****

**School of Medicine**

**Department of Health Sciences**

**Master’s degree in medical biotechnology**

**Name**: Tamegye Kouchou Boris

**Matricula**: 20053883

**Course**: MS1872 – Synthetic Biology and Proteomics: Synthetic Biology Project Proposal

**Project title**

Investigating The Dynamics of Gene Expression in Corynebacterium glutamicum Carotenoid Biosynthesis to Optimize and Scale Industrial Production.

**Project Acronym**

**"TRADRUEC"** (TRAnscriptome – Data - Re-Use – of Engineered Corynebacterium glutamicum).

**Professor**: Diego Cotella

**Academic** **year**: 2023/2024

Date: 24/01/2024

**Table of Contents**

1. Abstract2
2. Scientific background 3
3. Aims and rationale of the project 6
4. Experimental plan7
5. Methods and actions planned7
6. Tentative timeline – Ghant Chart11
7. Potential pitfalls and caveats, difficulties, and limitations 11
8. Alternative approaches to achieve the objectives 12
9. Project originality and significance 12
10. Impact on the community, the society, and ethical issues13
11. Bibliography15

Figures: 1- 10

Tables: 1 - 4

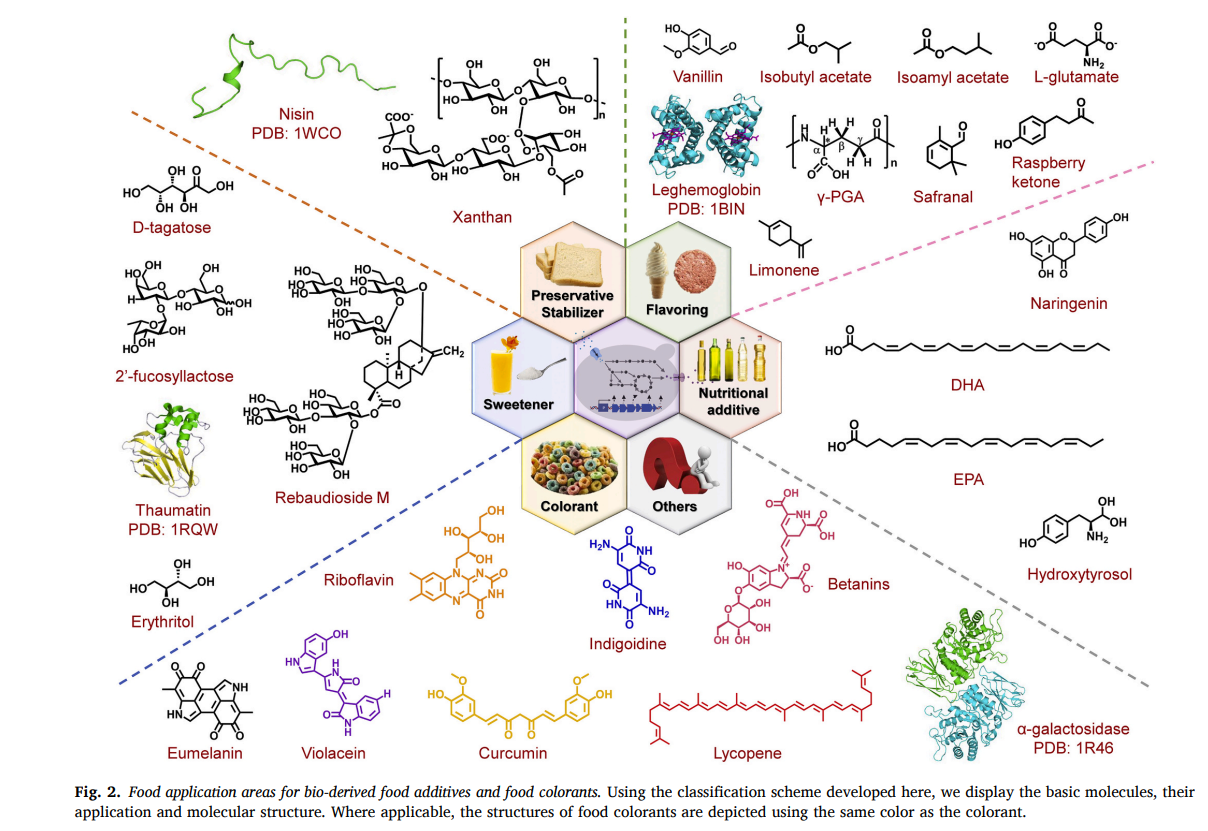
1. **Abstract**

Food additives and colorants are extensively used in the food industry to improve food quality and safety during processing, storage and packing. These food colorants are extracted from natural sources and chemical synthesis. However, growing demands for sustainability, safety and “natural” products have renewed interest in using bio-based production methods. Moreover, the waste product after extraction of food colorants when disposed leaves unfriendly environmental footprints. In addition, food colorants are likely to be detrimental to our health as some have been classified as potential risk factors for health conditions such as cancers, allergies, and asthma.

