mining, genome-scale modeling, and machine learning for predicting Yarrowia lipolytica bioproduction*. Metabolic Engineering 67, 227–236. <https://doi.org/10.1016/j.ymben.2021.07.003>
- Czech, L., Hermann, L., Stöveken, N., Richter, A.A., Höppner, A., Smits, S.H.J., Heider, J., Bremer, E. (2018). *Role of the Extremolytes Ectoine and Hydroxyectoine as Stress Protectants and Nutrients: Genetics, Phylogenomics, Biochemistry, and Structural Analysis*. Genes (Basel) 9, 177. <https://doi.org/10.3390/genes9040177>
- Daoud, L., Ben Ali, M. (2020). *Halophilic microorganisms: Interesting group of extremophiles with important applications in biotechnology and environment*, In: Salwan, R., Sharma, V. (Eds.), Physiological and Biotechnological Aspects of Extremophiles. Academic Press, pp. 51–64. <https://doi.org/10.1016/B978-0-12-818322-9.00005-8>
- DasSarma, P., Capes, M.D., DasSarma, S. (2019). *Comparative Genomics of Halobacterium Strains From Diverse Locations*, In: Das, S., Dash, H.R. (Eds.), Microbial Diversity in the Genomic Era. Academic Press, pp. 285–322. <https://doi.org/10.1016/B978-0-12-814849-5.00017-4>
- DasSarma, S., DasSarma, P. (2015). *Halophiles and their enzymes: Negativity put to good use*. Curr Opin Microbiol 25, 120–126. <https://doi.org/10.1016/j.mib.2015.05.009>
- de Castro, I., Mendo, S., Caetano, T. (2020). *Antibiotics from Haloarchaea: What Can We Learn from Comparative Genomics?* Mar Biotechnol 22, 308–316. <https://doi.org/10.1007/s10126-020-09952-9>
- De Lourdes Moreno, M., Pérez, D., García, M.T., Mellado, E. (2013). *Halophilic Bacteria as a Source of Novel Hydrolytic Enzymes*. Life 3, 38–51. <https://doi.org/10.3390/life3010038>
- Degtyarenko, K., de Matos, P., Ennis, M., Hastings, J., Zbinden, M., McNaught, A., Alcántara, R., Darsow, M., Guedj, M., Ashburner, M. (2008). *ChEBI: a database and ontology for chemical entities of biological interest*. Nucleic Acids Res 36, D344–D350. <https://doi.org/10.1093/nar/gkm791>

REFERENCES (contiuned)

- Deshmukh, S., Kanekar, P., Bhadekar, R. (2017). *Production And Characterization Of Exopolysaccharide From Marine Moderately Halophilic Bacterium Halomonas Smyrnensis Svd III*. International Journal of Pharmacy and Pharmaceutical Sciences 9, 146. <https://doi.org/10.22159/ijpps.2017v9i10.20935>
- Donio, M., Ronica, S., Viji, V.T., Velmurugan, S., Jenifer, J.A., Michaelbabu, M., Citarasu, T., (2013). *Isolation and characterization of halophilic Bacillus sp. BS3 able to produce pharmacologically important biosurfactants*. Asian Pacific Journal of Tropical Medicine 6, 876–883. [https://doi.org/10.1016/S1995-7645\(13\)60156-X](https://doi.org/10.1016/S1995-7645(13)60156-X)
- Dötsch, A., Severin, J., Alt, W., Galinski, E.A., Kreft, J.-U. (2008). *A mathematical model for growth and osmoregulation in halophilic bacteria*. Microbiology 154, 2956–2969. <https://doi.org/10.1099/mic.0.2007/012237-0>
- Ebrahim, A., Lerman, J.A., Palsson, B.O., Hyduke, D.R. (2013). *COBRApy: CONstraints-Based Reconstruction and Analysis for Python*. BMC Systems Biology 7, 74. <https://doi.org/10.1186/1752-0509-7-74>
- Enuh, B.M., Nural Yaman, B., Tarzi, C., Aytar Çelik, P., Mutlu, M.B., Angione, C. (2022). *Whole-genome sequencing and genome-scale metabolic modeling of Chromohalobacter canadensis 85B to explore its salt tolerance and biotechnological use*. MicrobiologyOpen 11, e1328. <https://doi.org/10.1002/mbo3.1328>
- Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., Haw, R., Jassal, B., Korninger, F., May, B., Milacic, M., Roca, C.D., Rothfels, K., Sevilla, C., Shamovsky, V., Shorser, S., Varusai, T., Viteri, G., Weiser, J., Wu, G., Stein, L., Hermjakob, H., D'Eustachio, P. (2018). *The Reactome Pathway Knowledgebase*. Nucleic Acids Research 46, D649–D655. <https://doi.org/10.1093/nar/gkx1132>
- Fang, X., Lloyd, C.J., Palsson, B.O. (2020). *Reconstructing organisms in silico: genome-scale models and their emerging applications*. Nat Rev Microbiol 18, 731–743. <https://doi.org/10.1038/s41579-020-00440-4>
- Fariq, A., Yasmin, A. (2020). *Production, characterization and bioactivities of biosurfactants from newly isolated strictly halophilic bacteria*. Process Biochemistry 98, 1–10. <https://doi.org/10.1016/j.procbio.2020.07.011>
- Fariq, A., Yasmin, A., Jamil, M. (2019). *Production, characterization and antimicrobial activities of bio-pigments by Aquisolibacillus elongatus MB592, Salinicoccus sesuvii MB597, and Halomonas aquamarina MB598 isolated from Khewra Salt Range, Pakistan*. Extremophiles 23, 435–449. <https://doi.org/10.1007/s00792-019-01095-7>

REFERENCES (continued)

