# STUDIES IN DIRECTED ENDOSYMBIOSIS

BY

JASON COURNOYER

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Chemistry
in the Graduate College of the
University of Illinois Urbana-Champaign, 2024

Urbana, Illinois

Doctoral Committee:

Assistant Professor Angad P. Mehta, Chair and Director of Research
Professor David Sarlah
Professor Zaida A. Luthey-Schulten
Professor Paul J. Hergenrother

**ABSTRACT**

The acquisition of membrane-bound organelles such as mitochondria and chloroplasts was a fundamental touchstone in eukaryotic cell evolution, and it has consequently shaped the course of evolution and ecology. With the advent of sequencing technology, there is now overwhelming evidence to support a once-controversial hypothesis: that mitochondria and chloroplasts evolved from previously free-living prokaryotes that were engulfed by larger cells and retained as endosymbionts. Over time, these endosymbionts ceded most of their own genomes, thereby making essential processes such as DNA replication dependent on peripheral factors: in eukaryotic organisms, essential host-encoded proteins must first localize to the organelle, meaning it is only within the host cell microenvironment that the organelle can persist. The survival of extant host cells, in turn, depends on central chemical processes carried out in the organelles which serve bioenergetic and catabolic functions, such as photosynthesis, the Calvin–Benson–Bassham (CBB) cycle, and the TCA cycle. Importantly, these processes are only known to have ever evolved in prokaryotes and their organellar descendants.

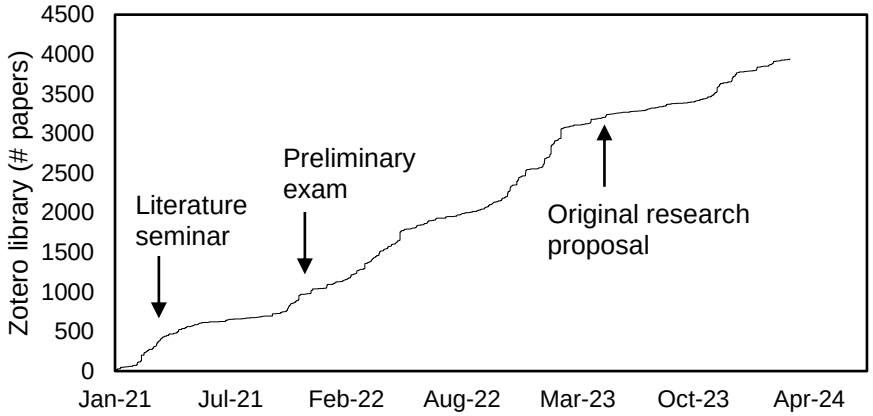
Endosymbiosis throughout nature is marked by several hallmarks: genome minimization, acquisition of protein import systems, metabolic interdependency, loss of peptidoglycan, and replication control. In principle, synthetic systems designed to mimic these hallmarks can serve to provide crucial insights into the process through which free-living bacteria transitioned into organelles, and test models derived from phylogenetic inferences. The first chapter of this dissertation is an introduction which briefly summarizes our current theory of eukaryotic cell evolution through endosymbiosis. The introduction also reviews advances made in engineering the two most important organisms in this dissertation, *Saccharomyces cerevisiae* and *Synechococcus elongatus* PCC7942 (hereafter Syn7942): these studies serve as precedence for the experimental approaches used in each of the subsequent chapters.

The second chapter of this dissertation is a reprint of a paper published in *Nature Communications*, which describes the first steps in engineering artificial, photosynthetic endosymbiosis using *S. cerevisiae* and Syn7942. In this system, Syn7942 cells are fused with mutant *S. cerevisiae* cells which are incapable of respiration; the cyanobacterial cells support the host bioenergetic functions by producing ATP through photosynthesis and exporting that ATP into the yeast from within. The yeast-cyanobacteria chimeras are characterized by analysis of their genomes and microscopy. The third chapter of this dissertation is a paper in preparation, detailing a study in which a variety of auxotrophic Syn7942 mutants were fused to yeast cells in order to demonstrate metabolic coupling between the host and endosymbiont. This study aims to mimic the loss of essential genes in endosymbiotic cyanobacteria during chloroplast evolution by iteratively deleting genes in Syn7942 which encode enzymes catalyzing steps in amino acid and coenzyme biosynthesis. In this system, the viability of these auxotrophic cyanobacteria as endosymbionts is measured in order to elucidate the need for acquisition of exogenous transport mechanisms. The fourth chapter of this study aims to use *in situ* hybridization analysis as a method to visualize how gene expression is altered in yeast-cyanobacteria chimeras. Using this method, established protocols can be used to perturb the abundance of transcripts corresponding to metabolic proteins in either the host or endosymbiont. On a cell-by-cell basis, adaptations on the transcriptomic level which may affect the viability of chimeras can be observed. This study could provide insights into to how to modulate environmental conditions (e.g., light) to favor sustained endosymbiosis using the engineered organisms.

**ACKNOWLEDGMENTS**

My doctoral work was completed over five exceptionally difficult years. I want to personally thank C.A.L., S.D.A., M.Y.O., S.J.B., M.N.P., my aunt and uncle, my siblings, and Stitch for keeping me grounded.

My Ph.D. took about 1,670 days to complete. During this time, I kept records of a few things. I don't know when I'll get another chance to preserve these data in writing, thus:




Cumulative size of my Zotero library. Data not shown prior to Jan 2021—my laptop had been destroyed, and all of my previously-saved Zotero data gone with it. The slope of the cumulative sum of papers is shown to increase directly before milestones of my degree, implying that I crammed.

Albums I listened to most between Nov 2019 and Apr 2024

| Artist | Album | Plays |
|---|---|---|
| Genesis | The Lamb Lies Down On Broadway | 1614 |
| Fleet Foxes | Shore | 548 |
| Bon Iver | i,i | 440 |
| Comus | First Utterance | 275 |
| Big Big Train | Folklore | 257 |
| Sufjan Stevens | The Ascension | 248 |
| Tally Hall | Marvin's Marvelous Mechanical Museum | 244 |
| Battles | Mirrored | 233 |
| Hiatus Kaiyote | Choose Your Weapon | 207 |
| Marillion | Misplaced Childhood | 205 |
| Sufjan Stevens | Illinois | 197 |
| Battles | Gloss Drop | 168 |

**TABLE OF CONTENTS**

## Chapter 1—Introduction

### Endosymbiosis and chloroplast evolution

Complex, multicellular life forms arose through the remarkable evolutionary transition of prokaryotes into eukaryotes. Through decades of research, it has come to be understood that this process was facilitated by the fusion of two or more taxonomically distinct lineages into a single organism, in what is known as the endosymbiotic theory[1–5]. It is in this manner that extant eukaryotes such as plants and algae must have acquired control of photosynthesis, the chemical process which feeds the biosphere of Earth: because photosynthesis evolved only once[6–10], the process belongs exclusively to the prokaryotes and their descendants in the form of plastids.

In the archetypal primary plastid endosymbiosis scenario, a β-cyanobacterium was engulfed by a heterotrophic eukaryote, forming the clade Arhcaeplastida[11] from which green algae (incl. plants), red algae, and glaucophyte algae diverged[12]. The acquisition of the primary plastids is estimated to have occurred 1-2 Gya[13–15]. Organisms in these lineages carry plastids bound by two membranes—one from the cyanobacterium and the other from the host. Red and green algal lineages have themselves been engulfed by eukaryotic cells (secondary endosymbiosis) and evolved into what are known as secondary plastids bound by three or more membranes[12] (**Figure 1.1**).

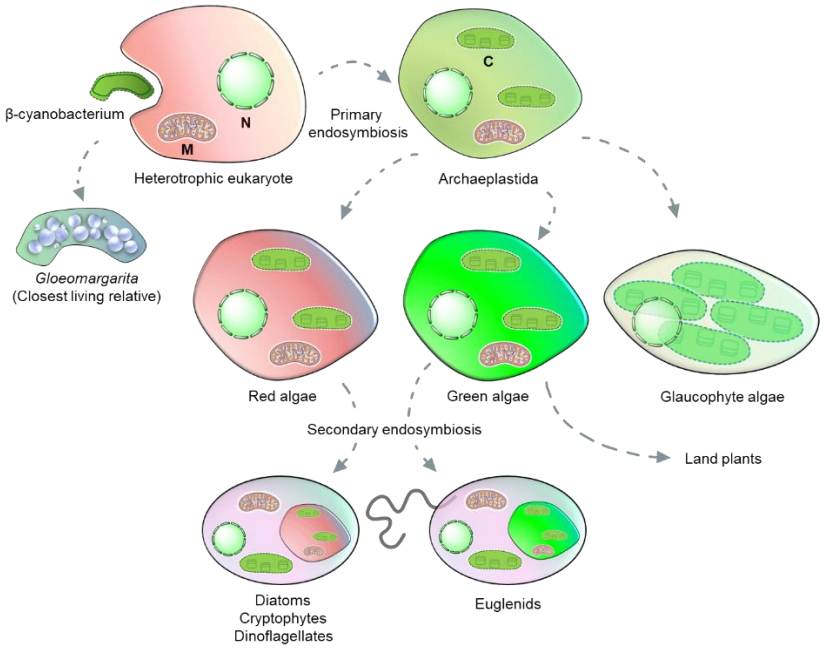



**Figure 1.1**. Plastid evolution through primary and secondary endosymbiosis. Reproduced from Cournoyer, J. E., De, B. C., Mehta, A. P. Molecular Insights from Endosymbiosis and their Implications on Evolution and Synthetic Biology (review article, in preparation).

There are several molecular characteristics observed consistently in both modern-day chloroplasts and photosynthetic endosymbionts. These characteristics are: 1) genome minimization, 2) acquisition of protein import/export mechanisms, 3) host-controlled replication of the endosymbiont/organelle, 4) loss of peptidoglycan, and 5) metabolic interdependency. In principle, each of these characteristics could be recapitulated in an artificial cell-

in-cell system. This dissertation will focus on the first, second, and fifth characteristics. The first characteristic, genome minimization, is highly distinct, as shown in **Figure 1.2**. Modern-day plastids encode about 200 proteins[16]. Predominantly, those proteins are either ribosomal subunits or components of the photosynthetic apparatus[2,17]. Non-coding RNAs are also highly retained[18]. This pattern of gene loss reflects a ceding of genetic control, by the endosymbiont, of numerous metabolic processes such as DNA replication and the biosynthesis of amino acids, nucleotides, and coenzymes. These critical metabolites are provided to the plastid either through direct import from the host cytosol, or through the import of biosynthetic enzymes expressed from the host nucleus. It has been shown that 2,000-5,000 unique proteins are localized to plastids, far outnumbering the number of protein-coding genes in their genomes[19,20]. Usually, nuclear-encoded proteins are marked for chloroplast localization with an N-terminal targeting peptide (TP) which is recognized and subsequently cleaved by TIC/TOC transport complexes, which facilitate GTP-mediated **t**ranslocation on the **i**nner **c**hloroplast membrane and **t**ranslocation on the **o**uter **c**hloroplast membrane, respectively[21]. Several studies have observed striking similarities between TPs and host-encoded antimicrobial peptides (AMPs) and implied an evolutionary link between the two: in these hypothesized scenarios, modern-day organelle targeting mediated by TPs evolved from "import-and-destroy" mechanisms used by endosymbionts to resist host-encoded AMPs[22–25].

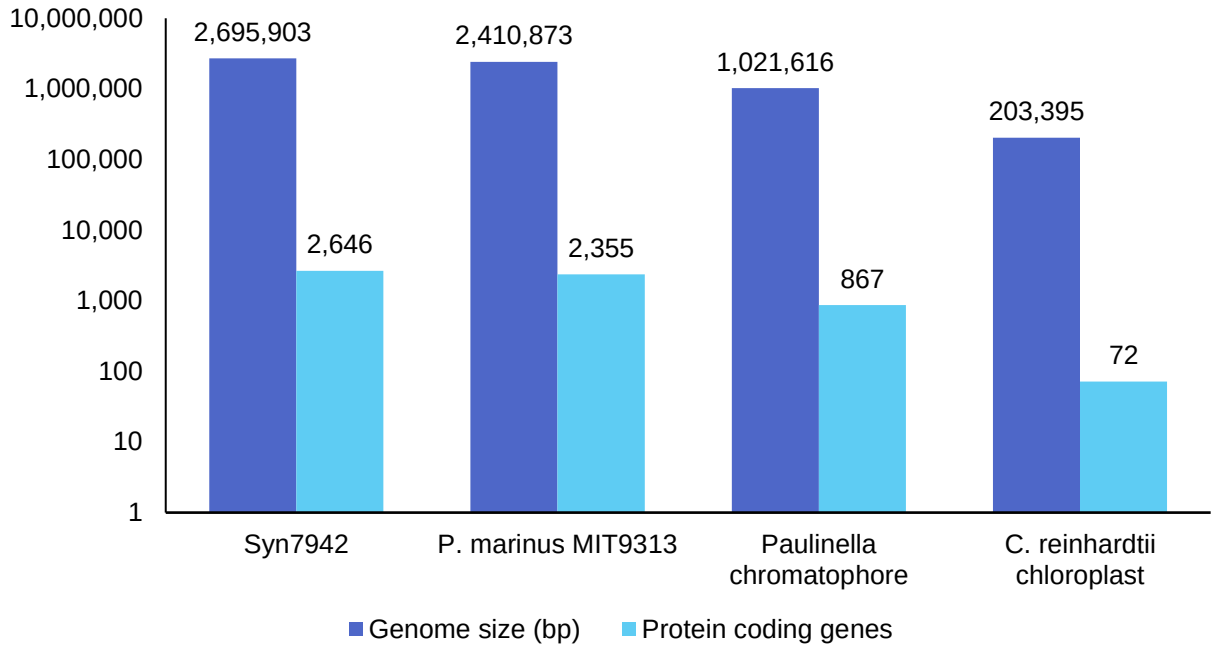



**Figure 1.2**. Genome sizes of photosynthetic cyanobacteria and organelles.

This pattern of gene loss and protein import is observed in modern-day photosynthetic endosymbionts as well as chloroplasts. This parallel is striking in studies of the amoeba *Paulinella chromatophora*, which acquired a primary α-cyanobacterial endosymbiont (called the "chromatophore") approximately 60-100 million years ago[26]. Due to the relatively recent acquisition of chromatophores in *Paulinella* as compared to the primary plastids of Archaeplastida, *P. chromatophora* has been described as an evolutionary "snapshot" in the process of evolving its endosymbiont into an organelle[27]. Similar to chloroplasts, hundreds of nuclear-enczoded proteins with N-terminal TPs

have been shown to localize to the chromatophore[26]. Many of these proteins are inferred, by gene ontology, to fill gaps in biosynthetic pathways (e.g., arginine and chlorophyll biosynthesis) which are incompletely encoded by the genome of the chromatophore[26].

Extensive metabolic interdependency (import and export of metabolites across endosymbiont or organelle membranes) arises as a necessary consequence of the adaptations described above. The end result of this integration by metabolic connectivity is observed in the inner and outer envelopes of chloroplasts, in which are embedded selective channels mediating transport of cations, amino acids, inorganic phosphate, and phosphorylated carbon compounds[28]. The majority of these solute transporters have been determined to originate from the host[29].

**Synthetic cell-in-cell systems**

Efforts to engineer artificial, chimeric cell-in-cell systems have persisted for a near-century[30–45]; in particular, there has been great interest in synthetically recapitulating mutualism between heterotrophic hosts and photosynthetic symbionts[40,46,47]. These "artificial chloroplast" or "synthetic lichen" systems would be powerful tools in a variety of synthetic biology applications, such as cell compartmentalization, bioremediation[48], food and biofuel production[49], natural product synthesis, and terraformation[47]. While several studies have demonstrated that artificially-introduced microalgae can thrive in a variety of eukaryotic cell types, including fungal mycelia[47], macrophage and zebrafish embryos[40,46], those studies have not yet shown intricate metabolic interconnectivity and genome remodeling as occurring in the evolution of naturally-occurring endosymbionts and organelles.

Studies in directed endosymbiosis[50,51] aim to build systems whereby the principles governing natural endosymbiosis are applied to synthetic systems. In a foundational study, auxotrophic *Escherichia coli* cells were fused with respiration-deficient *Saccharomyces cerevisiae* cells[52]. The *E. coli* cells expressed an ADP/ATP translocase, facilitating the transfer of ATP from the endosymbiont to the host, mimicking the essential bioenergetic function of the mitochondrion[52]. The same research group went on to delete multiple essential biosynthetic genes (totaling 46 kb) from the genome of *E. coli* and showed that those strains could also persist as endosymbionts in *S. cerevisiae*[53]. In this second study, the deletion of multiple genes encoding amino acid biosynthesis enzymes is designed to mimic the loss of most endosymbiont genes occurring as a necessary outcome of organelle evolution[53].

A variety of methods have been used to insert live bacteria and organelles into hosts. In the case of *S. elongatus* cells propagated in mammalian cells, macrophage and zebrafish embryos, the cyanobacteria invaded cells through the heterologous expression of invasin proteins, phagocytosis, and microinjection, respectively[40]. Importantly, in both this study and the *E. coli-S. cerevisiae* study, the bacterial cells expressed recombinant proteins (Listeriolysin O and SNARE-like proteins, respectively) in order to avoid degradation in the host lysosome[40,52,53]. Microfluidics and DNA nanostructures are also candidate methods for inserting these components into hosts[54]. Currently, PEG-induced fusion, which has long been used to transform DNA, introduce isolated mitochondria[55], and transfer whole bacterial genomes into yeast[45,56], has been shown to be highly effective in compartmentalizing bacterial cells within yeast.

**Synechococcus elongatus**

*Synechococcus elongatus* PCC7942 (hereafter Syn7942) is a unicellular, freshwater cyanobacterium possessing several features which make it a promising candidate for directed endosymbiosis studies. Briefly, it is an obligate photoautotroph[57] with a 2.7 Mbp genome and two endogenous plasmids[58,59] that is innately capable of

exogenous DNA uptake[60]. Compared to other unicellular cyanobacteria such as *Prochlorococcus*, axenic cultures of Syn7942 can be obtained with ease[61]. Finally, Syn7942 is fairly fast-growing, with a doubling time of about 7 h under ideal conditions[62]. Consequently, Syn7942 and its close relatives have been extensively studied in numerous contexts, such as to model the circadian clock[63], biophysics of photosynthesis[64], as a biosynthesis chassis[65,66], and in artificial microbial consortia[67,68]. Below, I summarize advances in engineering Syn7942 which are most relevant to the studies included in this dissertation.

**Physiology**

The mechanisms of photosynthesis are conserved in chloroplasts and algae, including cyanobacteria. Briefly, a pair of photosystem (PS) reaction center-localized chlorophyll *a* molecules, called P680 in PSII and P700 in PSI, are excited by light energy, initiating the downhill transfer of electrons through a series of donors and acceptors, ultimately yielding 2 mol equivalent of NADPH and 1 mol equivalent of $O_2$ per 8 photons (Figure 1.3). In oxygenic photosynthesis, water acts as the first electron donor (the water-splitting reaction). PSI and PSII generate protons as a product of the electron transport chain, providing the chemiosmotic potential that powers ATP synthase[69].

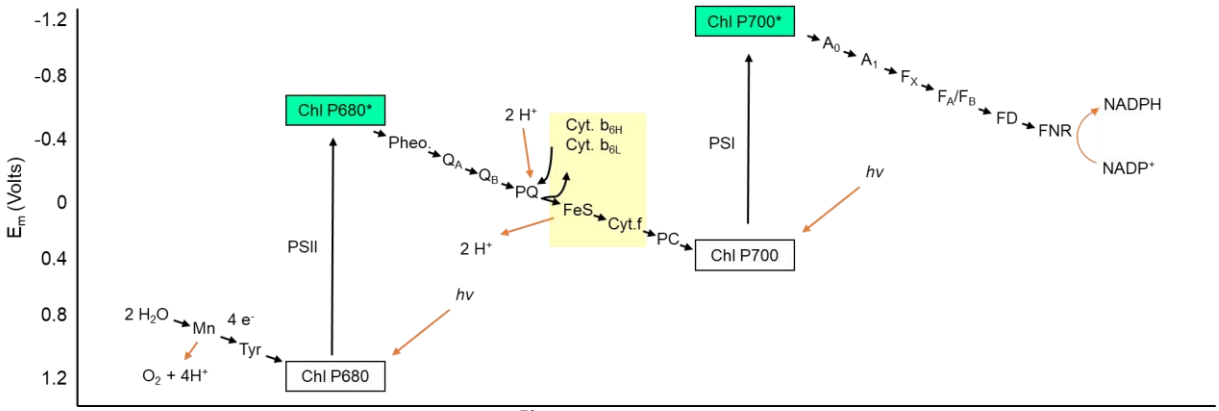



**Figure 1.3**. Z-scheme diagram of photosynthesis[70]. **Mn**, $Mn_4CaO_5$ cluster; **Tyr**, tyrosine; **Chl P680**, chlorophyll *a* dimer; **PSII**, photosystem II; **Pheo.**, pheophytin; **$Q_A$**, **$Q_B$**, **PQ**, plastoquinone; **FeS**, Rieske iron-sulfur protein; **Cyt.f**, cytochrome *f*; **Cyt. $b_{6H}$**, cytochrome *b* (high-energy); **Cyt. $b_{6L}$**, cytochrome *b* (low-energy); **PC**, plastocyanin; **Chl P700**, chlorophyll *a* dimer (primary donor); **PSI**, photosystem I; **$A_0$**, chlorophyll *a* (primary acceptor); **$A_1$**, phylloquinone; **$F_X$**, **$F_A$**, **$F_B$**, iron-sulfur protein centers; **FD**, ferredoxin; **FNR**, ferredoxin-NADP oxidoreductase

Cyanobacteria and red algae differ from chloroplasts in that they harvest light using phycobilisomes, rod-shaped, supramolecular protein complexes decorated with covalently-bound, open-chain tetrapyrrole chromophores termed phycobilins[71]. Excitation energy from light is transferred from these pigments to chlorophyll *a* in PSII[64] (**Figure 1.3**). The presence of different phycobilins and rod proteins significantly affects the spectral characteristics of cyanobacteria and algae (**Figure 1.4**): indeed, transfers of rod genes allow *Synechococcus* strains to adapt to different light niches[72,73]. Accordingly, these features must be considered when characterizing cyanobacteria using fluorescence microscopy approaches[74].

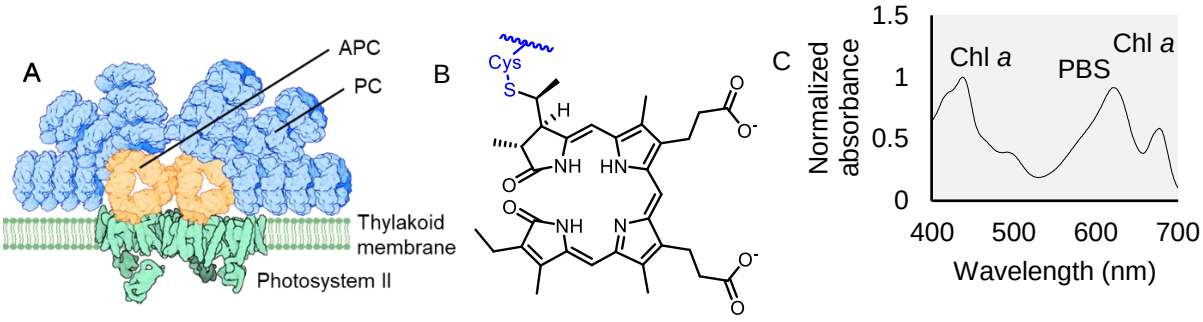



**Figure 1.4.** Summary of Syn7942 spectral characteristics. (A) Structure of the Syn7942 phycobilisome. APC, allophycocyanin; PC, phycocyanin. (B) Structure of (3Z)-phycocyanobilin, an open-chain tetrapyrrole chromophore attached to cysteine residues (blue) of phycobiliproteins in Syn7942, (C) Absorbance spectrum of Syn7942. Chl *a*, chlorophyll *a* absorbance peak; PBS, phycobilisome absorbance peak.

The only requirements for the sustained, axenic growth of Syn7942 are water, $CO_2$, light, and salts; exogenous nutrients such as organic carbon sources, amino acids, and coenzymes are not required. Previously, studies have generated auxotrophic mutants of Syn7942 (e.g., requiring Trp, Met, Phe, Cys, Orn, Leu, Ser, Glu, Gly, biotin, thiamin, nicotinic acid, Ura, Ade, $NH_4Cl$, $Na_2S_2O_3$, nitrate, acetate, *p*-aminobenzoic acid, or malate) and closely-related strains using random mutagenesis methods (e.g., ethidium bromide and nitrosoguanidine[75–79]) followed by ampicillin enrichment to select against mutants capable of growth under nutrient-poor conditions. These studies, conducted prior to the widespread adoption of DNA sequencing technology, did not reveal which specific gene mutations led to auxotrophic phenotypes. To this day, most of the Syn7942 genome is annotated based on gene ontology (GO) and not extensively functionally characterized: hence, auxotrophic Syn7942 mutants have not yet been generated through systematic, single gene deletions in the manner of other common laboratory microorganisms such as *E. coli*[80] and *S. cerevisiae*[81]. In principle, a series of single gene deletions could generate a wide variety of auxotrophic mutant strains derived from Syn7942, as long as the necessary metabolites are transported into the cells; it has been shown, however, that amino acid uptake in Syn7942 is highly asymmetrical[82,83] (**Table 1.1**). In general, neutral amino acids have been shown to efficiently enter Syn7942, while basic amino acids are imported poorly, if at all[82]. This may explain the bias observed in auxotrophic strains generated through random mutagenesis: if mutations in genes encoding biosynthesis enzymes for Lys or Arg are lethal even in the presence of those metabolites, then Lys or Arg auxotrophs would never be observed. Resources such as TransportDB[84,85] could be an invaluable tool in characterizing putative solute transporters in many microorganisms, including Syn7942, thereafter facilitating extensive engineering of solute import and export systems in cyanobacteria[86]. Parallel efforts are currently underway to engineer auxotrophic strains of *Synechocystis sp.* PCC 6803[87], a cyanobacterium closely related to Syn7942. Although these studies could help to provide insights in further engineering solute transport Syn7942, it is important to note that the two organisms have highly dissimilar amino acid uptake capabilities[82].

**Genome minimization**

A variety of genetic manipulation tools have been developed for Syn7942, such as self-replicating shuttle vectors[88], *in silico* modular plasmid assembly[89] and CRISPR-Cas mediated mutagenesis[90–93]. Golden and co-workers have also generated a library of Syn7942 transposon mutants and surveyed the essentiality of the whole Syn7942 genome through next-generation sequencing[94]. Delaye and co-workers identified large, contiguous “islands” of

dispensable genes as prime targets of deletion[95]. These surveys could serve as a valuable resources in efforts to engineer, in a top-down fashion, a minimal, "chloroplast-like" cyanobacterial endosymbiont: in principle, large tracts of nonessential genes could be deleted from the Syn7942 genome as done previously in a variety of other bacteria[96–100], and essential genes could be deleted in an iterative fashion under nutrient-rich conditions. Thus, the mutant Syn7942 genome would be pared down to a few hundred genes, and peripheral factors encoded by a host cell would control the remaining biological functions of the minimal cyanobacterium.

**Table 1.1**. Amino acid uptake activities in Syn7942 compared to *Synechocystis* sp. PCC 6803. Adapted from Montesinos et al., 1997[82]. Values are reported in units of nmol/mg chlorophyll per 10 min

| | Arg | His | Lys | Ala | Asn | Gln | Gly | Leu | Phe | Pro | Ser | Asp | Glu |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Syn7942 | 0.3 | 24.3 | 0.8 | 29.9 | 4.3 | 13.4 | 6.7 | 41.7 | 41.8 | 0.7 | 8.9 | 7.1 | 11.2 |
| *Synechocystis* sp. PCC 6803 | 3126 | 63.9 | 457.3 | 101.0 | 10.5 | 51.6 | 102.8 | 79.4 | 54.4 | 70.2 | 69.3 | 2.6 | 568.3 |

Until recently, efforts to engineer the Syn7942 genome in this manner have been limited. The most reliable methods of inserting foreign DNA into Syn7942 is through plasmid transformation or conjugation[101], and the overwhelming majority of vectors used for Syn7942 are non-replicating. These "suicide vectors" integrate foreign DNA into Syn7942 through homologous recombination. Thus, heterologous genes (gain of function) are usually expressed from "neutral sites" wherein those genes are integrated along with an antibiotic cassette[102], while deleted genes (loss of function) are usually replaced with the antibiotic cassette[103,104]. The latter scenario itself has drawbacks owing to the polyploid[103] nature of Syn7942. When Syn7942 is transformed with a homologous recombination plasmid, the possible outcomes of the transformation are 1) Double crossover recombination, successfully replacing a gene on a copy of the genome, 2) single crossover recombination, duplicating fragments of the target while the wild-type allele remains intact, and 3) nonspecific recombination[101]. In all three outcomes, mutants can be isolated through antibiotic selection, but in only one of these outcomes is the target gene deleted. Although single crossover recombination events are unlikely in Syn7942 compared to double crossovers, the likelihood increases when essential genes are targeted[101]. Moreover, the target gene must be deleted on all copies of the genome in order to reach extinction in a clonal population; otherwise, the resulting merodiploid bacteria survive antibiotic selection while retaining the intact gene. Thus, homozygous mutants must be segregated through persistent streaking of colonies onto new plates, a process which can take weeks[91,105,106].

Recently, Cas12a (previously Cpf1) nuclease has emerged as a potent tool for deleting genes in Syn7942[90,92,93]. Hou and co-workers used Cas12a to iteratively delete 10 kb fragments of non-essential genes from Syn7942 (including those previously identified by Delate et al.[95]), ultimately reducing the size of the 2.7 Mbp genome of Syn7942 by 3.8 %[93]. Together with methods to delete essential genes from Syn7942, these innovations will constitute a significant advance in engineering a minimal photosynthetic endosymbiont.

## Chapter 2—Engineering artificial photosynthetic life-forms through endosymbiosis

### Author contributions

Jason E. Cournoyer—conceptualization, methodology, validation, formal analysis, investigation, data curation, visualization, writing

Sarah D. Altman—validation, formal analysis, investigation

Yang-le Gao—validation, formal analysis, investigation

Catherine L. Wallace—methodology, formal analysis, investigation, visualization, writing

Dianwen Zhang—methodology, supervision, writing

Guo-Hsuen Lo—investigation

Noah T. Haskin—validation, formal analysis, investigation

Angad P. Mehta—conceptualization, methodology, formal analysis, investigation, resources, data curation, writing, supervision, project administration, funding acquisition


### Abstract

The evolutionary origin of the photosynthetic eukaryotes drastically altered the evolution of complex lifeforms and impacted global ecology. The endosymbiotic theory suggests that photosynthetic eukaryotes evolved due to endosymbiosis between non-photosynthetic eukaryotic host cells and photosynthetic cyanobacterial or algal endosymbionts. The photosynthetic endosymbionts, propagating within the cytoplasm of the host cells, evolved, and eventually transformed into chloroplasts. Despite the fundamental importance of this evolutionary event, we have minimal understanding of this remarkable evolutionary transformation. Here, we design and engineer artificial, genetically tractable, photosynthetic endosymbiosis between photosynthetic cyanobacteria and budding yeasts. We engineer various mutants of model photosynthetic cyanobacteria as endosymbionts within yeast cells where, the engineered cyanobacteria perform bioenergetic functions to support the growth of yeast cells under defined photosynthetic conditions. We anticipate that these genetically tractable endosymbiotic platforms can be used for evolutionary studies, particularly related to organelle evolution, and also for synthetic biology applications.


### Introduction

The origin of eukaryotic cells is a fundamental milestone in the evolution of complex life forms, and the evolution of organelles is one of the key steps in eukaryogenesis. Based on the endosymbiotic theory, eukaryotic organelles, like mitochondria and chloroplasts, are proposed to have originated and evolved from bacterial endosymbionts during an early stage of evolution (Fig. 1A) [5,107–110]. This cataclysmic evolutionary event resulted in the origin of eukaryotic cells, followed by subsequent evolution of versatile life forms that significantly impacted global ecology and ecosystems. Since organelles like mitochondria and chloroplasts possess their own genomes, DNA sequencing and biochemical studies spanning several decades have supported this hypothesis (Bonen et al., 1977, Zimorski et al., 2014). Despite the significant importance of endosymbiosis to the evolution of life and global ecology, we have little idea of how bacterial endosymbionts were established within host cells and how they evolved and transformed into organelles. Some of the key questions central to the transformation of bacterial endosymbionts into organelles are: (i) what are the minimal factors and mechanisms necessary for establishing endosymbiosis? (ii) how did endosymbiont genome minimization occur? (iii) how did the endosymbiont metabolism evolve once in a host cell?

It has been a long-standing quest to identify and characterize naturally existing endosymbiotic systems [112,113] for evolutionary studies and synthetic applications. Particularly for photosynthetic endosymbiosis, as early as 1930s, artificial endosymbiosis has been attempted with photosynthetic algae and freshly extracted mammalian cells [114,115]. Several subsequent studies have attempted to transiently propagate photosynthetic bacteria, algae or photosynthetic organelles within eukaryotic cells like zebrafish embryonic cells and animal cells [116,117]. While these systems suggest the possibility of a photosynthetic bacterial cell to survive within the eukaryotic cells, none of the photosynthetic bacteria were able to fully support the host bioenergetic functions or carbon-source requirements similar to chloroplasts in algal and plant cells. We believe that building genetically tractable model endosymbiotic systems that perform organelle-like functions (e.g., ATP synthesis and supply, carbon assimilation) will provide an endosymbiotic platform that can be metabolically manipulated, analytically studied and imaged, and computationally modeled and predicted. Such platforms will break the gridlock on our understanding of the evolutionary origin of photosynthetic eukaryotic cells. We (along with Schultz) had previously engineered *E. coli* endosymbionts within yeast cells to rescue compromised mitochondrial function.[118] This is a genetically tractable platform that can be used to recapitulate mitochondrial evolution.[119]

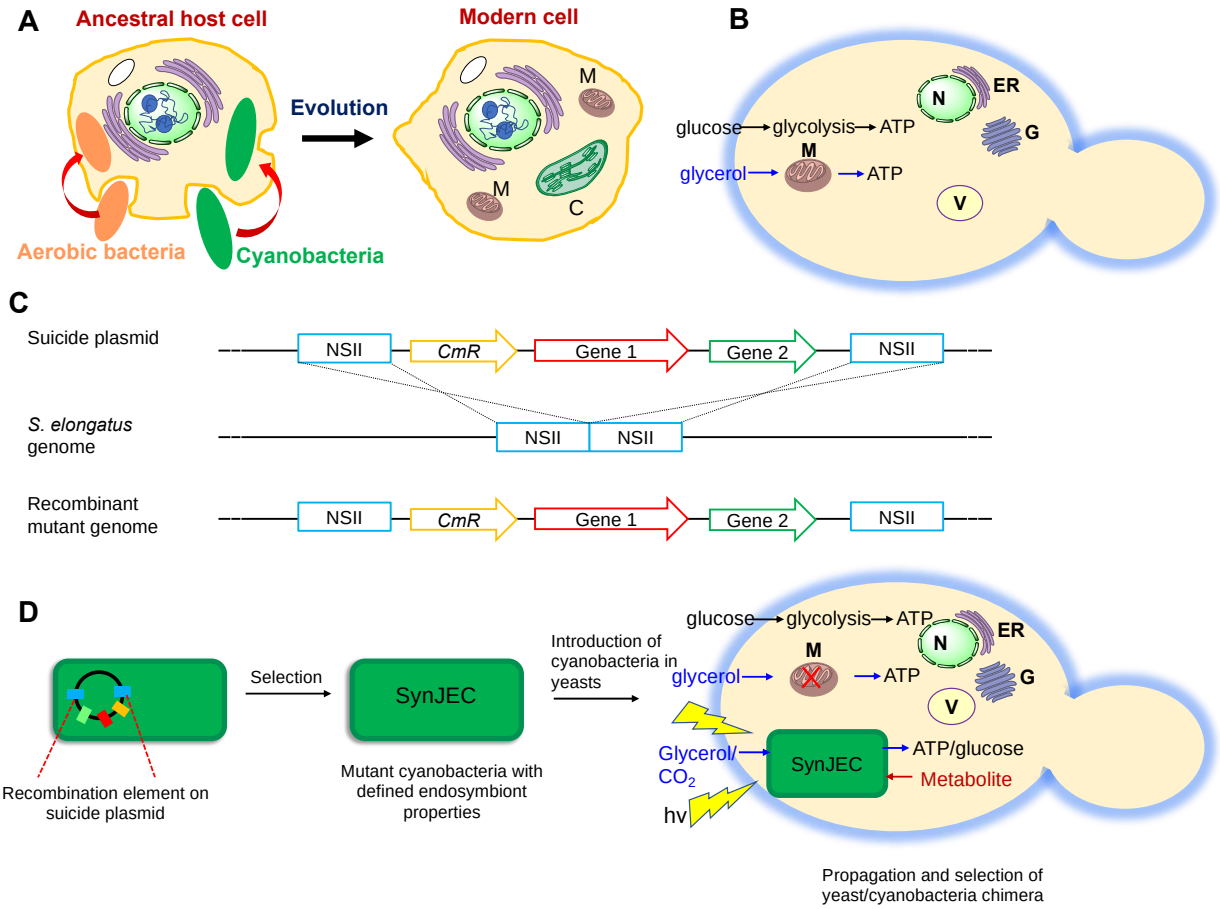



**Figure 2.1**: **Endosymbiotic theory and our platform to recapitulate the evolution of photosynthetic eukaryotic cells: (A)** The endosymbiotic theory – Mitochondria, M, are proposed to have evolved from a class of α-proteobacteria while the chloroplast, C, is proposed to have originated from cyanobacteria. Golgi apparatus—G, Endoplasmic reticulum—ER, Vacuole—V. **(B)** *S. cerevisiae* (budding yeast) cells produce ATP by glycolysis or mediated oxidative phosphorylation. **(C)** Suicide plasmid-based strategy used in this manuscript to engineer cyanobacterial mutants, SynJEC strains. **(D)** Our platform: We use suicide plasmid-based strategy to engineer cyanobacterial endosymbionts, SynJEC strains, such that they perform chloroplast-like functions. *S. cerevisiae* mutants, deficient in ATP synthesis by oxidative phosphorylation under defined photosynthetic selection conditions, are used as host strains. Engineered cyanobacteria strains, SynJEC, are then introduced into the yeast cells by a cell fusion process that is developed and optimized (see Methods). The yeast/cyanobacterial chimera are selected under defined photosynthetic selection conditions where the cyanobacterial endosymbionts provide ATP to the mutant *S. cerevisiae* host cells, and *S. cerevisiae* provide essential metabolites to the *S. elongatus* endosymbionts.