This project aims to provide a synthetic biology approach to food colorants production at an industrial scale by using RNAseq data of engineered corynebacterium glutamicum. It set to mine SRA RNA-Seq data of C. glutamicum sampled at different production phases, mine SRA RNA-Seq Datasets obtained from independent production conditions to serve as comparative benchmarks for gene expression changes. Unlike the traditional method of using the engineered bacterial strain which is often associated with many setbacks such as: off targeting during gene editing, random mutagenesis, limited understanding of operon structures regulating gene expression in C. glutamicum. Nevertheless, RNA seq data has revealed that 2/3 of genes transcribed in C. glutamicum are transcribed as operons, and some of these operons can be divided into sub-operons. Through RNAseq analysis, it is possible to spot the genes that are highly expressed in the carotenoid biosynthesis pathway. The identified genes could then be cloned into a vector to yield high titer during industrial production of biobased carotenoid food colorants.

1. **Scientific background**

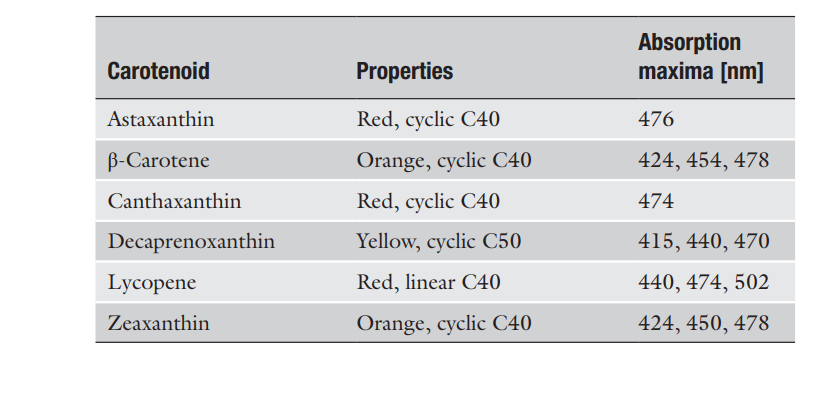
The soil microbe Corynebacterium glutamicumis a gram-positive, non-spore-forming facultative anaerobic bacterium, belonging to the phylum of actinobacteria. It is a workhorse of industrial amino acid production employed for the production of amino acids. Remarkable efforts in metabolic engineering have widened the product portfolio of C. glutamicum to over 70 different compounds. The microbe grows quickly to high cell densities, shows no autolysis, and can be easily propagated to a large scale (≥750 cubic meters). C. glutamicum produces no endotoxins, does not undergo phage lysis, and is generally recognised as safe (GRAS), allowing the synthesis of a range of commercial products. Moreover, metabolically engineered strains of C. glutamicumproduce more than 30 high value active ingredients, including signature molecules of raspberry, savoury, and orange flavours, sun blockers, anti-ageing sugars, and polymers for regenerative medicine.

**

*Figure 1: Food application for bio-derived food additives and colorants.*

This bacterium is naturally pigmented due to the biosynthesis of the C50 decaprenoxanthin (Henke *et al.*, 2018). Carotenoids are widely used in the industry as food colorants and additives. Food additives comprise a wide range of functional categories such as antioxidants, thickeners, stabilizers and emulsifiers, flavourings, sweeteners, nutritional additives, and colorants. They are broadly defined as substances added to food or food ingredients to preserve flavour, enhance taste, and/or alter appearance. Likewise, the move to more cultured foods and meat alternatives requires the production of new additives and colorants. Traditionally, these molecules are mainly produced through chemical synthesis or extraction from natural sources. In contrast, microbial production of food additives can provide advantages over chemical synthesis and natural extraction (Sun, Xin and Alper, 2021).