- Freed, E., Fenster, J., Smolinski, S.L., Walker, J., Henard, C.A., Gill, R., Eckert, C.A. (2018). *Building a genome engineering toolbox in nonmodel prokaryotic microbes*. Biotechnol. Bioeng. 115, 2120–2138. <https://doi.org/10.1002/bit.26727>
- Frikha-Dammak, D., Ayadi, H., Hakim-Rekik, I., Belbahri, L., Maalej, S. (2021). *Genome analysis of the salt-resistant Paludifilum halophilum DSM 102817T reveals genes involved in flux-tuning of ectoines and unexplored bioactive secondary metabolites*. World J Microbiol Biotechnol 37, 178. <https://doi.org/10.1007/s11274-021-03147-7>
- Fu, X.-Z., Tan, D., Aibaidula, G., Wu, Q., Chen, J.-C., Chen, G.-Q. (2014). *Development of Halomonas TD01 as a host for open production of chemicals*. Metabolic Engineering 23, 78–91. <https://doi.org/10.1016/j.ymben.2014.02.006>
- García, M.T., Mellado, E., Ostos, J.C., Ventosa, A. (2004). *Halomonas organivorans* sp. nov., a moderate halophile able to degrade aromatic compounds. International Journal of Systematic and Evolutionary Microbiology 54, 1723–1728. <https://doi.org/10.1099/ijs.0.63114-0>
- Gedikli, S., Aytar Çelik, P., Demirbilek, M., Mutlu, M., Denkbaz, E., Çabuk, A. (2019). *Experimental Exploration of Thermostable Poly (β -Hydroxybutyrates) by Geobacillus kaustophilus Using Box-Behnken Design*. Journal of Polymers and the Environment 27. <https://doi.org/10.1007/s10924-018-1335-z>
- Ghasemi, A., Asad, S., Kabiri, M., Dabirmanesh, B. (2017). *Cloning and characterization of Halomonas elongata L-asparaginase, a promising chemotherapeutic agent*. Appl Microbiol Biotechnol 101, 7227–7238. <https://doi.org/10.1007/s00253-017-8456-5>
- Gostinčar, C., Gunde-Cimerman, N. (2018). *Overview of Oxidative Stress Response Genes in Selected Halophilic Fungi*. Genes (Basel) 9, E143. <https://doi.org/10.3390/genes9030143>
- Grund, E., Denecke, B., Eichenlaub, R. (1992). *Naphthalene degradation via salicylate and gentisate by Rhodococcus sp. strain B4*. Applied and Environmental Microbiology 58, 1874–1877. <https://doi.org/10.1128/aem.58.6.1874-1877.1992>
- Guan, Y., Ngugi, D.K., Vinu, M., Blom, J., Alam, I., Guillot, S., Ferry, J.G., Stingl, U. (2019). *Comparative Genomics of the Genus Methanohalophilus, Including a Newly Isolated Strain From Kebrit Deep in the Red Sea*. Front. Microbiol. 0. <https://doi.org/10.3389/fmicb.2019.00839>
- Gunde-Cimerman, N., Plemenitaš, A., Oren, A. (2018). *Strategies of adaptation of microorganisms of the three domains of life to high salt concentrations*. FEMS Microbiology Reviews 42, 353–375. <https://doi.org/10.1093/femsre/fuy009>

REFERENCES (continued)

- Gupta, S., Sharma, P., Dev, K., Sourirajan, A. (2020). *Isolation of gene conferring salt tolerance from halophilic bacteria of Lunsu, Himachal Pradesh, India*. Journal of Genetic Engineering and Biotechnology 18, 57. <https://doi.org/10.1186/s43141-020-00070-6>
- Hamed, J., Mohammadipanah, F., Ventosa, A. (2013). *Systematic and biotechnological aspects of halophilic and halotolerant actinomycetes*. Extremophiles 17, 1–13. <https://doi.org/10.1007/s00792-012-0493-5>
- Hart, D.J., Vreeland, R.H. (1988). *Changes in the hydrophobic-hydrophilic cell surface character of Halomonas elongata in response to NaCl*. Journal of Bacteriology 170, 132–135. <https://doi.org/10.1128/jb.170.1.132-135.1988>
- Heller, S.R., McNaught, A., Pletnev, I., Stein, S., Tchekhovskoi, D. (2015). *InChI, the IUPAC International Chemical Identifier*. J Cheminform 7, 23. <https://doi.org/10.1186/s13321-015-0068-4>
- Henciya, S., Vengateshwaran, T.D., Gokul, M.S., Dahms, H.-U., James, R.A. (2020). *Antibacterial Activity of Halophilic Bacteria Against Drug-Resistant Microbes Associated with Diabetic Foot Infections*. Curr Microbiol 77, 3711–3723. <https://doi.org/10.1007/s00284-020-02190-1>
- Hobmeier, K., Goëss, M.C., Sehr, C., Schwaminger, S., Berensmeier, S., Kremling, A., Kunte, H.J., Pflüger-Grau, K., Marin-Sanguino, A. (2020). *Anaplerotic Pathways in Halomonas elongata: The Role of the Sodium Gradient*. Frontiers in Microbiology 11.
- Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J. (2019). *eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses*. Nucleic Acids Research 47, D309–D314. <https://doi.org/10.1093/nar/gky1085>
- Jadhav, K., Kushwaha, B., Jadhav, I., Shankar, P., Geethadevi, A., Kumar, G., Mittal, S., Sharma, G.P., Parashar, M., Parashar, D. (2021). *Genomic analysis of a novel species Halomonas shambharensis isolated from hypersaline lake in Northwest India*. Mol Biol Rep 48, 1045–1053. <https://doi.org/10.1007/s11033-020-06131-w>
- Jiang, X.-R., Yao, Z.-H., Chen, G.-Q. (2017). *Controlling cell volume for efficient PHB production by Halomonas*. Metabolic Engineering 44, 30–37. <https://doi.org/10.1016/j.ymben.2017.09.004>
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., Charpentier, E. (2012). *A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity*. Science 337, 816–821. <https://doi.org/10.1126/science.1225829>

REFERENCES (continued)