Inspired by the evolutionary observations and previous synthetic efforts, here, we sought to engineer genetically tractable platforms where the endosymbiotic bacteria perform chloroplast-like functions. To engineer artificial photosynthetic endosymbiosis, we used key observations from chloroplast evolution studies. First, we had to identify and engineer genetically tractable photosynthetic bacteria such that they could perform chloroplast-like functions within the cytoplasm of the host cell. In case of chloroplast evolution, there is a consensus that chloroplasts evolved from cyanobacterial endosymbionts; several sequencing and biochemical studies support this proposal. Based on this evidence, it is suggested that *Gloeomargarita* are the closest identified relatives of chloroplasts [120,121]. While strains of *Gloeomargarita* are not readily genetically manipulable, genetic tools have been developed to manipulate several strains of *Synechococcus*, which are relatives of *Gloeomargarita* [122,123]. The exact nature of the host cell is still

elusive [124], and it has been suggested that photosynthetic endosymbionts were established in various eukaryotic host cells during the evolution of chloroplasts [109,125]. Because of these observations, for our engineering efforts, we decided to engineer artificial endosymbiosis between genetically tractable *Synechococcus elongatus* PCC 7942 (Syn7942) and a model eukaryotic cell, *Saccharomyces cerevisiae*, budding yeast. Next, to determine the role that Syn7942 would play in our synthetic endosymbiotic system, we again sought inspiration from chloroplast function and evolution. Though the modern-day chloroplasts perform various functions like carbon assimilation, sulfate assimilation, nitrate assimilation, amino acid biosynthesis, fatty acids biosynthesis amongst others,[109] the key drivers of endosymbiotic evolution of chloroplast are still unclear due to extensive horizontal gene-transfer, secondary endosymbiosis and endosymbiotic adaptation.[126] However, it is widely suggested that bioenergetic considerations may have been the key drivers of organelle evolution.[127,126] This is highlighted by the fact that ATP/ADP translocases and transporters are widely conserved across organelles like mitochondria and chloroplasts, including organisms that are related to the endosymbiotic precursors of mitochondria and chloroplast.[128,129] Based on this premise, we designed our endosymbiosis platform such that the engineered cyanobacterial endosymbionts perform bioenergetic function for the host cells, i.e., endosymbionts provide ATP generated through photophosphorylation (Figure. 2.1B-D). Thorough series of cyanobacterial and yeast engineering efforts we were able to establish artificial photosynthetic endosymbiosis between yeast mutants and cyanobacterial mutants to generate yeast/cyanobacteria chimeras that were able to propagate through at least 15 to 20 generations of growth under optimal photosynthetic growth conditions. The yeast/cyanobacteria chimeras were characterized by analyzing their viability under photosynthetic selection conditions, analysis of the total genomic DNA, metabolic coupling of endosymbiont/host, dependence on photosynthesis, pseudo-total internal reflection fluorescence (pTIRF) microscopy, fluorescence confocal microscopy and Transmission Electron Microscopy (TEM). Our studies also highlight critical genetic elements that allowed us to engineer artificial photosynthetic endosymbiosis between model cyanobacteria and model eukaryotic cells. We anticipate that that such genetically tractable photosynthetic platforms, where the endosymbiont synthesizes ATP by photophosphorylation will have significant implications on synthetic biology applications.[130–132] Further, these photosynthetic endosymbiotic systems could also provide a platform to recapitulate various evolutionary trajectories related to the conversion of photosynthetic endosymbionts into photosynthetic organelles (i.e., chloroplasts).

**Results**

**Engineering Syn7942 to function as ATP-providing endosymbionts within yeast cells:**

Phylogenetic analysis of nucleotide transport proteins (NTT) suggests that cyanobacteria likely have uncharacterized nucleotide transport fusion proteins (similar to ADP/ATP translocases) [129]. However, chloroplasts are thought to possess NTTs that possibly originated from intracellular organisms such as chlamydial [128,129]. Therefore, to emulate bioenergetic functions of the chloroplast, our goal was to engineer Syn7942 to efficiently export ATP in the presence of ADP by expression of previously characterized, recombinant ADP/ATP translocase [133–135]. We began by generating a control mutant (SynJEC0) which was obtained by a transformation and selection process developed by Golden and coworkers[123]. Syn7942 was transformed with a plasmid, pCV0055, containing a chloramphenicol resistance cassette placed between two homologous recombination domains, known as neutral sites (NS) in the Syn7942 genome [122], and the mutant strain SynJEC0 was selected on BG-11 selection medium containing

chloramphenicol. The generation of this mutant was confirmed by PCR analysis, demonstrating the presence of gene(s) integrated into the Syn7942 genome at the NSII site while the vector was eliminated (Figure. 2.7). Having generated control SynJEC0 mutants, we then engineered Syn7942 mutants expressing ADP/ATP translocase from intracellular organisms. Starting with pCV0055, we constructed a plasmid (pML3) (Figure 2.6 for plasmid map) encoding: (i) codon-optimized ADP/ATP translocase gene, *ntt1*, from *Protochlamydia amoebophila* UWE25 [133], a bacterial endosymbiont of *Acanthamoeba spp*, (ii) a chloramphenicol selection marker and (iii) NSII homology region for genomic recombination. We transformed Syn7942 with pML3 to generate the strain SynJEC1. To determine if the *ntt1* was functional when expressed in SynJEC1, we assayed the ADP/ATP translocase activity (using luciferase assays) with SynJEC1 cells and compared it to the SynJEC0 control mutant. We observed that SynJEC1 released a significant amount of ATP when compared to control cells to which ADP was not added, and also as compared to the control strain SynJEC0 (Figure 2.2A) to which ADP was added. However, unlike *E. coli* cells [118,133], we observed substantial amount of ATP released when ADP was added to the SynJEC0 cells that did not contain ADP/ATP translocase gene from *Protochlamydia amoebophila* UWE25 (Figure 2.2A and Figure 2.12). This observation suggests the possibility of ADP/ATP translocase-like proteins that might be already encoded by the Syn7942 genome itself, consistent with previous phylogenetic diversity analysis of NTTs [129].

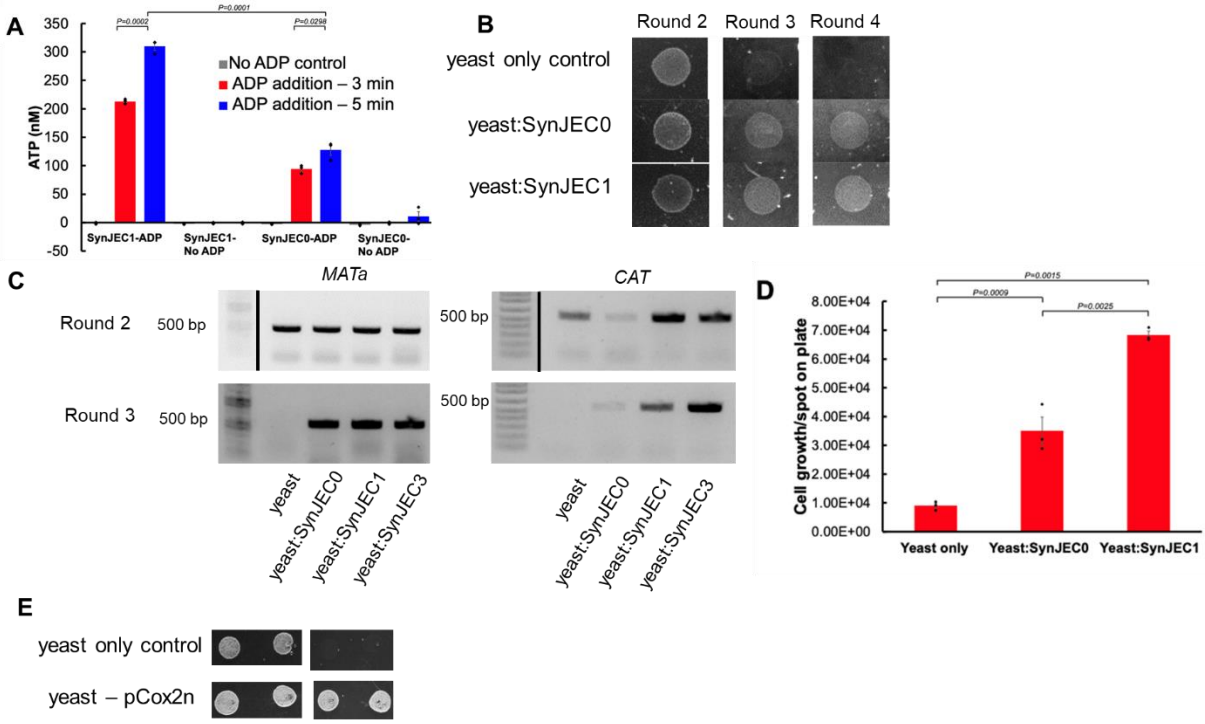



**Figure 2.2: *S. cerevisiae–SynJEC* chimeras have a partially rescued respiration-competent phenotype: (A)** Release of ATP by *SynJEC1* cells expressing the *UWE25* ADP/ATP translocase in the presence of 80 µM ADP in comparison to *SynJEC0* cells. ATP was released when SynJEC1 (expressing the ATP/ADP translocase) and SynJEC0 cells were challenged with extracellular ADP (80 µM), but not with a blank solution lacking ADP (N=3 biological replicates; data are presented as mean values +/- SEM). Two-sided *t*-tests were used to compare means without adjustments (95% CI, Cohen's *d*=10.6, DF=4, P=0.0002; 95% CI, Cohen's *d*=13.0, DF=4, P=0.0001; 95% CI, Cohen's *d*=2.7, DF=4, P=0.0002). **(B)** Growth of *S. cerevisiae* cox2-60–*SynJEC* chimeras on medium containing glycerol as the sole carbon source. No growth was observed in round 4 for yeast lacking intracellular *SynJEC*. The experiment was repeated independently six times with similar results. **(C)** Total DNA isolated from spots grown on selection medium III contain the yeast-encoded *MAT**a*** gene and *SynJEC*-encoded chloramphenicol acetyltransferase (*CAT)* gene. The experiment was repeated independently six times with similar results. **(D)** Growth rate of yeast-*SynJEC* chimeras on Selection Medium III. Cells ($3.00\times10^3$) were spotted on Selection Medium III and counted after 72 h growth. (N=3 technical replicates; data are presented as mean values +/- SEM. )*P*-values were calculated by two-tailed *t*-test comparing the two means. **(E)** Panel 1 describes the growth of *S. cerevisiae-cox2-60* and *S. cerevisiae-cox2-60*-pCOX2n under non-selection conditions. Panel 2 describes the growth of *S. cerevisiae-cox2-60* and *S. cerevisiae-cox2-60*-pCOX2n under selection conditions where the rescue in the growth of *S. cerevisiae-cox2-60*-pCOX2n is observed but no growth is observed for *S. cerevisiae-cox2-60*. The experiment was repeated independently three times with similar results.

**Generation and selection of the S. elongatus/ S. cerevisiae chimera:**

To introduce cyanobacterial cells into yeast cells, we optimized a polyethylene glycol (PEG)-induced fusion protocol which was used previously to install mitochondria [55] and *E. coli* cells [118] in yeast spheroplasts. For host cells, we used *S. cerevisiae cox2-60*, a strain that is incapable of assembling a functional cytochrome c oxidase complex and consequently has a respiration-deficient phenotype specifically due to lack of ATP synthesis under defined selection conditions.[136–139] Under selection conditions, the rescue of this phenotype can be observed on correcting this defect by plasmid (pML64) based expression of COX2 gene (Figure 2.2E) as described previously.[139] Our calculations suggest that the doubling time of *S. cerevisiae cox2*-60-pML64 is around 7 h (Figure 2,13). We next evaluated if ATP providing cyanobacterial endosymbionts could similarly rescue the phenotype of *S. cerevisiae cox2-60* under selection

conditions. We then fused SynJEC0 and SynJEC1 to the *S. cerevisiae cox2-60* cells and selected fusions by growing mixtures on partial selection conditions containing non-fermentable carbon source and low levels of fermentable carbon source (1% yeast extract, 2% peptone, 1 M sorbitol, 3 % glycerol, 0.1 % glucose, 1X BG-11; selection medium I). The fusions were propagated in 12 h light-dark cycles at 30ºC. Under these conditions, we did not observe yeast colonies with control *S. cerevisiae cox2*-60 cells but did observe small, distinct colonies for *S. cerevisiae cox2*-60-SynJEC0 and *S. cerevisiae cox2*-60-SynJEC1 fusions. Individual colonies were picked and re-plated for four consecutive rounds of regrowth: one round on selection medium II (1% yeast extract, 2% peptone, 1 M sorbitol, 3 % glycerol, 0.1 % glucose, 1X BG-11, 50 mg/ml carbenicillin; selection medium II) and two to three rounds on selection medium III (1% yeast extract, 2% peptone, 1 M sorbitol, 3 % glycerol, 50 mg/ml carbenicillin, 1X BG-11; selection medium III) (Figure. 2.2B). Note, in selection medium II and III, carbenicillin was added to eliminate any extracellular cyanobacteria. As expected, *S. cerevisiae cox2-60* cells failed to grow on selection medium III during subsequent rounds of regrowth when they were not fused to Syn7942. We observed higher growth for *S. cerevisiae cox2-60*-SynJEC0 chimeras as compared to the host strain by itself (i.e., *S. cerevisiae cox2*-60) on selection medium III through consecutive rounds of re-plating (Figure 2.2B). To calculate the doubling numbers in case of each of the fusions, we plated a defined number of starting cells and determined the endpoint cell-count. We observed that the host *S. cerevisiae cox2-60* propagated for only ~ 2 doublings on selection medium III and *S. cerevisiae cox2-60*-SynJEC0 chimeras propagated for ~10 doublings on selection medium III (Figure 2.2D). We observed robust phenotypic rescue for *S. cerevisiae cox2-60*-SynJEC1 (Figure. 2.2B and 2.2D); the *S. cerevisiae cox2-60*-SynJEC1 continued to propagate for ~14 doublings (Table 2.8) followed by a drop in the growth rate on selection medium III. These data suggested that the expression of the ADP/ATP translocase was important in restoring respiration competency in *S. cerevisiae cox2-60*. This phenotypic rescue of *S. cerevisiae cox2-60* indicated that the cyanobacterial endosymbionts, especially SynJEC1, could partially rescue the growth of the host *S. cerevisiae cox2-60* cells under selection conditions. It is possible that a weaker phenotypic rescue of the *S. cerevisiae cox2-60* cells when fused with SynJEC0 could be due to the background ADP/ATP translocase activity of SynJEC0 cells that we detected during our translocase activity studies described above. Our observations are consistent with the studies suggesting that cyanobacterial endosymbionts may have needed exogenous translocases to facilitate their bioenergetic functions[128,129].

To characterize the presence of cyanobacterial endosymbionts within yeast cells, we isolated total genomic DNA from fused yeast cells propagated for multiple generations under selection growth conditions to eliminate all the extracellular cyanobacteria (if any, see Figure 2.9, Figure 2.11), and performed PCR analysis to determine the presence of both yeast and cyanobacterial genomes. We detected the presence of both the yeast *MAT***a** gene and mutant Syn7942 chloramphenicol acetyltransferase (*CAT*) gene in the colonies by PCR (Figure 2.2C), suggesting the presence of both yeast and cyanobacterial genomes. Interestingly, this set of experiments suggested that unlike the engineered yeast/*E. coli* endosymbiosis studies [118], Syn7942 cells expressing ADP/ATP translocase were sufficient to transiently recover the growth phenotype of *S. cerevisiae cox2*-60 under selection conditions. In addition, PCR analysis for cyanobacterial genomes suggest that SynJEC0 cells that did not contain the ADP/ATP translocase gene, were also able to establish transient endosymbiosis with *S. cerevisiae cox2-60*. Again these results were unlike the yeast/*E. coli* endosymbiosis studies [118] where it was necessary to express ADP/ATP translocase and SNARE-like proteins for

establishing endosymbiosis with yeast cells. This set of experiments suggested therefore that cyanobacterial relatives of chloroplast precursors could have possessed innate ability to establish bioenergetically relevant endosymbiosis; however, as predicted [128,129], the cyanobacterial endosymbionts might have needed exogenous translocases to facilitate robust endosymbiosis.

**Improving the stability and growth of yeast/cyanobacteria chimera:**

It is hypothesized that the endosymbionts could have acquired a variety of genes through horizontal gene transfer which facilitated the establishment of robust endosymbiosis within the cytoplasm of the host cells [120,140]. Based on the previous observations with synthetic endosymbiotic systems,[118] we hypothesized that we could improve the stability of the cyanobacterial endosymbionts by engineering mechanisms that are known to evade intracellular degradation within the cytoplasm of the host cells. Literature suggests that one mechanism by which intracellular pathogenic bacteria avoid lysosomal degradation is through the expression of SNARE-like proteins [141,142], which are thought to inhibit SNARE-mediated membrane fusion through mimicry of the host SNAREs. Our previous studies suggested that expression of *Chlamydia trachomatis* genes *incA* and *CT_813* along with the *Chlamydia caviae* gene *incA* improved the stability of *E. coli* endosymbionts within yeast cells [118]. Therefore, we evaluated if expressing a combination of SNARE-proteins along with ADP/ATP translocase improves the stability of yeast/cyanobacteria chimeras.

To this end, we began by constructing a series of integrative suicide plasmids (pML14, pML17, pML28) starting from pCV0055; in addition to recombination/selection elements, each of these plasmids contained ADP/ATP translocase and a combination of gene fragments corresponding to one or more SNARE-like proteins driven by a promoter, Ptrc. Through transformation, recombination and selection methods we generated a series of *S. elongatus* mutants described in Table 1 (SynJEC2, SynJEC3, SynJEC4), each of which expressed ADP/ATP translocase and one or multiple SNARE-like proteins. Each of these mutants were confirmed by isolation of the genomic DNA, PCR amplification of the mutant locus and sequencing of the corresponding mutant locus (Figure 2.7). We then individually fused SynJEC2, SynJEC3, SynJEC4 to *S. cerevisiae cox2-60* to determine if the expression of SNARE-like proteins improved the stability of the yeast/cyanobacteria chimeras. As before, individual colonies were picked and continuously re-plated for four consecutive rounds of regrowth: two rounds on selection medium II and two rounds on selection medium III to identify optimal endosymbiosis that would result in higher stability and more generations of growth. Through this experiment, we found that *S. cerevisiae cox2-60*-SynJEC4 chimeras went extinct in the shortest amount of time compared to the *S. cerevisiae cox2-60*-SynJEC2 and *S. cerevisiae cox2-60*-SynJEC3 lineages (Figure 2.3A, 2.3C, 2.3D). The *S. cerevisiae cox2-60*-SynJEC3 chimeras were the longest-lived; in our hands, we typically observed around 20 doublings on selection medium III (Table 2.8). As per our genomic DNA analysis of the chimeras, after more than 20 rounds of doublings, we observe that the chimeras lose their viability possibly due to loss of cyanobacteria (Figure 2.16); our imaging studies described below confirm this observation (Figure 2.9). Our data suggests that the doubling times of *S. cerevisiae cox2-60*-SynJEC3 chimeras is around 10 h, possibly due to the slow growth rate of *S. elongatus strains* (doubling time ~ 6 to 8 h under optimal media conditions). Further our data indicate that during a single round or re-growth, 3000 cells of *S. cerevisiae cox2-60*-SynJEC3 chimera propagated to 3x10^5 cells under stringent selection conditions in 72 h, whereas *S. cerevisiae cox2-60* cells are not viable at all under

these conditions. This suggests that cyanobacteria are metabolically active and are able to provide ATP to the host cox2-60 strain; and hence endosymbiotic. Taken together, these data suggested that both the ADP/ATP translocase and *C. tr.* IncA and CT_813 SNARE-like proteins play a beneficial role for the cyanobacterial endosymbionts. In each of the above cases, we isolated total genomic DNA from fused yeast cells propagated for multiple generations under selection growth conditions (Figure 2.11) and PCR analyzed the presence of both yeast and cyanobacterial genomes, and as expected, we detected the presence of both yeast *MAT***a** and cyanobacterial *CAT* genes (Figure 3B).

**Table 2.1: A list of cyanobacterial strains engineered and used in this study.** See Figure 2.6 and Table 2.6 for detailed plasmid maps.

| Strain designation | Suicide plasmid | Plasmid described | Genotype |
|---|---|---|---|
| SynJEC0 | pCV0055 | Golden et al. | NSII::*CAT* |
| SynJEC1 | pML3 | This study | NSII::*CAT, UWE25-ntt1* |
| SynJEC2 | pML14 | This study | NSII::*CAT, UWE25-ntt1, Ctr-incA* |
| SynJEC3 | pML17 | This study | NSII::*CAT, UWE25-ntt1, Ctr-incA, CT_813* |
| SynJEC4 | pML28 | This study | NSII::*CAT, UWE25-ntt1, Ctr-incA, CT_813, Cca-incA* |
| SynJEC5 | pML14 and pML62 | This study | NSII::*CAT, UWE25-ntt1, Ctr-incA, CT_813; metA::KanR* |
| SynJEC9 | pML3 and pML58 | This study | NSII::*CAT, UWE25-ntt1; metA::KanR* |

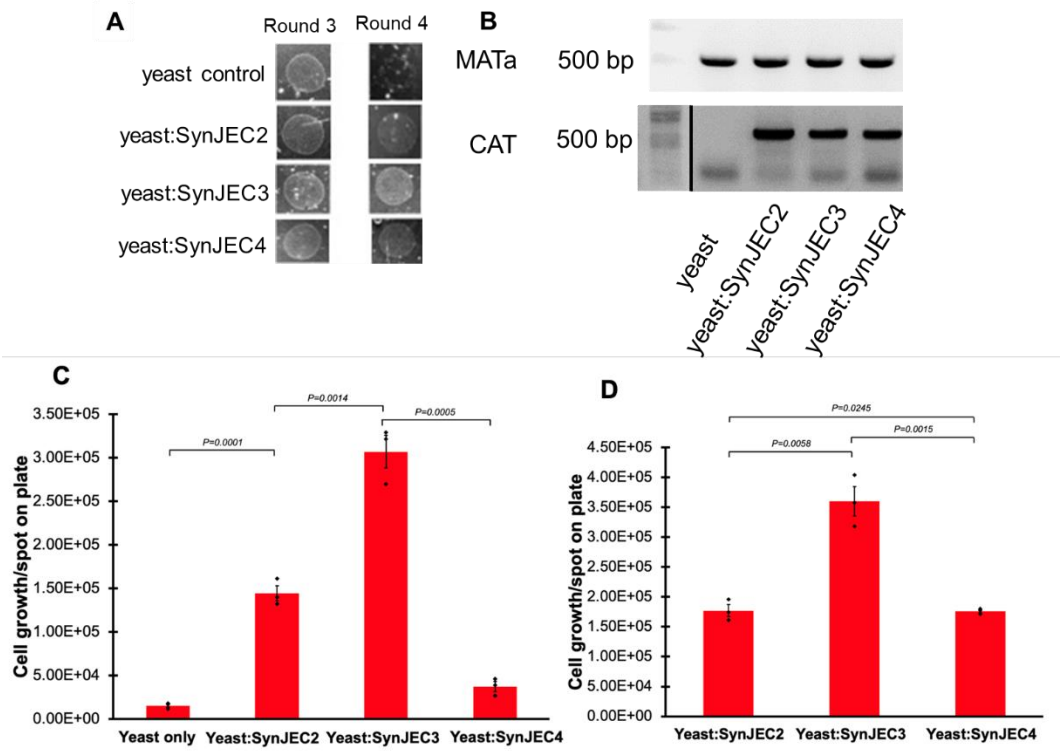
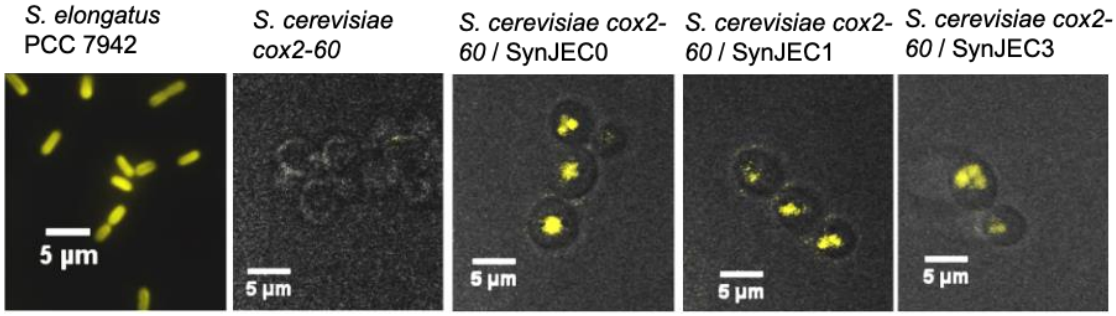
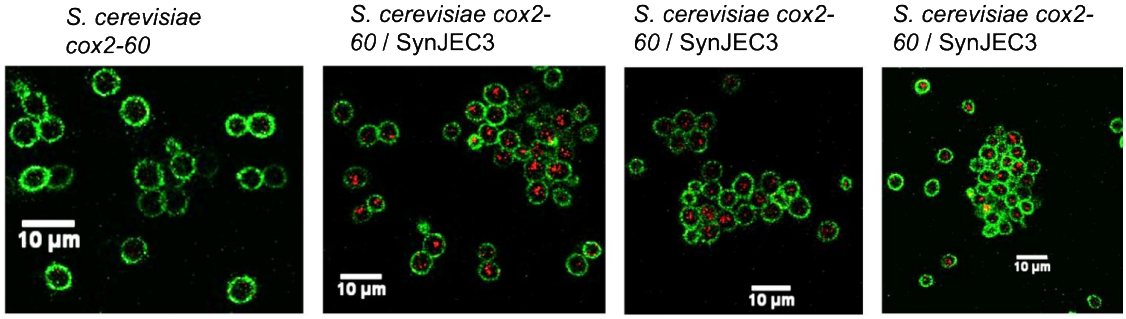



**Figure 2.3**: **Rescue of respiration deficient phenotype by yeast-*SynJEC* chimeras expressing an ATP/ADP translocase and SNARE-like proteins: (A)** Growth of *S. cerevisiae* cox2-60–*SynJEC2—4* chimeras on medium containing glycerol as the sole carbon source. The experiment was repeated three times independently with similar results. **(B)** Total DNA of yeast-*SynJEC* chimeras contains yeast *MAT**a*** and *SynJEC CAT* genes. The experiment was repeated three times independently with similar results. **(C-D)** Growth trends of *S. cerevisiae* cox2-60 (yeast only strain), *S. cerevisiae* cox2-60- SynJEC2 (yeast-SynJEC2), *S. cerevisiae* cox2-60- SynJEC3 (yeast-SynJEC3) and *S. cerevisiae* cox2-60- SynJEC4 (yeast-SynJEC4) chimeras on Selection Medium III. Cells ($3.00\times10^3$) were spotted on Selection Medium III the final number of cells/spot on plate were determined after 72 h of growth (N=3 technical replicates; data are presented as mean values +/- SEM). *P*-values were calculated by two-tailed *t*-test comparing the two means.

**(A) pTIRF**




**(B) Fluorescence confocal microscopy**




**(C) Transmission Electron Microscopy**

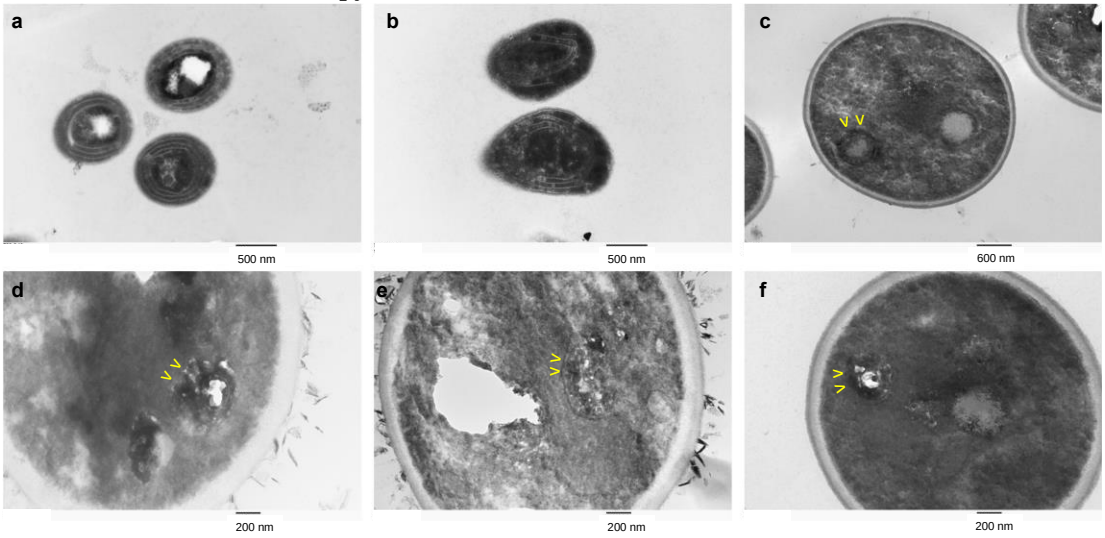



**Figure 2.4: Imaging intracellular endosymbiont *Synechococcus* by fluorescent microscopy**. **(A)** pTIRF microscopic images of *Synechococcus* cells, control yeast cells, and chimeric cells that were grown under selection conditions (Ex. = 561 nm; Em. = 653/95 nm). Panels are merged images of pTIRF (yellow) and brightfield microscopy (gray). The experiment was repeated three times independently with similar results. **(B)** Fluorescence confocal microscopy images of control yeast cells and chimeric cells, which were grown under selection conditions. The yeast cell wall was stained with Con A-FITC (green, Ex. = 488 nm; Em. = 510/20 nm) and the presence of cyanobacteria was monitored by cyanobacterial fluorescence (red, Ex. = 561 nm; Em. = 650/20 nm). Based on these images, it is possible that multiple cyanobacterial cells could be present in some of the yeast cells. The experiment was repeated three times independently with similar results. **(C)** Panel **a** and **b** are cyanobacterial samples imaged by Transmission Electron Microscopy (TEM) and Panels **c-f** are fusions imaged by TEM. Yellow arrows show characteristic cyanobacterial structures within the cytoplasm of the yeast cells. The experiment was repeated twice independently with similar results.

**Imaging cyanobacterial endosymbionts within yeast cells by TIRF microscopy:**

The phenotypic rescue of *S. cerevisiae cox2-60* and PCR analysis of total genomic DNA from our fusions suggested the presence of cyanobacterial endosymbionts present within yeast cells. To obtain more evidence supporting this possibility, we used fluorescence microscopy approaches detect the presence of cyanobacterial endosymbionts within yeast cells. Because Syn7942 cells possess natural fluorescence properties due to the presence of phycobilins and chlorophyll, we characterized our endosymbiotic chimeras using a home-built total internal reflection fluorescence (TIRF) microscope. In order to enable deep imaging inside live cells, the microscope was used in pTIRF mode, where the incident angle of the laser beam is tuned to be just below the critical angle at the top surface of coverslip, so the laser can travel through yeast cell wall and illuminate above the penetration depth of the excitation laser evanescent field. The microscope also has a Zeiss condenser for brightfield microscopy with white light illumination. In the experiment, yeast cells were first found and imaged in brightfield mode and then pTIRF mode was used to check and detect cyanobacterial autofluorescence. First, we imaged live Syn7942 cells with the TIRF microscope, and as expected, we detected the natural cyanobacterial fluorescence corresponding to the emission (661 nm) of phycobilins and chlorophyll [143,144] in cyanobacteria using 561 nm laser excitation (Figure 4A). Under the same conditions, no fluorescent signals were detected for the control *S. cerevisiae cox2-60* host cells (Figure 4A). This observation suggested that we should be able to detect the presence of Syn7942 derived cells within *S. cerevisiae cox2-60* with the TIRF microscopy without need for the expression of heterologous fluorescent entities. During each round of re-plating under selection conditions, we used the TIRF microscopy to image live chimeras. As shown in Figure 4A we observed prominent punctate signals corresponding to cyanobacteria in case of all the fusions that demonstrated phenotypic rescue of the host yeast cells, i.e., *S. cerevisiae cox2-60*-SynJEC0, *S. cerevisiae cox2-60*-SynJEC1 and *S. cerevisiae cox2-60*-SynJEC3 chimeras.

**Imaging cyanobacterial endosymbionts by confocal fluorescence microscopy:**

To obtain a better understanding of the fraction of yeast cells containing cyanobacterial endosymbionts, we next used confocal fluorescence microcopy approaches which allows us to precisely measure endosymbiont distribution on fixed cells. We first propagated the *S. cerevisiae cox2-60*-SynJEC3 chimeras for one round of growth on selection medium II and one round of growth on selection medium III. Following this selection, we confirmed the detection of cyanobacterial endosymbionts within yeast cells by PCR analysis of the total genomic DNA and pTIRF microscopy. Following this confirmation, the *S. cerevisiae cox2-60*-SynJEC3 cells were fixed with Karnovsky fixative [145], and stained with FITC-labeled concanavalin A (Con A-FITC). Similarly, the control yeast cells, *S. cerevisiae cox2-60*, were also fixed and stained with Con A-FITC. Samples were then analyzed by using a commercial Leica SP8 fluorescence confocal microscope. In the microscope, a 488 nm laser was used to excite Con A-FITC stained yeast while a 561 nm laser for naturally fluorescent cyanobacterial cells. The spectral detector in the microscope allows up to 3 spectral detection bands, each of which is independent from the others and continuously tunable from 400 nm to 800 nm. Two detection bands with an identical bandwidth of 20 nm were set at 510 nm and 650 nm respectively for the green fluorescent light from the Con A-FITC dyes on yeast cells and the red autofluorescence from cyanobacterial cells. Both yeast cells and cyanobacterial cells could therefore be spectrally separated and imaged simultaneously (see Figure 2.8). As shown in Figure 4B, we detected the presence of cyanobacterial endosymbionts within the host yeast

cells in confocal fluorescence microscopy further confirming the generation, selection and propagation of cyanobacterial endosymbionts within yeast cells after multiple rounds of growth under stringent selection conditions. In order to analyze the distribution of the cyanobacterial endosymbionts within yeast cells, we scanned a large number of cells by confocal microscopy. As shown in Figure 2.9, under the fusion, selection and detection conditions a significantly high fraction of yeast cells show signals specific to cyanobacteria in the early stages of chimera propagation. These observations were again fundamentally different from our previous investigations with the yeast/*E. coli* endosymbiotic platform [118].

**Imaging cyanobacterial endosymbionts by transmission electron microscopy:**

To further demonstrate the presence of intact cyanobacterial endosymbionts and to understand the localization of endosymbionts within host cells use analyzed the yeast/cyanobacteria chimera using Transmission Electron Microscopy (TEM). We first analyze SynJEC3 strains by TEM; characteristic cyanobacteria TEM images are shown in Figure 4C, panels a-b. Next, we generated and propagated *S. cerevisiae cox2-60*-SynJEC3 chimeras for one round of growth on selection medium II and one round of growth on selection medium III (at least 10 doubling detected by yeast cell count analysis). These samples were then fixed, treated and analyzed by TEM as described in the methods section. As shown in Figure 4C, panels c-f, we observe characteristic structures corresponding to cyanobacteria within the yeast cells. In addition to this, the cyanobacteria appear to be structurally intact and do not appear to be entrapped in any intracellular organelles within yeast cells.

**Metabolic coupling of the yeast/cyanobacterial endosymbiotic system:**

The endosymbiotic theory suggests that the host and the endosymbiont evolved metabolic interdependency which was crucial for retaining and optimizing the presence of endosymbionts within the host cells. These features are central to present-day organelle function; for example, organelle membranes possess specific or non-specific transporters that facilitate the uptake of necessary nucleotides, amino acids or amino acid precursors from the host cytosol [146]. Such uptake mechanisms are crucial for organelle genome replication and organelle protein synthesis. Therefore, as a next step, we generated an endosymbiotic system between *S. cerevisiae cox2-60* and *S. elongatus* where the *S. cerevisiae cox2-60* cells depended on *S. elongatus* endosymbionts for ATP and the *S. elongatus* endosymbionts depended on *S. cerevisiae cox2-60* for an essential amino acid.

Analysis of the Syn7942 genome suggests that this strain has several putative amino acid and cofactor transporters. Gene deletion of a key biosynthetic gene usually results in the generation of a corresponding auxotrophic strain (e.g., gene deletion of *metA* results in a methionine auxotroph). As a first step towards creating metabolite interdependency, we generated a SynJEC3 methionine auxotrophic mutant to test if it could establish an endosymbiotic relationship with *S. cerevisiae cox2-60*. To this end, we constructed another suicide plasmid pML58 containing homology regions corresponding to Syn7942 *metA* gene and a kanamycin resistance marker. We transformed pML58 into SynJEC3 and selected for cyanobacterial mutants that were able to grow in BG-11 medium in the presence of chloramphenicol, kanamycin and methionine. Individual colonies were grown in liquid selection medium and genomes were isolated and characterized by PCR analysis to confirm recombination at the *metA* locus. We confirmed that characterized colonies were methionine auxotrophs by growing them in the presence and absence of methionine (Figure 5A, Figure 2.10A); the corresponding strain was designated as SynJEC5. Next, SynJEC5 was

fused to *S. cerevisiae cox2-60* spheroplasts. Colonies were observed on selection medium II and were able to propagate in selection medium III for four rounds of re-plating (Figure 5B, Figure 2.10D-E). As before, total genomic DNA was isolated and PCR analysis was performed to confirm the presence of cyanobacterial endosymbionts (Figure 5C, Figure 2.10B). pTIRF microscopy further supported the presence of cyanobacterial endosymbionts in yeast cells (Figure S5C).

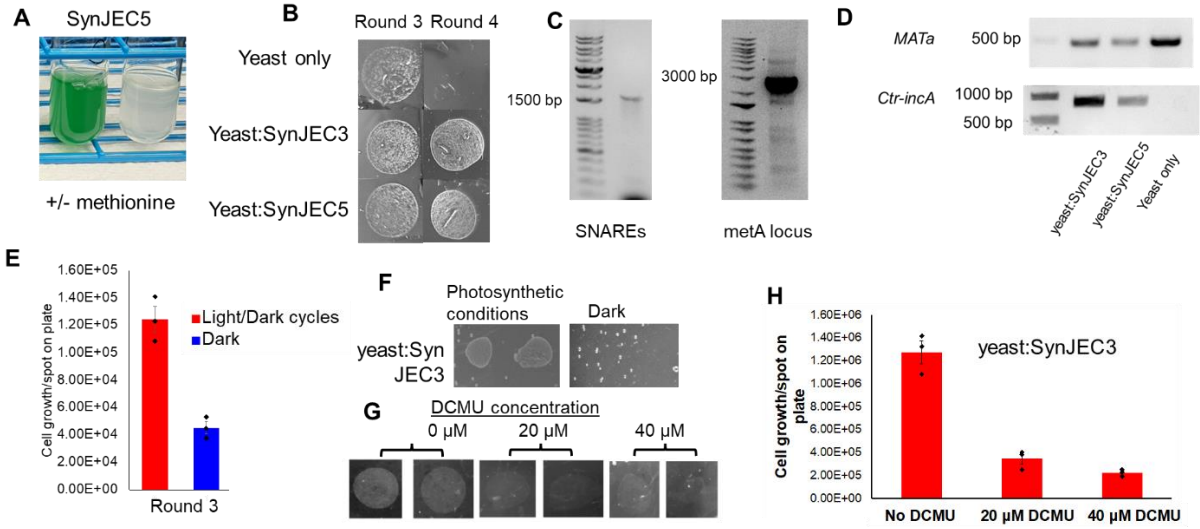



**Figure 2.5: Rescue of respiration-deficient phenotype by yeast-SynJEC chimeras containing auxotrophic SynJEC5 cells (A)** Growth of *SynJEC5* cells in BG-11 medium in the presence and absence of L-methionine **(B)** Growth of yeast-*SynJEC3* and yeast-*SynJEC5* chimeras on Selection Medium III. The experiment was repeated three times independently with similar results. **(C)** Total DNA analysis of *SynJEC5* DNA reveals the presence of genomic SNARE-like proteins *Ctr-incA* and *CT813* (left) and disruption of the *metA* gene with a kanamycin resistance marker (right). The experiment was repeated twice independently with similar results. **(D)** Total DNA of yeast-SynJEC chimeras contains yeast *MAT**a*** and *SynJEC Ctr-incA* genes. The experiment was repeated three times independently with similar results. **(E)** *S. cerevisiae cox2-60*-SynJEC3 chimeras are generated under and selected under photosynthetic conditions. After two rounds of growth *S. cerevisiae cox2-60*-SynJEC3 chimeras are propagated either under standard photosynthetic conditions or under dark; initial number of cells plated per spot is $10^4$ (N=3 technical replicates; data are presented as mean values +/- SEM). *P*-values were calculated by two-tailed *t*-test comparing the two means (95% CI, Cohen's *d*=6.2, P=0.0016). **(F)** SynJEC3 are starved for light for 72 h and the fused to *S. cerevisiae cox2-60* and selected under dark conditions (dark fusions). The *S. cerevisiae cox2-60*-SynJEC3 chimeras generated from dark fusions fail to grow whereas the control *S. cerevisiae cox2-60*-SynJEC3 chimeras propagated under photosynthetic conditions propagate as previously observed. The experiment was repeated twice independently with similar results. **(G)** Plate images to demonstrate the effect of (3-(3,4-dichlorophenyl)-1,1-dimethylurea), DCMU, on *S. cerevisiae cox2-60*-SynJEC3 chimeras. The experiment was repeated twice independently with similar results. **(H)** Cell count analysis to demonstrate the effect of (3-(3,4-dichlorophenyl)-1,1-dimethylurea), DCMU, on *S. cerevisiae cox2-60*-SynJEC3 chimeras (N=3 technical replicates; data are presented as mean values +/- SEM). *P*-values were calculated by two-tailed *t*-test comparing the two means.