Table 1. different types of carotenoids and their respective properties

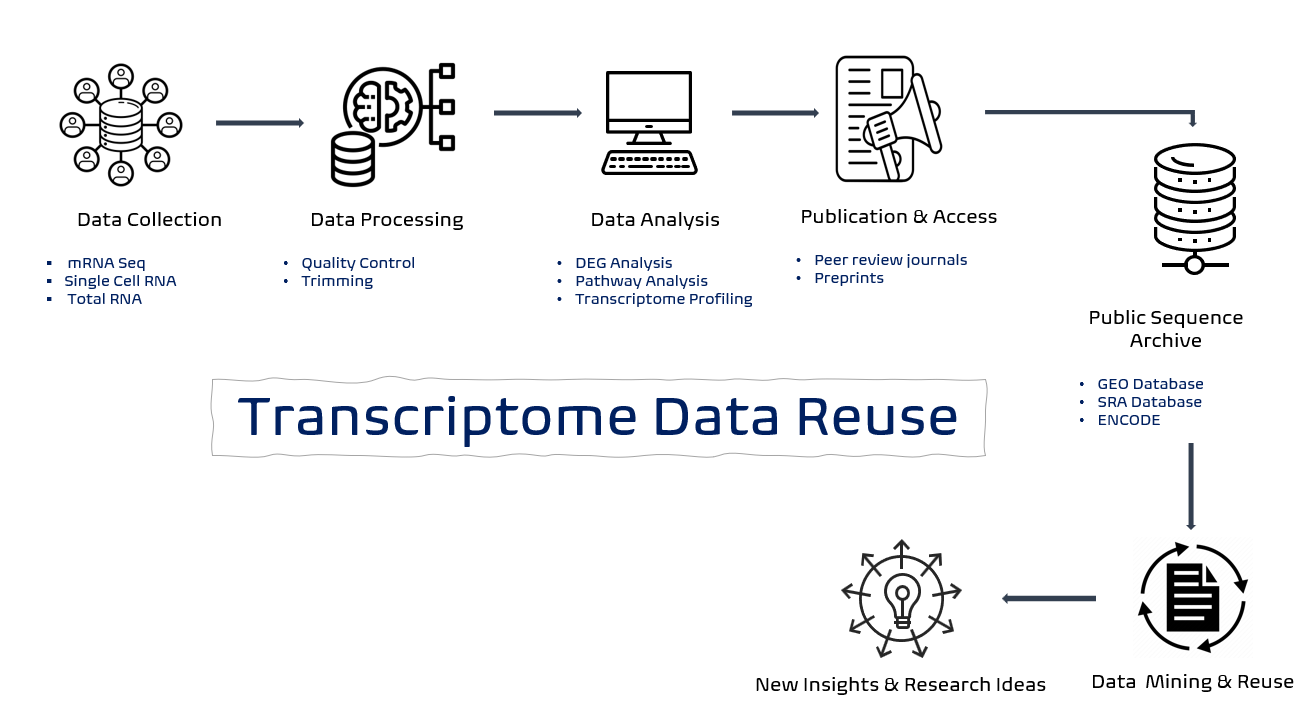


*Nadja A. Henke et al.*

This project aims to provide a bioinformatic approach to scaling the production of carotenoids industrially. Our objective is to analyse Corynebacterium glutamicum carotenoid biosynthesis data sampled at different stages of production and stored in public archives. We hope to use RNAseq analysis to identify differentially expressed genes involved in the pathway of carotenoid biosynthesis, i.e. the genes that are upregulated and those downregulated during carotenogenesis in C. glutamicum. Then, the use of synthetic biology techniques to clone the gene of interest (differentially expressed) into a suitable vector to scale the production of carotenoids industrially. We will consider genes with fold change >2 as differentially expressed.

To understand and decipher the transcriptome complexity of an organism and the underlying functionalities have become a major focus for post-genome research in recent years. This novel sequencing approach has been successfully applied for studying whole-genome transcription for various prokaryotes and eukaryotes and revealed an unexpected complexity of these transcriptomes, e.g. widespread antisense transcription, and an enormous amount of small and novel RNAs in bacterial genomes.

Transcriptome data reuse is an approach that is not quite common in the field of computational and molecular biology, even though the use of SRA RNAseq data stored in public archives has proven to be of significant value as it may be used to address many research problems. RNAseq analyses of this microbe have demonstrated that some transcripts are leaderless, imposing a different translation mechanism. The classical operon has multiple genes which are transcribed from a single promoter. Moreover, recent RNAseq data showed that various operons are divided into sub-operons due to internal transcription start sites which often respond to different conditions.

**

*Figure 2. shows the flow-chart of how data from public sequence archive can be used for research purposes.*

The SRA database contains increasing number of data sets on carotenoids biosynthesis, and this data could be used to address many research questions. Regarding the current state of the art in carotenoids synthesis:

* The carotenogenic genes are clustered in the genome (Kalinowski et al. 2003). The seven genes (crtE, cg0722, crtB, crtI, crtYe, crtYf, crtEb) (Krubasik et al. 2001a) that encode the enzymes needed for the conversion of the precursor molecules IPP and DMAPP to decaprenoxanthin are organized in an operon. RT-PCR revealed that the entire gene cluster represents a transcription unit (Heider et al. 2012), including one gene of unknown function (cg0722).
* C. glutamicum has been engineered for the production of C40 carotenoids (lycopene, β-carotene, astaxanthin) and for valencene (Frohwitter et al. 2014; Heider et al. 2012, 2014b, 2014d; Henke et al. 2016).
* Overproduction of its native pigment decaprenoxanthin and two non-native C50 carotenoids (sarcinaxanthin and C.p. 450) has been enabled by engineering of the terminal carotenoid pathway (Heider et al. 2014b) (Table 5.1).
* Titers for the C50 carotenoids of up to 10 m/g cell dry weight (CDW) have been reported using C. glutamicum as the production host (Heider et al. 2014a).
* The common production host and model organism E. coli is recognized as a suitable host for heterologous production of a wide range of both natural and novel C30 and C40 carotenoids (Wang et al. 2007).