- John, J., Siva, V., Kumari, R., Arya, A., Kumar, A. (2020). *Unveiling Cultivable and Uncultivable Halophilic Bacteria Inhabiting Marakkanam Saltpan, India and Their Potential for Biotechnological Applications*. Geomicrobiology Journal 37, 691–701. <https://doi.org/10.1080/01490451.2020.1764676>
- Joulak, I., Finore, I., Poli, A., Abid, Y., Bkhairia, I., Nicolaus, B., Di Donato, P., Dal Poggetto, G., Gharsallaoui, A., Attia, H., Azabou, S. (2020). *Hetero-exopolysaccharide from the extremely halophilic Halomonas smyrnensis K2: production, characterization and functional properties in vitro*. 3 Biotech 10, 395. <https://doi.org/10.1007/s13205-020-02356-5>
- Kanehisa, M., Goto, S. (2000). *KEGG: Kyoto Encyclopedia of Genes and Genomes*. Nucleic Acids Res 28, 27–30. <https://doi.org/10.1093/nar/28.1.27>
- Karp, P.D., Billington, R., Caspi, R., Fulcher, C.A., Latendresse, M., Kothari, A., Keseler, I.M., Krummenacker, M., Midford, P.E., Ong, Q., Ong, W.K., Paley, S.M., Subhraveti, P. (2019). *The BioCyc collection of microbial genomes and metabolic pathways*. Brief Bioinform 20, 1085–1093. <https://doi.org/10.1093/bib/bbx085>
- Kaur, J., Gulati, M., Singh, S.K., Kuppusamy, G., Kapoor, B., Mishra, V., Gupta, S., Arshad, M.F., Porwal, O., Jha, N.K., Chaitanya, M.V.N.L., Chellappan, D.K., Gupta, G., Gupta, P.K., Dua, K., Khursheed, R., Awasthi, A., Corrie, L. (2022). *Discovering multifaceted role of vanillic acid beyond flavours: Nutraceutical and therapeutic potential*. Trends in Food Science & Technology 122, 187–200. <https://doi.org/10.1016/j.tifs.2022.02.023>
- Kawata, Y., Kawasaki, K., Shigeri, Y. (2012). *Draft Genome Sequence of Halomonas sp. Strain KM-1, a Moderately Halophilic Bacterium That Produces the Bioplastic Poly(3-Hydroxybutyrate)*. Journal of Bacteriology 194, 2738–2739. <https://doi.org/10.1128/JB.00237-12>
- Keblbouche-Gana, S., Gana, M.L., Ferrioune, I., Khemili, S., Lenchi, N., Akmouci-Toumi, S., Bouanane-Darenfed, N.A., Djelali, N.-E. (2013). *Production of biosurfactant on crude date syrup under saline conditions by entrapped cells of Natrionalba sp. strain E21, an extremely halophilic bacterium isolated from a solar saltern (Ain Salah, Algeria)*. Extremophiles 17, 981–993. <https://doi.org/10.1007/s00792-013-0580-2>
- Kim, B., Kim, W.J., Kim, D.I., Lee, S.Y. (2015). *Applications of genome-scale metabolic network model in metabolic engineering*. J Ind Microbiol Biotechnol 42, 339–348. <https://doi.org/10.1007/s10295-014-1554-9>

REFERENCES (continued)

- Kindzierski, V., Raschke, S., Knabe, N., Siedler, F., Scheffer, B., Pflüger-Grau, K., Pfeiffer, F., Oesterhelt, D., Marin-Sanguino, A., Kunte, H.-J. (2017). *Osmoregulation in the Halophilic Bacterium Halomonas elongata: A Case Study for Integrative Systems Biology*. PLOS ONE 12, e0168818. <https://doi.org/10.1371/journal.pone.0168818>
- King, Z.A., Dräger, A., Ebrahim, A., Sonnenschein, N., Lewis, N.E., Palsson, B.O. (2015). *Escher: A Web Application for Building, Sharing, and Embedding Data-Rich Visualizations of Biological Pathways*. PLOS Computational Biology 11, e1004321. <https://doi.org/10.1371/journal.pcbi.1004321>
- Koller, M. (2018). *A Review on Established and Emerging Fermentation Schemes for Microbial Production of Polyhydroxyalkanoate (PHA) Biopolyesters*. Fermentation 4, 30. <https://doi.org/10.3390/fermentation4020030>
- Kraegeloh, A., Amendt, B., Kunte, H.J. (2005). *Potassium Transport in a Halophilic Member of the Bacteria Domain: Identification and Characterization of the K⁺ Uptake Systems TrkH and TrkI from Halomonas elongata DSM 2581T*. Journal of Bacteriology 187, 1036–1043. <https://doi.org/10.1128/JB.187.3.1036-1043.2005>
- Kucera, D., Pernicová, I., Kovalcik, A., Koller, M., Mullerova, L., Sedlacek, P., Mravec, F., Nebesarova, J., Kalina, M., Marova, I., Krzyzanek, V., Obruca, S. (2018). *Characterization of the promising poly(3-hydroxybutyrate) producing halophilic bacterium Halomonas halophila*. Bioresource Technology 256, 552–556. <https://doi.org/10.1016/j.biortech.2018.02.062>
- Kulski, J.K. (2016). *Next-Generation Sequencing — An Overview of the History, Tools, and “Omic” Applications, Next Generation Sequencing - Advances, Applications and Challenges*. IntechOpen. <https://doi.org/10.5772/61964>
- Kumar, S., Grewal, J., Sadaf, A., Hemamalini, R., Khare, S. (2016). *Halophiles as a source of polyextremophilic α -amylase for industrial applications*. AIMS Microbiology 2, 1–26. <https://doi.org/10.3934/microbiol.2016.1.1>
- Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K. (2018). *MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms*. Mol Biol Evol 35, 1547–1549. <https://doi.org/10.1093/molbev/msy096>
- Lang, S., Cressatti, M., Mendoza, K.E., Coumoundouros, C.N., Plater, S.M., Culham, D.E., Kimber, M.S., Wood, J.M. (2015). *YehZYXW of Escherichia coli Is a Low-Affinity, Non-Osmoregulatory Betaine-Specific ABC Transporter*. Biochemistry 54, 5735–5747. <https://doi.org/10.1021/acs.biochem.5b00274>

REFERENCES (continued)