**Growth dependence of the yeast/cyanobacteria chimeras on photosynthesis:**

Because light driven photosynthesis in *S. elongatus* is expected to be crucial towards ATP synthesis and therefore endosymbiosis in our platform, we investigated the importance of photosynthesis to the survival and propagation of yeast/cyanobacteria chimeras. To this end, we compared the growth of chimeras in a day/night cycle to those grown in the absence of light. As shown in Figure 5E, we saw significantly retarded growth for *S. cerevisiae cox2-60*-SynJEC3 chimeras when propagated under non-photosynthetic conditions as compared to the photosynthetic

conditions; particularly, we just observed two doublings for *S. cerevisiae cox2-60*-SynJEC3 chimeras under non-photosynthetic conditions (identical to *S. cerevisiae cox2-60* alone under identical selection conditions). Because *S. elongatus* strains have an extensive metabolic adaptation and stress response to adapt to dark environment, we also wanted to evaluate the effect of light starvation on cyanobacterial endosymbionts prior to fusion to the yeast cells. When light is no longer available, *S. elongatus* cells catabolize their glycogen stores through respiration and can remain viable for significant periods.[147,148] Therefore, in order to test the effect of light starvation prior to fusion on yeast/cyanobacteria endosymbiosis, we incubated SynJEC3 strains in darkness for 96 h prior to fusion with *S. cerevisiae cox2-60* spheroplasts. The recovered fusions were kept under strict darkness for during all rounds of selection and propagation. In all instances, we observed no growth for *S. cerevisiae cox2-60*-SynJEC3 chimeras generated from this experiment but observed normal growth when chimeras were generated under photosynthetic conditions described earlier (Figure 5F). This suggested that light is essential for the establishment and propagation of cyanobacterial endosymbionts within the yeast cells. We also tracked the propagation of *S. cerevisiae cox2-60*-SynJEC3 chimeras by pTIRF microscopy to determine the relative changes in the levels of photosynthetic apparatus, particularly phycobilisomes (PBS) and chlorophyll *a* (Chl). As indicated by the Figure 2.14, we observe alteration of internal ratio of PBS:Chl indicating possible changes to total composition of the photosynthetic apparatus in the endosymbiont over multiple rounds of growth.

Next, we evaluated the effects of photosynthesis inhibitor, 3-(3,4-dichlorophenyl)-1,1-dimethylurea (DCMU) on the viability of yeast/cyanobacteria chimera. We first demonstrated that DCMU completely arrested the growth of cyanobacteria but had no affect the viability of *S. cerevisiae cox2-60* (Figure 2.15). On the other hand, as expected, the *S. elongatus* strains completely lose their viability on DCMU treatment.[149] When *S. cerevisiae cox2-60/* SynJEC3 were treated with DCMU, we also observed loss of viability (Figure 5G-H) indicating again, the essential role of endosymbiont mediated photosynthesis on the viability of the yeast/cyanobacteria chimeras.

**Discussion**

*Premise for our studies*: In this study, we sought to engineer genetically tractable platforms where the endosymbiotic bacteria perform chloroplast-like functions for the host cell. To build such a system, we were inspired by the evolutionary origin of photosynthetic eukaryotic cells. The geological record places the advent of the earliest eukaryotes about 2.7—1.7 billion years ago (Gya) [150,151], coinciding with an explosion of cyanobacterial-derived oxygen ($O_2$) levels in the biosphere of the Earth known as the Great Oxidation Event (GOE) [152,153]. This fundamental shift from a reducing to oxidizing atmosphere placed pressure on cells to withstand oxidative damage; hence, many processes necessary for eukaryotic and multicellular life such as DNA repair [154] and meiotic sex [155,156] are thought to have originated as defense mechanisms against reactive oxygen species. The harsh conditions of GOE-era Earth are also thought to be connected to the evolution of bacterial endosymbiont-derived organelles such as mitochondria and chloroplasts.

Though primary photosynthetic endosymbiosis was once thought to be an unprecedented event which only occurred 2.7—1.7 Gya, phylogenetic analysis has revealed that the chromatophore of the amoeboid *Paulinella*, which contains cyanobacterial endosymbionts, diverged from its relatives about 140 million years ago through a separate primary endosymbiotic event [157–159]. Comparative genomic studies have provided significant insights into the genome

evolution of the chromatophore of *Paulinella*.[160,161] For modern-day chloroplasts, studies suggest that *Gloeomargarita* are the closest identified relatives of modern-day chloroplasts [120,121]. While strains of *Gloeomargarita* and *Paulinella* are not readily genetically tractable, genetic tools have been developed to manipulate several strains of *Synechococcus* [122,123], which are relatives of *Gloeomargarita* [158,159]. These observations were central to our choice of engineering Syn7942 endosymbionts as a step towards investigating the endosymbiotic transformation of cyanobacteria into modern-day chloroplasts. As is evidenced from our observations described in this manuscript, this choice turned out to be an important consideration due to several innate abilities of Syn7942 that facilitated endosymbiosis. Conversely, when it came to deciding an appropriate host strain for endosymbiosis, our choice was determined mostly by genetic tractability, because it has been suggested that photosynthetic endosymbionts were established in a variety of eukaryotic host cells during the chloroplast evolution [109,125]. Therefore, we decided to simply use a model eukaryotic cell, *S. cerevisiae*.

*Designing and engineering artificial endosymbiosis*: Though the modern day functions of chloroplasts are predominantly carbon assimilation, nitrate assimilation, sulfate assimilation, amino acid biosynthesis, fatty acid biosynthesis, it is widely suggested that bioenergetic considerations may have been the key drivers of organelle evolution.[127,126] Due to this wide range of chloroplast functions, it is still unclear what the exact drivers of endosymbiosis were. Through our investigations, we were able to engineer a yeast/cyanobacteria endosymbiotic platform where endosymbiotic cyanobacteria perform chloroplast-like functions for the host cells. Particularly, we demonstrated that SynJEC0 mutants lacking expression of any recombinant ADP/ATP translocase were able to transiently support the phenotypic rescue of *S. cerevisiae cox2-60* strains that were dependent on cyanobacterial endosymbionts for ATP. This set of experiments suggested that cyanobacterial relatives of chloroplast precursors already had an innate ability to establish bioenergetically relevant endosymbiosis with yeast mutants underdefined selection conditions. This hypothesis is supported by the presence of ATP/ADP translocase-like proteins or NTTs in Syn7942. Based on bioinformatics studies, proteins with homology to NTTs have been annotated as PBS lyase proteins in the cyanobacterial genome (Major et al., 2017). Our yeast/cyanobacteria endosymbiosis platform will facilitate systematic analysis of the importance of such cyanobacterial proteins to endosymbiosis. Although cyanobacteria are predicted to contain nucleotide transport fusion proteins (NTT, similar to ADP/ATP translocases), studies suggest that chloroplast NTTs were possibly derived from intracellular organisms like chlamydial[128,129]. We investigated this possibility by generating SynJEC1 strains that expressed recombinant *ntt1*, ADP/ATP translocase gene from an intracellular organism, *Protochlamydia amoebophila* UWE25, and observed that SynJEC1 endosymbionts were able to phenotypically rescue the growth of *S. cerevisiae cox2-60* strains that were respiration-deficient, and the chimeras had faster apparent growth rate and better stability. It is also suggested that precursors of organelles could have been parasitic in nature[162,163], and they may have genetic factors that allow the endosymbiont to establish a replicative niche within the host. To investigate this hypothesis, we generated a series of cyanobacterial mutants (SynJEC2, SynJEC3, SynJEC4 and SynJEC5) that recombinantly expressed various SNARE-like proteins from pathogenic intracellular bacteria along with ADP/ATP translocases and identified optimal cyanobacterial mutants that resulted in fastest apparent growth rate and stability. Under optimal conditions, the *S. cerevisiae*/Syn7942 chimera propagate through at least 15 to 20 generations of growth under defined photosynthetic growth conditions.

Though the SNARE-like proteins are not necessary for yeast/cyanobacteria chimeras, the exact mechanism of how the expression of SNARE-like proteins provides improved stability of synthetic endosymbiosis is still unclear and is an active area of investigation.[118,119] It is possible that synthetic endosymbionts expressing SNAREs utilize alternate mechanisms that pathogens which result in enhanced stability of the host/endosymbiont chimera. We were also able to characterize the *S. cerevisiae*/Syn7942 chimeras through total internal reflection fluorescence microscopy, fluorescence confocal microscopy and TEM. Our data suggested that the cyanobacteria were structurally intact and did not appear to be entrapped in any intracellular organelles within yeast cells. We also developed a platform that created a metabolite interdependency between the host and the cyanobacterial endosymbiont, where the engineered cyanobacterial endosymbiont provided ATP to the host yeast cells and the yeast cells provided methionine to the cyanobacterial endosymbiont. Such platforms will enable metabolic coupling of the endosymbiont to the host, in a manner similar to metabolic coupling between the chloroplast and its photosynthetic eukaryotic cells. Our subsequent studies will also use similar approaches to recapitulate genome minimization events that could have occurred during chloroplast evolution.

*Evolutionary implications*: Given that *Synechococcus* are close relatives of chloroplast precursors, we believe that our observations may have significant implications on the early stages of evolution of photosynthetic eukaryotic life forms. Such photosynthetic endosymbiotic systems could provide a platform to recapitulate various evolutionary trajectories related to the conversion of photosynthetic endosymbionts into photosynthetic organelles (i.e., chloroplasts). Here it is important to note that there is robust evidence to suggest that eukaryotic host already had a mitochondrion to provide ATP when cyanobacterial endosymbionts were established. However, it is possible that cyanobacteria could have provided added advantage due to its light driven synthesis of ATP (photophosphorylation) unlike the mitochondria which performs oxidative phosphorylation. As precedented by the present-day synergistic interactions between chloroplast and mitochondria[164], it is even possible that cyanobacterial photosynthesis derived ATP and oxygen synthesis could have complemented mitochondrial oxidative phosphorylation process that synthesis ATP. This endosymbiotic platform also has the potential to be further metabolically manipulated, analytically studied and imaged, and computationally modeled and predicted. Therefore, these platforms could facilitate laboratory evolution of artificial photosynthetic eukaryotic life forms by using a combination of rational engineering and laboratory evolution approaches; for example, this platform can be used to study: (i) minimization of the endosymbiont genome, (ii) transfer of genes from the endosymbiont genome to the host genome, (iii) developing strategies to facilitate protein exchange between the endosymbiont and host, (iv) mutation-based evolution and selection to identify and characterize metabolic adaptations that could have occurred during organelle evolution, (v) the role of light in establishing photosynthetic endosymbionts and (vi) investigating if multiple endosymbionts can be established in host yeast cells to mimic secondary endosymbiosis[125] that is proposed to have occurred during chloroplast evolution, amongst others. Such efforts to recreate organelle evolution in lab will likely provide molecular-level understanding of how bacterial endosymbionts were transformed into chloroplasts, thereby resulting in the origin of photosynthetic eukaryotic life. However, our current approach does not investigate the mechanisms that resulted in the internalization of cyanobacterial endosymbionts into eukaryotic cells and the establishment of these endosymbionts in the cytoplasm.

*Synthetic Biology implications*: In addition to the evolutionary insights, this genetically tractable photosynthetic endosymbiotic platform can be repurposed for various synthetic biology applications as well; for example, such artificial, genetically tractable photosynthetic platforms can be used for biosynthesis and bioproduction applications. These studies could also provide a genetically tractable experimental platform for synthetic biology efforts on designing minimal bacterial genomes to attain "minimal organism" to understand minimum requirements necessary to support photosynthetic life in privileged environments of the host cytosol [165]. In summary, our studies will provide a roadmap to use first principles of endosymbiotic theory to convert non-photosynthetic organisms into new photosynthetic life-forms.

**Materials and methods**

**Strains**

*Synechococcus elongatus* strains were derived from *S. elongatus* PCC 7942 (Syn7942). This strain was obtained from Prof. Susan Golden's lab (University of California San Diego, UCSD). We used *S. cerevisiae* ρ+ NB97 (*MAT***a** *leu2-3,112 lys2 ura3-52 his3ΔHindIII arg8Δ::URA3 [cox2-60]* as a host for *S. elongatus* endosymbionts. The *S. cerevisiae-cox2-60* strain was obtained from the Schultz lab (Scripps Research).

**Growth media**

*S. elongatus* cells were grown in sterile Erlenmeyer flasks containing liquid BG-11 medium. These cultures were shaken aerobically at 37 ºC and 250 rpm under 3000 lux. Yeast cells were shaken aerobically at 30 ºC and 250 rpm in YPD medium (1% yeast extract, 2% peptone, 2 % glucose) containing 50 mg/L carbenicillin.

List of all the fusion selection media:

Selection medium I: 1% yeast extract, 2 % peptone, 3 % glycerol, 0.1 % glucose, 1 M sorbitol, 2 % agar, 1X BG-11 salts

Selection medium II: 1% yeast extract, 2 % peptone, 3 % glycerol, 0.1 % glucose, 1 M sorbitol, 2 % agar, 1X BG-11 salts, 50 mg/ml carbenicillin

Selection medium III: 1% yeast extract, 2 % peptone, 3 % glycerol, 1 M sorbitol, 2 % agar, 1X BG-11 salts, 50 mg/ml carbenicillin

NSII-recombinant *S. elongatus* mutant cultures were supplemented with 7.5 mg/L chloramphenicol. Unless noted otherwise, kanamycin-resistant *S. elongatus* mutant cultures were supplemented with 50 mg/L kanamycin and 2 mg/L L-methionine.

When noted, light-starved Syn7942 cells were inoculated ($OD_{730}$ = 0.2) in BG-11 medium, wrapped in aluminum foil and incubated for 4 d, without shaking, at 30 °C. When noted, yeast media was supplemented with adenine sulfate (Alfa Aesar A16964-09, 20 mg/L), Antimycin A (Sigma A8674-25MG), or DCMU (A2B Chem AG00409). Yeast cells transformed with pML64 were selected on synthetic defined (SD) medium containing 0.67 % yeast nitrogen base without amino acids (Sigma Y0626-250G), L-lysine (Sigma L5501-5G, 60 mg/L), L-histidine (Sigma H8000-10G, 20 mg/L), L-arginine (Sigma A-3704, 20 mg/L), uracil (Sigma U1128-25G) and 3 % glycerol.

**Construction of plasmids**

Double-stranded and single-stranded DNA oligonucleotide fragments were purchased from Integrated DNA Technologies (IDT). Defined DNA fragments were amplified using PCR (Q5 Hot Start High-Fidelity 2X Master mix,

NEB catalog # M0494S) and inserted into defined sites in the host vectors using Gibson assembly [166]. Double-stranded oligonucleotide sequences (gblocks) used for Gibson assembly are listed in Table 2.2. Genomic DNA fragments used in cloning are listed in Supplementary Table 2.3. Cyanovectors were obtained from Prof. Susan Golden's lab (UCSD). Where noted, coding sequences were codon-optimized for *S. elongatus* expression using IDT codon optimization software (https://www.idtdna.com/CodonOpt). All vectors were transformed into One Shot® *ccd*B Survival™ 2 T1$^{R}$ Chemically Competent Cells (Invitrogen A10460) according to manufacturer's specifications. The oligonucleotides used in plasmid construction are listed in Table 2.4. The plasmids pML3-pML28 are derived from the Cyanovector pCV0055, pML58 and pML62 are derived from the CyanoVector pCV0049. Vector maps are included in Figure 2.6, and detailed vector map links are provided in Table 2.6.

pML3: pCV0055[123] was linearized by PCR using the oligonucleotides AM1216/AM1217. A gBlock of the *UWE25-ntt1* gene codon optimized for *S. elongatus* was amplified by PCR using the oligonucleotides AM1214/AM1215. The amplified DNA fragment was inserted into linearized pCV0055 by Gibson assembly to afford pML3.

pML14: pML3 was linearized by PCR using the oligonucleotides AM1195/AM1312. A gBlock of the *C. trachomatis incA* gene codon-optimized for *S. elongatus* (*Ctr-incA*) was amplified by PCR using the oligonucleotides AM1313/AM1314 and inserted into linearized pML3 by Gibson assembly to afford pML14.

pML17: pML14 was linearized by PCR using the oligonucleotides AM1312/AM1351. A gBlock of the *CT_813_CDO-1* gene was amplified by PCR using the oligonucleotides AM1352/AM1353 and inserted into linearized pML14 by Gibson assembly to afford pML17.

pML28: pML17 was linearized by PCR using the oligonucleotides AM1506/AM1507. A variant of the *C. caviae incA* gene codon-optimized for *S. elongatus* (*Cca-incA*) was amplified by PCR using the oligonucleotides AM1502/AM1503, purified and amplified once more using the oligonucleotides AM1504/AM1505. The amplicon was inserted into linearized pML17 by Gibson assembly to afford pML28.

pML62: pCV0049 was linearized by PCR using the oligonucleotides JC205/JC206. A 484-bp fragment of the 5'-end of the Syn7942 *metA* gene (*Synpcc7942_0370*) was amplified from purified Syn7942 genomic DNA using the oligonucleotides JC204/JC207. The amplicon was inserted into linearized pCV0049 by Gibson assembly and the construct was subsequently linearized using the oligonucleotides JC200/JC203. A 530-bp fragment of the 3'-end of *Syncpcc7942_0370* was amplified from genomic DNA using the oligonucleotides JC201/JC202 and the amplicon was inserted into the linearized vector by Gibson assembly to afford pML62.

pML58: pML62 was linearized by PCR using the oligonucleotides AM1531/AM1532. The PCR product was purified and treated with T4 Polynucleotide Kinase (NEB #M0201S), T4 Ligase (NEB #MO202S) and DpnI (NEB #R0176S) for 1 h at room temperature to produce the ligated plasmid. The plasmid was then linearized by PCR using the oligonucleotides JC151/JC152. An origin of transfer (oriT) domain was amplified by PCR from the commercial plasmid pSET152 using the oligonucleotides JC153/JC154. The amplicon was inserted into the linearized plasmid by Gibson assembly to afford pML58.

pML63 is derived from the commercial plasmid pRS425, which was donated by the van der Donk lab (UIUC). The 2×Su9-MTS-LP-*COX2*-W56R expression cassette was first reported by Supekova and Schultz (Supekova et al, 2010).

pML59: pML14 was linearized by PCR using the oligonucleotides JC221/JC233. The linear PCR product was purified, amplified with JC221/JC226 and ligated by KLD reaction to afford the FLAG-tagged *Ctr-incA* construct pML59.

pML60: pML17 was linearized by PCR using the oligonucleotides AM1312/JC235 and ligated by KLD reaction. The construct was then linearized by PCR using the oligonucleotides JC223/JC233. The linear PCR product was purified, amplified with JC223/JC226 and ligated by KLD reaction to afford the FLAG-tagged *CT_813* construct pML60.

pML63: pRS425 was linearized by PCR using the oligonucleotides JC249/JC248. The *S. cerevisiae ACT1* promoter was amplified from purified *S. cerevisiae* genomic DNA by PCR using the oligonucleotides JC246/JC247. The amplified DNA fragment was inserted into linearized pRS425 by Gibson assembly and the construct was subsequently amplified by PCR using the oligonucleotides JC250/JC251. A gBlock of 2×Su9-MTS-LP-*COX2*-W56R was amplified by PCR using the oligonucleotides JC254/JC255 and the amplified DNA fragment was inserted into the linearized plasmid by Gibson assembly to afford pML63.

pML64: pML63 was amplified by PCR using the oligonucleotides JC128/JC273. The *S. cerevisae TPI1* promoter was amplified from purified *S. cerevisiae* genomic DNA by PCR using the oligonucleotides JC272/JC127. The amplified DNA fragment was inserted into linearized pML63 by Gibson assembly to afford pML64.

**Site-directed mutagenesis in cyanobacteria**

Chromosomal integration of genes in Syn7942 was achieved using a modification of the method of Golden [167]. A portion of Syn7942 overnight culture (15 mL) was centrifuged for 10 min at 3,000×g and 24 ºC. The pellet was washed once with 10 mL NaCl (10 mM) and then resuspended in 0.3 mL BG-11 at room temperature. To this suspension was added purified plasmid (1.5 μL) containing genes flanked by NSII recombination sites. The mixture was added to a 1.5 mL microcentrifuge tube and shaken in the dark (12-16 h at 70 rpm). Cells transformed with NSII plasmids were added to glass test tubes containing liquid BG-11 (5 mL, room temperature) and antibiotic and cultured under normal conditions for 4 d, at which point surviving cells were added to Erlenmeyer flasks and grown under normal conditions. Cells transformed with pML58 were not viable when rescued in liquid BG-11 medium; instead, the cell-plasmid mixture was spread on BG-11-agar medium supplemented with kanamycin and L-methionine and incubated at 30 ºC under 3000 lux for ≤10 days. The quality of *S. elongatus* cultures was evaluated regularly by microscopy, streaking the cells onto BG-11-agar and by PCR analysis of recombinant loci.

SynJEC0 was generated by transformation of wild-type Syn7942 with pCV0055 to give a chloramphenicol-resistant recombinant mutant.

SynJEC1 was generated by transformation of wild-type Syn7942 with pML3 to give a recombinant mutant which ectopically expresses an ADP/ATP translocase from the NSII locus.

SynJEC2 was generated by transformation of wild-type Syn7942 with pML14 to give a recombinant mutant which ectopically expresses *Ctr-incA* and an ADP/ATP translocase from the NSII locus.

SynJEC3 was generated by transformation of wild-type Syn7942 with pML17 to give a recombinant mutant which ectopically expresses *Ctr-incA*, *CT813* and an ADP/ATP translocase from the NSII locus.

SynJEC4 was generated by transformation of wild-type Syn7942 with pML28 to give a recombinant mutant which ectopically expresses *Ctr-incA*, *CT813*, *Cca-incA* and an ADP/ATP translocase from the NSII locus.

SynJEC5 was generated by transformation of SynJEC3 cells with pML58. The cells were plated on BG-11-agar medium supplemented with chloramphenicol (5 mg/L), kanamycin (5 mg/L) and L-methionine (2 mg/L), with individual colonies appearing within 10 days. Colonies were extracted and spotted on fresh BG-11 agar supplemented with the same components. After 7 days, the resulting dark green spots were added to liquid BG-11 (1 mL) and incubated under normal culturing conditions for Syn7942. The cells were passaged every 4 days into fresh BG-11 medium (5 mL) with increasing concentrations of kanamycin (5, 25 and finally 50 mg/L) in order to eliminate all chromosomal copies with intact *Synpcc7942_0370*. During each passage, the *Synpcc7942_0370* locus was amplified by PCR using the oligonucleotides LL56/LL57. SynJEC5 cultures were considered "pure" after 5 rounds of passaging, at which point only recombinant amplicons were produced by PCR. Methionine auxotrophy of SynJEC5 was evaluated by washing the cells in BG-11 medium and inoculating cultures containing chloramphenicol and kanamycin in the presence and absence of L-methionine (Supplementary Figure 5a).

SynJEC9 was generated in a similar manner as SynJEC5 by transformation of SynJEC1 cells with pML58 instead of pML62. As with SynJEC5, the cells were plated on BG-11-agar medium supplemented with chloramphenicol (5 mg/L), kanamycin (5 mg/L) and L-methionine (2 mg/L). Individual colonies appeared within five days, which were then extracted and spotted on fresh BG-11 agar supplemented with kanamycin (5 mg/L) and L-methionine (2 mg/L). Afterwards, the cells were passaged through multiple rounds of growth in liquid BG-11 medium in the same manner as SynJEC5 until PCR analysis of genomic DNA with the oligonucleotides LL56/LL57 produced only a recombinant band at the *Syncpcc7942_0370* locus (Supplementary Figure 5a).

**Measurement of ATP release by cyanobacteria mutants expressing ATP/ADP translocase after ADP challenge**

In order to eliminate contaminant ATP, ADP solution (Sigma A2754) was treated with hexokinase according to a reported protocol [168]. ADP (80 mM, pH 7.5) was incubated with D-glucose (200 mM), $MgCl_2$ (2 mM) and hexokinase (Sigma H4502-500UN) (0.04 U/µL) at room temperature for 2 h. The solution was then filtered through an Amicon Ultra 0.5 column (14,000 ×g, 15 min) to eliminate hexokinase.

SynJEC0 and SynJEC1 cells were grown for 3 d to reach densities of ~30,000,000 cells/mL. For each assay, 300,000,000 cells were harvested by centrifugation (3,000×g, 5 min, room temperature), the supernatant was aspirated and the pellet was washed once with 20 mM Tris-HCl. The cells were then incubated with ATP solution (Sigma G8877) (10 mM, pH 7.5) for 30 min at 37 °C, washed three times with 20 mM Tris-HCl to eliminate extracellular ATP. ADP (80 µM final concentration) was added and the cells were incubated statically at 37 °C. The mixtures were then centrifuged (10,000×g, 5 min) and the supernatant ATP concentration was determined by luciferase assay (ATP determination kit, Life Technologies - #A22066). ATP standards (provided with the kit) were used obtain a calibration curve.

**Introduction of mutant cyanobacteria to *S. cerevisiae* cells**

We adapted a method for generating yeast-*E. coli* chimeras to be used with *S. elongatus* [118,119]. SynJEC0-9 mutants were grown under constant light, with shaking, for 4 d. After this time, the cells (30 mL) were harvested (3,000×g, 10 min, 24 ºC), washed twice with BG-11 and resuspended in BG-11 (500 µL). *S. cerevisiae cox2-60* cells were grown aerobically in YPD medium (120 mL) for 24 h. The yeast was harvested (4,696×g, 10 min, 24 ºC), washed twice with sterile water, twice more with SCEM (1 M sorbitol, 13 mM β-mercaptoethanol) and resuspended in ice-cold sterile-filtered SCEM solution (10 mL) containing Zymolyase 100T (15 mg per gram of yeast pellet). The suspension was incubated for 1 h at 37 ºC to give spheroplasts. The suspension was then cooled on ice for 30 min and centrifuged for 10 min at 1,500×g and 4 ºC. The pellet was washed twice, gently, with chilled SCEM and resuspended in SCEM (2.5 mL) The suspension would remain usable for fusions for at least 24 h if kept refrigerated.

The spheroplast suspension (750 µl) was mixed with chilled TSC buffer (10 mM Tris-HCl, 10 mM $CaCl_2$, 1 M sorbitol, pH 8) (750 µl) and incubated for 10 min at 30 ºC. This mixture was then centrifuged (1,500×g, 10 min) and the supernatant was carefully discarded. The spheroplasts were resuspended in room-temperature TSC buffer (120 µl) and sorbitol (4 M, 60 µl). Dense *S. elongatus* cell suspension (120 µl) was added quickly to the spheroplast suspension, mixed by tube inversion and incubated for 30 min at 30 ºC. This mixture was then decanted into a round-bottom polypropylene tube containing PEG buffer (20% PEG 8000, 10 mM Tris-HCl, 2.5 mM $MgCl_2$, 10 mM $CaCl_2$, pH 8) (2 mL) and incubated without shaking at 30 ºC for 45 min. The cells were centrifuged (1,500×g, 10 min, 24 ºC), the supernatant was discarded and YPDS (YPD with 1 M sorbitol added) was added gently without disrupting the pellet. The cells were incubated under light without shaking for 2 h at 30 ºC. After this time, the pellet was partially dislodged by flicking the tube. The mixture continued to incubate for 3 h with shaking (70 rpm), after which time the cells were harvested (1,500×g, 10 min, 24 ºC), resuspended in 1 M sorbitol (300 µL)and plated on Selection-I medium. After drying for 5 min, a second layer of Selection-I medium was overlaid on top of the cells. The plates were incubated at 30 ºC in a 12 h light-dark cycle for at least four days, until colonies appeared between the agar layers. The colonies were extracted from the agar, suspended in 1 M sorbitol and spotted on Selection-II medium. For subsequent rounds of propagation, cells were scraped from the surface of the agar, resuspended in 1 M sorbitol and spotted on Selection-III medium.

**Cell count of S. *cerevisiae cox2-60*/cyanobacteria chimeras**

Spots on Selection-III medium were removed manually from the agar plate and placed intact inside a 1.5-mL microcentrifuge tube. Sorbitol (1 M, 200 µL) was added to the surface of the agar and cells were removed by pipetting. The tube was briefly centrifuged (5 s) to remove the cell-containing flowthrough from the agar. Cells were then mounted on a reusable glass slide and counted in triplicate from a brightfield image using the Countess II FL Automated Cell Counter (Fisher cat. # AMQAF1000) per the manufacturer's instructions.

**Total genomic DNA isolation and PCR analysis**

Total DNA isolation of chimeras was achieved using the Yeast DNA Extraction Kit (Thermo Fisher 78870) using the manufacturer's protocol. Cells were scraped from the surface of the agar, harvested (5,000 ×g, 5 min), resuspended and incubated (65°C, 10 min) with Y-PER Reagent (8 µL/mg pellet). The mixture was centrifuged (13,000 × g, 5 minutes), the supernatant was discarded and the pellet was resuspended and incubated (65°C, 10 min)

with DNA Releasing Reagent A (16 μL) and DNA Releasing Reagent B (16 μL). Protein Removal Reagent (8 μL) was added to the mixture which was then centrifuged (13,000 × g, 5 minutes). The supernatant was then added to a clean microcentrifuge tube and DNA was precipitated by adding isopropyl alcohol (24 μL). The mixture was centrifuged (13,000 × g, 10 min) and the pellet was washed by adding 70 % ethanol (13,000 × g, 1 min). The supernatant was then removed by aspiration and the pellet was allied to dry inside a fume hood. TE buffer (10 mM Tris, 1 mM EDTA, 50 μL) was added to the pellet to afford the total isolated DNA.

**Growth curves for yeast and cyanobacteria cells**

Growth curve data were obtained by growing 200 uL of cells in sterile 96-well plates (Nunclon 167008). Yeast cells were grown aerobically with shaking at 30 °C and 240 rpm. Syn7942 cells were shaken at 30 °C and 70 rpm. Measurements were all obtained with the BioTek Synergy H1M Microplate Reader and Gen5 v.3.11.19 software. Pathlength correction for these measurements was determined manually. Water (200 μL) was added to a well and A975 and A900 were measured by plate reader. Water was also added to a quartz cuvette (Spectrocell R-3010-T) and the near-IR absorption spectrum was obtained using the Shimadzu UV-VIS-NIR Spectrophotometer UV-3600 and UVProbe v.2.34 software. Corrections to absorbance data were made with the following function:

$$A_{correct} = A_{raw} * \frac{A_{975}(cuvette) - A_{900}(cuvette)}{A_{975}(well) - A_{900}(well)} \quad (1)$$

**Analysis of yeast/cyanobacteria chimeras using pTIRF microscopy**

Samples were prepared by extracting the chimeric cells from plates and washing them once with 1 M sorbitol. They were analyzed with a home-built TIRF microscope based on a Zeiss Axiovert 200M stand. A Cobolt diode-pumped 561 nm laser was used in this work. The laser beam is combined with other two laser lines in the system by using a set of Semrock LaserMUX™ filters, and then sent through an acousto optic tunable filter (AOTF, Quanta Tech Inc). The AOTF is used through the microscope control software to simultaneously and independently control the power of each laser on sample. After the AOTF, a laser speckle-reducer (LSR-3005-17S-VIS, Optotune) is used with a set of achromatic lenses to provide homogeneous illumination across the entire field of view for TIRF or pTIRF microscopy. The dichroic beamsplitter used in the system is a Semrock LF405/488/561/635-B-000 and the emission filter is a Chroma bandpass filter HQ 653/95 nm. The pTIRF images were acquired with a Photometric 512 Evolve EMCCD camera. Samples were viewed and imaged using a 100X oil immersion objective lens with NA = 1.4. The software for microscope control and data acquisition was developed using C++ and Labview. All images were processed with Fiji, for example, to merge a qTIRF image and a brightfield image that were acquired from the same sample position. Fluorescence of chimeras was measured using ImageJ 1.53c. Images of chimeras excited with either 405 nm or 561 nm laser were concatenated and fluorescent regions of interest (ROIs) were identified by thresholding.

**Analysis of yeast/cyanobacteria chimeras using fluorescence confocal microscopy**

Stock solutions (40X) of Concanavalin A, Fluorescein Conjugate (ConA; Thermo Fisher C827) were prepared by dissolving the lyophilized powder in sodium bicarbonate (0.1 M) to a final concentration of 2 mg/mL. These solutions were stored at -20 °C without exposure to light. Directly prior to use in sample preparation, the stock solution was thawed and centrifuged for 10 s. Microscopy samples were prepared by extracting the chimeric cells

from plates and washing them once with Hank's Buffered Salt Solution (HBSS; NaCl (140 mM), KCl (5 mM), $CaCl_2$ (1 mM), $MgSO_4$ heptahydrate (0.4 mM), $MgCl_2$ hexahydrate (0.5 mM), $Na_2PO_4$ dihydrate (0.3 mM), $KH_2PO_4$ (0.4 mM), D-glucose (6 mM), $NaHCO_3$ (4 mM)). The cells were then incubated (37 °C, 10 min) in HBSS supplemented with ConA to a final concentration of 50 μg/mL, washed twice with HBSS, centrifuged and incubated with Karnovsky fixative (2 % glutaraldehyde, 2.5 % paraformaldehyde). Samples were analyzed with a commercial Leica SP8 fluorescence confocal microscope. Samples were viewed and imaged through a 63X/1.40 HC PL APO Oil CS2 lens and excited with 488 nm and 561 nm laser light. Emission wavelengths in the 510/20 nm range were detected with photomultiplier tube (PMT) detector, and emission wavelengths in the 650/20 range were detected using a high-sensitivity GaAsP HyD detector. Leica Application Suite (LASX) was used to collect raw data. All images were processed with ImageJ 1.53c to display an overlay of the two channels.

**Analysis of yeast/cyanobacteria chimeras using Transmission Electron Microscopy (TEM)**

Cells were extracted from agar plates in HBSS (700 μL) and pelleted by centrifugation (6,000×g, 5 min). The supernatant was removed, and the cells were resuspended in Karnovsky fixative (20 μL) and incubated for 30 min. The cells were then pelleted and resuspended in HBSS. The medium was removed and replaced with a fixative (2.5% EM-grade glutaraldehyde and 2.0% EM-grade formaldehyde in 0.1 M sodium cacodylate buffer, pH 7.4) for 3 hours at 4ºC. The fixative was then removed, replaced briefly with buffer, and then replaced with 1% osmium tetroxide in buffer for 90 minutes. Each sample was then subjected to 10-minute buffer rinse, after which it was placed in 1% aqueous uranyl acetate and left overnight. The next day, each sample was dehydrated via a graded ethanol series, culminating in propylene oxide. Following a graded propylene oxide; Epon812 series, the nuclear pellets were embedded in Epon812 prior to cutting. Ultrathin (ca. 90 nm) Epon sections on grids were stained with 1% aqueous uranyl acetate and lead citrate.[169] After the grids dried, areas of interest were imaged at 160 kV, spot 3 using a Philips/FEI (now Thermo Fisher FEI) Tecnai G2 F20 S-TWIN transmission electron microscope in the Microscopy Suite at the Beckman Institute of Advanced Science and Technology (University of Illinois at Urbana-Champaign).

**Growth dependence of chimeras on light**

For partial light starvation after the introduction on cyanobacteria, SynJEC3 cells were introduced to yeast spheroplasts in the manner described, plated on and overlaid with Selection-I medium. The plates were incubated at 30 °C for 4 d inside a box covered in aluminum foil placed into the incubator. After this time, the contents of the overlaid agar plate were extracted by stabbing and plated on Selection-II medium grown either in 12 h day/night cycle or under the same conditions described to remove light. The growth of chimeras under these different conditions was quantified in the manner described. For complete light starvation, a flask containing BG-11 medium was inoculated with log-phase SynJEC3 cells to $OD_{730}$ of 0.200 and incubated for 4 d without shaking inside a box covered in aluminum foil. These cells were introduced to yeast spheroplasts in the manner described, plated on Selection-I medium and incubated for 4 d at 30 °C inside a foil-covered box. After this time, the contents of the plate were extracted by stabbing the agar and plated on Selection-II medium. The plates were incubated for 4 d more in the absence of light.

**Supplementary information**

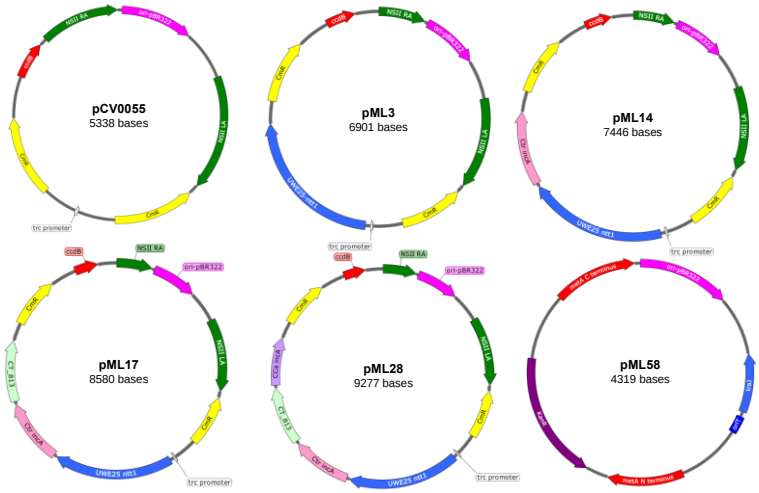



**Figure 2.6: Plasmids used in this study**: UWE25 ntt1 – ADP/ATP translocase; Ctr-incA, Cca_incA and CT_813 are genes encoding SNARE-like proteins from *C. trachomatis, C. caviae and C. trachomatis*, respectively; CmR – chloramphenicol acetyltransferase; KanR – aminoglycoside-3'-phosphotransferase; trc is a constitutive promoter; NSII, metA N-terminus and metA C terminus are all homologous recombination sites at the 5'-end and 3'-end of the *metA* gene in the Syn7942 genome.