This proposal represents a significant step forward in comparison to the current state of the art as it intends to make use of available RNAseq data sets of C. glutamicum wild type and engineered strains to decipher the protein-protein interactions, regulatory networks, metabolic pathways of differentially expressed genes involved in carotenogenesis. So far, previous studies focused on engineering highly productive strains and no studies have been conducted to critically evaluate the upregulated genes and transcriptional regulatory networks involved in carotenoid biosynthesis using RNAseq data sets. This study also aims to investigate the various culture conditions and how they affect the overall titre during the production of carotenoids.

1. **Aims and rationale of the project.**

**Aim**

* To investigate the dynamics of gene expression in biosynthesis of carotenoid by C. glutamicum through transcriptome data analysis.
* Test the identified top ten differentially expressed genes (DEGs) involved in carotenoid biosynthesis using qPCR.
* Evaluate the role of culture conditions in the dynamics of gene expression.

**Rationale**

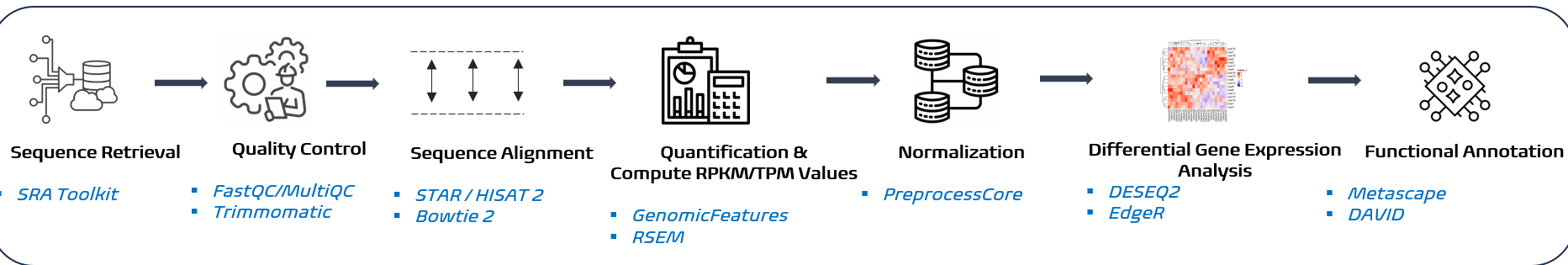
Transcriptome analysis is a robust and cost-efficient method which provides information about the internal biological processes, cellular biosynthesis, and metabolic functions of a cell.Carotenoids are naturally produced by endogenous genes in C. glutamicum, so, scientists have engineered strains of this microbe to induce overexpression of these genes to yield a high titre. However, it is insufficient to supply the market demands. This project seeks to address this defectt and provide promising solutions through analysis of already available RNAseq files. An increasing number of studies have found that computational approaches to investigating research questions could be cost-effective. Besides, the data needed to obtain preliminary results can be queried in public databases. Thus, we seek to evaluate the potential of SRA RNAseq data of C. glutamicum in bringing an innovative solution to the limitations faced by industries to produce carotenoids.

This approach sheds new light on an economically friendly methodology of research that could benefit the scientific community. Technical benefits could be achieved by critically evaluating the technical aspects of each RNAseq file such as culture conditions, strain type, carbon source etc.

**Hypothesis**

To investigate the dynamics of gene expression in C. glutamicum, we look forward to identifying the top ten genes upregulated in the carotenoid biosynthesis pathway, amplify these genes, clone them in a suitable vector. To determine its efficacy, we will compare samples (from RNAseq sequence files) grown under industrial conditions (controls) with the test samples (bacterium with inserts of DEGs) grown in the laboratory. The test sample will have a higher yield (fold change>2) with respect to the control.

1. **Experimental plan**
2. **Methods and actions planned.**

**

*Figure 3. illustrates the workflow of RNAseq analysis.*

* To begin with, we query C. glutamicum and its wild type (WT) in the SRA run selector tool of the NCBI database and filter organism then check RNA.

Table 2: results of search obtain from the SRA run selector tool.

|  |  |
| --- | --- |
| Organism | Entries |
| Corynebacterium glutamicum | 353 |
| Corynebacterium glutamicum ATCC 13032 (WT) | 157 |
| Total | 510 |

* The retrieve sequences are screened using *Trimmomatic* software for poor-quality reads.