- Lee, H., Subburaj, S., Tu, L., Lee, K.-Y., Park, G., Lee, G.-J. (2020). *Overview of CRISPR/Cas9: a chronicle of the CRISPR system and application to ornamental crops*. Korean Journal of Agricultural Sciences, 47, 903–920. <https://doi.org/10.7744/KJOAS.20200075>
- Liebisch, G., Fahy, E., Aoki, J., Dennis, E.A., Durand, T., Ejsing, C.S., Fedorova, M., Feussner, I., Griffiths, W.J., Köfeler, H., Merrill, A.H., Murphy, R.C., O'Donnell, V.B., Oskolkova, O., Subramaniam, S., Wakelam, M.J.O., Spener, F. (2020). *Update on LIPID MAPS classification, nomenclature, and shorthand notation for MS-derived lipid structures*. J Lipid Res 61, 1539–1555. <https://doi.org/10.1194/jlr.S120001025>
- Lieven, C., Beber, M.E., Olivier, B.G., Bergmann, F.T., Ataman, M., Babaei, P., Bartell, J.A., Blank, L.M., Chauhan, S., Correia, K., Diener, C., Dräger, A., Ebert, B.E., Edirisinghe, J.N., Faria, J.P., Feist, A.M., Fengos, G., Fleming, R.M.T., García-Jiménez, B., Hatzimanikatis, V., van Helvoirt, W., Henry, C.S., Hermjakob, H., Herrgård, M.J., Kaafarani, A., Kim, H.U., King, Z., Klamt, S., Klipp, E., Koehorst, J.J., König, M., Lakshmanan, M., Lee, D.-Y., Lee, S.Y., Lee, S., Lewis, N.E., Liu, F., Ma, H., Machado, D., Mahadevan, R., Maia, P., Mardinoglu, A., Medlock, G.L., Monk, J.M., Nielsen, J., Nielsen, L.K., Nogales, J., Nookaew, I., Palsson, B.O., Papin, J.A., Patil, K.R., Poolman, M., Price, N.D., Resendis-Antonio, O., Richelle, A., Rocha, I., Sánchez, B.J., Schaap, P.J., Malik Sheriff, R.S., Shoaie, S., Sonnenschein, N., Teusink, B., Vilaça, P., Vik, J.O., Wodke, J.A.H., Xavier, J.C., Yuan, Q., Zakhartsev, M., Zhang, C. (2020). *MEMOTE for standardized genome-scale metabolic model testing*. Nat Biotechnol 38, 272–276. <https://doi.org/10.1038/s41587-020-0446-y>
- Lin, Y., Fan, H., Hao, X., Johnstone, L., Hu, Y., Wei, G., Alwathnani, H.A., Wang, G., Rensing, C. (2012). *Draft Genome Sequence of Halomonas sp. Strain HAL1, a Moderately Halophilic Arsenite-Oxidizing Bacterium Isolated from Gold-Mine Soil*. Journal of Bacteriology 194, 199–200. <https://doi.org/10.1128/JB.06359-11>
- Lin, Z., Zhang, Y., Yuan, Q., Liu, Q., Li, Y., Wang, Z., Ma, H., Chen, T., Zhao, X. (2015). *Metabolic engineering of Escherichia coli for poly(3-hydroxybutyrate) production via threonine bypass*. Microbial Cell Factories 14, 185. <https://doi.org/10.1186/s12934-015-0369-3>
- Liu, C., Baffoe, D.K., Zhan, Y., Zhang, M., Li, Y., Zhang, G. (2019). *Halophile, an essential platform for bioproduction*. Journal of Microbiological Methods 166, 105704. <https://doi.org/10.1016/j.mimet.2019.105704>
- Ma, H., Zhao, Y., Huang, W., Zhang, L., Wu, F., Ye, J., Chen, G.-Q. (2020). *Rational flux-tuning of Halomonas bluephagenesis for co-production of bioplastic PHB and ectoine*. Nat Commun 11, 3313. <https://doi.org/10.1038/s41467-020-17223-3>

REFERENCES (continued)

- Machado, D., Andrejev, S., Tramontano, M., Patil, K.R. (2018). *Fast automated reconstruction of genome-scale metabolic models for microbial species and communities*. Nucleic Acids Res, 46, 7542–7553. <https://doi.org/10.1093/nar/gky537>
- Mata, J.A., Béjar, V., Llamas, I., Arias, S., Bressollier, P., Tallon, R., Urdaci, M.C., Quesada, E. (2006). *Exopolysaccharides produced by the recently described halophilic bacteria Halomonas ventosae and Halomonas anticariensis*. Research in Microbiology 157, 827–835. <https://doi.org/10.1016/j.resmic.2006.06.004>
- Matsutani, M., Wakinaka, T., Watanabe, J., Tokuoka, M., Ohnishi, A. (2021). *Comparative Genomics of Closely Related Tetragenococcus halophilus Strains Elucidate the Diversity and Microevolution of CRISPR Elements*. Frontiers in Microbiology 12.
- McAdam, B., Brennan Fournet, M., McDonald, P., Mojicevic, M. (2020). *Production of Polyhydroxybutyrate (PHB) and Factors Impacting Its Chemical and Mechanical Characteristics*. Polymers (Basel) 12, 2908. <https://doi.org/10.3390/polym12122908>
- Mekanik, M., Fotovat, R., Motamedian, E., Jafarian, V. (2022). *Improvement of Lutein Production in Auxenochlorella protothecoides Using Its Genome-Scale Metabolic Model and a System-Oriented Approach*. Appl Biochem Biotechnol. <https://doi.org/10.1007/s12010-022-04186-y>
- Mohite, O.S., Weber, T., Kim, H.U., Lee, S.Y. (2019). *Genome-Scale Metabolic Reconstruction of Actinomycetes for Antibiotics Production*. Biotechnology Journal 14, 1800377. <https://doi.org/10.1002/biot.201800377>
- Moreno, M. de L., Sánchez-Porro, C., Piubeli, F., Frias, L., García, M.T., Mellado, E. (2011). *Cloning, Characterization and Analysis of cat and ben Genes from the Phenol Degrading Halophilic Bacterium Halomonas organivorans*. PLOS ONE 6, e21049. <https://doi.org/10.1371/journal.pone.0021049>
- Moretti, S., Tran, V.D.T., Mehl, F., Ibberson, M., Pagni, M. (2021). *MetaNetX/MNXref: unified namespace for metabolites and biochemical reactions in the context of metabolic models*. Nucleic Acids Research 49, D570–D574. <https://doi.org/10.1093/nar/gkaa992>
- Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A.C., Kanehisa, M. (2007). *KAAS: an automatic genome annotation and pathway reconstruction server*. Nucleic Acids Res 35, W182–W185. <https://doi.org/10.1093/nar/gkm321>
- Mormile, M.R., Edwards, T., Frank, R., Geurin, Z., Haendiges, J., Hoffmann, M., Miller, J. (2019). *Whole-Genome Analysis of Halomonas sp. Soap Lake #7 Reveals It Possesses Putative Mrp Antiporter Operon Groups 1 and 2*. Genome Biol Evol 11, 1706–1709. <https://doi.org/10.1093/gbe/evz117>

REFERENCES (continued)