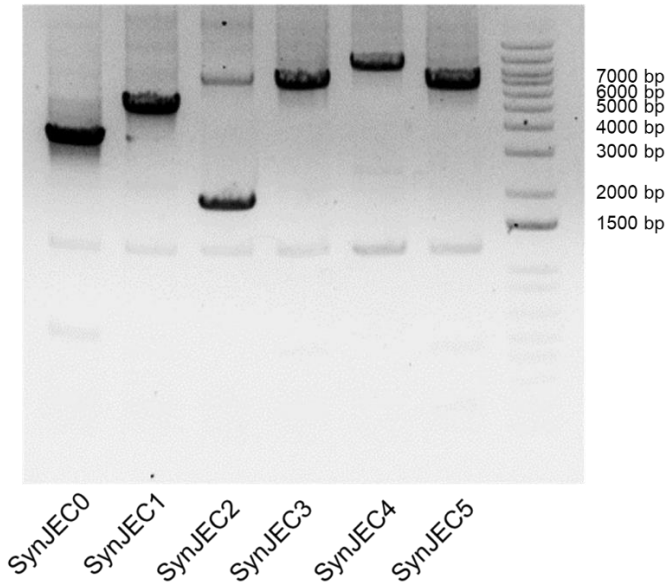



**Figure 2.7: PCR amplification of the *NSII* locus of Syn7942 to confirm recombination:** Recombination was verified by DNA sequencing analysis of amplified and gel purified DNA fragment. The experiment was repeated twice independently with similar results.

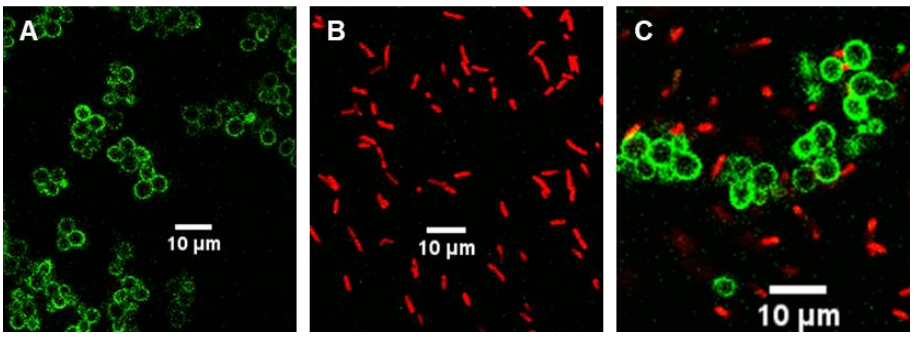



**Figure 2.8: Confocal image for controls: (A)** yeast only (pseudo-color: green), **(B)** cyanobacteria only (pseudo-color: red), **(C)** yeast/cyanobacteria mix. The experiment was repeated twice independently with similar results.

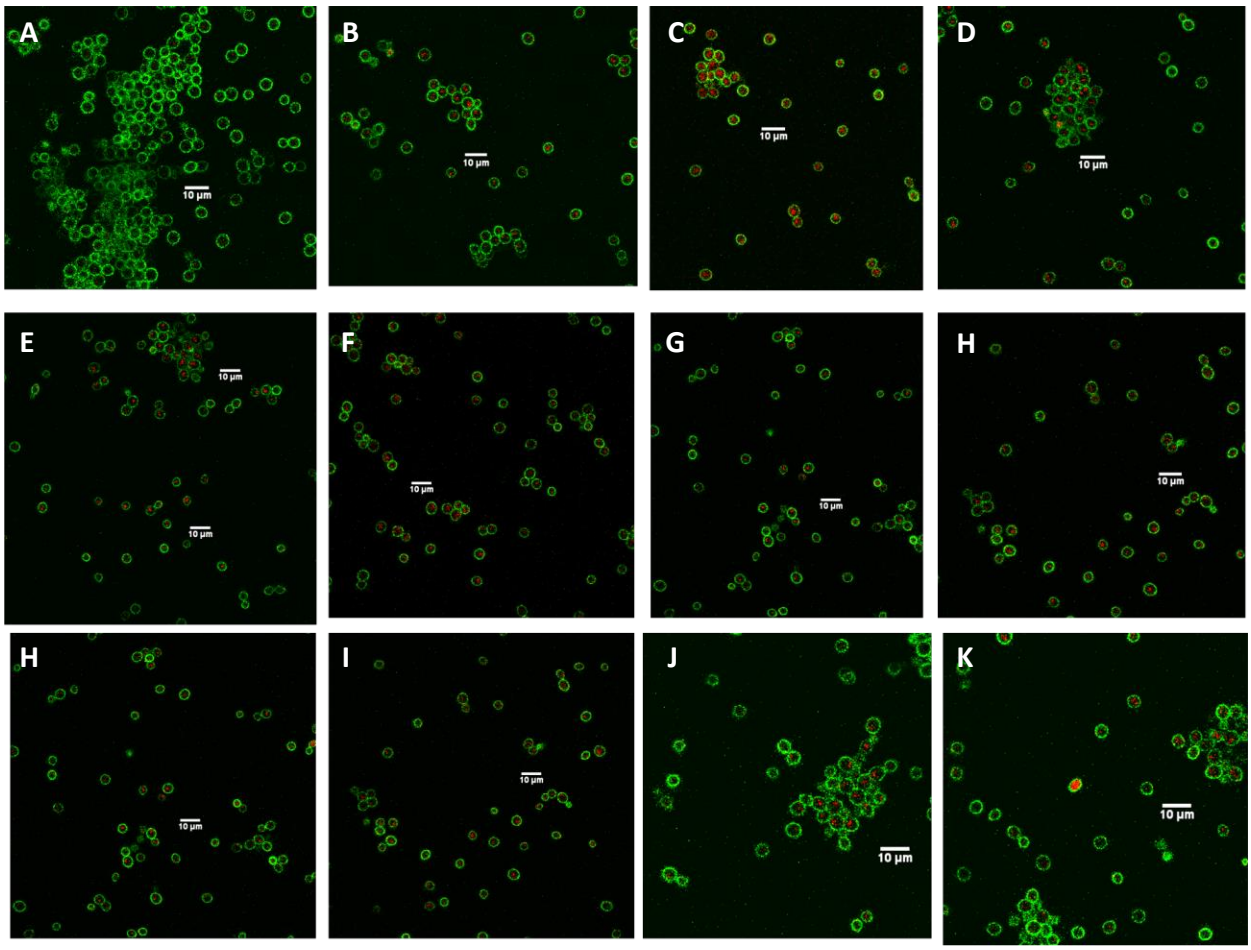



**Figure 2.9: Tracking the yeast/cyanobacteria chimeras during various stages of selection using fluorescence confocal microscopy: (A)** Control yeast cells **(B)-(K)** Images of early stages (5-10 doublings) to late stages Yeast/SynJEC3 (15-20 doublings) chimeras propagated for multiple rounds of selection imaged by fluorescence confocal microscopy. The yeast cell wall was stained with Con A-FITC (pseudo-color: green, Ex. = 488 nm; Em. = 510/20) and presence of cyanobacteria was monitored by cyanobacterial fluorescence (pseudo-color: red, Ex. = 561 nm; Em. = 650/20). The experiments were repeated twice independently with similar results.

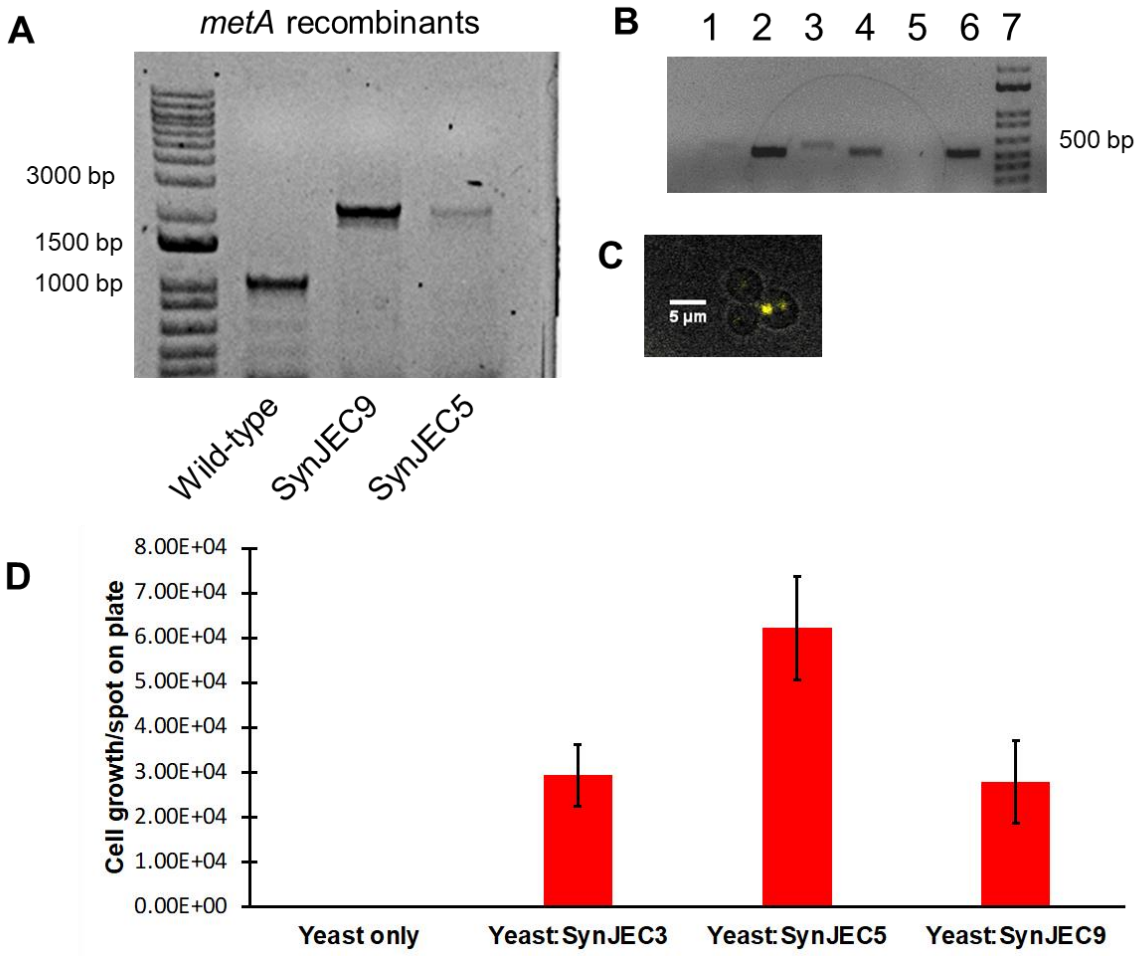



**Figure 2.10: Yeast-SynJEC chimeras containing endosymbionts with methionine auxotrophs: (A)** PCR amplification of the metA locus to confirm *metA* deletion. The experiment was repeated twice independently with similar results. **(B)** Analysis of total genomic DNA from Yeast-SynJEC chimeras. Lane 1: PCR amplification of *CAT* from yeast:SynJEC9 chimeras; Lane 2: PCR amplification of *MATa* from yeast:SynJEC9 chimeras; Lane 3: PCR amplification of *CAT* from yeast:SynJEC5 chimeras; Lane 4: PCR amplification of *MATa* from yeast:SynJEC5 chimeras; Lane 5: PCR amplification of *CAT* from yeast only; Lane 2: PCR amplification of *MATa* from yeast only; Lane 7: Standards. The experiment was repeated twice independently with similar results. **(C)** TIRF microscopy image of yeast:SynJEC5 chimeras (pseudo-color: yellow corresponds to pTIRF/excitation with 561 nm laser with open shutter; black and white image corresponds to pTIRF with closed laser shutter. The experiment was repeated three times independently with similar results. **(D)** Growth trends (at round IV of repropagation) of *S. cerevisiae cox2-60* (yeast only strains), *S. cerevisiae cox2-60*- SynJEC3 (yeast-SynJEC3), *S. cerevisiae cox2-60- SynJEC5* (yeast-SynJEC3) and *S. cerevisiae cox2-60- SynJEC9* (yeast-SynJEC4) chimeras on Selection Medium III. Cells ($3.00\times10^3$) were spotted on Selection Medium III and the final number of cells/spot on plate were determined after 48 h of growth ($n$ = 3 measurements per sample; error bars represent standard error of the mean).

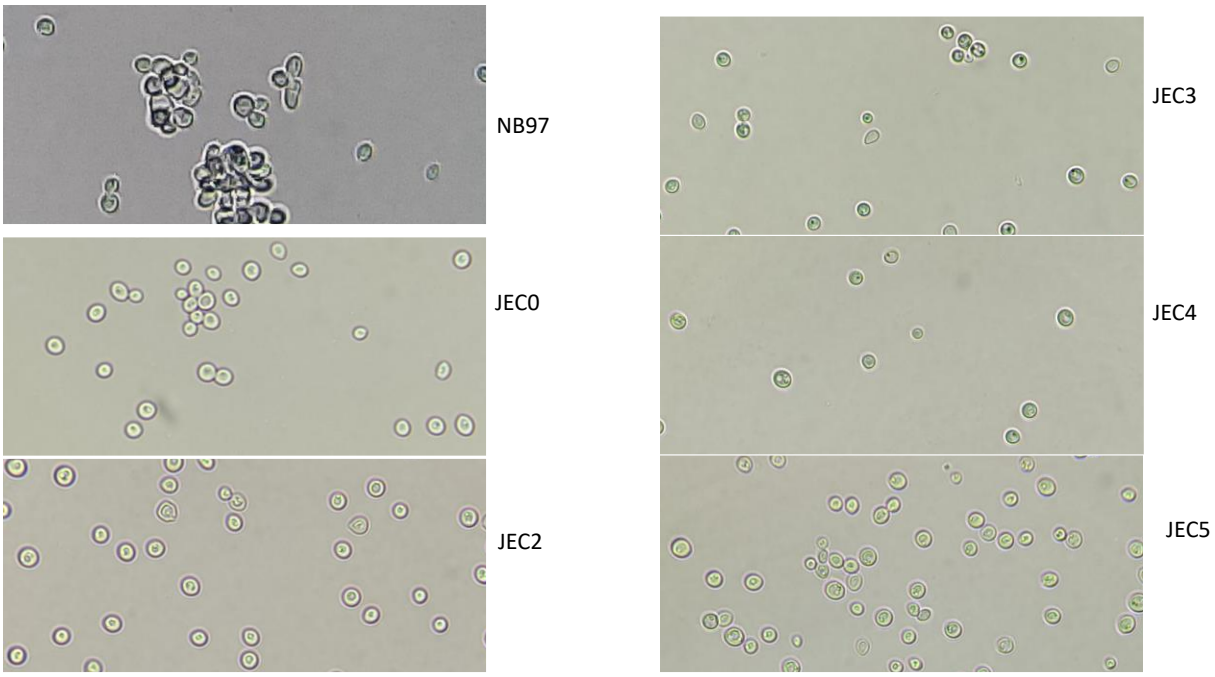



**Figure 2.11: Representative microscopic images of the of the yeast/cyanobacteria chimera that were used for the isolation of the total genomic DNA and PCR analysis:** Using a combination of these images and TIRF images, we do not detect the presence of extracellular cyanobacterial cells under these conditions. The experiment was repeated six times independently with similar results.

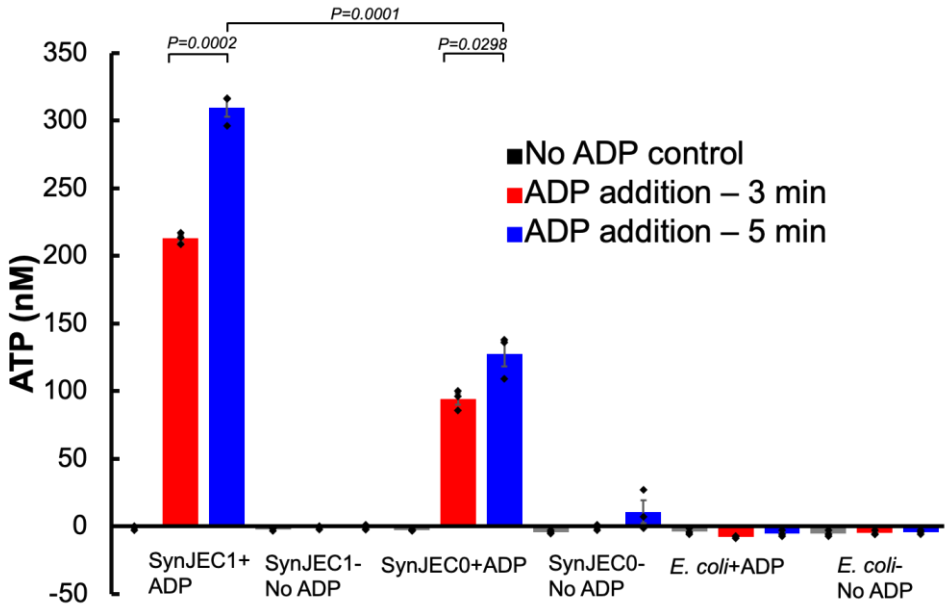



**Figure 2.12: ADP/ATP translocase assays: Comparison of SynJEC strains to *E. coli* DH10B cells.** Release of ATP by SynJEC1 cells expressing the UWE25 ADP/ATP translocase in the presence of 80 µM ADP in comparison to SynJEC0 cells. ATP was released when SynJEC1 (expressing the ATP/ADP translocase) and SynJEC0 cells were challenged with extracellular ADP (80 µM), but not with a blank solution lacking ADP. ATP was not released when *E. coli* DH10B were challenged with extracellular ADP (80 µM). N=3 biologically independent experiments. Data are presented as mean values +/- SEM. Two-sided t-tests were used to compare means without adjustments (95% CI, Cohen's d=10.6, DF=4, P=0.0002; 95% CI, Cohen's d=13.0, DF=4, P=0.0001; 95% CI, Cohen's d=2.7, DF=4, P=0.0002).

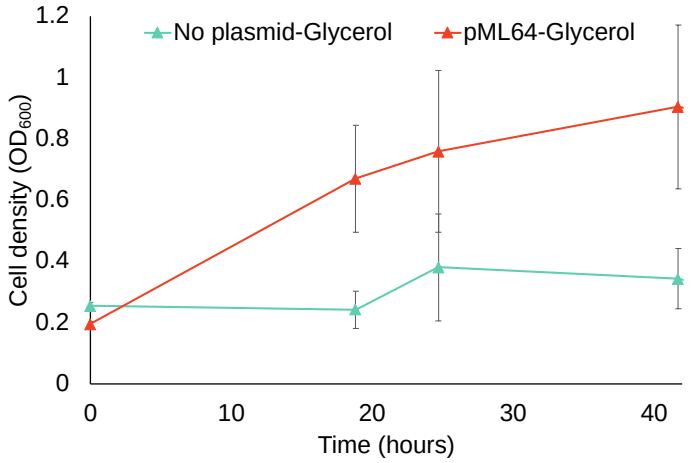



**Figure 2.13: Growth rate of *S. cerevisiae cox2-60* under selection and non-selection conditions:** Red trace - *S. cerevisiae cox2-60-pML64* growth rate under selection conditions containing glycerol as the carbon source. Cyan trace - *S. cerevisiae cox2-60* growth rate under selection conditions containing glycerol as the carbon source. N=24 biological replicates. The experiment was repeated twice independently with similar results. Data are presented as mean values +/- standard deviation.

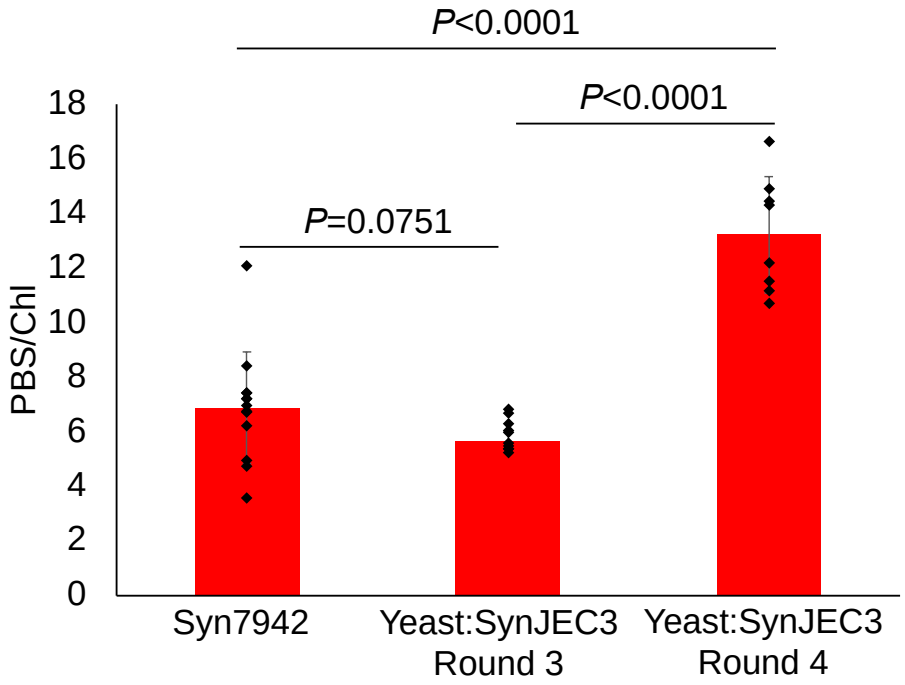



**Figure 2.14: pTIRF microscopy to determine the relative changes in phycobilisomes and chlorophyll *a*.** The relative levels of phycobilisomes (PBS, Ex. 561 nm) and chlorophyll (Chl, Ex. 405 nm) a as plotted as an internal ratio (PBS/Chl) with rounds of repropagation under selection conditions. N=13 (Syn7942), 10 (Yeast:SynJEC3 Round 3), 8 (Yeast:SynJEC3 Round 4) measurements of different ROIs taken from a single captured pTIRF image. The experiment was repeated twice independently with similar results. Data are presented as mean values +/- standard deviation. Two-sided t-tests were used to compare means without adjustments (95% CI, Cohen's d=0.8, DF=21, P=0.0751; 95% CI, Cohen's d=3.0, DF=19, P<0.0001; 95% CI, Cohen's d=4.9, DF=16, P<0.0001).

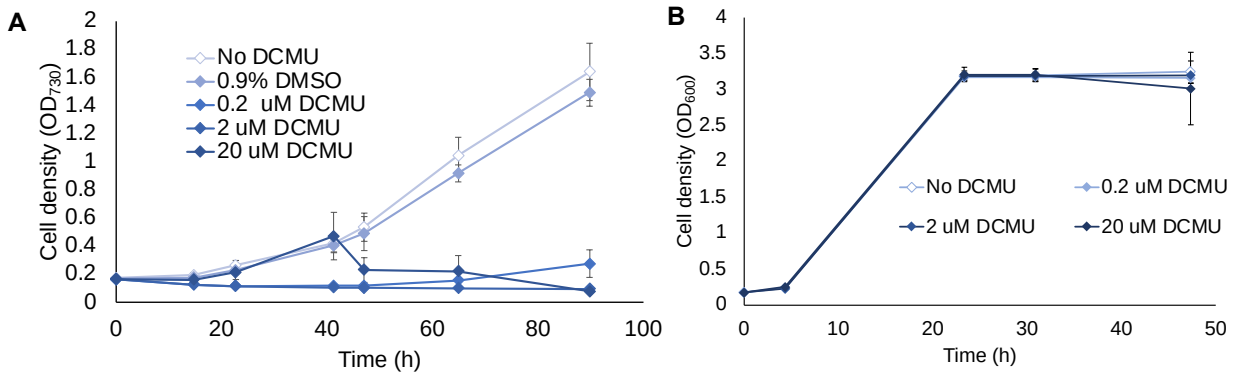



**Figure 2.15: Effect of DCMU on the yeast and cyanobacteria growth: (A)** Effect of DCMU on the growth of Syn7942. (B) Effect of DCMU on the growth of *S. cerevisiae cox2-60*. N=12 biological replicates. The experiment was repeated twice independently with similar results. Data are presented as mean values +/- standard deviation.

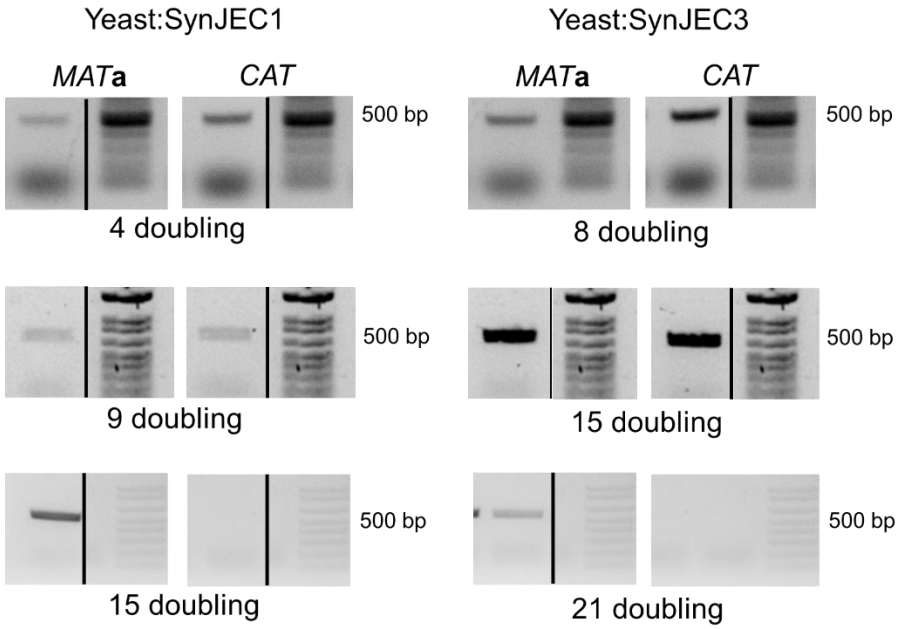



**Figure 2.16: PCR analysis of total genomic DNA isolated from chimeras at various stages of growth:** To characterize the presence of cyanobacterial endosymbionts within yeast cells, we isolated total genomic DNA from chimeras propagated under selection conditions for various rounds of growth. Presence of yeast genome is detected by amplification of yeast *MAT***a** gene and presence of SynJEC strains is detected by SynJEC chloramphenicol acetyltransferase (*CAT*) gene. After more than 20 rounds of doublings, we observe that the chimeras lose their viability possibly due to loss of cyanobacteria. The experiment was repeated three times independently with similar results.

**Table 2.2. Genomic DNA fragments used in the construction of pML58**

| Fragment name | Fragment sequence |
|---|---|
| *metA*-N term | GAAGAACAAGCCGCAGCTGGCGTCAGCCCTGACTTTGTGCGGCTTTCGGTTGG<br>CTTGGAGCATTTAGACGACATCTTGGCTGACCTTGACCAAGCCCTGCAGGCTT<br>CGCAAGTCGCCTAGGTTGTTTCGTACTGGGCTTGCGATCGCCGACCCTCCCAG<br>CGTTTTGCCATGCCGATCATCATTCCGCAAGATCTGCCTGCTCGTCGCATTCTG<br>GATGCTGAGTCCGTATTCACGCTGGGGGATGCGGATGCGCGGCGGCAGGATA<br>TCCGCGCTTTGCAAGTGGTTGTTTTGAACCTAATGCCGACCAAGGTGACAACC<br>GAAACGCAGATTGCTCGGGTACTCGCCAATACGCCGCTACAAGTAGAACTCA<br>CTTTAATTCACACGGCTAGCTATCAACCAACCCACACCGACCCCGAGCATCTT<br>CGTAACTTCTACAGCACCTTTGATCAAATTCGCGATCGCCAGTTTGATGGCTT<br>GATTGTGA |
| *metA*-C term | TGACGACTATTCTGGATTGGAGTCGTGAGGCGGTGCGCTCCAGTTTGTTCATC<br>TGCTGGGGAGCCCAAGCCGCGCTTCAGCATTTTCATGGCATTGAAAAGCAAA<br>CCTTGCCAGCCAAGCGCTTTGGCGTTTTTTGGCATCATCTCCGCGATCGCAGTT<br>CTCCCTTGGTACGCGGCCACGATGATGATTTTCTGGTGCCGGTCAGTCGCCAT<br>ACGGAGGTAATTGCGGCTGAGGTATTGGCTCAATCACAGTTGCAAATTCTGGC<br>AGAAAGCTCAGAGGCTGGACTCCATCTCCTCTGGGATGCAGACCAACATCGC<br>ACCTATCTGTTCAACCATCCGGAATACGATGCAGACACCCTCGATCGCGAATA<br>TCGACGCGATCGCGAGAAAGGGTTGCCGATTCAGTTACCTCTCAACTACTACC<br>CTAATGATGACCCGAATCAAGTGCCGAGAGTGCGTTGGCGTAGCCATGCTCA<br>ACTGCTTTACACTAACTGGCTAAACTACGAGGTTTATCAACCACTGTCACGCT<br>AA |

**Table 2.3. Oligonucleotides used for analysis of the total genomic DNA isolated from the yeast/cyanobacteria chimera.**

| Oligonucleotide name | Oligonucleotide sequence |
|---|---|
| AM965 | AGTCACCATCAAGATCGTTTATGG |
| AM966 | GCACGGAATATGGGACTACTTCG |
| AM967 | ACTCCACTTCAAGTAAGAGTTTG |
| AM1224 | CAAAATGGAGAGTTTGATCCTGGCTCAGG |
| AM1225 | AAAGGAGGTGATCCAGCCACACCT |
| AM1351 | TCAAGACGACTTGGTACTAGGACTCG |
| AM1456 | ATGCCAATGACAACCCCTACGTTG |
| C2F | GCCTTGCCATTCATCACAGTAC |
| C9R | CCGCAAAGAACACTTTGAAGCC |

**Table 2.4: Vector map links for maps listed in this manuscript**

| Plasmid name | Benchling link |
|---|---|
| pCV0049 | https://benchling.com/s/seq-P2Wv1DvB3F5HeufjL2IH |
| pCV0055 | https://benchling.com/s/seq-QlPIEWddeBGaxRumfGBm |
| pML3 | https://benchling.com/s/seq-XugIBKLDymnz5C6tVHm8 |
| pML14 | https://benchling.com/s/seq-OYZhEqEUAERMC4k36bqg |
| pML17 | https://benchling.com/s/seq-BnSbm61OdiyAK1pvOXcg |
| pML28 | https://benchling.com/s/seq-bl7pi8OvkcpczeVMVgNn |
| pML58 | https://benchling.com/s/seq-lRLmm9gNLVt0jO3h4Smk |
| pML59 | https://benchling.com/s/seq-gY35CjgwmBJEBfxfooAh?m=slm-j56wvCpY7Q3qazIi36ZL |
| pML60 | https://benchling.com/s/seq-sT6JgJbNyzsiPXVxcVHy?m=slm-16xb853pom6UG8po4LPd |
| pML62 | https://benchling.com/s/seq-CGCepat4VSOKikZ1vPHy |
| pML64 | https://benchling.com/s/seq-P2Wv1DvB3F5HeufjL2IH |

**Table 2.5. Oligonucleotides used for analysis of genomic DNA isolated from recombinant Syn7942 mutants.**

| Oligonucleotide name | Oligonucleotide sequence |
|---|---|
| LL56 | GAAGAACAAGCCGCAGC |
| LL57 | TTAGCGTGACAGTGGTTGA |
| JC192 | TCTGGATGCGGTGACTTGGCA |

**Table 2.6. Comparison of total doublings detected for the key yeast/cyanobacteria chimera under optimal growth conditions in selection medium III:**

| Yeast/Cyanobacteria chimera | Total doubling detected |
|---|---|
| *S. cerevisiae cox2-60* | 2 |
| *S. cerevisiae cox2-60*-SynJEC0 | 11 |
| *S. cerevisiae cox2-60*-SynJEC1 | 14 |
| *S. cerevisiae cox2-60*-SynJEC3 | 22 |

**Table 2.7. Oligonucleotides used in the construction of plasmids**

| Oligonucleotide name | Oligonucleotide sequence |
|---|---|
| AM1195 | TTAACTAGTTGCGATCTCACTAGTCTTGTCAATATTTTTCTCA |
| AM1214 | CAGAGATCTGATGTCACAAGATGCAAAGCAAGACTTTG |
| AM1215 | TTTGTACAAACTTGTTTAACTAGTTGCGATCTCACTAGTCTTGTCA |
| AM1216 | TGCATCTTGTGACATCAGATCTCTGTTTCCTGTGTGAAATTGTTATC |
| AM1217 | CGCAACTAGTTAAACAAGTTTGTACAAAAAAGCTGAACGAGA |
| AM1312 | ACAAGTTTGTACAAAAAAGCTGAACGAGAAAC |
| AM1313 | AAGACTAGTGAGATCGCAACTAGTTAATCTC |
| AM1314 | TCGTTCAGCTTTTTTGTACAAACTTGTTC |
| AM1351 | TCAAGACGACTTGGTACTAGGACTCG |
| AM1352 | AGTCCTAGTACCAAGTCGTCTTGAGT |
| AM1353 | TCGTTCAGCTTTTTTGTACAAACTTGTTCA |
| AM1502 | TTTCCGAATAAAGTAGGAGAATTGTACAATGACAGTATCCACAGACAACACAA( |
| AM1503 | GTACAAACTTGTTTACAAAGACTGAGCTAATTTCTCAAATGGAGA |
| AM1504 | GGTAGCATCTGAACATATTTTCCGAATAAAGTAGGAGAATTGTACAATGACA |
| AM1505 | GTACAAACTTGTTTACAAAGACTGAGCTAATTTCT |
| AM1506 | ATTCGGAAAATATGTTCAGATGCTACCGCGCCGCC |
| AM1507 | AGCTCAGTCTTTGTAAACAAGTTTGTACAAAAAAGCTGAACGAGAAAC |
| AM1531 | CTGTTATCTGGCTTTTAGTAAGCCGG |
| AM1532 | ATGTCAGGCTCCCTTATACACAGC |
| JC151 | TCTAGCGATTTGATGCGGTATTTTCTCCTTACGCATCT |
| JC152 | CGGTGAGTTCTCTTCCGCTTCCTCGCTCAC |
| JC153 | TACCGCATCAAATCGCTAGAGCTTGCATGCCTG |
| JC154 | AAGCGGAAGAGAACTCACCGCGACGTATCGG |
| JC200 | GTCACGCTAAAGCGTCAGACCCCGTAGAAAAGA |
| JC201 | GTCTGACGCTTTAGCGTGACAGTGGTTGATAAACCT |
| JC202 | TGACCCGCTTTGACGACTATTCTGGATTGGAGTCGT |
| JC203 | ATAGTCGTCAAAGCGGGTCACTACTTGGTAGCA |
| JC204 | GAGGGACGAAGAAGAACAAGCCGCAGCTGG |
| JC205 | CTTGTTCTTCTTCGTCCCTCAAATCGTGGCC |
| JC206 | TTGATTGTGATCTAGCGCAGCTGCTTGGAAC |
| JC207 | CTGCGCTAGATCACAATCAAGCCATCAAACTGGC |

**Table 2.8. gBlocks purchased from IDT**

| gBlock name | gBlock sequence |
|---|---|
| Ctr-incA | AAGACTAGTGAGATCGCAACTAGTTAATCTCATACCGGTACGCCCGCATGAAGGAGatc<br>GAGCTCATGCCAATGACAACCCCTACGTTGATCGTTACTCCTCCTTCGCCTCCAGCACC<br>TAGCTATAGTGCGAACCGTGTTCCTCAGCCCAGCTTAATGGACAAAATCAAAAAGATC<br>GCTGCTATCGCAAGTTTAATCTTGATTGGGACCATCGGATTTCTCGCGCTCCTCGGTCA<br>CTTGGTCGGGTTCTTGATTGCTCCTCAGATCACAATCGTACTCCTCGCCTTATTTATCAT<br>CAGTTTGGCCGGAAACGCGCTGTACTTACAAAAGACTGCAAATCTGCACTTATACCAG<br>GATCTCCAACGGGAGGTTGGTTCGTTAAAAGAGATCAACTTTATGCTCTCGGTGCTGCA<br>AAAGGAATTCCTGCATTTAAGTAAGGAATTTGCTACCACCTCTAAGGACTTATCGGCG<br>GTCTCCCAAGACTTCTACTCCTGTTTGCAAGGTTTCCGCGATAATTACAAGGGCTTTGA<br>GAGTCTGTTGGATGAATATAAGAACAGTACTGAAGAAATGCGTAAACTCTTTAGCCAA<br>GAGATTATCGCAGACCTCAAGGGTTCGGTAGCAAGTTTGCGGGAAGAAATCCGTTTCC<br>TGACCCCCTTAGCAGAGGAAGTACGTCGGTTAGCACACAATCAGCAGAGTTTGACAGT<br>AGTCATCGAAGAACTGAAAACTATCCGTGACTCTCTGCGGGATGAAATTGGGCAGCTG<br>AGTCAACTCAGTAAGACATTAACGTCGCAGATCGCGTTACAGCGTAAGGAGAGTAGTG<br>ATCTGTGTAGCCAAATTCGTGAAACTTTAAGTAGCCCCCGCAAGTCTGCGAGTCCTAGT<br>ACCAAGTCGTCTTGAACAAGTTTGTACAAAAAAGCTGAACGA |
| CT_813 | AGTCCTAGTACCAAGTCGTCTTGAGTAGCCGCATCCGGTATTCACACAGGAAACAGAG<br>ATCTGATGACTACGTTGCCGAACAATTGCACGTCTAACAGCAATTCCATTAACACCTTT<br>ACAAAAGATATTGAAATGGCAAAGCAAATTCAAGGAAGCCGGAAAGACCCGTTAGCA<br>AAGACCTCCTGGATTGCAGGCTTGATTTGCGTGGTGGCAGGGGTTCTGGGATTACTGGC<br>GATTGGGATCGGTGGCTGTTCCATGGCGAGTGGTTTAGGCTTAATCGGTGCTGTGGTGG<br>CGGCAGTTATCGTCGCGGTGGGTTTGTGCTGCCTGGTTTCGGCTTTGTGCCTGCAAGTG<br>GAGAAGTCGCAGTGGTGGCAAAAGGAATTCGAAAGTTGGATCGAACAAAAGTCGCAA<br>TTCCGTATTGTGATGGCTGATATGCTCAAGGCGAACCGTAAGTTACAGTCGGAAGTAG<br>AATTTCTCTCCAAAGGTTGGTCTGACGACACTGCCGTTCACAAAGAAGATGTTACTAAA<br>TATGAGCAAGTGGTCGAGGAATATGCTGAAAAAATTATGGAGCTGTACGAGGAGACA<br>GGGGTATTAACTATCGAAAAGATCAACCTGCAAAAAGAGAAAAAGGCTTGGCTCGAG<br>GAGAAGGCCGAAATGGAGCAGAAACTGACAACTGTCACGGACCTGGAGGCTGCGAAA<br>CAACAGCTGGAGGAGAAGGTGACTGATTTAGAGTCCGAGAAGCAGGAGCTCCGGGAG<br>GAGTTAGATAAGGCAATCGAAAATCTCGACGAAATGGCCTATGAGGCAATGGAATTCG<br>AAAAGGAAAAACACGGTATCAAGCCTGGGCGGCGCGGTAGCATCTGAACAAGTTTGT<br>ACAAAAAAGCTGAACGA |
| Cca-incA | ATGACAGTATCCACAGACAACACAAGTCCTGTAATATCGAGAGCGTCCTCACCTACTTT<br>TGGAGATCATGGTAAGGATTTCGACAACAATAAAATTATACCCATTTCAATAGAAGCT<br>CCAACTTCTTCAGCTGCTGCTGTAGGGGCTAAAACGGCTATCGAGCCTGAAGGAAGAA<br>GCCCACTACTTCAAAGGATTTGCTATCTTGTTAAAATTATCGCTGCCATCGCCCTCTTTG<br>TTGTTGGTATCGCAGCCTTAGTTTGCTTATATCTCGGTAGCGTTATCTCAACGCCTTCTC<br>TTATTCTTATGCTTGCGATCATGCTTGTATCCTTTGTGATCGTTATTACGGCAATTCGAG<br>ATGGCACACCGTCTCAAGTGGTCCGTCACATGAAACAGCAAATTCAGCAATTTGGCGA<br>AGAAAACACGCGTTTACATACCGCAGTAGAAAATCTAAAAGCTGTTAACGTTGAGCTC<br>TCAGAGCAAATTAACCAACTTAAACAACTACATACTAGATTATCGGATTTTGGTGATA<br>GGCTTGAAGCGAATACCGGTGATTTTACTGCACTTATTGCGGATTTCCAACTCAGTCTG<br>GAAGAGTTTAAGTCTGTTGGTACTAAAGTTGAAACCATGCTCTCTCCATTTGAGAAATT<br>AGCTCAGTCTTTGTAA |