Table 3: classification of files according to type and protocol used.

|  |  |  |  |
| --- | --- | --- | --- |
| Paired end data sets | Single end data sets | ILLUMINA | BGISEQ |
| 433 | 77 | 498 | 12 |

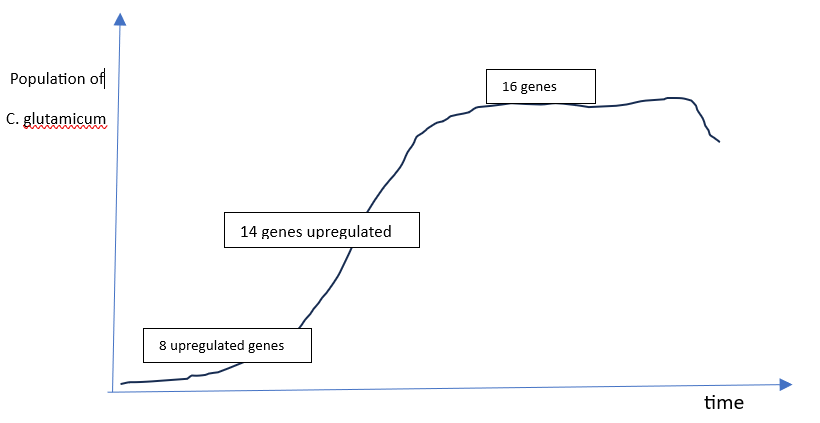
We may have a total of 350 samples of sequences that meet our inclusion criteria after screening.

* Sequence alignment and quantification:

We intend to use *HISAT2* for sequence alignment and *RSEM* to quantify the samples of interest**.**

* DEGs analysis and functional annotations:

In this step, we first normalize our data before proceeding to identify the top ten differentially expressed genes upregulated with the DESEG2 software. Then, we conduct a functional analysis to determine the selected genes belonging to the carotenoid biosynthesis pathway.

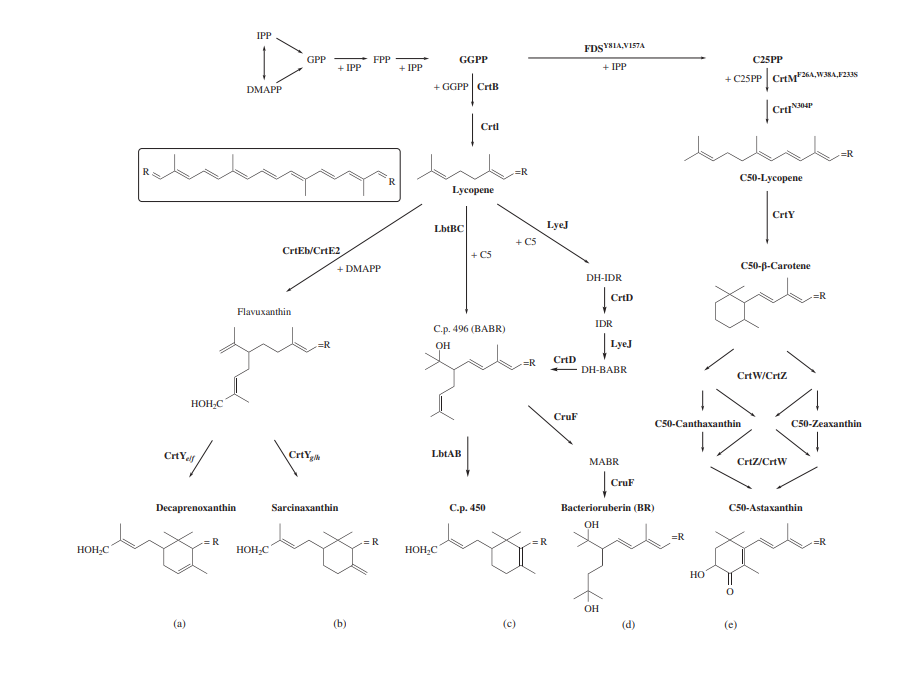
 *Figure 4.* *DEGs involved in the carotenoid biosynthesis pathway according to the different growth stages of C. glutamicum.*

RNAseq analysis enables the identification of DEGs at the different growth stages of the microbe. From the sample of thirty-eight upregulated genes, a list of the top 10 DEGs could be CrtZ, CrtW, CrtX, CrtF, CrtY, CrtK, CrtH, CrtQ, CrtP, and CrtV belonging to the pathway of interest (verify with metascape). We pick two genes from phase 1, seven genes from phase 2 and one gene from phase 3 with a fold change > 2 for cloning.

*Figure 5. vector construction and verification.*

Plasmids harbouring a carotenogenic gene (general abbreviation crt), the vector pECXT99A allows inducible overexpression of crt. Alternatively, the vector pSH1 expresses crt constitutively under the tuf -promoter (Ptuf). The expression vector is constructed based on pECXT99A with a carriage capacity optimal to incorporate our gene of interest. Our amplified products are cloned into the BamHI restricted pECXT-crt plasmid DNA by Gibson assembly. The resultant vector is trans-conjugated with C. glutamicum and cultured under standard laboratory conditions. Extractions could be envisaged after 36h with all lab protocols respected.

Carotenoid biosynthesis is regulated by the MarR-like regulator CrtR. CrtR binds the promoter of the crt operon to repress its transcription. This regulation indicates a feedforward regulatory mechanism of carotenogenesis: first, isoprenoid pyrophosphates are synthesized that act as effector molecules on the repressor CrtR and therefore induce transcription of the terminal carotenoid biosynthetic genes. Even though some studies have shown that deleting the repressor gene CrtR leads to a three-fold increase in titer of certain carotenoids.