- Mothes, G., Schubert, T., Harms, H., Maskow, T. (2008). *Biotechnological Coproduction of Compatible Solutes and Polyhydroxyalkanoates using the Genus Halomonas*. Engineering in Life Sciences 8, 658–662. <https://doi.org/10.1002/elsc.200800097>
- Mougiakos, I., Bosma, E.F., Vos, W.M. de, Kranenburg, R. van, Oost, J. van der. (2016). *Next Generation Prokaryotic Engineering: The CRISPR-Cas Toolkit*. Trends in Biotechnology 34, 575–587. <https://doi.org/10.1016/j.tibtech.2016.02.004>
- Nawaz, A., Chaudhary, R., Shah, Z., Dufossé, L., Fouillaud, M., Mukhtar, H., ul Haq, I. (2021). *An Overview on Industrial and Medical Applications of Bio-Pigments Synthesized by Marine Bacteria*. Microorganisms 9, 11. <https://doi.org/10.3390/microorganisms9010011>
- Neifar, M., Chouchane, H., Najjari, A., El Hidri, D., Mahjoubi, M., Ghedira, K., Naili, F., Soufi, L., Raddadi, N., Sghaier, H., Ouzari, H.I., Masmoudi, A.S., Cherif, A. (2019). *Genome analysis provides insights into crude oil degradation and biosurfactant production by extremely halotolerant Halomonas desertis G11 isolated from Chott El-Djerid salt-lake in Tunisian desert*. Genomics 111, 1802–1814. <https://doi.org/10.1016/j.ygeno.2018.12.003>
- Oberhardt, M.A., Chavali, A.K., Papin, J.A. (2009). *Flux Balance Analysis: Interrogating Genome-Scale Metabolic Networks*, In: Maly, I.V. (Ed.), Systems Biology, Methods in Molecular Biology. Humana Press, Totowa, NJ, pp. 61–80. https://doi.org/10.1007/978-1-59745-525-1_3
- O'Brien, E.J., Monk, J.M., Palsson, B.O. (2015). *Using Genome-scale Models to Predict Biological Capabilities*. Cell 161, 971–987. <https://doi.org/10.1016/j.cell.2015.05.019>
- Occhipinti, A., Eyassu, F., Rahman, T.J., Rahman, P.K.S.M., Angione, C. (2018). *In silico engineering of Pseudomonas metabolism reveals new biomarkers for increased biosurfactant production*. PeerJ 6. <https://doi.org/10.7717/peerj.6046>
- Occhipinti, A., Hamadi, Y., Kugler, H., Wintersteiger, C.M., Yordanov, B., Angione, C. (2021). *Discovering Essential Multiple Gene Effects Through Large Scale Optimization: An Application to Human Cancer Metabolism*. IEEE/ACM Transactions on Computational Biology and Bioinformatics 18, 2339–2352. <https://doi.org/10.1109/TCBB.2020.2973386>
- Ongagna-Yhombi, S.Y., McDonald, N.D., Boyd, E.F. (2015). *Deciphering the Role of Multiple Betaine-Carnitine-Choline Transporters in the Halophile Vibrio parahaemolyticus*. Applied and Environmental Microbiology 81, 351–363. <https://doi.org/10.1128/AEM.02402-14>

REFERENCES (continued)

- Oren, A. (2002). *Diversity of halophilic microorganisms: Environments, phylogeny, physiology, and applications*. J Ind Microbiol Biotech 28, 56–63. <https://doi.org/10.1038/sj/jim/7000176>
- Orth, J.D., Conrad, T.M., Na, J., Lerman, J.A., Nam, H., Feist, A.M., Palsson, B.Ø. (2011). *A comprehensive genome-scale reconstruction of Escherichia coli metabolism*. Mol Syst Biol 7, 535. <https://doi.org/10.1038/msb.2011.65>
- Orth, J.D., Thiele, I., Palsson, B.Ø. (2010). *What is flux balance analysis?* Nature Biotechnology 28, 245–248. <https://doi.org/10.1038/nbt.1614>
- Park, Y.-L., Choi, T.-R., Han, Y.-H., Song, H.-S., Park, J.-Y., Bhatia, S.K., Gurav, R., Choi, K.-Y., Kim, Y.-G., Yang, Y.-H. (2020). *Effects of osmolytes on salt resistance of Halomonas socia CKY01 and identification of osmolytes-related genes by genome sequencing*. Journal of Biotechnology 322, 21–28. <https://doi.org/10.1016/j.jbiotec.2020.07.006>
- Pastor, J.M., Bernal, V., Salvador, M., Argandoña, M., Vargas, C., Csonka, L., Sevilla, Á., Iborra, J.L., Nieto, J.J., Cánovas, M. (2013). *Role of Central Metabolism in the Osmoadaptation of the Halophilic Bacterium Chromohalobacter salexigens*. J Biol Chem 288, 17769–17781. <https://doi.org/10.1074/jbc.M113.470567>
- Peek, J., Roman, J., Moran, G.R., Christendat, D. (2017). *Structurally diverse dehydroshikimate dehydratase variants participate in microbial quinate catabolism*. Mol Microbiol 103, 39–54. <https://doi.org/10.1111/mmi.13542>
- Piubeli, F., Salvador, M., Argandona, M., Nieto, J.J., Bernal, V., Pastor, J.M., Canovas, M., Vargas, C. (2018). *Insights into metabolic osmoadaptation of the ectoines-producer bacterium Chromohalobacter salexigens through a high-quality genome scale metabolic model*. Microb. Cell. Fact. 17, 2. <https://doi.org/10.1186/s12934-017-0852-0>
- Poli, A., Moriello, V.S., Esposito, E., Lama, L., Gambacorta, A., Nicolaus, B. (2004). *Exopolysaccharide production by a new Halomonas strain CRSS isolated from saline lake Cape Russell in Antarctica growing on complex and defined media*. Biotechnol Lett 26, 1635–1638. <https://doi.org/10.1007/s10529-004-3187-y>
- Poli, A., Nicolaus, B., Denizci, A.A., Yavuzturk, B., Kazan, D. (2013). *Halomonas smyrnensis sp. nov., a moderately halophilic, exopolysaccharide-producing bacterium*. International Journal of Systematic and Evolutionary Microbiology 63, 10–18. <https://doi.org/10.1099/ijs.0.037036-0>

REFERENCES (continued)