**Table 2.8 (cont.): gBlocks purchased from IDT**

| | |
|---|---|
| UWE25ntt1-Se-codon-opt | ATGTCGCAGGACGCAAAACAAGATTTTGGAAAGTGGCGGGCGTTCTTTTGGCCTGTT<br>CATGGCTACGAACTGAAAAAGCTGCTGCCGATGTTTTTCATGTTCTTTTTTATTTCTT<br>TTAACTATACGATTTTGCGGGACACGAAAGACACCCTGATTGTAACATCTGCCGGTG<br>CCGAGGCCATCCCTTTTTTTAAAGTCGTTCGGAGTAGTTCCAGCCGCGATCTTGTTCAT<br>GATTATTTACGCGAAATTAAGCAACACCCTGTCCCGCGAAAACTTGTTTTACGTCAC<br>ATTGCTGCCTTTTATCATCTTCTTTGGTTTGTTCGCCTTTGTAATGTATCCAGCCCGTG<br>AAGTTCTGATGCCTCACGCGTCTGCGGAAGCGCTCAAGGCGTACTTGCCAGGGGGG<br>TGGACTGGTTTAGCCGCCGCTTATGAGAATTGGATGTATTCGATTTTCTATATCTTGG<br>CAGAATTGTGGGGTAGTGTAGTGCTCTCGTTACTGTTCTGGGGCTTTGCGAATCAGA<br>TTACACGCGTAAACGAGGCTAAGCGTTTTTATAGCTTGTTCGGTCTGGGGGCCAACC<br>TCGCATTGTTAGTCTCTGGCCCTGCAATTGTCTACGTTTCGGATATTCGGAAGCACTT<br>ACCGGCTAATGTAGATGCTTGGCAGATCAGCCTCAATTATCTGATGGGAATGGTGGT<br>TATTGCTGGCTTAGCCATCCTGGCAATCTATTGGTGGATCAATCGGGCGGTTCTGAC<br>GGACCCACGCTTCTACGATTTGAACCAAGAAAAGGCGCCCGGTAAAAAAAAAAAAG<br>CAAAAATGAGCTTAGGGGAATCCTTCAAATTTTTATTTACGAGCAAGTATATCTTAT<br>GCCTCGCTATCTTAGTCATCGCTTACGGTATCAGCATTAACCTCGTTGAAATTACTTG<br>GAAGAGTCTGGTTAAGCTGCAATACCCTAACCCTAATGATTATTCCACATTTATGGG<br>TTGGTTCAGTACTATGACAGGAGCAGTAACCATCCTCATGATGTTATTTGTAGGGGG<br>AAATGTGATTCGCCACAAGGGATGGGGTTTTGCTGCCCTGATTACACCTGTGGTTTT<br>GCTGGTAACAGGAATTGCATTCTTCAGTTTCGTCATCTTTAAAGATCACTTGGCAGG<br>ATACATCGCCGCGCTGGGAACGACACCCCTGTTCCTCGCAGTAATCTTCGGCGCCGC<br>GCAGAATATTATGTCTAAGTCGGCGAAATACTCCTTGTTTGACCCCACTAAAGAGAT<br>GGCGTATATTCCATTAGATGACGAGTCTAAAGTAAAAGGCAAGGCCGCCGTTGACG<br>TAGTCGGGGCTCGGCTCGGAAAGAGCGGCGGTTCTATTATTCAAATGGGACTCCTGG<br>CGTTTGGAACCTTGGCGACCATTACGCCCTATATTGGTGCGATCCTGATGGTTATCA<br>TCGCAGCATGGATCGTTGCCGCACGTTCTCTCAGCAAACAGTTCACGCAACTCACGG<br>CGGAACAAAACATCGAGAAAAACATTGATAAAACTTCGGAGATTGCTACTTCCTAG |

## Chapter 3—Organelle-like metabolic coupling in synthetic endosymbiosis and its implication on organelle evolution

### Author contributions

Jason E. Cournoyer—conceptualization, methodology, formal analysis, investigation, data curation, writing, visualization

Yang-le Gao—formal analysis, investigation

Sarah D. Altman——formal analysis, investigation

Bidhan C. De——methodology, investigation, writing, visualization

Jason Gao——investigation

Guo-Hsuen Lo——investigation

Zachary De Leon——investigation

Angad P. Mehta——conceptualization, methodology, formal analysis, investigation, data curation, writing, visualization, supervision, project administration, funding acquisition


### Abstract

Organelles like mitochondria and chloroplasts evolved from free-living bacteria that were established as endosymbionts within the host cell. One of the hallmarks of naturally existing host/endosymbiont systems as well as organelle evolution is the extensive metabolic coupling between the endosymbiont or organelle and the host cell. This intricate metabolic coupling is facilitated by the acquisition of exogenous transporters, the loss of non-essential genes from the endosymbiont genomes resulting in the endosymbiont genome minimization. For example, amino acid biosynthetic genes are lost from the endosymbiont genome and host cell genome encodes the corresponding biosynthetic genes and maintains the supply of amino acids which are imported into the endosymbiont or the organelles by the transporters that are localized to the endosymbionts or the organelles. However, the order and the molecular details of these evolutionary events are unclear. Here, we engineered synthetic yeast/cyanobacteria endosymbiosis to experimentally investigate the requirements for organelle-like metabolic coupling. We deleted key biosynthetic enzymes corresponding to 5 coenzyme and 19 amino acids to generate a series of endosymbiotic cyanobacterial mutants that were auxotrophic for defined coenzymes, amino acids, or their derivatives. This resulted in the identification of novel auxotrophies in our endosymbiotic cyanobacterial mutants which uptake previously unidentified metabolites, e.g., a series of dipeptides were identified as auxotrophs of deleting biosynthetic genes corresponding to a single amino acid. We then introduced each of the endosymbiotic cyanobacterial auxotrophic mutants within the yeast cells and investigated if the yeast metabolome was able to sustain the growth of endosymbiotic cyanobacterial mutants under photosynthetic growth conditions by providing deficient metabolites to the endosymbiont. Our results suggested that most of these cyanobacterial auxotrophs can be readily engineered as endosymbionts within yeast cells, indicating that the nutrient rich environment of the yeast cytoplasm is able to provide most of the lacking metabolites to the cyanobacterial endosymbiont and the cyanobacteria is able to uptake these metabolites without the acquisition of exogenous transporters. As a proof-of-concept experiment, demonstrated that we could delete entire biosynthetic pathways corresponding to one of the metabolites and the generated auxotrophic cyanobacterial endosymbionts can be metabolically coupled to the yeast cells. The accomplishment of this organelle-

like metabolic coupling in synthetic endosymbiotic system where yeast depend on cyanobacteria for ATP and cyanobacteria depend on yeast of amino acids and coenzymes suggests that endogenous metabolite uptake pathways in endosymbionts are sufficient to sustain endosymbiosis. Our observations suggest that during organelle evolution, it may have been possible for the endosymbiont to start losing non-essential genes and begin genome minimization process without the acquisition of exogenous transporters, and subsequent acquisition of exogenous transporters would have improved the metabolic integration of the host and endosymbiont.

**Introduction**

Eukaryotic organelles like mitochondria and chloroplasts evolved from once free-living bacteria that were established as endosymbionts (symbionts inside of the host cell) within the host cell. Because these organelles possess their own genomes, phylogenetic studies and biochemical analysis have provided insights into the precursor free-living bacteria which could have evolved and transformed into modern day mitochondria and chloroplasts. To obtain insights into how a free-living bacteria evolved and transformed into organelles, wide range of organelle genomes have been sequenced and analyzed to predict organelle biochemistry. Further, the organelle genomes have been compared to their possible free-living bacterial relatives to predict the modification that occurred to the bacteria as it became and obligate endosymbiont and eventually into organelles. Organelle biochemistry has also been studied to understand the organelle biochemistry and the role that the modern-day organelles play in the host cell. In addition to this, our observations from naturally existing endosymbiotic systems have provided us with a few snapshots in the evolutionary transformation of a free-living bacteria into an obligate endosymbiont. Few hallmarks of this remarkable evolutionary transformation are (i) metabolic coupling of the host and the endosymbiont, (ii) endosymbiont genome minimization, (iii) loss of endosymbiont peptidoglycan cell-wall, (iv) acquisition of protein exchange systems, amongst others. While we are often able to predict these outcomes of endosymbiosis, we still have limited understanding of the molecular and mechanistic details of this fundamentally important evolutionary transformation.

The focus of these studies is to experimentally investigate and understand the molecular details of how metabolic coupling can be achieved between the host and the endosymbiont. It is important to note that the endosymbionts and organelles possess their own genomes and are transcriptionally and translationally active. Further their genomes are replicated within the endosymbionts or organelles by using a variety of proteins that are encoded by the host genome but are imported into the organelles. All these processes require extensive metabolites like nucleotides, amino acids, and coenzymes. However, the genomes of modern-day organelles and several of the naturally existing endosymbionts like *Paulinella chromatophore* have lost the genes encoding the biosynthesis of amino acids, coenzymes, and nucleotides. They depend on the host genome for these metabolites that they cannot biosynthesize. In case of some amino acids, they are directly imported into the organelle from the metabolic pool in the host cytoplasm and in other cases, the biosynthetic proteins are imported into the organelles and the biosynthesis takes place within the organelle. All these processes in organelles are facilitated by the acquisition of metabolite or protein import systems that are localized to the organelle membrane. It is suggested that the acquisition of the transporters rendered endosymbiont/organelle biosynthetic genes non-essential and eventually resulted in the loss of these biosynthetic pathways from the organelle genomes.

Here, we engineered synthetic endosymbiosis between yeast and cyanobacteria to experimentally investigate how such metabolic coupling can be accomplished between the host (yeast) and their endosymbionts (cyanobacteria). We had previously developed a directed endosymbiosis approach that allows us to engineer cyanobacterial endosymbionts within yeast cells to generate an artificial photosynthetic life-form that is a chimera of yeast and cyanobacteria. Building off this system, here we investigated if we could generate a series of novel auxotrophies in our endosymbiotic cyanobacteria (similar to natural endosymbionts or organelles), introduce them within the yeast cells and generate a yeast/cyanobacteria chimera where the yeast cells provided all of these deficient metabolites to the cyanobacterial endosymbionts and cyanobacterial endosymbionts perform bioenergetic functions for the yeast cells. Our studies suggested that that organelle-like metabolic coupling, which is crucial for stable endosymbiosis, remarkably can be accomplished in synthetic yeast/cyanobacteria endosymbiosis without the acquisition of exogenous endosymbiont transporters and without extensive genome minimization. We also demonstrated that endogenous metabolite uptake pathways in endosymbionts facilitate the metabolic coupling between the host and the endosymbiont. As a proof-of-concept experiment, demonstrated that we could delete entire biosynthetic pathways corresponding to one of the metabolites and the generated auxotrophic cyanobacterial endosymbionts can be metabolically coupled to the yeast cells. Our observations suggest that during organelle evolution, it may have been possible for the endosymbiont to start losing non-essential genes and begin genome minimization process without the acquisition of exogenous transporters, and subsequent acquisition of exogenous transporters would have improved the metabolic integration of the host and endosymbiont.

**Results**

**Engineering single amino acid and coenzyme auxotrophs in cyanobacteria.**

Our first goal was to identify key genes corresponding to the biosynthesis of amino acids and coenzymes in *Synechococcus elongatus* PCC7942 (Syn7942). Although orthologs are annotated in the Syn7942 genome, most have not been characterized. Previously-isolated auxotrophic strains derived from Syn7942 were generated through random mutagenic methods (e.g., UV radiation or nitrosoguanidine treatment)[78], so most auxotrophic phenotypes have not yet been ascribed to the loss of specific gene product(s). The genes we identified, and their corresponding biosynthetic pathways are listed in **Table 3.1**. Next, we constructed integrative plasmids (**Table 3.1**) that would allow us to delete each of the defined genes through homology-directed recombination. Briefly, in each plasmid, 400-1100 bp of homology arms adjacent to or within each of the genes at both 5' and 3' ends were cloned into deletion plasmids. A kanamycin resistance marker was cloned in between the 5' and 3' homology arms to generate a deletion plasmid. Importantly, these deletion plasmids lack an origin of replication, and Syn7942 is immune to most synthetic replicons[92,170]. Therefore, once transformed into Syn7942, resistant colonies were expected to have kanamycin markers incorporated within their genomes in place of targeted genes. As expected, when we transformed our deletion plasmids into Syn7942, we observed colonies that were resistant to kanamycin. In each case, individual colonies were repropagated in growth medium supplemented with kanamycin and defined metabolite(s) complementing the deleted gene. We complemented the growth medium with individual amino acids when a gene corresponding to that amino acid biosynthesis was deleted, e.g., phenylalanine complementation for *pheA* deletion (see **Table 3.1**). Similarly, we complemented the growth medium with defined coenzymes when a gene corresponding to coenzyme biosynthesis

was deleted, e.g., thiamin complementation for *thiC* deletion. After each round of regrowth, we isolated the genomic DNA and PCR-amplified the locus corresponding to the deletion to determine if we had obtained desired recombination at the desired site, e.g., homology-mediated double recombination to incorporate the kanamycin resistance marker at the *pheA* locus when *pheA* deletion plasmid was transformed into Syn7942. Importantly, Syn7942 is polyploid, so we often observed heterozygous mutants lacking the double crossover recombination on all copies of the genome. We segregated these mutant genomes in increasing kanamycin concentrations until we obtained homozygous recombinant mutants. This allowed us to generate cyanobacterial mutants that were auxotrophic to 4 amino acids and 3 coenzymes (**Figure 3.1**, **Figure 3.3**). For all the other tested metabolites, including riboflavin, pyridoxine, and rest of the 15 proteogenic amino acids, we were unable to generate homozygous mutants that were auxotrophic to these individual metabolites (**Figure 3.1**). This was consistent with a previous study which suggested that the uptake of these metabolites might be minimal (if any) due to lack of transporters[82,171]. In fact, it is also suggested that cyanobacterial endosymbionts may have acquired protein or metabolite transport systems that may have facilitated the loss of endosymbiont genome encoded biosynthetic pathways and endosymbiont genome minimization.

**Table 3.1**: Non-replicating plasmids used for single gene deletions in Syn7942

| Name | Target gene | Gene ID | Biosynthetic pathway | Link to Benchling map |
|---|---|---|---|---|
| pML66 | *lysA* | Synpcc7942_0262 | Lysine | pML66 |
| pML67 | *argJ* | Synpcc7942_1896 | Arginine | pML67 |
| pML69 | *hisB* | Synpcc7942_0125 | Histidine | pML69 |
| pML71 | *thiC* | Synpcc7942_1096 | Thiamin | pML71 |
| pML75 | *nadA* | Synpcc7942_1448 | NAD+ | pML75 |
| pML76 | *glyA* | Synpcc7942_0282 | Glycine | pML76 |
| pML77 | *ilvD* | Synpcc7942_0626 | Ile, Leu, Val | pML77 |
| pML78 | *aspC* | Synpcc7942_2545 | Aspartic acid | pML78 |
| pML79 | *thrC* | Synpcc7942_1782 | Threonine | pML79 |
| pML80 | *proA* | Synpcc7942_2265 | Proline | pML80 |
| pML81 | *trpC* | Synpcc7942_1197 | Tryptophan | pML81 |
| pML83 | *bioB* | Synpcc7942_0419 | Biotin | pML83 |
| pML84 | *pdxA* | Synpcc7942_0804 | PLP | pML84 |
| pML85 | *ribF* | Synpcc7942_0492 | FAD | pML85 |
| pML90 | *hisC* | Synpcc7942_1030 | Histidine | pML90 |
| pML91 | *proC* | Synpcc7942_2058 | Proline | pML91 |
| pML92 | *hisS* | Synpcc7942_0813 | Histidine | pML92 |
| pML95 | *glsF* | Synpcc7942_0890 | Glutamic acid | pML95 |
| pML96 | *bioD* | Synpcc7942_0030 | Biotin | pML96 |
| pML101 | *bioF*, *bioH*, *bioC*, *bioD*, *bioA* | Synpcc7942_0027<br>Synpcc7942_0028<br>Synpcc7942_0029<br>Synpcc7942_0030<br>Synpcc7942_0031 | Biotin | pML101 |
| pML102 | *thrB* | Synpcc7942_1440 | Threonine | pML102 |
| pML104 | *cysE* | Synpcc7942_2420 | Cysteine | pML104 |
| pML110 | *gatB* | Synpcc7942_0118 | Asp, Asn | pML110 |
| pML111 | *argF* | Synpcc7942_2514 | Arginine | pML111 |
| pML112 | *asd* | Synpcc7942_1848 | Lysine | pML112 |
| pML113 | *dapA* | Synpcc7942_1847 | Lysine | pML113 |
| pML114 | *argD* | Synpcc7942_0943 | Arginine | pML114 |
| pML115 | *hisD* | Synpcc7942_1519 | Histidine | pML115 |
| SA01 | *pheA* | Synpcc7942_0881 | Phenylalanine | SA01 |

**Table 3.1 (cont.):** Non-replicating plasmids used for single gene deletions in Syn7942

| | | | | |
|---|---|---|---|---|
| SA02 | *aroH* | Synpcc7942_1915 | Tyrosine | SA02 |
| SA03 | *serA* | Synpcc7942_1501 | Serine | SA03 |
| SA14 | *leuB* | Synpcc7942_1505 | Leucine | SA14 |

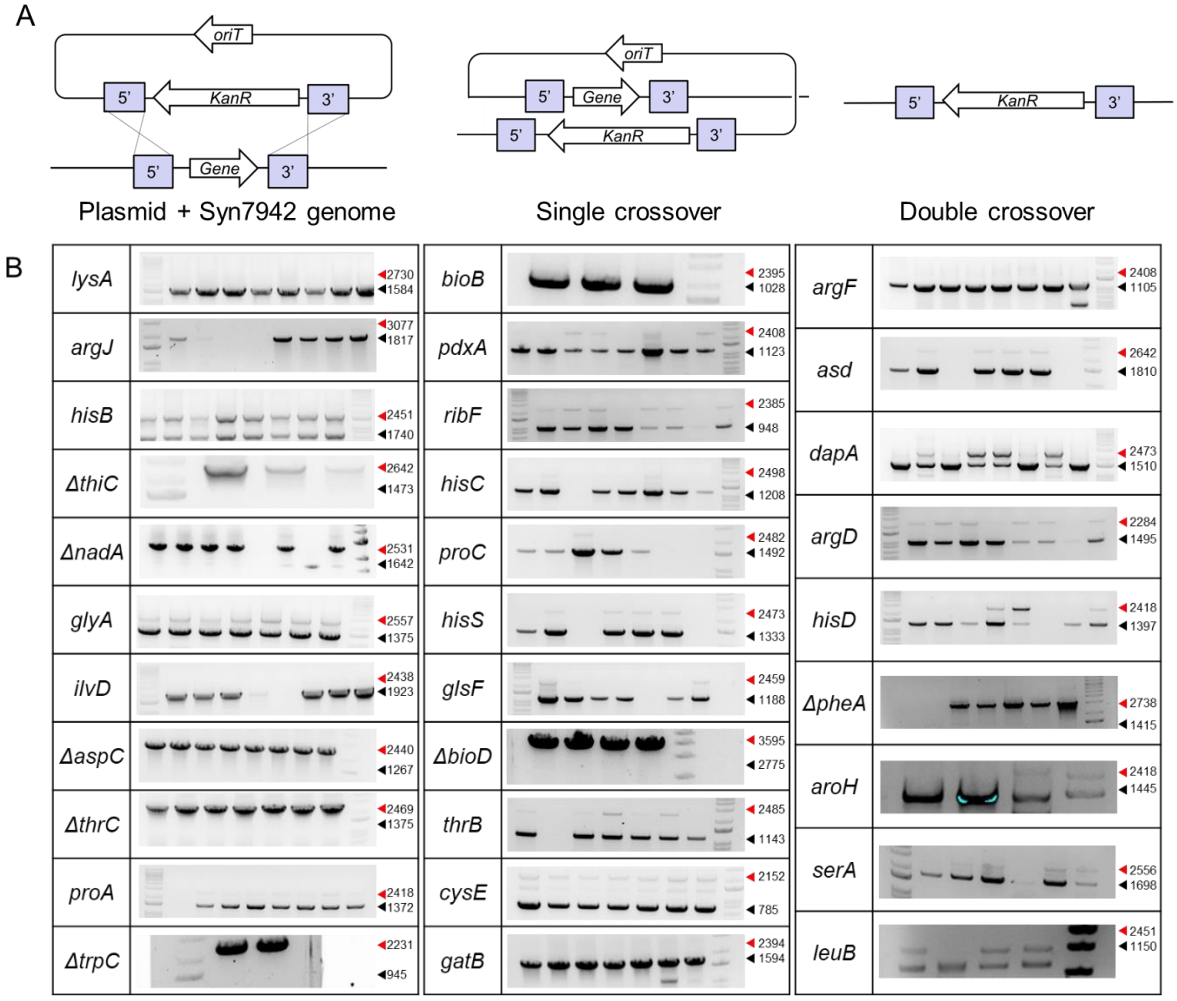



**Figure 3.1**: Single gene deletions of Syn7942 mutants complemented with amino acids or coenzymes. A) Recombination scheme for replacing functional biosynthetic genes with the *kanR* marker B) PCR amplification of Syn7942 mutant genomes. The black arrow indicates the expected bp of the PCR product when amplifying the wild-type allele. The red arrow indicates the expected size of the double crossover recombinant allele. The genes *thiC, nadA, aspC, thrC, trpC, bioD,* and *pheA* are marked with Δ indicating that only the recombinant allele is present in a population of cells tested.

**Organelle-like metabolic coupling in synthetic yeast/cyanobacteria endosymbiosis.**

Modern-day organelles typically have minimal genomes where they have lost the genes encoding the biosynthesis of amino acids, coenzymes, and nucleotides. They depend on the host genome for these metabolites that they cannot biosynthesize and in return the organelles perform bioenergetic functions for the host cells. This extensive metabolic coupling of the host and the organelle is one of the hallmarks of the evolutionary transformation of the endosymbiont into an organelle. We investigated if this organelle-like metabolic coupling could be achieved in synthetic endosymbiosis between cyanobacterial auxotrophs of amino acids and coenzymes and engineered yeast

strains, and if any additional metabolite transporters would be needed for a successful endosymbiosis. We had previously developed a directed endosymbiosis approach that allowed us to engineer wild type Syn7942 strains as endosymbionts (SynJEC3 strains) within yeast mutants, *S. cerevisiae cox2-60*. Under these conditions the SynJEC3 strains provided photosynthetically generated ATP to the yeast mutants deficient in ATP production under selection conditions, thereby generating a cyanobacteria chimera that propagated under photosynthetic selection conditions. We used this directed endosymbiosis approach to investigate if a series of cyanobacterial auxotrophs listed in **Table 3.1** were able to establish endosymbiotic relationship with yeast mutants. Under these conditions, the cyanobacterial endosymbionts would be expected to provide ATP to the yeast cells and the yeast cells provide the auxotrophic amino acids or coenzymes to the cyanobacterial endosymbionts. This would allow us to determine if the endogenous metabolite uptake systems in cyanobacterial endosymbionts are sufficient to uptake metabolites from the host cytoplasm or if additional transporters would be needed to accomplish this.

To investigate this, our first step was to engineer cyanobacterial auxotrophs listed in **Table 3.1** to express a functional ADP/ATP translocase and SNARE-like proteins that are needed for endosymbiosis between Syn7942-derived strains and yeasts, *S. cerevisiae cox2-60*. We had previously generated an integrative plasmid, pML17, that contains an NSII site integrative elements, a pTrc promoter, *Protochlamydia amoebophila* UWE25 ADP/ATP translocase, and SNARE-like proteins *C. tr. incA* and *CT813*. We transformed this plasmid into all the auxotrophic cyanobacterial mutants listed in **Table 3.1**. In each case, we observed individual colonies in BG-11 medium plates containing auxotrophic metabolite (**Table 3.1**) and antibiotics (kanamycin [5-50 mg/L] and chloramphenicol [2-25 mg/L]). Next, we used our fusion protocol to introduce cyanobacteria inside of the yeast cells and select for the generation of yeast/cyanobacteria chimera.

Briefly, we generated spheroplasts of *S. cerevisiae cox2-60* cells and fused engineered cyanobacterial auxotrophs to the *S. cerevisiae cox2-60* cells and selected fusions by growing mixtures on partial selection conditions containing non-fermentable carbon source and low levels of fermentable carbon source (1% yeast extract, 2% peptone, 1 M sorbitol, 3 % glycerol, 0.1 % glucose, 1X BG-11; selection medium I). The fusions were propagated in 12 h light-dark cycles at 30 $^{o}$C. In each case, distinct colonies from *S. cerevisiae cox2-60*-cyanobacteria fusions were picked and re-plated for four consecutive rounds of regrowth: one round on selection medium II (1% yeast extract, 2% peptone, 1 M sorbitol, 3 % glycerol, 0.1 % glucose, 1X BG-11, 50 mg/ml carbenicillin; selection medium II) and two to three rounds on selection medium III (1% yeast extract, 2% peptone, 1 M sorbitol, 3 % glycerol, 50 mg/ml carbenicillin, 1X BG-11; selection medium III) (Figure 2B). Selection medium II and III contained carbenicillin to eliminate any extracellular cyanobacteria. As expected, *S. cerevisiae cox2-60* cells failed to grow on selection medium III during subsequent rounds of regrowth when they were not fused to Syn7942. On the other hand, we observed growth for *S. cerevisiae cox2-60*-cyanobacteria chimeras on selection medium III when most engineered auxotrophic cyanobacteria when fused individually to yeast cells. Only the engineered phenylalanine auxotrophic strain of Syn7942 were unable to establish endosymbiosis with yeast, thereby resulting in lack of growth of fusions under photosynthetic conditions in selection medium III. To characterize the presence of cyanobacterial endosymbionts within yeast cells, in each of the cases we isolated total genomic DNA from fused yeast cells propagated for multiple generations under selection growth conditions and performed PCR analysis to determine the presence of both yeast and cyanobacterial genomes.

We detected the presence of both the yeast *MAT***a** gene and mutant Syn7942 SNARE gene in the colonies by PCR (**Figure 3.4**), suggesting the presence of both yeast and cyanobacterial genomes. We also verified the presence of cyanobacterial endosymbionts within the yeast cells by fluorescence microscopy including total internal reflection fluorescence (TIRF) and fluorescence confocal microscopy (**Figure 3.2**). The number of generations for which each of the endosymbiotic chimera were stable in our hands is listed in **Table 3.2**. These set of experiments suggested that organelle-like metabolic coupling could be achieved in synthetic endosymbiosis between cyanobacterial auxotrophs of four out of five amino acids and five of the coenzymes and engineered yeast strains, without any additional metabolite transporters.

**Table 3.2**: Doublings of yeast-cyanobacteria chimera.

| Endosymbiont genotype | Doublings |
|---|---|
| ***ΔthiC::kan**, NSII::CAT, ntt1, Ctr-incA, CT_813* | 24.8 |
| ***ΔnadA::kan**, NSII::CAT, ntt1, Ctr-incA, CT_813* | 20.1 |
| ***ΔilvD::kan**, NSII::CAT, ntt1, Ctr-incA, CT_813* | 22.8 |
| ***ΔaspC::kan**, NSII::CAT, ntt1, Ctr-incA, CT_813* | 24.9 |
| ***ΔthrC::kan**, NSII::CAT, ntt1, Ctr-incA, CT_813* | 21.0 |
| ***ΔtrpC::kan**, NSII::CAT, ntt1, Ctr-incA, CT_813* | 18.4 |
| ***ΔproC::kan**, NSII::CAT, ntt1, Ctr-incA, CT_813* | 17.9 |
| ***ΔglsF::kan**, NSII::CAT, ntt1, Ctr-incA, CT_813* | 19.8 |
| ***ΔbioFHCDA::kan**, NSII::CAT, ntt1, Ctr-incA, CT_813* | 18.4 |
| ***ΔcysE::kan**, NSII::CAT, ntt1, Ctr-incA, CT_813* | 20.2 |
| ***ΔgatB::kan**, NSII::CAT, ntt1, Ctr-incA, CT_813* | 23.4 |
| ***ΔargF::kan**, NSII::CAT, ntt1, Ctr-incA, CT_813* | 14.5 |
| ***Δasd::kan**, NSII::CAT, ntt1, Ctr-incA, CT_813* | 18.1 |
| ***ΔhisD::kan**, NSII::CAT, ntt1, Ctr-incA, CT_813* | 23.0 |
| ***ΔpheA::kan**, NSII::CAT, ntt1, Ctr-incA, CT_813* | N/A |
| ***ΔaroH::kan**, NSII::CAT, ntt1, Ctr-incA, CT_813* | 19.8 |
| ***ΔserA::kan**, NSII::CAT, ntt1, Ctr-incA, CT_813* | 21.7 |
| ***ΔleuB::kan**, NSII::CAT, ntt1, Ctr-incA, CT_813* | 20.9 |

**Generating novel auxotrophies in cyanobacteria by deleting biosynthetic genes that were previously identified as essential genes.**

As mentioned in the results section above, we were able to generate amino acid auxotrophs for 4 of the 20 proteogenic amino acids and not for the other 15 proteogenic amino acids under the complementation of single amino acids. This was likely due to the lack of transporters to efficiently uptake these metabolites. Our next goal was to investigate if we could develop approaches that would allow us to determine if the biosynthetic pathways corresponding to these other 15 proteogenic amino acids can be rendered non-essential if a right combination of a metabolite and transporter is identified that complements this deficiency. An important observation we made from previous studies was that while 100-150 transporters involved in plastid metabolite exchange in the *Archaeplastida*, solute transporters are nearly absent in the cyanobacterial endosymbiont (*chromatophore*) in a naturally existing endosymbiotic system *Paulinella chromatophora*. Importantly, amino acid and coenzyme biosynthetic pathways are lost in both the chloroplast genomes of *Archaeplastida* and the endosymbiont genome of the *Paulinella chromatophore*. These observations led us to speculate that it might be possible that cyanobacteria may have alternate mechanisms to complement the lack of amino acid biosynthesis. Previous studies have suggested that cyanobacteria

often exchange nutrients via exchanging short peptides, e.g., arginine-derived dipeptides are used as a source of nitrogen by *Anabaena* strains of cyanobacteria. Based on the genome sequence of Syn7942, we observed that Syn7942 genome encodes a putative dipeptide permease annotated as *appB* (**Figure 3.7**) . With this premise, we investigated if dipeptides could complement lack of amino acid biosynthesis for the remaining proteogenic amino acids (listed in **Table 1**). Since lysine-based dipeptides have been previously shown to be transported by other gram-negative bacteria expressing dipeptide permeases, we synthesized a series of lysine-containing dipeptides (**Table 3.3**, **Figure 3.8**). Next, we transformed our deletion plasmids (corresponding to the biosynthetic genes of remaining 14 proteogenic amino acids, excluding alanine) into Syn7942 and propagated the transformants in BG-11 media containing supplement groups of closely synthetically related amino acids[77] (**Table 3.4**) and kanamycin. Individual colonies were repropagated in this medium, total genomic DNA was isolated and characterized by PCR to identify Syn7942 that contained desired double recombinant. Under these conditions, all the Syn7942 mutants were heterozygous, i.e., mutants where one or more, but not all, copies of chromosome had a desired, double crossover recombination. We then propagated these mutant strains in BG-11 medium containing kanamycin and a corresponding lysine-containing dipeptide, for example, a heterozygous mutant containing *thrC* deletion corresponding to threonine biosynthesis was complemented with lysine-threonine dipeptide. We then investigated if this approach resulted in a homozygous strain where all copies of chromosome had a desired double crossover recombination. Remarkably, this approach allowed us to obtain homozygous mutants corresponding to deletions in key biosynthetic genes corresponding to 14 out of the remaining 15 amino acids (**Figure 3.9**). We then propagated these mutant strains in BG-11 medium containing either single amino acid or a corresponding lysine-containing dipeptide, for example, homozygous *thrC* deletion mutant (corresponding to threonine biosynthesis) was complemented was propagated either in BG-11 medium containing threonine or BG-11 medium containing lysine-threonine dipeptide. We only observed significant growth in BG-11 medium complemented with a corresponding lysine-containing dipeptide and not on growth BG-11 medium complemented with a single amino acid (**Figure 3.9**). This suggests that dipeptides are being transported by these Syn7942-derived homozygous mutant strains, but the corresponding single amino acid are not efficiently up taken by Syn7942 derived strains. This novel observation allowed us to delete a series of biosynthetic genes that were previously rendered as essential in Syn7942.

**Table 3.3**: Dipeptides and corresponding single gene deletions.

| Dipeptide(s) | Complemented gene deletion(s) |
|---|---|
| $NH_2$-Lys-Leu-OH, $NH_2$-Lys-Val-OH | *ilvD* |
| $NH_2$-Lys-Thr-OH | *thrC*, *asd* |
| $NH_2$-Lys-Pro-OH | *proA* |
| $NH_2$-Lys-Glu-OH | *glsF* |
| $NH_2$-Lys-Cys-OH | *cysE* |
| $NH_2$-Lys-Gln-OH, $NH_2$-Lys-Asn-OH | *gatB* |
| $NH_2$-Lys-Arg-OH | *argF*, *argD* |
| $NH_2$-Lys-His-OH | *hisD* |
| $NH_2$-Lys-Tyr-OH | *aroH* |
| $NH_2$-Lys-Ser-OH | *serA* |

**Testing dipeptide auxotrophs of cyanobacteria as endosymbionts in yeast.**

We next investigated if we could engineer these cyanobacterial mutants that are dependent on dipeptides for growth can be engineered as endosymbionts within yeast cells without adding any exogenous amino acid or protein transporters to the cyanobacterial mutants. To investigate this, we first transformed an integrative plasmid, pML17 (containing an NSII site integrative elements, a Ptrc promoter, *Protochlamydia amoebophila* UWE25 ADP/ATP translocase, and SNARE-like proteins *C. tr. incA* and *CT813*) into homozygous cyanobacterial mutants listed in **Table 3.1**. In each case, we observed individual colonies in BG-11 medium plates containing auxotrophic dipeptides (**Table 3.2**) and antibiotics (kanamycin [5-50 mg/L] and chloramphenicol [2-25 mg/L]). Next, we used our fusion process to introduce cyanobacteria inside of the yeast cells and then propagate them under selection conditions as before. Remarkably, all the yeast cyanobacteria chimera were able to propagate under the most stringent selection conditions (selection medium III under photosynthetic conditions). We confirmed the presence of cyanobacterial endosymbionts within these yeast cells as before by performing PCR amplification of defined loci from total genomic DNA and pTIRF microscopy. These experiments suggested that components of the yeast cytoplasm were able to sustain the growth of the endosymbiotic cyanobacteria that were unable to uptake single amino acids but were dependent of dipeptides under independent culture conditions. Our observations suggest that during organelle evolution, it may have been possible for the endosymbiont to start losing non-essential genes and begin genome minimization process without the acquisition of exogenous transporters, and subsequent acquisition of exogenous transporters would have improved the metabolic integration of the host and endosymbiont. This also sheds insights into possible reasons for why 100-150 transporters involved in plastid metabolite exchange in the *Archaeplastida*, but in the evolutionarily more recent endosymbiotic system of *Paulinella chromatophore* the solute transporters are nearly absent in the cyanobacterial endosymbiont (*chromatophore*).

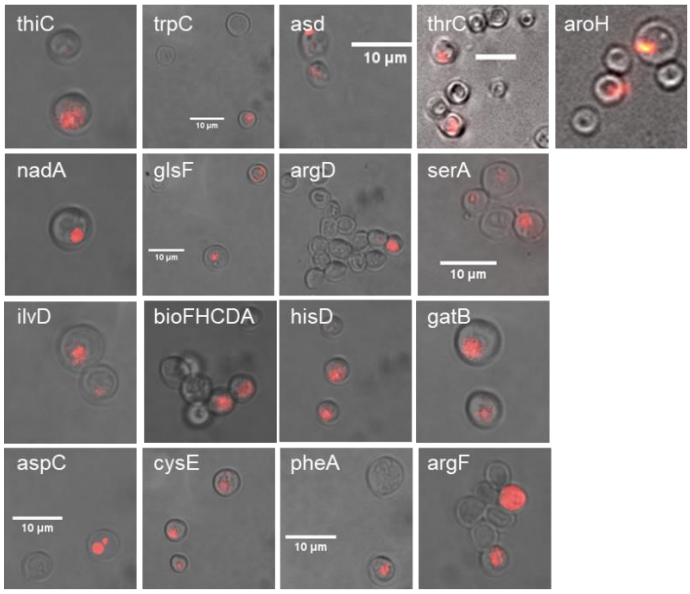



**Figure 3.2**. pTIRF characterization of auxotrophic cyanobacterial cells fused with *S. cerevisiae cox2-60* cells. The images are merged brightfield (B&W) and cyanobacterial auto-fluorescence (PBS, red) channels.

**Deleting entire biosynthetic operons in cyanobacteria and testing them as endosymbionts within yeasts.**

As a proof-of-concept experiment, we next investigated if we could delete entire biosynthetic pathways corresponding to one of the above metabolites and demonstrate that the mutant strain was compatible for metabolic coupling in our synthetic yeast/cyanobacteria endosymbiotic system. Given that the biosynthetic operon for biotin biosynthesis appeared to be clustered in Syn7942, we designed homology directed deletion plasmids which contained homology arms on both the 5'-end and the 3'-end of the biotin biosynthetic operon. We transformed this plasmid into Syn7942, picked individual colonies and propagated them in BG-11 medium containing biotin and kanamycin until we observed homozygous mutants with desired double crossover recombinant where the entire biotin biosynthetic operon was deleted and replaced by kanamycin resistant cassette (**Figure 3.10**). We then transformed pML17 into these mutants. We then fused these strains to yeast mutants as before and were able to generate yeast/cyanobacteria chimera (**Figure 3.2**, **Table 3.2**). This serves as a proof-of-concept experiment that demonstrates that we can use this approach to systematically test the genome minimization of our synthetic endosymbiont genome.