**

*Figure 6. highlights the biochemical pathways and genes involved in the biosynthesis of carotenoids. With a-b-c-d-e representing the different carotenoids.*

The table below illustrates a sample comparison of titres in g/l of carotenoids per round of extraction per growth stage obtain under industrial conditions from bioreactors (control) with those from cloning of DEGs in the lab (test).

Table 4.

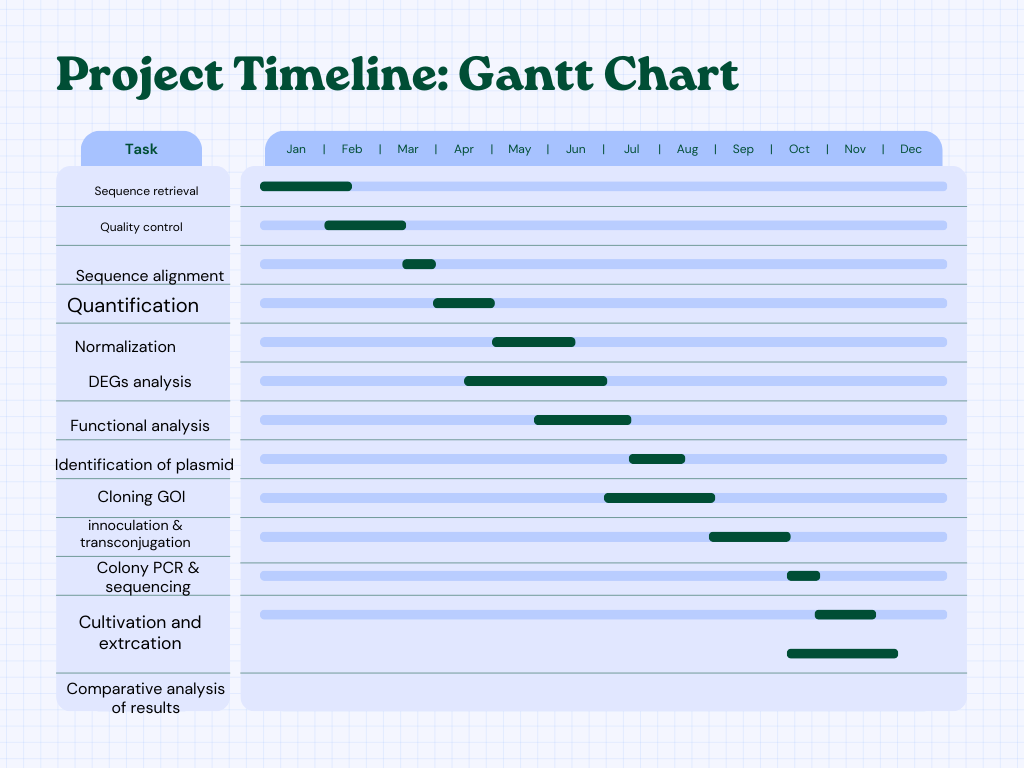
|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Lab | 200 | 240 | 280 | 365 | 405 | 479 | 548 | 613 | 835 | 918 |
| Bioreactor | 210 | 225 | 240 | 274 | 293 | 319 | 356 | 391 | 416 | 436 |

**FC >2.1**

*Figure 7. shows a comparison between cloned DEGs cultured in the lab and results from industry*.

The above results suggest that DEGs identified through RNAseq analysis when cloned in an expression vector, inserted into C. glutamicum and grown in the lab (test), yields twice more than the engineered strains of the microbe cultured industrially (control) in bioreactors. In addition, a fold change >2 in the test sample compared to the control per round of extraction indicates that we might now be able to produce two times the current maximum volumetric productivity of carotenoids. This supports our hypothesis of using computational approaches to optimizing carotenoid biosynthesis. Moreso, the ideal growth conditions in the lab could be a contributing factor to the high yield implying that though the growth of C. glutamicum is less affected by culture conditions, but dynamics in these conditions may disrupt operon structures which deregulates the expression of carotenogenic genes.

Furthermore, adapting our test sample to an industrial setting will compete economically with chemical synthesis, overcoming the challenge of supplying the growing demands of biobased food colorants and additives in the market.