- Qin, Q., Ling, C., Zhao, Y., Yang, T., Yin, J., Guo, Y., Chen, G.Q. (2018a). *CRISPR/Cas9 editing genome of extremophile Halomonas spp.* Metabolic Engineering 47, 219–229. <https://doi.org/10.1016/j.ymben.2018.03.018>
- Quillaguamán, J., Delgado, O., Mattiasson, B., Hatti-Kaul, R. (2006). *Poly(β -hydroxybutyrate) production by a moderate halophile, Halomonas boliviensis LC1.* Enzyme and Microbial Technology 38, 148–154. <https://doi.org/10.1016/j.enzmictec.2005.05.013>
- Radchenkova, N., Hasköylü, M.E., Vassilev, S., Yıldız, S.Y., Boyadzhieva, I., Oner, E.T., Kambourova, M. (2020). *Improved Exopolymer Production by Chromohalobacter canadensis Cultures for Its Potential Cosmeceutical Applications.* Microorganisms 8, 1935. <https://doi.org/10.3390/microorganisms8121935>
- Ravikumar, S., Uma, G., Gokulakrishnan, R. (2016). *Antibacterial Property of Halobacterial Carotenoids against Human Bacterial Pathogens.* Journal of Scientific and Industrial Research 75, 253–257.
- Rodríguez-Sáiz, M., Sánchez-Porro, C., De La Fuente, J.L., Mellado, E., Barredo, J.L. (2007). *Engineering the halophilic bacterium Halomonas elongata to produce β -carotene.* Appl Microbiol Biotechnol 77, 637–643. <https://doi.org/10.1007/s00253-007-1195-2>
- Römer, M., Eichner, J., Dräger, A., Wrzodek, C., Wrzodek, F., Zell, A. (2016). ZBIT Bioinformatics Toolbox: A Web-Platform for Systems Biology and Expression Data Analysis. PLoS ONE 11, e0149263. <https://doi.org/10.1371/journal.pone.0149263>
- RStudio Team, 2015. RStudio: Integrated Development Environment for R.
- Sakai, M., Miyauchi, K., Kato, N., Masai, E., Fukuda, M. (2003). *2-Hydroxypenta-2,4-dienoate metabolic pathway genes in a strong polychlorinated biphenyl degrader, Rhodococcus sp. strain RHA1.* Appl Environ Microbiol 69, 427–433. <https://doi.org/10.1128/AEM.69.1.427-433.2003>
- Sander, J.D., Joung, J.K. (2014). *CRISPR-Cas systems for editing, regulating and targeting genomes.* Nature Biotechnology 32, 347–355. <https://doi.org/10.1038/nbt.2842>
- Schellenberger, J., Park, J.O., Conrad, T.M., Palsson, B.Ø. (2010). *BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions.* BMC Bioinformatics 11, 213. <https://doi.org/10.1186/1471-2105-11-213>
- Schultenkämper, K., Brito, L.F., Wendisch, V.F. (2020). *Impact of CRISPR interference on strain development in biotechnology.* Biotechnology and Applied Biochemistry, 67, 7–21.

REFERENCES (continued)

- Schwibbert, K., Marin-Sanguino, A., Bagyan, I., Heidrich, G., Lentzen, G., Seitz, H., Rampp, M., Schuster, S.C., Klenk, H.-P., Pfeiffer, F., Oesterheld, D., Kunte, H.J. (2011). *A blueprint of ectoine metabolism from the genome of the industrial producer Halomonas elongata DSM 2581T*. Environmental Microbiology 13, 1973–1994. <https://doi.org/10.1111/j.1462-2920.2010.02336.x>
- Seaver, S.M.D., Liu, F., Zhang, Q., Jeffryes, J., Faria, J.P., Edirisinghe, J.N., Mundy, M., Chia, N., Noor, E., Beber, M.E., Best, A.A., DeJongh, M., Kimbrel, J.A., D’haeseleer, P., McCorkle, S.R., Bolton, J.R., Pearson, E., Canon, S., Wood-Charlson, E.M., Cottingham, R.W., Arkin, A.P., Henry, C.S. (2020). *The ModelSEED Biochemistry Database for the integration of metabolic annotations and the reconstruction, comparison and analysis of metabolic models for plants, fungi and microbes*. Nucleic Acids Res 49, D575–D588. <https://doi.org/10.1093/nar/gkaa746>
- Sirohi, R., Prakash Pandey, J., Kumar Gaur, V., Gnansounou, E., Sindhu, R. (2020). *Critical overview of biomass feedstocks as sustainable substrates for the production of polyhydroxybutyrate (PHB)*. Bioresource Technology 311, 123536. <https://doi.org/10.1016/j.biortech.2020.123536>
- Slatko, B.E., Gardner, A.F., Ausubel, F.M. (2018). *Overview of Next-Generation Sequencing Technologies*. Current Protocols in Molecular Biology 122. <https://doi.org/10.1002/cpmb.59>
- Sohn, S.B., Kim, T.Y., Park, J.M., Lee, S.Y. (2010). *In silico genome-scale metabolic analysis of Pseudomonas putida KT2440 for polyhydroxyalkanoate synthesis, degradation of aromatics and anaerobic survival*. Biotechnology Journal 5, 739–750. <https://doi.org/10.1002/biot.201000124>
- Stothard, P., Grant, J.R., Van Domselaar, G. (2019). *Visualizing and comparing circular genomes using the CGView family of tools*. Briefings in Bioinformatics 20, 1576–1582. <https://doi.org/10.1093/bib/bbx081>
- Sushmitha, T.J., Rajeev, M., Toleti, S.R., Pandian, S.K. (2021). *Complete genome sequence of Halomonas boliviensis strain knpp38, a chlorine-resistant bacterium isolated from the early-stage marine biofilm*. Marine Genomics 100890. <https://doi.org/10.1016/j.margen.2021.100890>
- Suthindhiran, K., Sarath Babu, V., Kannabiran, K., Ishaq Ahmed, V.P., Sahul Hameed, A.S. (2011). *Anti-fish nodaviral activity of furan-2-yl acetate extracted from marine Streptomyces spp.* Natural Product Research 25, 834–843. <https://doi.org/10.1080/14786419.2010.530599>

REFERENCES (continued)