**Discussion**

Metabolic coupling of the host and endosymbiont is critical for endosymbiosis and is a key feature of the evolution of organelles which are essentially one of the key end points of endosymbiosis. Metabolic coupling allows renders several biosynthetic pathways, including those corresponding to amino acids, coenzymes, and nucleotides, in the endosymbiont to be non-essential. Such non-essential biochemical pathways are lost as the endosymbiont evolves into organelles. Along this evolutionary trajectory, the endosymbiont also acquires transporters which facilitate the uptake of metabolites from the host cells. Further, certain proteins acquire a signal sequence that allows them to be imported into the endosymbionts/organelles. All of this results in the loss of non-essential genes from the endosymbiont genome and transfer of genes to the host nuclear genome. This extensively decreases the metabolic burden of the endosymbiont, eventually transforming it into an organelle that specializes in supporting the bioenergetic functions of the host cell and nutrient assimilation (e.g., carbon assimilation by chloroplasts) for the host cell. While we can predict some of these key outcomes of endosymbiosis, the order and the molecular details of these evolutionary events are unclear. For example, it is unclear at which step non-essential genes were lost, transporters were acquired, and genome minimization was attained. As described in this manuscript, we have begun to experimentally test how metabolic coupling can be achieved between the endosymbiont and the host.

For our studies, we engineered synthetic yeast/cyanobacteria endosymbiosis to test how metabolic coupling can be achieved between the host (yeast) and the endosymbiont (cyanobacteria). As a first step, we generated a series of single amino acid auxotrophs of cyanobacterial endosymbionts and tested if these strains were able to establish endosymbiosis with the yeast cells. Under selection conditions, the cyanobacterial mutants provided photosynthetically generated ATP to the yeast cells and the yeast cells provided defined amino acids to the endosymbiont, essentially metabolically coupling the endosymbiont to the host cell. We were able to engineer coenzyme auxotrophs of cyanobacterial endosymbionts and demonstrate that the yeast cells were able to provide 3 coenzymes to the cyanobacterial endosymbiont resulting in the generation of yeast/cyanobacteria chimera that was stable for >20 generations. Similarly, we were able to demonstrate that yeast cells were able to provide 15 amino acids to the cyanobacterial endosymbiont resulting in the generation of yeast/cyanobacteria chimera that was stable for >20

generations. Despite our extensive efforts we were unable to generation auxotrophs for all the other amino acids under corresponding single amino acid complementation conditions. Therefore, we investigated if there are other transport mechanisms for amino acid derivatives that could be potentially leveraged to generate homozygous cyanobacterial mutants that were deficient in the biosynthesis of these amino acids. Particularly, based on previous studies, we noted that while 100-150 transporters involved in plastid metabolite exchange in the *Archaeplastida*, solute transporters are nearly absent in the cyanobacterial endosymbiont (*chromatophore*) in a naturally existing endosymbiotic system *Paulinella chromatophora*. Importantly, amino acid and coenzyme biosynthetic pathways are lost in both the chloroplast genomes of *Archaeplastida* and the endosymbiont genome of the *Paulinella chromatophore*. These observations led us to speculate that it might be possible that cyanobacteria may have alternate mechanisms to complement the lack of amino acid biosynthesis. Previous studies have suggested that cyanobacteria often exchange nutrients via exchanging short peptides, e.g., arginine-derived dipeptides are used as a source of nitrogen by *Anabaena* strains of cyanobacteria. Based on genome sequence of Syn7942, we observed that the Syn7942 have a putative dipeptide permease (annotated as *appB*). Further, certain strains of cyanobacteria like *Anabaena* use dipeptides for nutrient exchange. Therefore, we tested if we could delete essential genes corresponding to the biosynthesis of the other 15 amino acids and complement this deficiency by using dipeptides. This approach proved to be successful in deleting key biosynthetic enzymes corresponding to these 15 amino acids to generate cyanobacterial mutants that were now auxotrophic for dipeptides instead of single amino acids. When we introduced these cyanobacterial mutants as endosymbionts within the yeast cells, remarkably we were able to generate yeast/cyanobacteria chimera. This implied the yeast were able to provide a dipeptide (or some form of an oligopeptide) that was able to complement the endosymbiont deficiency under defined photosynthetic growth conditions. Additionally, we were able to delete entire representative biosynthetic operon in cyanobacteria and demonstrate that the yeast cells were able to provide 3 coenzymes to the cyanobacterial endosymbiont resulting in the generation of yeast/cyanobacteria chimera that was stable for >20 generations. As a proof-of-concept experiment, demonstrated that we could delete entire biosynthetic pathways corresponding to one of the metabolites and the generated auxotrophic cyanobacterial endosymbionts can be metabolically coupled to the yeast cells.

Our studies suggested that that organelle-like metabolic coupling, which is crucial for stable endosymbiosis, remarkably can be accomplished in synthetic yeast/cyanobacteria endosymbiosis without the acquisition of exogenous endosymbiont transporters and without extensive genome minimization. The endogenous metabolite uptake pathways in endosymbionts were able to facilitate the metabolic coupling between the host and the endosymbiont. Our observations suggest that during organelle evolution, it may have been possible for the endosymbiont to start losing non-essential genes and begin genome minimization process without the acquisition of exogenous transporters, and subsequent acquisition of exogenous transporters would have improved the metabolic integration of the host and endosymbiont. Further, we believe that these studies set up a platform to achieve cyanobacterial genome minimization now systematically and artificially generate chloroplast-like organelles in yeast cells.

**Methods**

**Strains**

*S. elongatus* PCC 7942 was obtained from the lab of Prof. Susan Golden (UCSD). *S. cerevisiae* ρ + NB97 (MAT**a** *leu2-3,112 lys2 ura3-52 his3Δ HindIII arg8Δ::URA3* [*cox2-60*])[172] was obtained from the lab of Peter Schultz (Scripps Research). *E. coli* BL21(DE3) and 5-alpha strains were purchased from New England Biolabs. *E. coli* One Shot® *ccdB* Survival™ 2 T1R Chemically Competent Cells were purchased from Invitrogen.

**Growth medium:**

Cyanobacteria were cultured in BG-11 medium (18 mM $NaNO_3$, 0.2 mM $CaCl_2$, 23 μM $C_6H_9FeNO_7$, 31 μM $C_6H_8O_7$, 3.0 μM EDTA disodium salt, 0.2 mM $K_2HPO_4$, 0.6 mM $MgSO_4$, 0.2 mM $Na_2CO_3$, 46.3 μM $H_3BO_3$, 9.1 μM $MnCl_2$, 0.8 μM $ZnSO_4$, 1.6 μM $Na_2MoO_4$, 0.3 μM $CuSO_4$, 0.2 μM $Co(NO_3)$, pH 7.5). Liquid cultures were shaken aerobically (250 rpm) in sterile Erlenmeyer flasks at 37 °C under 70-90 μmol photosynthetically active photons·$m^{-2}s^{-1}$ emitted by fluorescent bulbs. Agar plates were supplemented with $Na_2S_2O_3$ (100 μM) and grown at 30-34 °C under equivalent PPFD emitted by an LED grow lamp. Unless otherwise noted, single amino acids and dipeptides were supplemented to a final concentration of 1 mM, and the pH of the medium was adjusted back to 7.5 before sterilization. Complete-I medium was prepared as previously reported[77]. Stock solutions (100X) were made containing 1) 100 mM Glu, 100 mM Pro, and 100 mM Arg; 2) 100 mM Asp, 100 mM Lys, 100 mM Thr, and 100 mM Met; 3) 5 mM Tyr, 5 mM Trp, 5 mM Phe, and 5 mM His; and 4) 100 mM Ser, 100 mM Gly, and 50 mM Cys. The four stock solutions were added to 2X BG-11 medium, and the pH was adjusted to 8.0. $NaHCO_3$ (1 g/L) was added to liquid medium only, and the volume was adjusted to reach 1X concentrations of the salts and amino acids. For solid medium, 2X BG-11 was supplemented with the stock solutions without $NaHCO_3$. Agar suspension in water (8% w/v) and the 2X Complete-I medium were autoclaved separately and then combined before pouring into plates.

Yeast cells were shaken aerobically at 30 °C and 250 rpm in YPD medium (1% yeast extract, 2% peptone, 2% glucose) containing carbenicillin (50 mg/L). Selection medium I contained 1% yeast extract (Y), 2 % peptone (P), 3 % glycerol (G), 0.1 % glucose (D), 1 M sorbitol, 2 % agar, and 1X BG-11. Selection medium II contained 1% Y, 2 % P, 3 % G, 0.1 % D, 1 M sorbitol, 2 % agar, 1X BG-11 salts, and carbenicillin. Selection medium III contained 1% Y, 2 % P, 3 % G, no D, 1 M sorbitol, 2 % agar, 1X BG-11 salts, and carbenicillin. Plates containing Selection medium I, II and III were grown at 30 °C under 70-90 μmol photosynthetically active photons·$m^{-2}s^{-1}$ emitted by fluorescent bulbs.

**Construction of plasmids:**

pET28-FLUC: A DNA fragment encoding the *Photinus pyralis* luciferase was codon-optimized for *E. coli* using IDT codon optimization software (https://www.idtdna.com/CodonOpt). The double-stranded DNA was purchased from IDT. The gene fragment was amplified by PCR (PrimeSTAR polymerase, Takara) using the primers BD3A and BD3B. The commercial vector pET28 was amplified by PCR using the primers SB125A and SB125B. The PCR products were cycle-purified after DpnI (NEB R0176S) digestion at 37 °C for 1 hour. The fragments were cloned into pET28-FLUC by Gibson assembly[173] and transformed into to *E. coli* 5-alpha chemically competent cells.

Gene deletion plasmids were all cloned by Gibson assembly of four DNA fragments amplified by PCR. Homology regions were amplified from purified Syn7942 genomic DNA template using Q5 Hot Start High-Fidelity

2X Master mix (NEB M0494S). Plasmid backbone regions containing the origin and antibiotic selection marker were amplified by PCR in the same fashion using the plasmid pML58[75] as the template. The amplicons from plasmid DNA were treated with DpnI and purified afterwards. After assembly, the plasmids were transformed into One Shot *ccdB* Survival cells. Plasmid clones were isolated from *E. coli* and purified to concentrations of 25-150 ng/μL and confirmed by Sanger sequencing.

| | 5' homology | | 3' homology | | Backbone *ori* | | Backbone *kanR* | |
|---|---|---|---|---|---|---|---|---|
| Plasmid | Primer 1 | Primer 2 | Primer 1 | Primer 2 | Primer 1 | Primer 2 | Primer 1 | Primer 2 |
| pML66 | JC284 | JC287 | LL36 | JC315 | JC314 | JC285 | JC286 | LL34 |
| pML67 | JC289 | JC291 | JC261 | JC263 | JC262 | JC288 | JC290 | JC260 |
| pML69 | ZD15 | ZD13 | ZD9 | ZD10 | ZD16 | ZD11 | ZD14 | ZD12 |
| pML71 | JC280 | JC282 | JC276 | JC277 | JC281 | JC279 | JC283 | JC278 |
| pML75 | JC379 | JC380 | JC376 | JC375 | JC378 | JC382 | JC381 | JC377 |
| pML76 | JC385 | JC386 | JC389 | JC390 | JC391 | JC392 | JC387 | JC388 |
| pML77 | JC393 | JC394 | JC397 | JC398 | JC399 | JC432 | JC395 | JC396 |
| pML78 | JC424 | JC425 | JC427 | JC428 | JC429 | JC430 | JC426 | JC431 |
| pML79 | JC416 | JC417 | JC420 | JC421 | JC422 | JC423 | JC418 | JC419 |
| pML80 | JC408 | JC409 | JC412 | JC413 | JC414 | JC415 | JC410 | JC411 |
| pML81 | JC400 | JC401 | JC404 | JC405 | JC406 | JC407 | JC402 | JC403 |
| pML83 | JC475 | JC476 | JC479 | JC480 | JC481 | JC482 | JC477 | JC478 |
| pML84 | JC467 | JC468 | JC471 | JC472 | JC474 | JC473 | JC469 | JC470 |
| pML85 | JC459 | JC460 | JC463 | JC464 | JC465 | JC466 | JC461 | JC462 |
| pML90 | JC518 | JC520 | JC523 | JC524 | JC519 | JC525 | JC521 | JC522 |
| pML91 | JC510 | JC512 | JC515 | JC516 | JC511 | JC517 | JC513 | JC514 |
| pML92 | JC502 | JC504 | JC507 | JC508 | JC503 | JC509 | JC505 | JC506 |
| pML95 | JC543 | JC544 | JC547 | JC548 | JC549 | JC550 | JC545 | JC546 |
| pML96 | YG156 | YG158 | YG161 | YG162 | YG157 | YG163 | YG159 | YG160 |
| pML101 | JC556 | JC557 | JC560 | JC561 | JC555 | JC562 | JC558 | JC559 |
| pML102 | JC563 | JC565 | JC568 | JC569 | JC564 | JC570 | JC566 | JC567 |
| pML104 | JC575 | JC576 | JC579 | JC580 | JC581 | JC582 | JC577 | JC578 |
| pML110 | JC637 | JC638 | JC641 | JC642 | JC643 | JC644 | JC639 | JC640 |
| pML111 | JC677 | JC678 | JC681 | JC682 | JC683 | JC684 | JC679 | JC680 |
| pML112 | JC655 | JC656 | JC659 | JC660 | JC661 | JC662 | JC657 | JC658 |
| pML113 | JC647 | JC648 | JC649 | JC650 | JC653 | JC654 | JC651 | JC652 |
| pML114 | JC669 | JC670 | JC673 | JC674 | JC675 | JC676 | JC671 | JC672 |
| pML115 | JC685 | JC686 | JC689 | JC690 | JC691 | JC692 | JC687 | JC688 |
| SA01 | JC210 | JC208 | JC213 | JC214 | JC215 | JC211 | JC209 | JC212 |
| SA02 | ZD3 | ZD4 | ZD7 | ZD2 | ZD6 | ZD1 | ZD5 | ZD8 |
| SA03 | SA40 | SA41 | SA42 | SA43 | SA45 | SA38 | SA39 | SA44 |
| SA14 | SA62 | SA64 | SA60 | SA57 | SA63 | SA59 | SA58 | SA61 |

**Site-directed mutagenesis in cyanobacteria:**

Syn7942 were transformed according to Clerico, Golden, and co-workers[101]. Syn7942 cultures were grown to mid-log phase ($OD_{750}$ of 0.2—0.6) and 15 mL portions were collected by centrifugation (10 min, 3000 × g, rt). The pellet was washed once with NaCl (10 mL, 10 mM) and resuspended in BG-11 medium (0.3 mL, rt) in a microcentrifuge tube. For targeted gene deletions, the BG-11 medium was supplemented with amino acid(s) or coenzyme(s) corresponding to the presumed pathway which was to be disrupted; e.g., 1 mM L-tryptophan was added to the *trpC* transformation suspension. Purified plasmid (1.5 μL) was added to the cell suspension, and it was shaken

overnight in the dark (12–20 h, 70 RPM). Agar plates containing kanamycin (5-50 mg/L) were prepared and allowed to sit at ambient temperature for at least 12 h before selection or re-streaking of the cells.

Luciferase overexpression:

pET28-FLUC was transformed into *E. coli* BL21(DE3) cells. The culture (1.5 L) was shaken (37 °C, 200 RPM) until reaching optical density ($OD_{600}$) of 0.6. IPTG was added to a final concentration of 0.1 mM and the culture was incubated longer (20 h, 37 °C, 180 RPM). The cells were then collected by centrifugation (10,000×g, 15 min). The pellet was resuspended in 50 mL of lysis buffer (50 mM $Na_2HPO_4$, 300 mM NaCl, 20 mM imidazole, 1mM DTT, 10% glycerol, pH 7.5), and sonicated for three cycles (40% amplitude, pulse on for 1 sec, pulse off for 1 sec, total time 30 sec, rest in ice for 10 min). After sonication, the crude protein mixture was centrifuged (25,000×g , 20 min). The ice-cold supernatant was filtered through a 0.45 µm syringe filter into a separate 50 mL tube. A Ni-NTA affinity column was washed with 40 mL water, followed by washing with 40 mL lysis buffer. The ice-cold crude protein supernatant was loaded onto the preactivated Ni-NTA affinity column. Enzyme-bound affinity column was then washed with 20 mL binding buffer to wash off the partial unbound protein. The bound protein was eluted with 3-5 mL elution buffer (50 mM $Na_2HPO_4$, 300 mM NaCl, 250 mM imidazole, 1mM DTT, 10% glycerol, pH 7.5). The total protein was quantified ($A_{280}$) by NanoDrop.

For desalting, a new PD-10 column was washed twice with deionized water and then with 10 mL storage buffer. The protein fraction containing luciferase was desalted through the activated PD-10 column and eluted in storage buffer (50 mM $Na_2HPO_4$, 100 mM NaCl, 10% glycerol, 1 mM DTT, pH 7.5). Desalted fractions were flash-frozen using liquid nitrogen and stored at – 80 °C.

**Growth curves of Syn7942 mutants auxotrophic for single amino acids or coenzymes**

Log-phase auxotrophic Syn7942 strains were diluted in BG-11 medium ($OD_{750}$<0.05) supplemented with kanamycin (50 mg/L) and nutrient(s) listed below. Amino acids were added to a final concentration of 1 mM, while all other nutrients were added to a final concentration of 2 mg/L. Aliquots of the dilute culture (200 µL) were seeded into clear 96-well plates and incubated statically at 30 °C under 70-90 µmol photosynthetically active photons·$m^{-2}s^{-1}$ emitted by fluorescent bulbs. $OD_{730}$ measurements were taken periodically using a BioTek Synergy H1 spectrophotometer.

**Luciferase standard curve:**

The activity and sensitivity of purified luciferase was evaluated by generating a standard curve. D-luciferin potassium salt was dissolved to a concentration of 50 µM in commercial buffer (Invitrogen A22066) containing tricine (25 mM, pH 7.8), $MgSO_4$ (5 mM), EDTA (100 µM) and DTT (1 mM). Purified luciferase enzyme was added to a final concentration of 1.25 µg/mL and the assay mixture was stored on ice and shielded from light. Serial dilutions of ATP, ADP and GTP were prepared in TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0) in an opaque-walled 96-well plate. Luciferase reactions were carried out with 10 µL of substrate and 90 µL of reaction buffer containing the enzyme. Luminescence was measured using a BioTek Synergy H1 spectrophotometer.

**Measurement of ATP release by cyanobacteria mutants expressing ATP/ADP translocase:**

ADP solution (Sigma A2754) was treated with hexokinase to remove contaminating ATP[174]. ADP (80 mM, pH 7.5) was incubated with D-glucose (200 mM), $MgCl_2$ (2 mM) and hexokinase (Sigma H4502-500UN) (0.04 U/µL)

at rt for 2 h. Hexokinase was then removed from the solution by centrifugation through an Amicon Ultra 0.5 column (14,000 × g, 15 min).

Cyanobacteria transformed with pML17 were grown to a density of ~$3\times10^7$ cells/mL. For each assay, 10 mL of cells were harvested by centrifugation (3000 × g, 5 min, rt) and washed once with 20 mM Tris-HCl. ADP (80 μM) was added, and the cells were incubated statically at 37 °C. The cell suspensions were centrifuged (10,000 × g, 5 min) and the supernatant was removed carefully to prevent any carryover of the cell pellet. Supernatant fractions (10 μL) were transferred to an opaque-walled 96-well plate and ATP concentrations were determined by luciferase assay (10 μL sample, 90 μL enzyme/reaction buffer) against the standard curve. Luminescence was measured using the Synergy H1 spectrophotometer.

**Introduction of mutant cyanobacteria to S. cerevisiae cells:**

Syn7942 cells were introduced to *S. cerevisiae* spheroplasts as reported previously[75]. Mutant cyanobacteria cultures (30 mL) were harvested (3000 × g, 10 min, rt) in mid-log phase, washed twice with BG-11 medium and resuspended in BG-11 medium (500 μL). *S. cerevisiae cox2-60* cells were grown aerobically in YPD medium (300 mL) overnight. The optical density ($OD_{600}$) of the yeast culture was monitored closely as to not exceed 1.0 because *S. cerevisiae* cells in stationary phase are resistant to lytic enzymes[175,176]. The yeast was harvested (4696 × g, 10 min, rt), washed twice with water, twice with SCEM (1 M sorbitol, 13 mM β-mercaptoethanol) and resuspended in ice-cold SCEM (10 mL) containing 8 mg of Zymolyase 100 T (Amsbio 120493-1). The suspension was incubated for 30-60 min at 37 °C. The formation of spheroplasts was monitored by diluting 1 μL of cell suspension into 7 μL of water; Zymolyase treatment was ended when all the yeast cells were shown by light microscopy to rupture, indicating that the cell wall digestion was complete. The spheroplast suspension was placed on ice for 15 min and centrifuged (1500 × g, 10 min, 4 °C). The pellet was washed twice with chilled SCEM, resuspended in SCEM (2.0 mL) and kept refrigerated for up to 24 h.

The spheroplasts (750 μL) were mixed with chilled TSC buffer (10 mM Tris-HCl, 10 mM $CaCl_2$, 1 M sorbitol, pH 8) (750 μL) and incubated for 10 min at 30 °C. The mixture was centrifuged (1500 × g, 10 min) and resuspended in room-temperature TSC buffer (120 μL) in a microcentrifuge tube. Sorbitol (4 M, 60 μL) was then added. Syn7942 cells (120 μL) were added to the spheroplast suspension, mixed by inversion of the tube, and incubated statically (30 min, 30 °C). The mixture was decanted into a round-bottom polypropylene tube containing PEG buffer (20% PEG 8000, 10 mM Tris-HCl, 2.5 mM $MgCl_2$, 10 mM $CaCl_2$, pH 8) (2 mL) and incubated statically (45 min, 30 °C). The cells were centrifuged (1500 × g, 10 min, rt), the supernatant was discarded and YPDS (YPD with 1 M sorbitol added) was added on top without disturbing the pel let. The cells were incubated statically under light (2 h, 30 °C). The cell pellet was then dislodged by flicking the side of the tube and allowed to incubate longer (3 h, with shaking at 70 rpm). The cells were harvested (1500 × g, 10 min, rt), resuspended in 1 M sorbitol (300 μL) and spread on Selection-I bottom agar medium. After drying for 5 min, a top layer of Selection-I medium was overlaid on the cells. The plates were incubated at 30 °C in a 12 h light-dark cycle for 4 d, until colonies appeared between the agar layers. The colonies were extracted from the agar, suspended in 1 M sorbitol, and spotted on Selection-II medium. For subsequent rounds of propagation, cells were scraped from the surface of the agar, resuspended in 1 M sorbitol, and spotted on Selection-III medium.

**Cell count of chimeras:**

Sections of Selection-III medium containing individual spots were extracted and placed in a microcentrifuge tube. The cells were resuspended in sorbitol (1 M, 200 µL) on the agar surface and separated by brief centrifugation (5 s). The cells were mounted on a reusable glass slide (Invitrogen A25750) and counted with a Countess II FL Automated Cell Counter (Fisher cat. AMQAF1000).

**Genomic DNA isolation and PCR analysis:**

Double crossover recombination in Syn7942 colonies was routinely checked by colony PCR (GoTaq Green Master Mix, Promega M7123). Colonies which showed only one band corresponding to double crossover recombination were used to inoculate flasks containing supplemented BG-11 medium. Genomic DNA was isolated from these flasks using the Purelink Genomic DNA Mini Kit (Thermo Fisher K182001) according to manufacturer's protocol following treatment with lysozyme (20 mg/mL). The same primers were used to check for recombination in purified genomic DNA as with colony PCR, listed in **Table 3.6**.

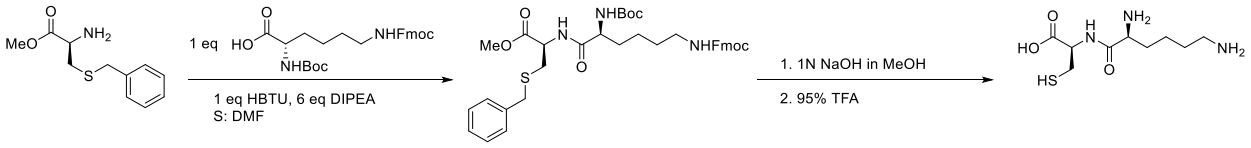
Total genomic DNA of yeast-cyanobacteria chimeras was isolated with the Yeast DNA Extraction Kit (Thermo Fisher 78870) according to the manufacturer's protocol. PCR of the yeast *MAT* allele and cyanobacterial *C. tr. incA* gene was performed as previously reported[75].

**pTIRF microscopy:**

Cells were scraped from Selection-III agar plates and washed once with PBS. They were analyzed with a TIRF microscope based on a Zeiss Axiovert 200 M stand as reported previously[75]. The sample was excited with a Cobolt diode-pumped 561 nm laser filtered through an acousto-optic tunable filter (AOTF, Quanta Tech Inc). The emission filter was a Chroma bandpass filter HQ 653/95 nm. The pTIRF images were acquired with a Photometric 512 Evolve EMCCD camera. Samples were viewed and imaged using a 100X oil immersion objective lens with NA = 1.4. Brightfield and fluorescence images were taken at the same sample position, and the images were merged using ImageJ 1.53c.

**Dipeptide synthesis:**

**$NH_2$-Lys-Cys-OH**

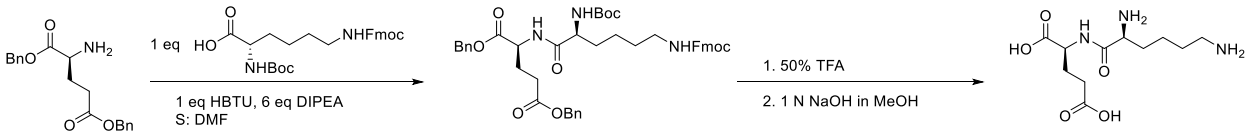



H-Cys(Bzl)-OMe.HCl (2.13 mmol, 481 mg), Boc-Lys(Fmoc)-OH (1 g, 2.13 mmol), HBTU (1.07 g, 2.84 mmol) and DIPEA (1.1 mL, 11.5 mmol) were dissolved in dry DMF in an oven-dried round bottom flask and stirred under argon overnight at rt. The reaction was partitioned with 1 M HCl (20 mL) and EtOAc (20 mL). The organic layer was washed with 1 M HCl (20 mL), saturated $NaHCO_3$ (20 mL) and brine (20 mL). The organic layer was dried with anhydrous $MgSO_4$, filtered, and concentrated *in vacuo*. Boc-Lys(Fmoc)-Cys(Bzl)-OMe was purified by flash chromatography (1:3 EtOAC:hexanes→100% EtOAc) as a crystalline white solid.

Boc-Lys(Fmoc)-Cys(Bzl)-OMe was dissolved in 1 N NaOH in methanol (10 mL) and stirred for 4 h at rt. The pH of the solution was adjusted to 6 by adding 5% acetic acid dropwise to form a precipitate. The suspension was concentrated *in vacuo* to give a white solid used immediately in the next step. The solid was treated with 95% TFA in

water (10 mL) and stirred for 1 h, rt. The reaction was concentrated *in vacuo* and washed with dry diethyl ether to give a white solid. The product $NH_2$-Lys-Cys-OH was purified by reverse-phase MPLC using gradient elution of 100% water→100% acetonitrile.

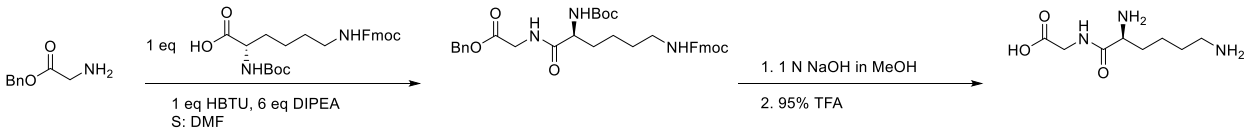
**$NH_2$-Lys-Glu-OH**




H-Glu(Bzl)-OBn.HCl (2.13 mmol, 697 mg), Boc-Lys(Fmoc)-OH (1 g, 2.13 mmol), HBTU (1.07 g, 2.84 mmol) and DIPEA (1.1 mL, 11.5 mmol) were dissolved in dry DMF in an oven-dried round bottom flask and stirred under argon overnight at rt. The reaction was partitioned with 1 M HCl (20 mL) and EtOAc (20 mL). The organic layer was washed with 1 M HCl (20 mL), saturated $NaHCO_3$ (20 mL) and brine (20 mL). The organic layer was dried with anhydrous $MgSO_4$, filtered, and concentrated *in vacuo*. Boc-Lys(Fmoc)-Glu-OBn was purified by flash chromatography (1:3 EtOAC:hexanes→100% EtOAc) as a crystalline white solid.

Boc-Lys(Fmoc)-Glu-OBn was dissolved $CH_2Cl_2$ (5 mL) and TFA (5 mL) and stirred for 30 min at rt. The reaction was concentrated *in vacuo* and washed with dry diethyl ether to give a white solid. 1 N NaOH in methanol (10 mL) was added and the reaction was stirred for 1 h at rt. The pH of the solution was adjusted to 6 by adding 5% acetic acid dropwise to form a precipitate. The suspension was concentrated *in vacuo* to give a white solid. The product $NH_2$-Lys-Glu-OH was purified by reverse-phase MPLC using gradient elution of 100% water→100% acetonitrile. *m/z* 276 (M+)

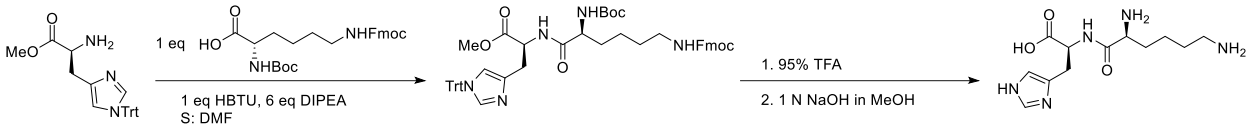
**$NH_2$-Lys-Gly-OH**




H-Gly-OBn (2.13 mmol, 353 mg), Boc-Lys(Fmoc)-OH (1 g, 2.13 mmol), HBTU (1.07 g, 2.84 mmol) and DIPEA (1.1 mL, 11.5 mmol) were dissolved in dry DMF in an oven-dried round bottom flask and stirred under argon overnight at rt. The reaction was partitioned with 1 M HCl (20 mL) and EtOAc (20 mL). The organic layer was washed with 1 M HCl (20 mL), saturated $NaHCO_3$ (20 mL) and brine (20 mL). The organic layer was dried with anhydrous $MgSO_4$, filtered, and concentrated *in vacuo*. Boc-Lys(Fmoc)-Gly-OBn was purified by flash chromatography (1:3 EtOAC:hexanes→100% EtOAc) as a crystalline white solid.

Boc-Lys(Fmoc)-Gly-OBn was dissolved in 1 N NaOH in methanol and stirred for 1 h at rt. The pH of the solution was adjusted to 6 by adding 5% acetic acid dropwise to form a precipitate. The suspension was concentrated *in vacuo* to give a white solid. $CH_2Cl_2$ (5 mL) was added, followed by TFA (5 mL) and the mixture was stirred for 30 min at rt. The mixture was concentrated *in vacuo* and washed with dry diethyl ether to give a white solid $NH_2$-Lys-Gly-OH. The product was purified by reverse-phase MPLC using gradient elution of 100% water→100% acetonitrile. *m/z* 204 (M+)

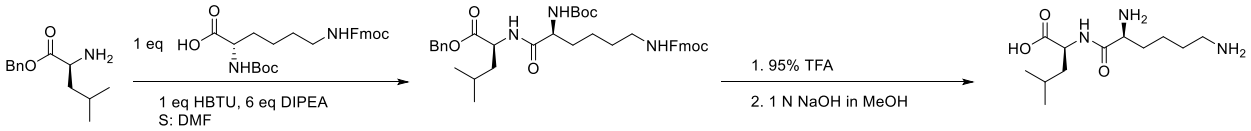
**$NH_2$-Lys-His-OH**




H-His(Trt)-OMe.HCl (1.07 mmol, 479 mg), Boc-Lys(Fmoc)-OH (500 mg, 1.07 mmol), HBTU (530 mg, 1.42 mmol) and DIPEA (0.55 mL, 5.7 mmol) were dissolved in dry DMF in an oven-dried round bottom flask and stirred under argon overnight at rt. The reaction was partitioned with 1 M HCl (20 mL) and EtOAc (20 mL). The organic layer was washed with 1 M HCl (20 mL), saturated NaHCO3 (20 mL) and brine (20 mL). The organic layer was dried with anhydrous MgSO4, filtered, and concentrated in vacuo. Boc-Lys(Fmoc)-His(Trt)-OMe was purified by flash chromatography (1:3 EtOAC:hexanes→100% EtOAc) as a crystalline white solid.

Boc-Lys(Fmoc)-His(Trt)-OMe was dissolved $CH_2Cl_2$ (5 mL) and TFA (5 mL) and stirred for 30 min at rt. The reaction was concentrated *in vacuo* and washed with dry diethyl ether to give a white solid. 1 N NaOH in methanol (10 mL) was added and the reaction was stirred for 1 h at rt. The pH of the solution was adjusted to 6 by adding 5% acetic acid dropwise to form a precipitate. The suspension was concentrated *in vacuo* to give a white solid. The product $NH_2$-Lys-His-OH was purified by reverse-phase MPLC using gradient elution of 100% water→100% acetonitrile. *m/z* 284 (M+)

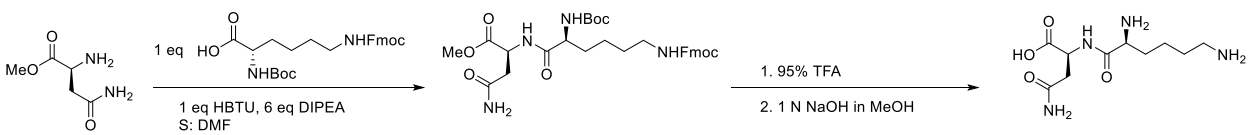
**$NH_2$-Lys-Leu-OH**




H-Leu-OBn.HCl (2.13 mmol, 549 mg), Boc-Lys(Fmoc)-OH (1 g, 2.13 mmol), HBTU (1.07 g, 2.84 mmol) and DIPEA (1.1 mL, 11.5 mmol) were dissolved in dry DMF in an oven-dried round bottom flask and stirred under argon overnight at rt. The reaction was partitioned with 1 M HCl (20 mL) and EtOAc (20 mL). The organic layer was washed with 1 M HCl (20 mL), saturated $NaHCO_3$ (20 mL) and brine (20 mL). The organic layer was dried with anhydrous $MgSO_4$, filtered, and concentrated *in vacuo*. Boc-Lys(Fmoc)-Leu-OBn was purified by flash chromatography (1:3 EtOAC:hexanes→100% EtOAc) as a crystalline white solid.

Boc-Lys(Fmoc)-Leu-OBn was dissolved in 1 N NaOH in methanol and stirred for 1 h at rt. The pH of the solution was adjusted to 6 by adding 5% acetic acid dropwise to form a precipitate. The suspension was concentrated *in vacuo* to give a white solid. $CH_2Cl_2$ (5 mL) was added, followed by TFA (5 mL) and the mixture was stirred for 30 min at rt. The mixture was concentrated *in vacuo* and washed with dry diethyl ether to give a white solid $NH_2$-Lys-Leu-OH. The product was purified by reverse-phase MPLC using gradient elution of 100% water→100% acetonitrile. *m/z* 260 (M+)

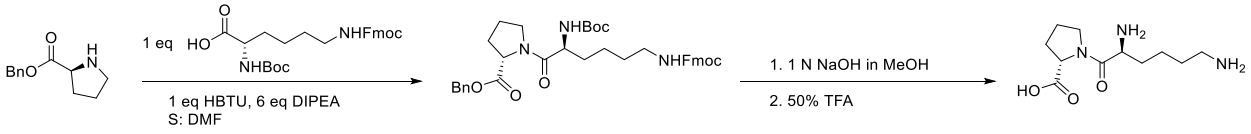
**$NH_2$-Lys-Asn-OH**



H-Asn-OMe.HCl (2.13 mmol, 311 mg), Boc-Lys(Fmoc)-OH (1 g, 2.13 mmol), HBTU (1.07 g, 2.84 mmol) and DIPEA (1.1 mL, 11.5 mmol) were dissolved in dry DMF in an oven-dried round bottom flask and stirred under argon overnight at rt. The reaction was partitioned with 1 M HCl (20 mL) and EtOAc (20 mL). The organic layer was washed with 1 M HCl (20 mL), saturated $NaHCO_3$ (20 mL) and brine (20 mL). The organic layer was dried with anhydrous $MgSO_4$, filtered, and concentrated in vacuo. Boc-Lys(Fmoc)-Asn-OMe was purified by flash chromatography (1:3 EtOAC:hexanes→100% EtOAc) as a crystalline white solid.

Boc-Lys(Fmoc)-Asn-OMe was dissolved $CH_2Cl_2$ (5 mL) and TFA (5 mL) and stirred for 30 min at rt. The reaction was concentrated *in vacuo* and washed with dry diethyl ether to give a white solid. 1 N NaOH in methanol (10 mL) was added and the reaction was stirred for 1 h at rt. The pH of the solution was adjusted to 6 by adding 5% acetic acid dropwise to form a precipitate. The suspension was concentrated in vacuo to give a white solid. The product $NH_2$-Lys-Asn-OH was purified by reverse-phase MPLC using gradient elution of 100% water→100% acetonitrile.

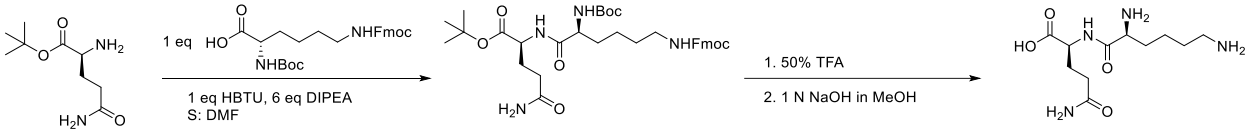
**$NH_2$-Lys-Pro-OH**




H-Pro-OBn.HCl (2.13 mmol, 515 mg), Boc-Lys(Fmoc)-OH (1 g, 2.13 mmol), HBTU (1.07 g, 2.84 mmol) and DIPEA (1.1 mL, 11.5 mmol) were dissolved in dry DMF in an oven-dried round bottom flask and stirred under argon overnight at rt. The reaction was partitioned with 1 M HCl (20 mL) and EtOAc (20 mL). The organic layer was washed with 1 M HCl (20 mL), saturated $NaHCO_3$ (20 mL) and brine (20 mL). The organic layer was dried with anhydrous $MgSO_4$, filtered, and concentrated *in vacuo*. Boc-Lys(Fmoc)-Pro-OBn was purified by flash chromatography (1:3 EtOAC:hexanes→100% EtOAc) as a crystalline white solid.