**

**b.**

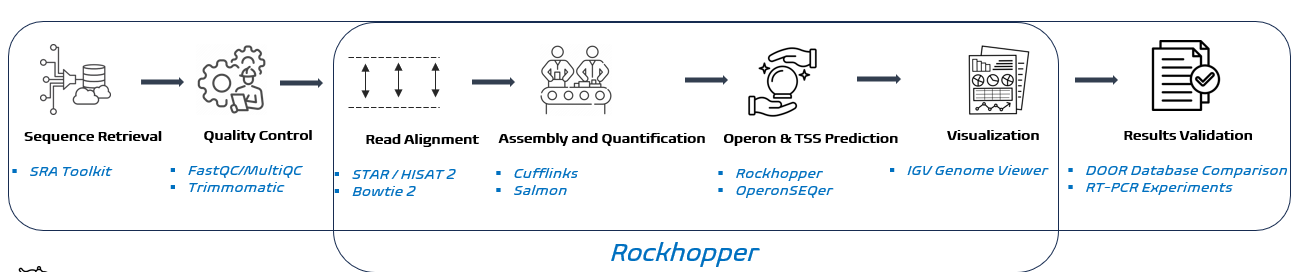
*Figure 8. tentative timeline to realize the project (minimum of 1 year).*

1. **Potential pitfalls and caveats**

* Data availability: sufficient sequences might not be available yet in public archives.
* Functional annotation is critical for obtaining a reliable result.
* Data Quality: We might not be fortunate to retrieve enough sequences with good quality reads plenty enough to support the sample size needed for the study.
* Analysis of gene expression is a very sensitive phenomenon since it may vary depending on the platform and library protocol.
* Although C. glutamicum is resistant to toxic compounds and stress, studies suggest that there are considerable variations in the response of its operons with changing conditions. This dynamic affects gene expression and the likelihood of identifying differentially expressed genes.
* A vector with a large carrying capacity is required to clone our gene of interest which may incur cost.
* Constitutive overexpression of our clone gene of interest could lead to random mutations and short live the bacteria.
* Genes upregulated in the early phase could be downregulated in the final phase.

1. **Alternative methods to this proposal.**

* Another approach to achieving similar results is through resolving operon structures with the use of Rockhopper. RNA seq data has shown that 2/3 of genes transcribed in Corynebacterium glutamicum are transcribed as operons, with certain operons divided into sub-operons (Pfeifer-Sancar et al., 2013). Therefore, we could use RNA-Seq datasets to determine and fully define C. glutamicum operon structures, interchanging native and synthetic promoters might yield similar results.

**

*Figure 9. represents a flow diagram of operon structure analysis.*

* Metabolic engineering of activator enzymes and enzymes with negative effects on inhibitors of our pathway of interest.
* Omics based approach to identify the regulatory networks and protein-protein interactions involved in the biosynthesis of carotenoid.

1. **Project originality and significance**

In this report, we outlined the use of RNAseq data to scale the production titre of carotenoids. This project is unique to the best of my knowledge because similar publications that uses genomic data for research purposes were different from carotenoid. We may highlight:

* Enabling inverse metabolic engineering through genomics (Gill, 2003)
* Understanding the high l-valine production in Corynebacterium glutamicum VWB-1 using transcriptomics and proteomics (Zhang *et al.*, 2018)
* Transcriptome Analysis and Genetic Engineering (Qaisar *et al.*, 2017)
* Comprehensive analysis of the Corynebacterium glutamicum transcriptome using an improved RNAseq technique (Pfeifer-Sancar *et al.*, 2013).

Conversely, previous studies on amino acids, carotenoids and other high value ingredients used as colorants and food additives considered mainly genetic engineering as means to boost production. Already published data include Carotenoid Production by Recombinant Corynebacterium glutamicum: Strain Construction, Cultivation, Extraction, and Quantification of Carotenoids and Terpenes (Henke et al., 2018); (Bio-synthesis of food additives and colorants-a growing trend in future food - ScienceDirect, no date)**;** and Functional food additives/ingredients production by engineered Corynebacterium glutamicum (Cankar, Henke and Wendisch, 2023).

The deliverables from this project could be adopted for biobased biosynthesis of indigo in E. coli. The waste from the chemical synthesis of indigo constitutes chemical pollution to the environment. Also, this method has a promising application in the production of biofuels, limiting carbon emissions from the combustion of fossils.

Summarily, this literature presents a method of scaling production of carotenoids by spotting DEGs from RNAseq data sets, applying synthetic biology techniques to clone the spotted genes into a plasmid, then insert the vector into C. glutamicum for culture and extraction. Till now, this is the first comprehensive proposal which combines genetic engineering and transcriptome analysis for high yield production of carotenoids, supporting the originality of this project.

1. **Impact on the scientific community, the society, and ethical issues.**

The bacterium C. glutamicum, one of the workhorses of industrial biotechnology, has been used for decades for the safe production of food and feed additives in particular amino acids (Becker and Wittmann 2012; Eggeling and Bott 2005; Heider and Wendisch 2015; Mitsuhashi, 2014; Zahoor et al. 2014). The total global commercial value of this microbe was estimated to be $15 billion in 2022, while the market value of carotenoids was reported to be $1.8 billion in 2019 (Marz 2015). Commercially available carotenoids are primarily used as food and feed colorants, with more than 100 tons of astaxanthin used annually in the aquaculture and poultry industries (Gassel et al. 2013; Johnson and Schroeder 1996), in coloration of beverages, and in the cosmetics industry (Dembitsky 2005; Downham and Collins 2000). Synthetic colorants have predominated over natural colorants mainly due to their relatively more stable colour, larger synthesis scale, and lower manufacturing costs. However, the environmental and health impacts arising from the excess use of these chemicals have brought up increasing demand for natural colorants.