- Sztukowska, M., Bugno, M., Potempa, J., Travis, J., Kurtz, D.M. (2002). *Role of rubrerythrin in the oxidative stress response of Porphyromonas gingivalis*. Mol Microbiol 44, 479–488. <https://doi.org/10.1046/j.1365-2958.2002.02892.x>
- Tan, D., Xue, Y.-S., Aibaidula, G., Chen, G.-Q. (2011). *Unsterile and continuous production of polyhydroxybutyrate by Halomonas TD01*. Bioresource Technology 102, 8130–8136. <https://doi.org/10.1016/j.biortech.2011.05.068>
- Tanimura, K., Nakayama, H., Tanaka, T., Kondo, A. (2013). *Ectoine production from lignocellulosic biomass-derived sugars by engineered Halomonas elongata*. Bioresource Technology 142, 523–529. <https://doi.org/10.1016/j.biortech.2013.05.004>
- Tao, W., Lv, L., Chen, G.-Q. (2017). *Engineering Halomonas species TD01 for enhanced polyhydroxyalkanoates synthesis via CRISPRi*. Microbial Cell Factories 16, 48. <https://doi.org/10.1186/s12934-017-0655-3>
- Tatar, D. (2021). *Isolation, phylogenetic analysis and antimicrobial activity of halophilic actinomycetes from different saline environments located near Çorum province*. Biologia 76, 773–780. <https://doi.org/10.2478/s11756-020-00612-w>
- Terzer, M., Maynard, N.D., Covert, M.W., Stelling, J. (2009). *Genome-scale metabolic networks*. WIREs Systems Biology and Medicine 1, 285–297. <https://doi.org/10.1002/wsbm.37>
- Thiele, I., Palsson, B.Ø. (2010). *A protocol for generating a high-quality genome-scale metabolic reconstruction*. Nat Protoc 5, 93–121. <https://doi.org/10.1038/nprot.2009.203>
- Thomas, T., Elain, A., Bazire, A., Bruzard, S. (2019). *Complete genome sequence of the halophilic PHA-producing bacterium Halomonas sp. SF2003: insights into its biotechnological potential*. World J Microbiol Biotechnol 35, 50. <https://doi.org/10.1007/s11274-019-2627-8>
- Tohme, S., Hacısmanoğlu, G.G., Eroğlu, M.S., Kasavi, C., Genç, S., Can, Z.S., Toksoy Oner, E. (2018). *Halomonas smyrnensis as a cell factory for co-production of PHB and levan*. International Journal of Biological Macromolecules 118, 1238–1246. <https://doi.org/10.1016/j.ijbiomac.2018.06.197>
- Tsallagov, S.I., Sorokin, D.Y., Tikhonova, T.V., Popov, V.O., Muyzer, G. (2019). *Comparative Genomics of Thiohalobacter thiocyanaticus HRh1T and Guyparkeria sp. SCN-R1, Halophilic Chemolithoautotrophic Sulfur-Oxidizing Gammaproteobacteria Capable of Using Thiocyanate as Energy Source*. Frontiers in Microbiology 10.

REFERENCES (continued)

- Ungerer, J., Pakrasi, H.B. (2016). *Cpf1 Is A Versatile Tool for CRISPR Genome Editing Across Diverse Species of Cyanobacteria*. Sci Rep 6, 39681. <https://doi.org/10.1038/srep39681>
- Vandrich, J. (2019). *Metabolic engineering in Halomonas elongata*, PhD dissertation, Universität Potsdam.
- Vandrich, J., Pfeiffer, F., Alfaro-Espinoza, G., Kunte, H.J. (2020). *Contribution of mechanosensitive channels to osmoadaptation and ectoine excretion in Halomonas elongata*. Extremophiles 24, 421–432. <https://doi.org/10.1007/s00792-020-01168-y>
- Vauclare, P., Natali, F., Kleman, J.P., Zaccai, G., Franzetti, B. (2020). *Surviving salt fluctuations: stress and recovery in Halobacterium salinarum, an extreme halophilic Archaeon*. Sci Rep 10, 3298. <https://doi.org/10.1038/s41598-020-59681-1>
- Vijayakumar, R., Panneer Selvam, K., Muthukumar, C., Thajuddin, N., Panneerselvam, A., Saravanamuthu, R. (2012). *Antimicrobial potentiality of a halophilic strain of Streptomyces sp. VPTSA18 isolated from the saltpan environment of Vedaranyam, India*. Ann Microbiol 62, 1039–1047. <https://doi.org/10.1007/s13213-011-0345-z>
- Vijayakumar, S., Angione, C. (2021). *Protocol for hybrid flux balance, statistical, and machine learning analysis of multi-omic data from the cyanobacterium Synechococcus sp. PCC 7002*. STAR protocols 2, 100837.
- Vreeland, R.H., Anderson, R., Murray, R.G. (1984). *Cell wall and phospholipid composition and their contribution to the salt tolerance of Halomonas elongata*. Journal of Bacteriology 160, 879–883. <https://doi.org/10.1128/jb.160.3.879-883.1984>
- Vreeland, R.H., Litchfield, C.D., Martin, E.L., Elliot, E.Y. (1980). *Halomonas elongata, a New Genus and Species of Extremely Salt-Tolerant Bacteria*. International Journal of Systematic and Evolutionary Microbiology 30, 485–495. <https://doi.org/10.1099/00207713-30-2-485>
- Wang, T., Li, Y., Li, J., Zhang, D., Cai, N., Zhao, G., Ma, H., Shang, C., Ma, Q., Xu, Q., Chen, N. (2019). *An update of the suicide plasmid-mediated genome editing system in Corynebacterium glutamicum*. Microbial Biotechnology 12, 907–919. <https://doi.org/10.1111/1751-7915.13444>
- Wattam, A.R., Brettin, T., Davis, J.J., Gerdes, S., Kenyon, R., Machi, D., Mao, C., Olson, R., Overbeek, R., Pusch, G.D., Shukla, M.P., Stevens, R., Vonstein, V., Warren, A., Xia, F., Yoo, H. (2018). *Assembly, Annotation, and Comparative Genomics in PATRIC, the All Bacterial Bioinformatics Resource Center*. Methods Mol Biol 1704, 79–101. https://doi.org/10.1007/978-1-4939-7463-4_4

REFERENCES (continued)