Boc-Lys(Fmoc)-Pro-OBn was dissolved in 1 N NaOH in methanol and stirred for 1 h at rt. The pH of the solution was adjusted to 6 by adding 5% acetic acid dropwise to form a precipitate. The suspension was concentrated *in vacuo* to give a white solid. $CH_2Cl_2$ (5 mL) was added, followed by TFA (5 mL) and the mixture was stirred for 30 min at rt. The mixture was concentrated *in vacuo* and washed with dry diethyl ether to give a white solid $NH_2$-Lys-Pro-OH. The product was purified by reverse-phase MPLC using gradient elution of 100% water→100% acetonitrile. *m/z* 244 (M+)

**$NH_2$-Lys-Gln-OH**




H-Gln-OtBu.HCl (2.13 mmol, 430 mg), Boc-Lys(Fmoc)-OH (1 g, 2.13 mmol), HBTU (1.07 g, 2.84 mmol) and DIPEA (1.1 mL, 11.5 mmol) were dissolved in dry DMF in an oven-dried round bottom flask and stirred under argon overnight at rt. The reaction was partitioned with 1 M HCl (20 mL) and EtOAc (20 mL). The organic layer was washed with 1 M HCl (20 mL), saturated $NaHCO_3$ (20 mL) and brine (20 mL). The organic layer was dried with anhydrous $MgSO_4$, filtered, and concentrated *in vacuo*. Boc-Lys(Fmoc)-Gln-OtBu was purified by flash chromatography (1:3 EtOAC:hexanes→100% EtOAc) as a crystalline white solid.

Boc-Lys(Fmoc)-Gln-OtBu was dissolved $CH_2Cl_2$ (5 mL) and TFA (5 mL) and stirred for 30 min at rt. The reaction was concentrated *in vacuo* and washed with dry diethyl ether to give a white solid. 1 N NaOH in methanol (10 mL) was added and the reaction was stirred for 1 h at rt. The pH of the solution was adjusted to 6 by adding 5% acetic acid dropwise to form a precipitate. The suspension was concentrated *in vacuo* to give a white solid. The product $NH_2$-Lys-Gln-OH was purified by reverse-phase MPLC using gradient elution of 100% water→100% acetonitrile. *m/z* 275 (M+)

**$NH_2$-Lys-Arg-OH**

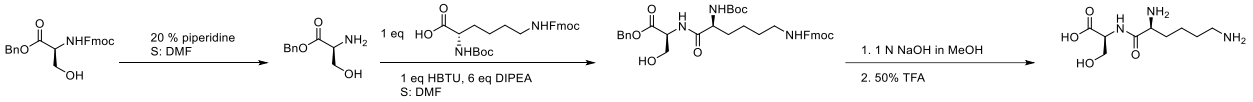
Fmoc-Arg(Pbf)-OH (2 g, 3.1 mmol), 4-DMAP (0.75 g, 6.17 mmol), and EDCI · HCl (0.88 g, 4.6 mmol) were added to an oven-dried round-bottom flask and dissolved in $CH_2Cl_2$ (30 mL) over ice. Benzyl alcohol (2.8 mL, 20.2 mmol) was added dropwise under argon, and the contents were allowed to warm to room temperature and stirred overnight. The reaction was washed with DI water (2 × 20 mL), $NH_4Cl$ (2 × 10 mL), and brine (10 ml). The organic layer was dried over anhydrous $MgSO_4$, filtered, and concentrated *in vacuo*, yielding a yellow oil. Fmoc-Arg(Pbf)-OBn was purified by flash chromatography (1:1→1:0 EtOAc/hexanes → flush with isopropanol) as a crystalline white solid.

Fmoc-Arg(Pbf)-OBn was dissolved in 20% piperidine in DMF under argon. The reaction was stirred for 45 min and concentrated *in vacuo*. H-Arg(Pbf)-OBn was purified by flash chromatography (5%→9% methanol in DCM), giving a gummy solid.

H-Arg(Pbf)-OBn (0.83 mmol, 430 mg), Boc-Lys(Fmoc)-OH (390 mg, 0.83 mmol), HBTU (419 mg, 1.1 mmol) and DIPEA (1.1 mL, 11.5 mmol) were dissolved in dry DMF in an oven-dried round bottom flask and stirred under argon overnight at rt. The reaction was partitioned with 1 M HCl (20 mL) and EtOAc (20 mL). The organic layer was washed with 1 M HCl (20 mL), saturated $NaHCO_3$ (20 mL) and brine (20 mL). The organic layer was dried with anhydrous $MgSO_4$, filtered, and concentrated *in vacuo*. Boc-Lys(Fmoc)-Arg(Pbf)-OBn was purified by flash chromatography (1:3 EtOAC:hexanes→100% EtOAc) as a crystalline white solid.

Boc-Lys(Fmoc)-Arg(Pbf)-OBn was treated with 1 N NaOH in methanol for 1 h, rt. The pH of the solution was adjusted to ~6 by adding 5% AcOH dropwise, forming a precipitate. The product was concentrated *in vacuo* and used directly in the next step. The solid was treated with 95% TFA in water for 4 h at rt and concentrated *in vacuo*. The remaining product was washed with cold diethyl ether. The product $NH_2$-Lys-Arg-OH was purified by reverse-phase MPLC using gradient elution of 100% water→100% acetonitrile. *m/z* 303 (M+)

**$NH_2$-Lys-Ser-OH**



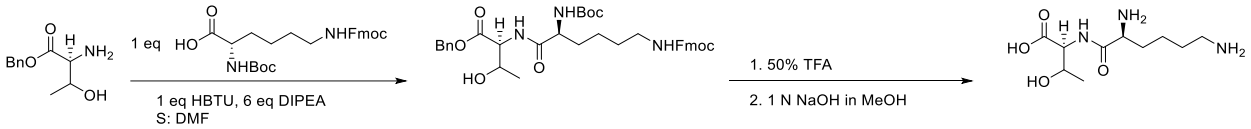

Fmoc-Ser-OBn (300 mg) was dissolved in 20% piperidine in DMF under argon. The reaction was stirred for 45 min and concentrated *in vacuo*. H-Ser-OBn was purified by flash chromatography (5% methanol in DCM), giving a gummy solid.

H-Ser-OBn (0.28 mmol, 54 mg), Boc-Lys(Fmoc)-OH (130 mg, 0.28 mmol), HBTU (141 g, 0.37 mmol) and DIPEA (0.1 mL, 0.56 mmol) were dissolved in dry DMF in an oven-dried round bottom flask and stirred under argon overnight at rt. The reaction was partitioned with 1 M HCl (2 mL) and EtOAc (2.5 mL). The organic layer was washed

with 1 M HCl (2 mL), saturated $NaHCO_3$ (2 mL) and brine (1.5 mL). The organic layer was dried with anhydrous $MgSO_4$, filtered, and concentrated *in vacuo*. Boc-Lys(Fmoc)-Ser-OBn was purified by flash chromatography (1:3 EtOAC:hexanes→100% EtOAc) as a crystalline white solid.

Boc-Lys(Fmoc)-Ser-OBn was dissolved in 1 N NaOH in methanol and stirred for 1 h at rt. The pH of the solution was adjusted to 6 by adding 5% acetic acid dropwise to form a precipitate. The suspension was concentrated *in vacuo* to give a white solid. $CH_2Cl_2$ (5 mL) was added, followed by TFA (5 mL) and the mixture was stirred for 30 min at rt. The mixture was concentrated *in vacuo* and washed with dry diethyl ether to give a white solid $NH_2$-Lys-Ser-OH. The product was purified by reverse-phase MPLC using gradient elution of 100% water→100% acetonitrile.

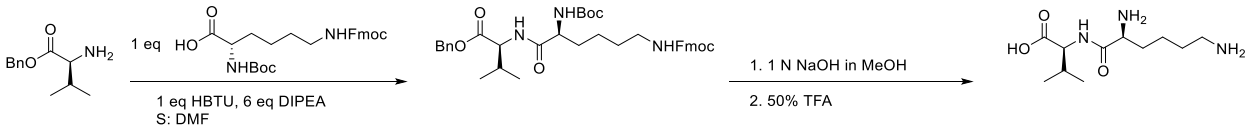
**$NH_2$-Lys-Thr-OH**




H-Thr-OBn.HCl (2.13 mmol, 447 mg), Boc-Lys(Fmoc)-OH (1 g, 2.13 mmol), HBTU (1.07 g, 2.84 mmol) and DIPEA (1.1 mL, 11.5 mmol) were dissolved in dry DMF in an oven-dried round bottom flask and stirred under argon overnight at rt. The reaction was partitioned with 1 M HCl (20 mL) and EtOAc (20 mL). The organic layer was washed with 1 M HCl (20 mL), saturated $NaHCO_3$ (20 mL) and brine (20 mL). The organic layer was dried with anhydrous $MgSO_4$, filtered, and concentrated *in vacuo*. Boc-Lys(Fmoc)-Thr-OBn was purified by flash chromatography (1:3 EtOAC:hexanes→100% EtOAc) as a crystalline white solid.

Boc-Lys(Fmoc)-Thr-OBn was dissolved $CH_2Cl_2$ (5 mL) and TFA (5 mL) and stirred for 30 min at rt. The reaction was concentrated *in vacuo* and washed with dry diethyl ether to give a white solid. 1 N NaOH in methanol (10 mL) was added and the reaction was stirred for 1 h at rt. The pH of the solution was adjusted to 6 by adding 5% acetic acid dropwise to form a precipitate. The suspension was concentrated *in vacuo* to give a white solid. The product $NH_2$-Lys-Thr-OH was purified by reverse-phase MPLC using gradient elution of 100% water→100% acetonitrile. *m/z* 248 (M+)

**$NH_2$-Lys-Val-OH**

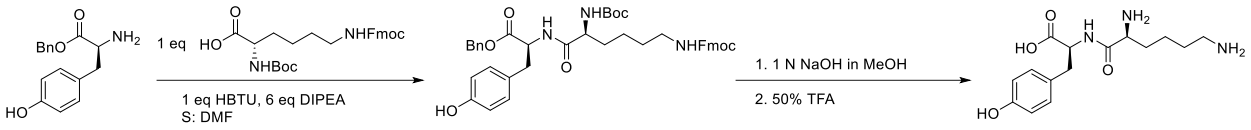



H-Val-OBn.HCl (0.65 mmol, 158 mg), Boc-Lys(Fmoc)-OH (300 mg, 0.65 mmol), HBTU (325 mg, 0.86 mmol) and DIPEA (0.35 mL, 1.9 mmol) were dissolved in dry DMF in an oven-dried round bottom flask and stirred under argon overnight at rt. The reaction was partitioned with 1 M HCl (4 mL) and EtOAc (4 mL). The organic layer was washed with 1 M HCl (4 mL), saturated $NaHCO_3$ (4 mL) and brine (4 mL). The organic layer was dried with anhydrous $MgSO_4$, filtered, and concentrated *in vacuo*. Boc-Lys(Fmoc)-Val-OBn was purified by flash chromatography (1:3 EtOAC:hexanes→100% EtOAc) as a crystalline white solid.

Boc-Lys(Fmoc)-Val-OBn was dissolved in 1 N NaOH in methanol and stirred for 1 h at rt. The pH of the solution was adjusted to 6 by adding 5% acetic acid dropwise to form a precipitate. The suspension was concentrated *in vacuo* to give a white solid. $CH_2Cl_2$ (5 mL) was added, followed by TFA (5 mL) and the mixture was stirred for 30

min at rt. The mixture was concentrated *in vacuo* and washed with dry diethyl ether to give a white solid $NH_2$-Lys-Val-OH. The product was purified by reverse-phase MPLC using gradient elution of 100% water→100% acetonitrile. *m/z* 246 (M+)

**$NH_2$-Lys-Tyr-OH**




H-Tyr-OBn (0.65 mmol, 166 mg), Boc-Lys(Fmoc)-OH (300 mg, 0.65 mmol), HBTU (325 mg, 0.86 mmol) and DIPEA (0.35 mL, 1.9 mmol) were dissolved in dry DMF in an oven-dried round bottom flask and stirred under argon overnight at rt. The reaction was partitioned with 1 M HCl (4 mL) and EtOAc (4 mL). The organic layer was washed with 1 M HCl (4 mL), saturated $NaHCO_3$ (4 mL) and brine (4 mL). The organic layer was dried with anhydrous $MgSO_4$, filtered, and concentrated *in vacuo*. Boc-Lys(Fmoc)-Tyr-OBn was purified by flash chromatography (1:3 EtOAC:hexanes→100% EtOAc) as a crystalline white solid.

Boc-Lys(Fmoc)-Tyr-OBn was dissolved in 1 N NaOH in methanol and stirred for 1 h at rt. The pH of the solution was adjusted to 6 by adding 5% acetic acid dropwise to form a precipitate. The suspension was concentrated *in vacuo* to give a white solid. $CH_2Cl_2$ (5 mL) was added, followed by TFA (5 mL) and the mixture was stirred for 30 min at rt. The mixture was concentrated *in vacuo* and washed with dry diethyl ether to give a white solid $NH_2$-Lys-Tyr-OH. The product was purified by reverse-phase MPLC using gradient elution of 100% water→100% acetonitrile.

**Supplementary information**

Given the non-viability of Syn7942 cells auxotrophic for phenylalanine as endosymbionts, we identified an aromatic amino acid transporter expressed in extant red algae. In principle, a plastid-localized aromatic amino acid permease encoded by red algae could be codon-optimized for Syn7942 and expressed in order to enhance import of exogenous amino acids such as phenylalanine. Given the lack of functional biochemical characterization of most hydroxy/aromatic amino acid permease (HAAAP) family proteins, we instead extracted 96 HAAAP family protein sequences from TransportDB[84,85] (see **Appendix**) and used those sequences as basic local alignment search tool (BLAST) queries against all red algal taxa in the NCBI nr/nt collection (see **Supplementary methods**). Through this BLAST search, we identified 37 putative HAAAP family proteins in a variety of red algae (**Table 3.9**). Next, those 37 FASTAs were used as queries for the PredSL[177], Predotar[178], and TargetP[179] algorithms in order to predict their subcellular localization based on their N-terminal sequences. From this set of 37 candidates, we identified one putative red algal HAAAP protein (KAA8497193.1) which was predicted by all three tools to localize to the chloroplast without a lumenal TP. Next, the gene encoding KAA8497193.1 can be cloned and integrated into the Syn7942 genome. If expression of KAA8497193.1 results in viable endosymbiosis, then that result would suggest that similar host-encoded factors could have facilitated integration of plastids in the early stages of endosymbiosis.

The queries and search strategy used in BLAST analysis are listed in the **Appendix**.

**Supplementary information**

**Table 3.4**: Complete media[77]

| Supplement group | Tested insertional mutations |
|---|---|
| Glu, Pro, Arg, Asp, Thr, Met, Lys, Tyr, Trp, Phe, His, Ser, Gly, Cys | *lysA, argJ, glyA, aspC, thrC, proA, trpC, hisC, proC, hisS, glsF, thrB, cysE, argF, asd, dapA, argD, hisD, pheA, aroH, serA* |
| Asp, Thr, Met, Lys, Val, Leu, Ile, Gln, thiamin, 4-aminobenzoic acid, biotin, succinic acid | *lysA, thrC, thiC, ilvD, bioB, bioD, bioFHCDA, asd, dapA, gatB, leuB* |

**Table 3.5**: Overexpression plasmids.

| Name | Link to Benchling map |
|---|---|
| pET28[180] | pET28 |
| pET28-FLUC | pET28-FLUC |
| pML17[75] | pML17 |

**Table 3.6**: Oligonucleotides used for PCR analysis of recombinant Syn7942 mutants.

| Name | Sequence (5'→3') | Target gene |
|---|---|---|
| JC266 | GGCGCAGGATCGTACACTGAT | *lysA* |
| JC267 | AGCCGCTGCTAAGCCATGTT | |
| JC328 | GCAGCATCGCCGCTTTTATCG | *argJ* |
| JC329 | GACGATCACTGGGAGCATCGA | |
| JC354 | AGCCTTGAGCAAAGCGATCG | *hisB* |
| JC355 | CTTCGACATCCTGATCCACTGC | |
| JC332 | TCCCCCAGTGCCACTAGAGG | *thiC* |
| JC333 | GTTGCGAATTGCCACTCATCTCG | |
| JC445 | GCAGCTTTGAACATTGAGCATG | *nadA* |
| JC446 | CAGCCTCCGCAACCTAATAC | |
| JC433 | TGAGGATGGTTTTGTTGCGG | *glyA* |
| JC434 | GTCATGATGGCAAGACGATAGG | |
| JC439 | AATTCGGTTAGAAGACGACGC | *ilvD* |
| JC440 | CTCTCGCACAGTGAAACCCT | |
| JC441 | CGTGGAGTGAGCCAAGACTTTA | *aspC* |
| JC442 | CAAGGACGACTAATGCAGCC | |
| JC437 | GCTAACATCCGTGTCATTGCAG | *thrC* |
| JC438 | AGCTGCCGCTCATTTGTATC | |
| JC435 | CGATCGCTCTTGGTTGATCC | *proA* |
| JC436 | GCTTAGAACAGCAGGGTTCTGA | |
| JC443 | CGATCGCAACCCTGTCATTT | *trpC* |
| JC444 | TAGTGAAGGCTGTTGCAGGA | |
| JC487 | TGTTTGGTCTTGAGGGAGTGATGAC | *bioB* |
| JC488 | GTATGATCCTAAGGTCGCGATCGC | |
| JC485 | ACAGCGCCAGAATTTTTAGGTAGC | *pdxA* |
| JC486 | CCAGCTGTTGGCACCGTAATAAC | |
| JC483 | AAATTCCTCTGAGCAATCGCGG | *ribF* |
| JC484 | GGTTAGGGCAAGAACGCTCTATG | |
| JC536 | TCGCTCTTCGTTAATTTCTGCTCC | *hisC* |
| JC537 | AAAACTGGGCTCTTGACCAGC | |
| JC538 | CAACGGCTATCAACCAATCGTTAAG | *proC* |
| JC539 | GCTGGACTACGACGAAGAGTAC | |
| JC534 | AAGAACAAGTAAGACAGACTGAGGAC | *hisS* |
| JC535 | CAACAGCAAGTCAGCGAGATAAC | |
| JC600 | CAGTGCCGCTGTGTTTAGCG | *glsF* |
| JC601 | CTAGGGCGTTGTCGAGGTTG | |

**Table 3.6 (cont.):** Oligonucleotides used for PCR analysis of recombinant Syn7942 mutants.

| | | |
|---|---|---|
| JC591 | CGATCGCGGCTTTACGATCAGAG | *bioFHCDA* |
| JC561 | GTCTGACGCTATCGCAGCATAAATATCCTGCAAC | |
| JC602 | AGGTGCAAGCCTGGTCGATAAG | *thrB* |
| JC603 | GGGCATTAGAAACCCTGCCG | |
| JC635 | GAACTGTTTACCCTGCTGCCAC | *cysE* |
| JC636 | CCTGACTAGGTCGAAGTGACTGG | |
| JC645 | AAATGCCTGACAGCGACAGG | *gatB* |
| JC646 | CCCAGCTTGACCGACTGTC | |
| JC693 | CGCAAGTCTCTCGCGGTAAAG | *argF* |
| JC694 | CGATCAAACTCCTTGTCCGGC | |
| JC695 | ACAGCTAAGGACTGAGCAATGTCG | *asd* |
| JC696 | TGTAAGTCCTCGGTAAAGCCCAG | |
| JC697 | GCAGAAGTCCAAGAGCAAAGCC | *dapA* |
| JC698 | AGAAGGCATAGGCTTGAGAGTCAAC | |
| JC699 | CGCTGAGGTTCGGTCAAGATTC | *argD* |
| JC700 | AGATCAACGCCTCGCTGAGC | |
| JC701 | GCATTGATTGGGAAGGCGCA | *hisD* |
| JC702 | GCCCTTGTCATTCTGGGTTCC | |
| YG203 | TCACAAGCGGCAATACCGCCA | *bioD* |
| YG204 | TGGCAGATTTGGAACAGCTCG | |
| LL50 | GCAGCCAATAAATCTAAGCCCC | *pheA* |
| LL51 | GGTCGCAGGCCATCTCTAC | |
| SA55 | ATGCAAGAGCTGACAATTA | *aroH* |
| SA56 | CAACCTTGAGCGCCTGCTCG | |
| SA48 | CTAGAGCACCACCGTGTAGGC | *serA* |
| SA49 | ATGATCGGGCGAATCTGCGGA | |
| SA65 | AAGAAGCGGGGACTACAAAT | *leuB* |
| SA66 | AAATCGCTGACTTCAGGCGA | |

**Table 3.7**: Oligonucleotides used for plasmid cloning.

| Name | Sequence (5'→3') |
|---|---|
| BD3A | AGCTCGAATTCGGATCCGCGTCACAGCTTCGATTTTCCCCCCTTCT |
| BD3B | TTCTTGATGTTCTTCGCATCTTCCATGCCGCTGCTGTGATGATGAT |
| SB125A | GCCGCTGCTGTGATGATGAT |
| SB125B | AGCTCGAATTCGGATCCGCG |
| JC208 | GCTGCGCTAGAAAAACTAATCAGGGCATGGGCAATCG |
| JC209 | CTGATTAGTTTTTCTAGCGCAGCTGCTTGGAA |
| JC210 | GAGGGACGAACTTGCAACTGGTCGCGGTTG |
| JC211 | CAGTTGCAAGTTCGTCCCTCAAATCGTGGCC |
| JC212 | TGCAACCACTAAGCGGGTCACTACTTGGTAGCA |
| JC213 | TGACCCGCTTAGTGGTTGCAGCGGCATCTG |
| JC214 | GTCTGACGCTCTAGCGAGCCTGTGCAGGTACTA |
| JC215 | GGCTCGCTAGAGCGTCAGACCCCGTAGAAAAGA |
| JC260 | TCATTCGTGCAAGCGGGTCACTACTTGGTAGCA |
| JC261 | TGACCCGCTTGCACGAATGACTGTGTGTTTGCCC |
| JC262 | CACCACCTAAAGCGTCAGACCCCGTAGAAAAGA |
| JC263 | GGGTCTGACGCTTTAGGTGGTGTATTCAGCATTGATGCGG |
| JC276 | TGACCCGCTTAGCAAATGGAAGAGTGCTCAGAAGC |
| JC277 | GGTCTGACGCTGGTTAGATCTTCCTCGGTGATCTTGGT |
| JC278 | TCCATTTGCTAAGCGGGTCACTACTTGGTAGC |
| JC279 | AGATCTAACCAGCGTCAGACCCCGTAGAAA |
| JC280 | GAGGGACGAAATCTAATTCGGGATGAAGTGGCACG |
| JC281 | CCCGAATTAGATTTCGTCCCTCAAATCGTGGCC |
| JC282 | GCTGCGCTAGAAAGTGGGTATAGAGCGGGTTTTGC |
| JC283 | ATACCCACTTTCTAGCGCAGCTGCTTGGAA |

**Table 3.7 (cont.):** Oligonucleotides used for plasmid cloning.

| | |
|---|---|
| JC284 | GAGGGACGAACTAAAGGGTCGGGATCTTCGGTG |
| JC286 | TCGCTCTCAATCTAGCGCAGCTGCTTGGAA |
| JC287 | CTGCGCTAGATTGAGAGCGACGCCAGTGATC |
| JC285 | CGACCCTTTAGTTCGTCCCTCAAATCGTGGCC |
| JC288 | GCAATCCGTTTTCGTCCCTCAAATCGTGGC |
| JC289 | GAGGGACGAAAACGGATTGCCGATTGTGCG |
| JC290 | TCTTAGTTCAATTCTAGCGCAGCTGCTTGGAA |
| JC291 | GCTGCGCTAGAATTGAACTAAGACTCCCGCAGGC |
| JC314 | TCGGTTACCAAGCGTCAGACCCCGTAGAAA |
| JC315 | GTCTGACGCTTGGTAACCGATCCTGTCGCAA |
| JC375 | TCTGACGCTTCAAGCGCTCATTGCCAGCA |
| JC376 | TGACCCGCTTTTATCTTTGCCCCCGATCGCA |
| JC377 | GGCAAAGATAAAAGCGGGTCACTACTTGGTAGC |
| JC378 | GAGCGCTTGAAGCGTCAGACCCCGTAGAAAAG |
| JC379 | GAGGGACGAATGCTCAAGCTCAATCAGCTGCA |
| JC380 | CTGCGCTAGACTAGCTAGCATTGCCGCGATAAAT |
| JC381 | TGCTAGCTAGTCTAGCGCAGCTGCTTGGAA |
| JC382 | AGCTTGAGCATTCGTCCCTCAAATCGTGGCC |
| JC385 | GAGGGACGAAGAACACCTTGAACTGATCGCCAG |
| JC386 | CTGCGCTAGACCGACTTCATCAGCAATTTCGCG |
| JC387 | ATGAAGTCGGTCTAGCGCAGCTGCTTGGAA |
| JC388 | GCAATTACGTAAGCGGGTCACTACTTGGTAGC |
| JC389 | GTGACCCGCTTACGTAATTGCTGCGAAAGCCG |
| JC390 | TCTGACGCTCTAAACCGCGACTGGCGTTG |
| JC391 | CGCGGTTTAGAGCGTCAGACCCCGTAGAAAA |
| JC392 | CAAGGTGTTCTTCGTCCCTCAAATCGTGGC |
| JC393 | GAGGGACGAAATGCCTCAGTACCGATCGCG |
| JC396 | TTCATCTTGGAAGCGGGTCACTACTTGGTAGC |
| JC397 | TGACCCGCTTCCAAGATGAAGCCGTCAACTGG |
| JC398 | GTCTGACGCTCTAGCGAGATCCGAGAATTTGGC |
| JC399 | ATCTCGCTAGAGCGTCAGACCCCGTAGAAAAG |
| JC400 | GAGGGACGAAATGGAAGTCCGCCGCCG |
| JC401 | TGCGCTAGATCAGCACCGAAATGCAAGCC |
| JC402 | CGGTGCTGATCTAGCGCAGCTGCTTGGAA |
| JC403 | CTTTGTAGATCAAGCGGGTCACTACTTGGTAGC |
| JC404 | GTGACCCGCTTGATCTACAAAGCCCGCTTACTGG |
| JC405 | GTCTGACGCTTTAGGCTGAGGGGCGATCG |
| JC406 | CTCAGCCTAAAGCGTCAGACCCCGTAGAAAA |
| JC407 | GGACTTCCATTTCGTCCCTCAAATCGTGGC |
| JC408 | GAGGGACGAAGTGGCTACTCTCTCCGTCGATC |
| JC409 | CTGCGCTAGACCCTTGAGAATGGCACCATTGC |
| JC410 | TTCTCAAGGGTCTAGCGCAGCTGCTTGGAA |
| JC411 | CTCATGAATCAAAGCGGGTCACTACTTGGTAGC |
| JC412 | TGACCCGCTTTGATTCATGAGGCGATCGCCC |
| JC413 | GTCTGACGCTTTAGCGATCGCGGTGTAGGAA |
| JC414 | CGATCGCTAAAGCGTCAGACCCCGTAGAAAA |
| JC415 | GAGTAGCCACTTCGTCCCTCAAATCGTGGCC |
| JC416 | TGAGGGACGAAATGACCCAAACGACACTCAGCA |
| JC417 | CTGCGCTAGATCAGCAGGAATGAAGACGCAG |
| JC418 | TTCCTGCTGATCTAGCGCAGCTGCTTGGAA |
| JC419 | CATCGACAAAAGCGGGTCACTACTTGGTAGC |
| JC420 | TGACCCGCTTTTGTCGATGACAAAGCCGTGC |
| JC421 | GGTCTGACGCTCTAGATCAGCACTTGCTGCCATTC |
| JC422 | GCTGATCTAGAGCGTCAGACCCCGTAGAAAAG |
| JC423 | TTTGGGTCATTTCGTCCCTCAAATCGTGGCC |

**Table 3.7 (cont.):** Oligonucleotides used for plasmid cloning.

| | |
|---|---|
| JC424 | GAGGGACGAAATGAAACTATCCGAGCGTGTGGG |
| JC425 | CTGCGCTAGAGCTGGAGTTTAAAGCCGTCGC |
| JC426 | AAACTCCAGCTCTAGCGCAGCTGCTTGGAA |
| JC427 | GACCCGCTTTGATCAGCAACGGCTTCGC |
| JC428 | GTCTGACGCTTCAAACCGAGACACCAAGCAGG |
| JC429 | CTCGGTTTGAAGCGTCAGACCCCGTAGAAAAG |
| JC430 | CGGATAGTTTCATTTCGTCCCTCAAATCGTGGC |
| JC431 | TTGCTGATCAAAGCGGGTCACTACTTGGTAGC |
| JC432 | ACTGAGGCATTTCGTCCCTCAAATCGTGGC |
| JC449 | AATGGTGAAGATCTAGCGCAGCTGCTTGGAA |
| JC450 | CTGCGCTAGATCTTCACCATTGAGGATGACTTTG |
| JC459 | TGAGGGACGAAGTGAGAATTCTTGATGACCTAGACCC |
| JC460 | TGCGCTAGAAATCGCTTGCAGATCCTGCG |
| JC461 | GCAAGCGATTTCTAGCGCAGCTGCTTGGAA |
| JC462 | TGCTTAAACAAAGCGGGTCACTACTTGGTAGC |
| JC463 | TGACCCGCTTTGTTTAAGCACCGTGACTTGCG |
| JC464 | GTCTGACGCTTCAAGCCAAGCTCGATGACTCTG |
| JC465 | CTTGGCTTGAAGCGTCAGACCCCGTAGAAAAG |
| JC466 | AGAATTCTCACTTCGTCCCTCAAATCGTGGC |
| JC467 | AGGGACGAAATGCTTTCGCCGTCACTTTCAC |
| JC468 | CTGCGCTAGACACAAACAACATCCCTGATCGCG |
| JC469 | TGTTGTTTGTGTCTAGCGCAGCTGCTTGGAA |
| JC470 | TGAGTTTCTGAAGCGGGTCACTACTTGGTAGC |
| JC471 | TGACCCGCTTCAGAAACTCAGCCTCTTGATCGACTG |
| JC472 | TCTGACGCTTCACACCGTTTCCAGTTGAAGAGC |
| JC473 | AACGGTGTGAAGCGTCAGACCCCGTAGAAAA |
| JC474 | GCGAAAGCATTTCGTCCCTCAAATCGTGGCC |
| JC475 | TGAGGGACGAAATGGAAGCTATCCGCCACGA |
| JC476 | CTGCGCTAGATAGAACTCGGGGCTGGTGTC |
| JC477 | CCCGAGTTCTATCTAGCGCAGCTGCTTGGAA |
| JC478 | ACACTAATGCAAGCGGGTCACTACTTGGTAGC |
| JC479 | TGACCCGCTTGCATTAGTGTTTGCTGCGGCG |
| JC480 | GTCTGACGCTCTAAGCCGTCACCACTTCAAGC |
| JC481 | TGACGGCTTAGAGCGTCAGACCCCGTAGAAAAG |
| JC482 | ATAGCTTCCATTTCGTCCCTCAAATCGTGGC |
| JC502 | GAGGGACGAAATGGTGCATCAGCCTCCAGC |
| JC503 | GATGCACCATTTCGTCCCTCAAATCGTGGC |
| JC504 | CTGCGCTAGAACGACTAGATGCCAGCTCCC |
| JC505 | CATCTAGTCGTTCTAGCGCAGCTGCTTGGAAC |
| JC506 | CTAGCGTTATAAAGCGGGTCACTACTTGGTAGC |
| JC507 | GTGACCCGCTTTATAACGCTAGCGCTGGCCC |
| JC508 | GGTCTGACGCTCTAACAAGTTGTGGTGGCTCGC |
| JC509 | CAACTTGTTAGAGCGTCAGACCCCGTAGAAAAG |
| JC510 | GAGGGACGAAACTCTTACAACGGTTATAACGACGC |
| JC511 | CGTTGTAAGAGTTTCGTCCCTCAAATCGTGGC |
| JC512 | CTGCGCTAGACTATTGAGCAGCGTAGCGGC |
| JC513 | CTGCTCAATAGTCTAGCGCAGCTGCTTGGAAC |
| JC514 | GATCAAACGATAAGCGGGTCACTACTTGGTAGC |
| JC515 | TGACCCGCTTATCGTTTGATCCTGTCGATCTTGG |
| JC516 | GTCTGACGCTCTAGCCCAATTGTTGCGATCGC |
| JC517 | ATTGGGCTAGAGCGTCAGACCCCGTAGAAAAG |
| JC518 | GAGGGACGAATGCCGTTTTTGCGTTCTGAGC |
| JC519 | AAAAACGGCATTCGTCCCTCAAATCGTGGC |
| JC520 | CTGCGCTAGAAAACAGAGGCGAACTGGCG |
| JC521 | GCCTCTGTTTTCTAGCGCAGCTGCTTGGAAC |

**Table 3.7 (cont.):** Oligonucleotides used for plasmid cloning.

| | |
|---|---|
| JC522 | CTGAATTCAAAATAAAGCGGGTCACTACTTGGTAGC |
| JC523 | TGACCCGCTTTATTTTGAATTCAGCGGCGAAACCC |
| JC524 | GTCTGACGCTGGTAATGATTGAGTGATTGCTGTTTGC |
| JC525 | TCAATCATTACCAGCGTCAGACCCCGTAGAAAAG |
| JC543 | GAGGGACGAAAGAAATCGTGAGAGGGCCCTAAAC |
| JC544 | CTGCGCTAGATGTCATCAAGCCAGCGCCATC |
| JC545 | GCTTGATGACATCTAGCGCAGCTGCTTGGAAC |
| JC546 | CAATTTGCTCAAGCGGGTCACTACTTGGTAGC |
| JC547 | TGACCCGCTTGAGCAAATTGTTGCTGAGGAAAACC |
| JC548 | GGTCTGACGCTATCATCCAGTTGATGTTGCCCAG |
| JC549 | ACTGGATGATAGCGTCAGACCCCGTAGAAAAG |
| JC550 | CTCACGATTTCTTTCGTCCCTCAAATCGTGGC |
| JC555 | GCTCACCATTTCGTCCCTCAAATCGTGGC |
| JC556 | GAGGGACGAAATGGTGAGCGATCGCGGG |
| JC557 | AGCTGCGCTAGAAATAACTGTTCCAAATGGTCGAAGTCG |
| JC558 | AACAGTTATTTCTAGCGCAGCTGCTTGGAAC |
| JC559 | GTAATCGACAAAGCGGGTCACTACTTGGTAGC |
| JC560 | TGACCCGCTTTGTCGATTACGGTGGCAACG |
| JC561 | GTCTGACGCTATCGCAGCATAAATATCCTGCAAC |
| JC562 | ATGCTGCGATAGCGTCAGACCCCGTAGAAAAG |
| JC563 | GAGGGACGAATCTACAACCACTTCTGGTTTGCG |
| JC564 | AGTGGTTGTAGATTCGTCCCTCAAATCGTGGC |
| JC565 | TGCGCTAGAATCGCCCATTGATTCTGGCG |
| JC566 | AATGGGCGATTCTAGCGCAGCTGCTTGGAA |
| JC567 | CGAAAGCTCAAGCGGGTCACTACTTGGTAGC |
| JC568 | TGACCCGCTTGAGCTTTCGACCGAAGCGG |
| JC569 | GTCTGACGCTTTCGGTTTCCTGCTCAATCTGC |
| JC570 | GGAAACCGAAAGCGTCAGACCCCGTAGAAAAG |
| JC575 | GAGGGACGAAGTGTTTAAAACGCTGGCTGCC |
| JC576 | CTGCGCTAGAGATATTGCCGAGTACCTTCGCTC |
| JC577 | CGGCAATATCTCTAGCGCAGCTGCTTGGAA |
| JC578 | AGCACCAATGAAGCGGGTCACTACTTGGTAGC |
| JC579 | GACCCGCTTCATTGGTGCTGGTTCGGTGG |
| JC580 | GTCTGACGCTTTAGATGCCCGCGCCGTCT |
| JC581 | GGGCATCTAAAGCGTCAGACCCCGTAGAAAA |
| JC582 | CGTTTTAAACACTTCGTCCCTCAAATCGTGGCC |
| JC637 | GAGGGACGAAATGACAGCGACGGCTCCTG |
| JC638 | CTGCGCTAGAACGCCAATTTTCTTGGTGTAGGG |
| JC639 | AAATTGGCGTTCTAGCGCAGCTGCTTGGAAC |
| JC640 | TAGGCCGAAAAAGCGGGTCACTACTTGGTAGC |
| JC641 | TGACCCGCTTTTTCGGCCTACGATGCGCG |
| JC642 | GGTCTGACGCTCTAGCCTTTCAGCTTTTGGTTGAGG |
| JC643 | GAAAGGCTAGAGCGTCAGACCCCGTAGAAAAG |
| JC644 | TCGCTGTCATTTCGTCCCTCAAATCGTGGC |
| JC647 | TTTGAGGGACGAAATTCTTGCCGGTCAACCCG |
| JC648 | CTGCGCTAGATTAGTGAGTCAGAGCCGGCG |
| JC649 | TGACCCGCTTATGCTGTACAACATCCCTGGG |
| JC650 | GGTCTGACGCTCTAAGCGGTGTAAAGACCGAGTTC |
| JC651 | CTGACTCACTAATCTAGCGCAGCTGCTTGGAAC |
| JC652 | TGTACAGCATAAGCGGGTCACTACTTGGTAGC |
| JC653 | CCGCTTAGAGCGTCAGACCCCGTAGAAAAG |
| JC654 | GGCAAGAATTTCGTCCCTCAAATCGTGGC |
| JC655 | TGAGGGACGAATCAACGCTGAGTTTGCATTGTTG |
| JC656 | CTGCGCTAGACGATAGCGAAAGCAGACAACTCTG |
| JC657 | TTCGCTATCGTCTAGCGCAGCTGCTTGGAA |