*A diagram of food additives and colorants

Description automatically generated*

*Figure 10. shows a summary of the use of food colorants and additives, from production to consumption.*

This project may stimulate the development of economically feasible production titres of carotenoids as food colorants and additives. Since bio-based additives are in high demand, exploiting this project will boost the market value of carotenoids from the current worth, enabling financial benefits to the society. Some merits of this project include:

1. Addressing some sustainable development goals (SDGs). It could directly tackle SDG number 12 (responsible production and consumption). In addition, consuming bio-based products has potential health benefits over synthetic products, this points to SDG number 3 (good health and wealth being).
2. Health benefits from bio-based colorants and additives: several C50 carotenoids have been reported to have superior antioxidative and radical-quenching properties.

The field of synthetic biology might be impacted positively by the lessons from this literature. With the bypass of the time consuming and cost of strain engineering, many cloning experiments could easily be done. For instance, transcriptome analysis of a microbe could lead to the discovery of upregulated genes capable of synthesizing biofuels at an industrial scale. This will span the field of synthetic biology as many companies will kick-start with the cloning of these genes in an industrial workhorse to scale its production. As a result, reducing carbon emissions and mitigating climate change (SDGs #13). It might be a tentative solution to providing affordable and clean energy (SDG #7).

Although manufacturing food additives and colorants from microbial factories shows potential merits, so long as synthetic biology and genetic engineering techniques are involved, ethical concerns remain. We cannot totally undermine the risk associated with the consumption of products obtained from genetically modified systems.

1. **Bibliography**

* Henke, N.A. and Wendisch, V.F. (2019). Improved Astaxanthin Production with Corynebacterium glutamicum by Application of a Membrane Fusion Protein. *Marine Drugs*, 17(11), p.621. doi:https://doi.org/10.3390/md17110621.
* Heider, S.A.E., Peters-Wendisch, P., Wendisch, V.F., Beekwilder, J. and Brautaset, T. (2014). Metabolic engineering for the microbial production of carotenoids and related products with a focus on the rare C50 carotenoids.

*Applied Microbiology and Biotechnology*, 98(10), pp.4355–4368. doi:https://doi.org/10.1007/s00253-014-5693-8.

* ‌Wolf, S., Becker, J., Tsuge, Y., Kawaguchi, H., Kondo, A., Marienhagen, J., Bott, M., Wendisch, Volker F. and Wittmann, C. (2021). Advances in metabolic engineering of Corynebacterium glutamicum to produce high-value active ingredients for food, feed, human health, and well-being. *Essays in Biochemistry*, [online] 65(2), pp.197–212. doi:https://doi.org/10.1042/EBC20200134.
* Bio-synthesis of food additives and colorants-a growing trend in future food. (2021). *Biotechnology Advances*, [online] 47, p.107694. doi:https://doi.org/10.1016/j.biotechadv.2020.107694.
* Katarina Cankar, Henke, N.A. and Wendisch, V.F. (2022). Functional food additives/ingredients production by engineered Corynebacterium glutamicum. *Systems Microbiology and Biomanufacturing*, 3(1), pp.110–121. doi:https://doi.org/10.1007/s43393-022-00141-4.
* Gill, R.T. (2003). Enabling inverse metabolic engineering through genomics. *Current Opinion in Biotechnology*, 14(5), pp.484–490. doi:https://doi.org/10.1016/s0958-1669(03)00116-2.
* Zhang, H., Li, Y., Wang, C. and Wang, X. (2018). Understanding the high l-valine production in Corynebacterium glutamicum VWB-1 using transcriptomics and proteomics. *Scientific Reports*, [online] 8(1), p.3632. doi:https://doi.org/10.1038/s41598-018-21926-5.
* Lv, Y., Wu, Z., Han, S., Lin, Y. and Zheng, S. (2011). Genome Sequence of Corynebacterium glutamicum S9114, a Strain for Industrial Production of Glutamate. *Journal of Bacteriology*, 193(21), pp.6096–6097. doi:https://doi.org/10.1128/jb.06074-11.
* Henke, N., Heider, S., Peters-Wendisch, P. and Wendisch, V. (2016). Production of the Marine Carotenoid Astaxanthin by Metabolically Engineered Corynebacterium glutamicum. *Marine Drugs*, 14(7), p.124. doi:https://doi.org/10.3390/md14070124.
* ‌Kim, H.-I., Nam, J.-Y., Cho, J.-Y., Lee, C.-S. and Park, Y.-J. (2013). Next-generation sequencing-based transcriptome analysis of l-lysine-producing Corynebacterium glutamicum ATCC 21300 strain. *Journal of Microbiology*, 51(6), pp.877–880. doi:https://doi.org/10.1007/s12275-013-3236-0.