- Wattam, A.R., Davis, J.J., Assaf, R., Boisvert, S., Brettin, T., Bun, C., Conrad, N., Dietrich, E.M., Disz, T., Gabbard, J.L., Gerdes, S., Henry, C.S., Kenyon, R.W., Machi, D., Mao, C., Nordberg, E.K., Olsen, G.J., Murphy-Olson, D.E., Olson, R., Overbeek, R., Parrello, B., Pusch, G.D., Shukla, M., Vonstein, V., Warren, A., Xia, F., Yoo, H., Stevens, R.L. (2017). *Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center*. Nucleic Acids Research 45, D535–D542. <https://doi.org/10.1093/nar/gkw1017>
- Wells, T., Ragauskas, A.J. (2012). *Biotechnological opportunities with the β -ketoadipate pathway*. Trends in Biotechnology 30, 627–637. <https://doi.org/10.1016/j.tibtech.2012.09.008>
- Wickham, H. (2009). Introduction, in: Wickham, H. (Ed.), *Ggplot2: Elegant Graphics for Data Analysis*, Use R. Springer, New York, NY, pp. 1–7. https://doi.org/10.1007/978-0-387-98141-3_1
- Wishart, D.S., Tzur, D., Knox, C., Eisner, R., Guo, A.C., Young, N., Cheng, D., Jewell, K., Arndt, D., Sawhney, S., Fung, C., Nikolai, L., Lewis, M., Coutouly, M.-A., Forsythe, I., Tang, P., Shrivastava, S., Jeroncic, K., Stothard, P., Amegbey, G., Block, D., Hau, D.D., Wagner, J., Miniaci, J., Clements, M., Gebremedhin, M., Guo, N., Zhang, Y., Duggan, G.E., Macinnis, G.D., Weljie, A.M., Dowlatabadi, R., Bamforth, F., Clive, D., Greiner, R., Li, L., Marrie, T., Sykes, B.D., Vogel, H.J., Querengesser, L. (2007). *HMDB: The Human Metabolome Database*. Nucleic Acids Res 35, D521–526. <https://doi.org/10.1093/nar/gkl923>
- Xavier, J.C., Patil, K.R., Rocha, I. (2017). *Integration of Biomass Formulations of Genome-Scale Metabolic Models with Experimental Data Reveals Universally Essential Cofactors in Prokaryotes*. Metab Eng 39, 200–208. <https://doi.org/10.1016/j.ymben.2016.12.002>
- Xu, L., Dong, Z., Fang, L., Luo, Y., Wei, Z., Guo, H., Zhang, G., Gu, Y.Q., Coleman-Derr, D., Xia, Q., Wang, Y. (2019). *OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species*. Nucleic Acids Res 47, W52–W58. <https://doi.org/10.1093/nar/gkz333>
- Yao, R., Liu, D., Jia, X., Zheng, Y., Liu, W., Xiao, Y. (2018). *CRISPR-Cas9/Cas12a biotechnology and application in bacteria*. Synthetic and Systems Biotechnology 3, 135–149. <https://doi.org/10.1016/j.synbio.2018.09.004>
- Yaşar Yildiz, S., Nikerel, E., Toksoy Öner, E. (2019). *Genome-Scale Metabolic Model of a Microbial Cell Factory (Brevibacillus thermoruber 423) with Multi-Industry Potentials for Exopolysaccharide Production*. OMICS: A Journal of Integrative Biology 23, 237–246. <https://doi.org/10.1089/omi.2019.0028>

REFERENCES (continued)

- Ye, J., Hu, D., Che, X., Jiang, X., Li, T., Chen, J., Zhang, H.M., Chen, G.-Q. (2018). *Engineering of Halomonas bluephagenesis for low cost production of poly(3-hydroxybutyrate-co-4-hydroxybutyrate) from glucose*. Metab. Eng. 47, 143–152. <https://doi.org/10.1016/j.ymben.2018.03.013>
- Yin, J., Fu, X.-Z., Wu, Q., Chen, J.-C., Chen, G.-Q. (2014). *Development of an enhanced chromosomal expression system based on porin synthesis operon for halophile Halomonas sp.* Appl Microbiol Biotechnol 98, 8987–8997. <https://doi.org/10.1007/s00253-014-5959-1>
- Yu, J., Wang, Z., Wang, J., Mohisin, A., Liu, H., Zhang, Y., Zhuang, Y., Guo, M. (2022). *Physiological metabolic topology analysis of Halomonas elongata DSM 2581T in response to sodium chloride stress*. Biotechnology and Bioengineering 119, 3509–3525. <https://doi.org/10.1002/bit.28222>
- Yu, L.-P., Wu, F.-Q., Chen, G.-Q. (2019). *Next-Generation Industrial Biotechnology-Transforming the Current Industrial Biotechnology into Competitive Processes*. Biotechnology Journal 14, 1800437. <https://doi.org/10.1002/biot.201800437>
- Zhai, L., Xie, J., Lin, Y., Cheng, K., Wang, L., Yue, F., Guo, J., Liu, J., Yao, S. (2018). *Genome sequencing and heterologous expression of antiporters reveal alkaline response mechanisms of Halomonas alkalicola*. Extremophiles 22, 221–231. <https://doi.org/10.1007/s00792-017-0991-6>
- Zhang, J., Petersen, S.D., Radivojevic, T., Ramirez, A., Pérez-Manríquez, A., Abeliuk, E., Sánchez, B.J., Costello, Z., Chen, Y., Fero, M.J., Martin, H.G., Nielsen, J., Keasling, J.D., Jensen, M.K. (2020). *Combining mechanistic and machine learning models for predictive engineering and optimization of tryptophan metabolism*. Nat Commun 11, 4880. <https://doi.org/10.1038/s41467-020-17910-1>
- Zhang, K., Zhao, W., Rodionov, D.A., Rubinstein, G.M., Nguyen, D.N., Tanwee, T.N.N., Crosby, J., Bing, R.G., Kelly, R.M., Adams, M.W.W., Zhang, Y. (2021). *Genome-Scale Metabolic Model of Caldicellulosiruptor bescii Reveals Optimal Metabolic Engineering Strategies for Bio-based Chemical Production*. mSystems 6, e01351-20. <https://doi.org/10.1128/mSystems.01351-20>

APPENDIX

The data and codes that were used to generate the results in this study have been provided in a github repository accessible through this link below.

<https://github.com/EnuhBlaise/Halomonas-elongata-Genome-Scale-Model/tree/master>