**Table 3.7 (cont.):** Oligonucleotides used for plasmid cloning.

| | |
|---|---|
| JC658 | CGGCAAGAATAAGCGGGTCACTACTTGGTAGC |
| JC659 | TGACCCGCTTATTCTTGCCGGTCAACCCG |
| JC660 | GGTCTGACGCTTTAGTGAGTCAGAGCCGGCG |
| JC661 | TGACTCACTAAAGCGTCAGACCCCGTAGAAAAG |
| JC662 | CAGCGTTGATTCGTCCCTCAAATCGTGGCC |
| JC669 | GAGGGACGAAATTGCATGAGATGAAGGCGATCG |
| JC670 | CTGCGCTAGAAGAGATTGGAGACGTGGTGGAG |
| JC671 | TCCAATCTCTTCTAGCGCAGCTGCTTGGAA |
| JC672 | AGCTTGCAGAAAGCGGGTCACTACTTGGTAGC |
| JC673 | TGACCCGCTTTCTGCAAGCTGGCGAACATG |
| JC674 | GGTCTGACGCTAACAATGCGCGTTCGGTCATC |
| JC675 | GCGCATTGTTAGCGTCAGACCCCGTAGAAAAG |
| JC676 | CTCATGCAATTTCGTCCCTCAAATCGTGGCC |
| JC677 | GAGGGACGAATACTGACGCTAGCGGACCTTTC |
| JC678 | CTGCGCTAGAGACATTATTGCCATCCCCGAGG |
| JC679 | GGCAATAATGTCTCTAGCGCAGCTGCTTGGAAC |
| JC680 | TGGCAACAATAAGCGGGTCACTACTTGGTAGC |
| JC681 | TGACCCGCTTATTGTTGCCAAAGCGCAGCAG |
| JC682 | GGTCTGACGCTCTAGATAGCGCCAAGTAGTAAGGC |
| JC683 | CGCTATCTAGAGCGTCAGACCCCGTAGAAAAG |
| JC684 | AGCGTCAGTATTCGTCCCTCAAATCGTGGC |
| JC685 | GAGGGACGAAGAATTGTTTCTCACCTCGCTGAAGC |
| JC686 | CTGCGCTAGACATTCATCAGCACAGTGCTGGG |
| JC687 | GCTGATGAATGTCTAGCGCAGCTGCTTGG |
| JC688 | GTAATCAGAATTGAAGCGGGTCACTACTTGGTAGC |
| JC689 | TGACCCGCTTCAATTCTGATTACGCCCGATCTGG |
| JC690 | GTCTGACGCTCTAGCGATCCAAAGATTGCTGGC |
| JC691 | GGATCGCTAGAGCGTCAGACCCCGTAGAAAAG |
| JC692 | GTGAGAAACAATTCTTCGTCCCTCAAATCGTGGC |
| LL34 | GCACCCATTCAAGCGGGTCACTACTTGGTAGCA |
| LL36 | TGACCCGCTTGAATGGGTGCGGCAGGTGAG |
| ZD1 | CACCCAACGAGCGTCAGACCCCGTAGAAAAGATCAA |
| ZD2 | GGTCTGACGCTCGTTGGGTGCTTTGTTTGCGGG |
| ZD3 | GATTTGAGGGACGAAAGGTGATCAAGTCTGGCCCCTATAAGGAC |
| ZD4 | GCTGCGCTAGAGGCCTATCCTCCTGGGCAATCATCC |
| ZD5 | AGGATAGGCCTCTAGCGCAGCTGCTTGGAACTG |
| ZD6 | ACTTGATCACCTTTCGTCCCTCAAATCGTGGCCC |
| ZD7 | AGTGACCCGCTTACGCAATGACTTAGACCTGATTCGCTGC |
| ZD8 | GTCATTGCGTAAGCGGGTCACTACTTGGTAGCAACTC |
| ZD9 | TGACCCGCTTAGTTCTGAACTCCATCGTTCTGAG |
| ZD10 | GTCTGACGCTATCATGGCACCAGAACAGTCCC |
| ZD11 | GTGCCATGATAGCGTCAGACCCCGTAGAAAAG |
| ZD12 | AGTTCAGAACTAAGCGGGTCACTACTTGGTAGC |
| ZD13 | GCTGCGCTAGACTACATTCCCATAATGCTGTAGCCG |
| ZD14 | GGGAATGTAGTCTAGCGCAGCTGCTTGGAA |
| ZD15 | GAGGGACGAAGTAAAGAAGAACTGTCGGGTGAC |
| ZD16 | CAGTTCTTCTTTACTTCGTCCCTCAAATCGTGGC |
| SA38 | GCCCGATCATTTCGTCCCTCAAATCGTGGC |
| SA39 | CTACGATCCCTCTAGCGCAGCTGCTTGGAA |
| SA40 | CTGCGCTAGAGGGATCGTAGGCGAGCAATTTCA |
| SA41 | GAGGGACGAAATGATCGGGCGAATCTGCGG |
| SA42 | TGACCCGCTTAAGGGGAATTAGCCGCGCAA |
| SA43 | GTCTGACGCTCTAGAGCACCACCGTGTAGGC |
| SA44 | AATTCCCCTTAAGCGGGTCACTACTTGGTAGCA |
| SA45 | GGTGCTCTAGAGCGTCAGACCCCGTAGAAAAG |

**Table 3.7 (cont.):** Oligonucleotides used for plasmid cloning.

| | |
|---|---|
| SA57 | GCTGCGCTAGATCGAGGTCTCCCAACTCTGGC |
| SA58 | GAGACCTCGATCTAGCGCAGCTGCTTGGAA |
| SA59 | TCTCTGCCTAGTTCGTCCCTCAAATCGTGGCC |
| SA60 | GAGGGACGAACTAGGCAGAGACGGGAGATTCTAAA |
| SA61 | ACCATGGTCTAAAGCGGGTCACTACTTGGTAGC |
| SA62 | TGACCCGCTTTAGACCATGGTGTTGAAAGCTCGC |
| SA63 | GGTAATCCGATAAGCGTCAGACCCCGTAGAAAAG |
| SA64 | GGTCTGACGCTTATCGGATTACCCTGCTTCCTG |
| YG156 | TGAGGGACGAAGCAGTCAGCTTGCCATTGAAC |
| YG157 | AAGCTGACTGCTTCGTCCCTCAAATCGTGGCC |
| YG158 | GCTGCGCTAGACAGGTTAAGTCAGCAGAGAAC |
| YG159 | GACTTAACCTGTCTAGCGCAGCTGCTTGGAAC |
| YG160 | CGGTTCTAGCTAAGCGGGTCACTACTTGGTAG |
| YG161 | GTGACCCGCTTAGCTAGAACCGTTACCCGCAC |
| YG162 | GGTCTGACGCTTGAGCTAATAGCTCTAAGCCG |
| YG163 | CTATTAGCTCAAGCGTCAGACCCCGTAGAAAAG |

**Table 3.8**: Firefly luciferase gBlock gene fragment:

ATGGAAGATGCGAAGAACATCAAGAAGGGCCCCGCACCCTTCTATCCCCTGGAGGACGGAACAGCTG
GCGAGCAGCTGCACAAGGCGATGAAGCGTTACGCGCTTGTCCCTGGGACGATCGCGTTCACGGATGCC
CACATCGAGGTGAATATCACGTATGCAGAATACTTTGAGATGTCTGTGCGTTTAGCCGAGGCGATGAA
GCGCTACGGTTTAAACACAAATCACCGTATCGTTGTTTGCTCCGAGAATTCACTGCAATTTTTTATGCC
GGTACTTGGAGCTCTGTTTATCGGGGTTGCTGTAGCTCCGGCCAACGACATTTATAACGAACGTGAATT
GTTAAACAGCATGAATATTTCACAACCAACCGTAGTTTTTGTAAGCAAAAAGGGACTTCAGAAAATTC
TGAATGTTCAGAAAAAACTTCCCATTATTCAAAAAATTATCATTATGGACTCAAAGACCGATTACCAG
GGGTTCCAGAGCATGTACACTTTTGTAACCAGCCACTTACCACCCGGCTTTAATGAGTATGATTTCGTG
CCTGAAAGCTTCGACCGCGATAAAACCATCGCCTTAATTATGAACTCGTCGGGTTCGACAGGCTTACC
CAAGGGCGTGGCCTTGCCGCATCGTACAGCCTGTGTCCGTTTCTCTCACGCACGTGACCCCATTTTCGG
CAACCAAATTATCCCAGACACCGCAATTCTGAGTGTGGTACCTTTCCATCATGGATTCGGAATGTTCAC
TACGTTAGGATACTTGATTTGCGGTTTTCGCGTAGTATTGATGTATCGTTTTGAGGAGGAATTATTTTTG
CGTAGTTTGCAGGACTATAAGATTCAGAGTGCTTTGTTAGTGCCGACTTTGTTTTCCTTTTTTGCGAAGT
CAACACTTATCGACAAGTATGACTTAAGCAATTTGCACGAGATCGCATCGGGGGGAGCACCTTTGAGC
AAAGAGGTGGGGGAAGCGGTTGCGAAGCGTTTCCATTTGCCAGGCATCCGCCAGGGGTACGGATTAA
CGGAAACAACTTCAGCAATTTTAATTACGCCCGAGGGGGATGACAAGCCCGGTGCTGTAGGTAAGGTG
GTTCCGTTTTTTGAAGCGAAGGTTGTGGATTTGGATACCGGCAAGACACTTGGCGTGAACCAACGTGG
GGAGCTGTGCGTTCGCGGACCAATGATTATGAGCGGTTATGTTAATAACCCGGAGGCGACTAACGCCC
TTATCGACAAAGATGGCTGGCTGCACTCAGGGGATATTGCGTACTGGGATGAGGATGAGCATTTCTTC
ATCGTGGATCGCCTTAAATCATTGATCAAATATAAAGGATACCAAGTTGCACCAGCGGAGTTGGAGAG
CATCTTGTTACAGCATCCAAACATTTTCGACGCAGGTGTTGCGGGACTGCCCGATGACGATGCTGGGG
AGCTGCCGGCAGCCGTAGTGGTTCTGGAACATGGCAAAACTATGACAGAGAAAGAGATCGTCGATTA
CGTGGCCTCGCAAGTCACCACTGCAAAAAAACTTCGTGGTGGCGTAGTGTTCGTTGATGAGGTTCCGA
AAGGACTGACTGGCAAACTTGACGCGCGCAAGATTCGTGAAATTCTTATTAAAGCAAAGAAGGGGGG
AAAATCGAAGCTGTGA

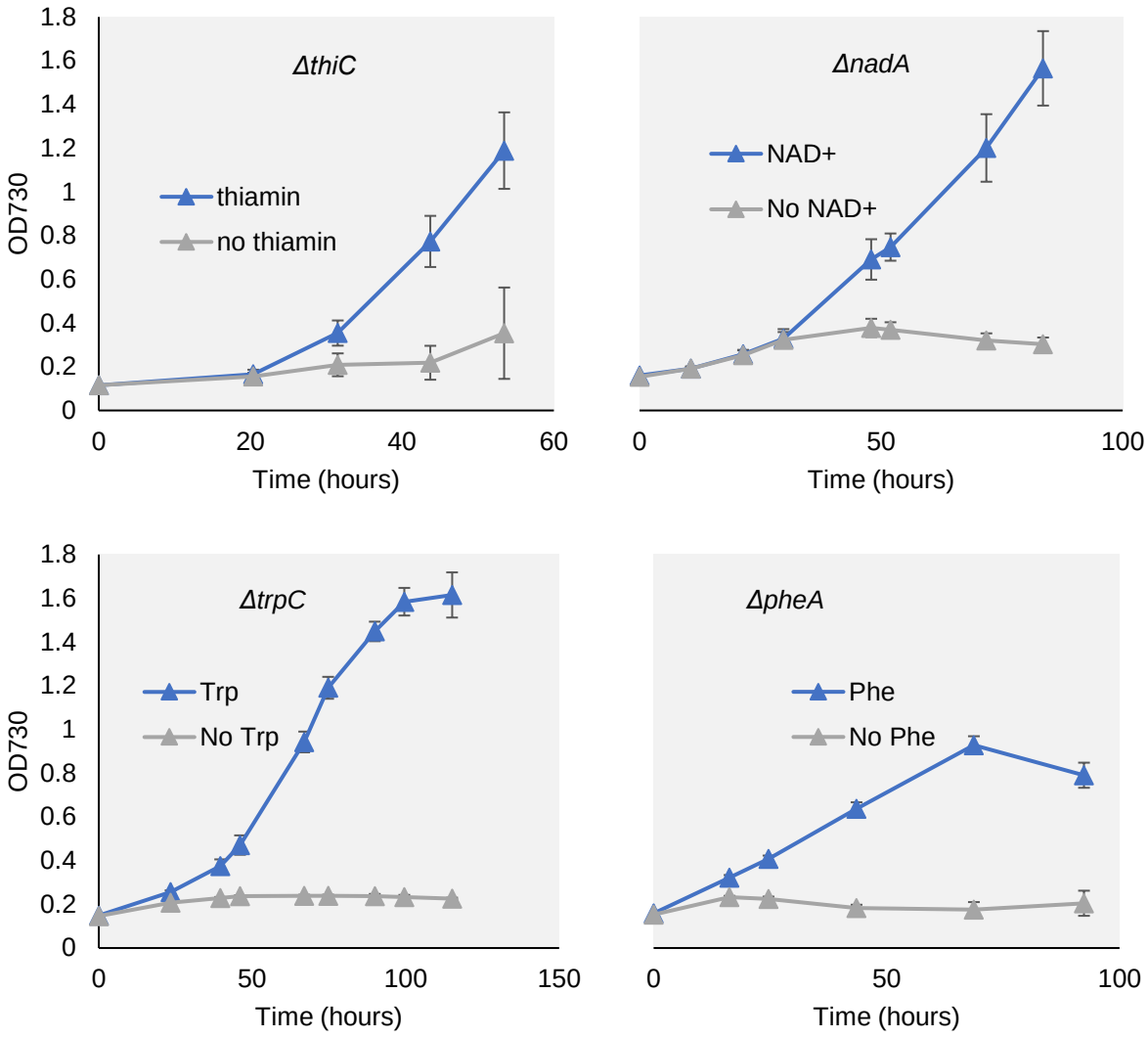



**Figure 3.3**. Growth curves of auxotrophic Syn7942 mutants complemented with 1 mM single amino acids.

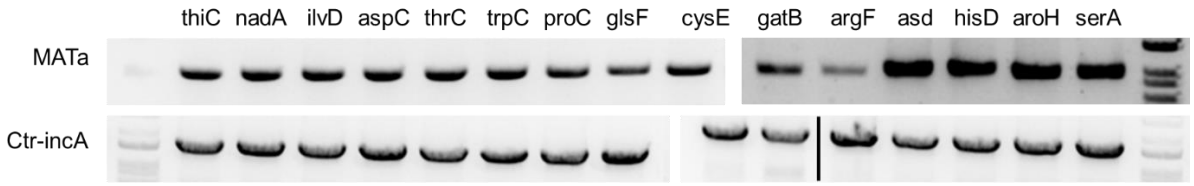



**Figure 3.4**: PCR detection of yeast and cyanobacterial DNA in yeast-cyanobacteria chimeras

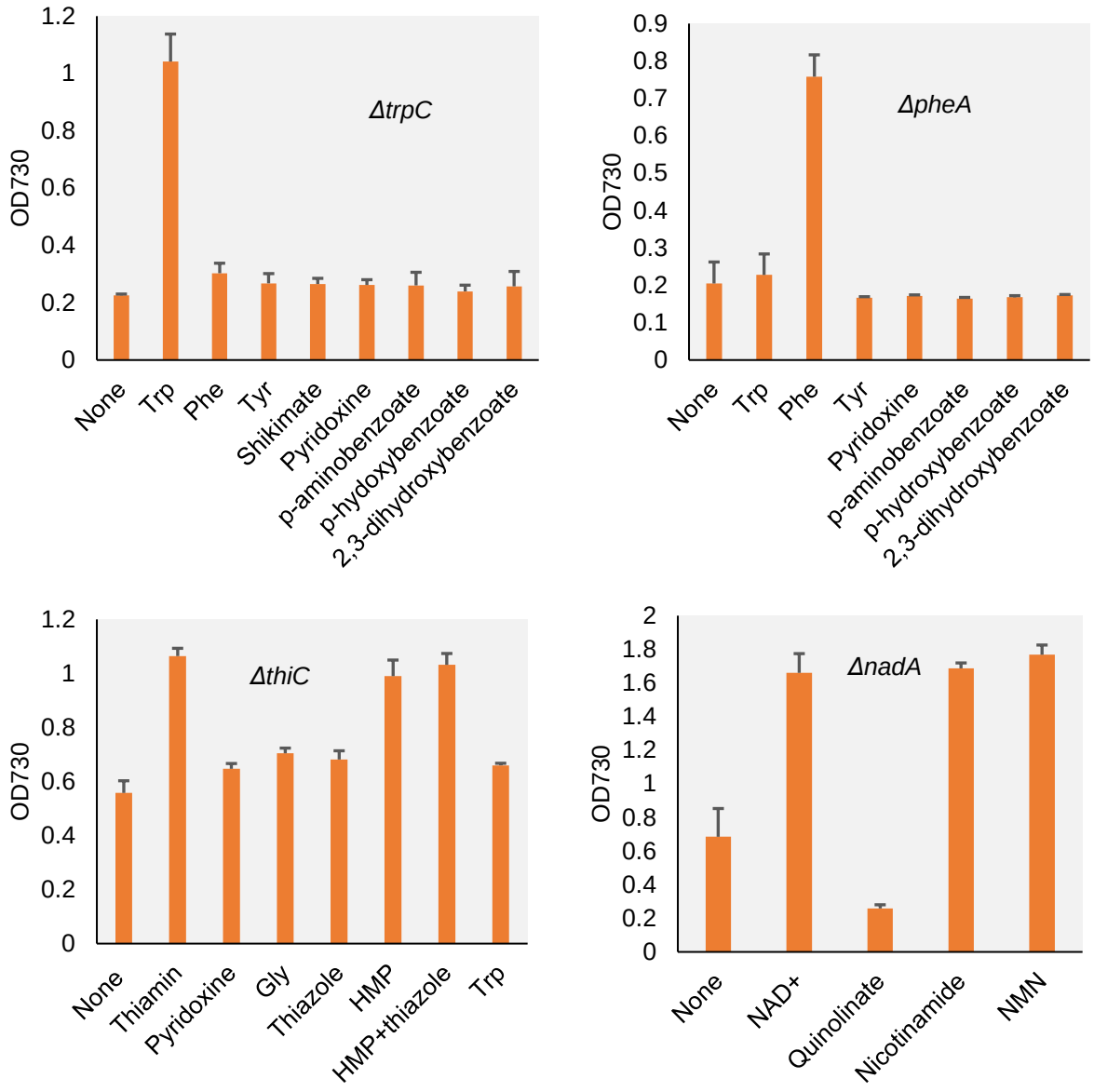
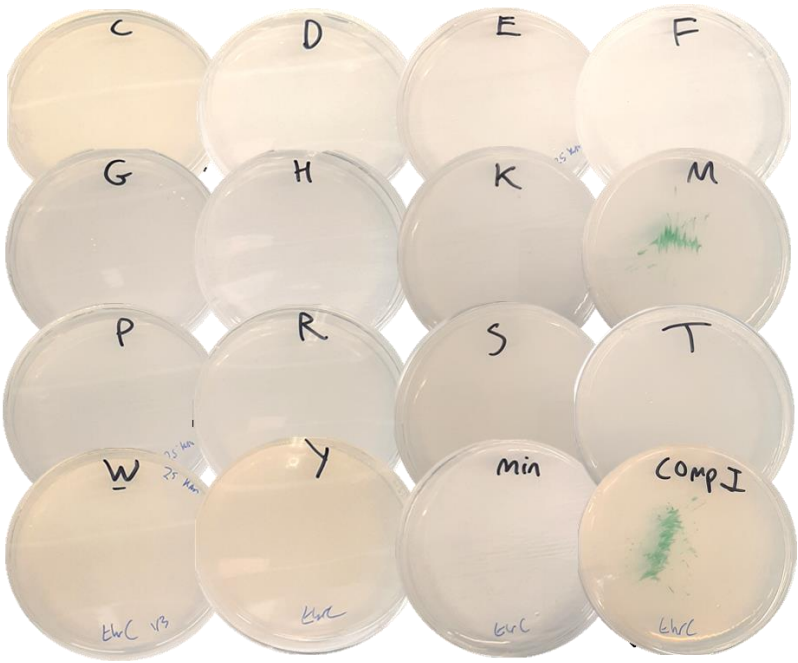



**Figure 3.5**. Endpoint growth of auxotrophic Syn7942 mutants cultured with media supplemented with biosynthetically-related metabolites.

A




B

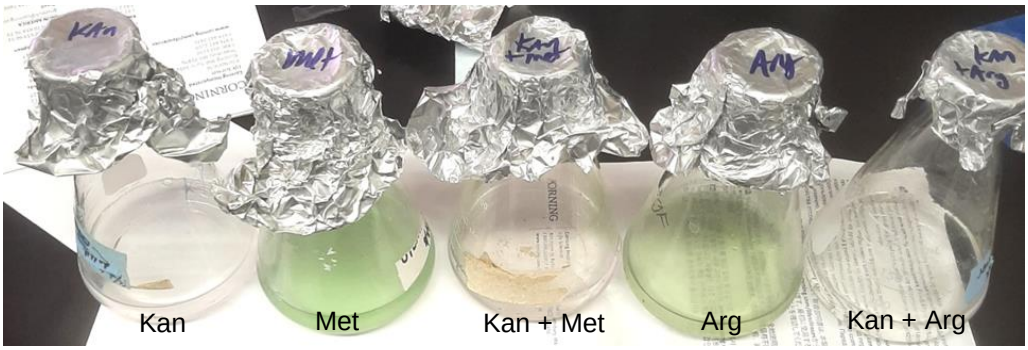



**Figure 3.6**. (A) *ΔthrC* cells were streaked on BG-11 medium supplemented with single amino acids (indicated by single letter codes). Comp-I—Complete I medium; min—BG-11 medium without supplementation. (B) Wild-type Syn7942 cells are challenged with 25 mg/L kanamycin in the presence or absence of 1 mM L-methionine or L-arginine. Growth was only observed in the flasks lacking kanamycin, indicating that the presence of amino acids did not inhibit kanamycin toxicity.

```
Query  9    RYLLARLLLLPLMLWTIATLIFLLLRATPGDPIDAILGPKAPAAARLA-LRSQLGLDRPL  67
            +++L RL L+      I  L F  +   PGDP+  + G +  +A R A L +++GLD+PL
Sbjct  3    QFILRRLGLVIPTFIGITLLTFAFVHMIPGDPVTIMAGERGISAERHAQLMAEMGLDKPL  62

Query  68   WQQYFDYLGQLLKGDLGQSLSSQGQSVRQIIGEYFPATAELAIASLAIAIAIGLGLGLMA  127
            +QQYF Y+  +L GDLG SL S+  SV       F AT EL   ++  A+ +G+ +G++A
Sbjct  63   YQQYFSYVSNVLHGDLGTSLKSR-ISVWSEFVPRFQATLELGFCAMLFAVLVGIPVGVLA  121

Query  128  AAKPNSRRETASRLFGILTYALPTFWAAMLAQLLFAVDLGWFPVGTRYP--VTLDPPLGP  185
            A K  S  +  +    +  Y++P FW  M+  +L +V L   PV  R    V LD
Sbjct  122  AVKRGSVFDHTAVGISLTGYSMPIFWWGMMLIMLVSVQLNLTPVSGRISDTVFLDDSQPL  181

Query  186  TGLYVLDALLKADWRAAGLALRYLALPALTLGLLLSGVFERLVRVNLGQSLQADYIDAGR  245
            TG  ++D L+  +      A+ ++ LPA+ LG +   V  R+ R ++ + L  DYI   R
Sbjct  182  TGFMLIDTLIWGEPGDFIDAVMHMILPAIVLGTIPLAVIVRMTRSSMLEVLGEDYIRTAR  241

Query  246  SRGLSERRLLLNHALRNALIPVVTLLGLTLASLLGGALLTEVAFSWPGLANRLYEAISLR  305
            ++G+S  R+++ HALRNAL+PVVT++GL + ++L GA+LTE  FSWPGL   L +A+  R
Sbjct  242  AKGVSRMRVIVVHALRNALLPVVTVIGLQVGTMLAGAILTETIFSWPGLGRWLIDALQRR  301

Query  306  DYSTVQGIIVFFGCLVVLASIVVDFINALIDPRIRY  341
            DY  VQG ++   C+++L +++VD +  +++PRIR+
Sbjct  302  DYPVVQGGVLLVACMIILVNLLVDVLYGVVNPRIRH  337
```

**Figure 3.7**: Bioinformatic identification of dipeptide/oligopeptide permeases in Syn7942

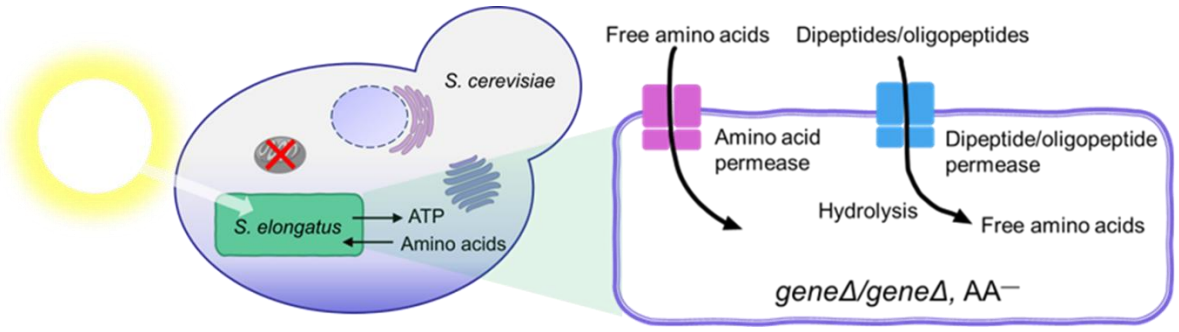



**Figure 3.8**: Rationale and synthetic scheme of Lys-based dipeptides.

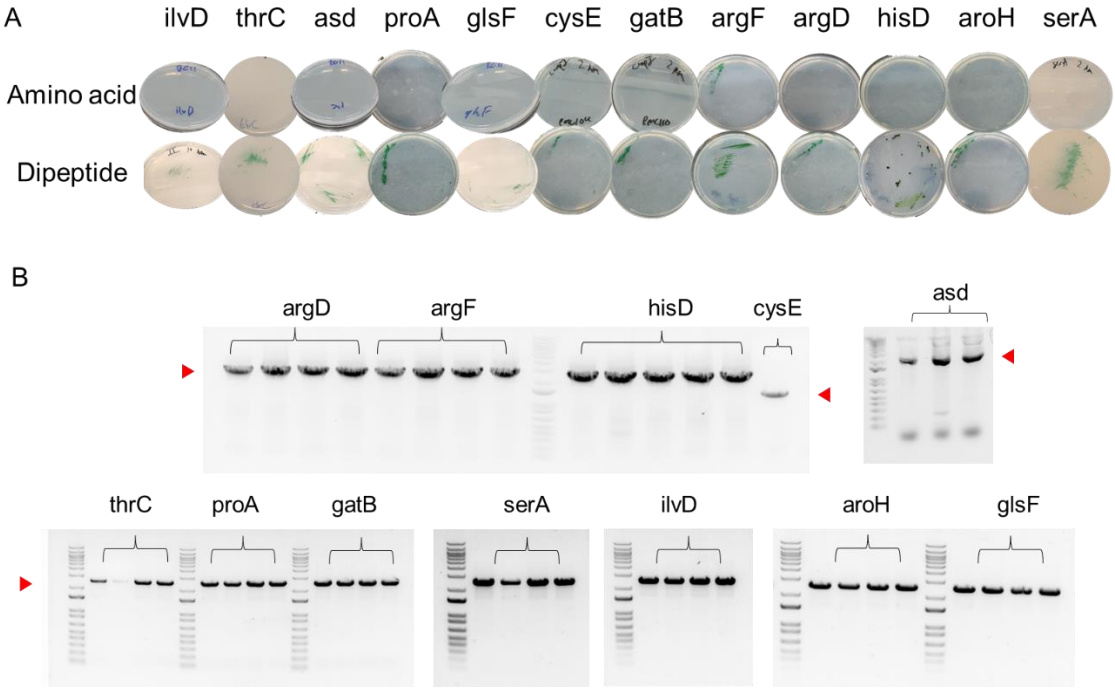



**Figure 3.9**: Auxotrophic strains grown on BG-11 medium supplemented with Lys dipeptides. (A) Growth phenotype on plates supplemented with single amino acids or Lys-containing dipeptides (B) PCR analysis of genomes of strains selected from dipeptide-containing media.

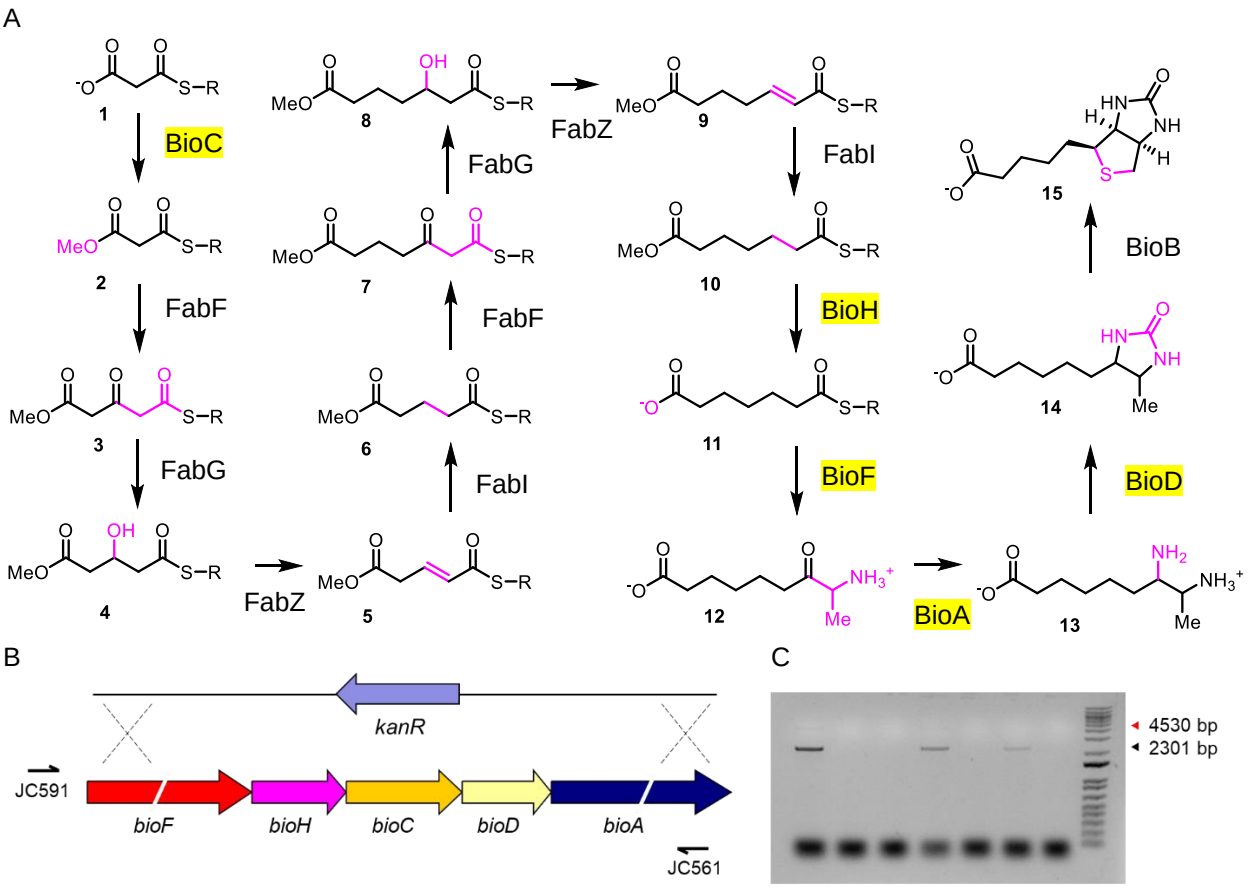



**Figure 3.10**: Deletion of *bioFHCDA* operon. (A) Inferred biotin biosynthesis in Syn7942 based on gene ontology. 1, malonyl-[acyl carrier protein, acp]; 2, malonyl-[acp]-OMe; 3, 3-ketoglutaryl-[acp]-OMe; 4, 3-hydroxyglutaryl-[acp]-OMe; 5, enoylglutaryl-[acp]-OMe; 6, glutaryl-[acp]-OMe; 7, 3-ketopimeloyl-[acp]-OMe; 8, 3-hydroxypimeloyl-[acp]-OMe; 9, enoylpimeloyl-[acp]-OMe; 10, pimeloyl-[acp]-OMe; 11, pimeloyl-[acp]; 12, 8-amino-7-oxononanoate; 13, 7,8-diaminononanoate; 14, dethiobiotin; 15, biotin. Biosynthetic transformations catalyzed by enzymes encoded by the *bioFHCDA* operon are highlighted. (B) Recombination scheme for knockout of the *bioFHCDA* operon and PCR primers, (C) Agarose gel of recombinant *bioFHCDA* operon in mutant Syn7942. The red arrow indicates the size of the wt allele, and the black arrow indicates the size of the fully-deleted mutant.

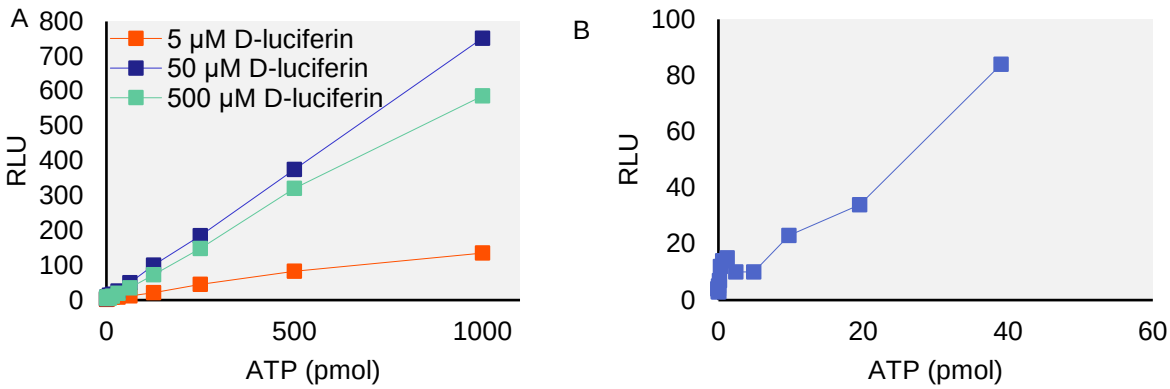



**Figure 3.11**: Validation of recombinant luciferase enzyme for translocase assays. (A) Standard curves of recombinant luciferase with carrying concentrations of D-luciferin substrate in order to maximize the amount of signal per mol of substrate. (B) Sensitivity of recombinant luciferase with diluted ATP standards. The graph suggests that the recombinant luciferase can detect accurately >10 pmol of ATP under optimized assay conditions.

**Table 3.9**. Identification of red algal aromatic amino acid permeases in red algae.

| | Predotar score × 100 | | | | TargetP score × 100 | | | | PredSL score × 100 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | mTP | cTP | ER | Other | mTP | cTP | Other | SP | mTP | cTP | SP | lTP |
| XP_005702608.1 | 1 | 0 | 0 | 99 | 0 | 0 | 100 | 0 | 59 | 0 | 80 | 0 |
| XP_005702902.1 | 1 | 0 | 0 | 99 | 0 | 0 | 100 | 0 | 0 | 0 | 100 | 0 |
| XP_005703914.1 | 4 | 0 | 0 | 96 | 0 | 0 | 100 | 0 | 48 | 0 | 0 | 0 |
| XP_005704447.1 | 1 | 0 | 0 | 99 | 0 | 0 | 100 | 0 | 54 | 0 | 86 | 0 |
| XP_005705149.1 | 1 | 0 | 0 | 99 | 0 | 0 | 100 | 0 | 6 | 0 | 8 | 0 |
| XP_005706910.1 | 1 | 0 | 0 | 99 | 0 | 0 | 100 | 0 | 12 | 0 | 0 | 0 |
| XP_005707288.1 | 4 | 0 | 0 | 96 | 0 | 0 | 100 | 0 | 3 | 0 | 3 | 0 |
| XP_005707586.1 | 2 | 0 | 0 | 98 | 0 | 0 | 100 | 0 | 4 | 0 | 4 | 0 |
| XP_005708981.1 | 1 | 0 | 0 | 99 | 0 | 0 | 100 | 0 | 0 | 0 | 97 | 0 |
| XP_005712768.1 | 1 | 1 | 0 | 98 | 0 | 0 | 100 | 0 | 0 | 0 | 100 | 0 |
| XP_005713394.1 | 1 | 0 | 40 | 59 | 0 | 2 | 96 | 1 | 98 | 0 | 4 | 0 |
| OSX70155.1 | 1 | 37 | 44 | 35 | 0 | 1 | 91 | 8 | 0 | 0 | 98 | 0 |
| OSX80505.1 | 2 | 29 | 32 | 47 | 0 | 80 | 20 | 0 | 100 | 0 | 0 | 0 |
| PXF42804.1 | 4 | 32 | 1 | 64 | 1 | 77 | 1 | 0 | 100 | 0 | 0 | 22 |
| KAA8497193.1 | 1 | 35 | 0 | 65 | 1 | 87 | 12 | 0 | 6 | 99 | 0 | 0 |
| KAI0560662.1 | 9 | 1 | 3 | 87 | 4 | 0 | 94 | 2 | 2 | 0 | 100 | 1 |
| GJD11683.1 | 1 | 1 | 14 | 85 | 0 | 0 | 100 | 0 | 1 | 0 | 100 | 0 |
| GJD09231.1 | 4 | 0 | 0 | 96 | 0 | 0 | 100 | 0 | 44 | 0 | 0 | 0 |
| GJD07843.1 | 1 | 0 | 2 | 98 | 0 | 0 | 100 | 0 | 80 | 0 | 60 | 0 |
| GJD06884.1 | 2 | 0 | 0 | 98 | 0 | 0 | 100 | 0 | 4 | 0 | 4 | 0 |
| GJD06867.1 | 1 | 0 | 0 | 99 | 0 | 0 | 100 | 0 | 59 | 0 | 80 | 0 |
| GJQ15736.1 | 1 | 0 | 1 | 98 | 0 | 0 | 100 | 0 | 8 | 0 | 86 | 0 |
| GJQ13988.1 | 2 | 0 | 0 | 98 | 0 | 0 | 100 | 0 | 3 | 0 | 2 | 0 |
| GJQ13387.1 | 1 | 0 | 0 | 99 | 0 | 0 | 100 | 0 | 0 | 5 | 96 | 0 |
| GJQ13279.1 | 1 | 0 | 0 | 99 | 0 | 0 | 100 | 0 | 1 | 0 | 0 | 0 |
| GJQ13164.1 | 1 | 2 | 99 | 1 | 0 | 0 | 100 | 0 | 52 | 0 | 87 | 0 |
| GJQ12168.1 | 2 | 0 | 0 | 98 | 0 | 0 | 100 | 0 | 2 | 0 | 1 | 0 |
| GJQ11816.1 | 2 | 0 | 0 | 98 | 0 | 0 | 100 | 0 | 6 | 0 | 5 | 0 |
| GJQ10514.1 | 1 | 0 | 0 | 99 | 0 | 0 | 100 | 0 | 6 | 0 | 2 | 0 |
| GJQ09156.1 | 1 | 0 | 0 | 99 | 0 | 0 | 100 | 0 | 0 | 0 | 94 | 0 |
| GJQ08530.1 | 2 | 0 | 1 | 97 | 0 | 0 | 100 | 0 | 1 | 0 | 4 | 0 |
| KAJ8900761.1 | 7 | 4 | 1 | 89 | 1 | 2 | 90 | 0 | 2 | 0 | 1 | 7 |
| KAJ8905642.1 | 1 | 0 | 0 | 99 | 0 | 0 | 100 | 0 | 1 | 0 | 1 | 0 |
| KAJ8905646.1 | 1 | 0 | 0 | 99 | 0 | 0 | 100 | 0 | 1 | 0 | 1 | 0 |
| KAJ8909182.1 | 1 | 0 | 0 | 99 | 0 | 0 | 100 | 0 | 0 | 0 | 57 | 0 |
| KAK1863313.1 | 9 | 41 | 2 | 53 | 0 | 38 | 62 | 0 | 99 | 1 | 0 | 0 |
| KAK1867025.1 | 2 | 17 | 99 | 1 | 1 | 31 | 61 | 7 | 3 | 71 | 31 | 0 |

## Chapter 4—Monitoring metabolic fluctuations in yeast-cyanobacteria chimera using single molecule fluorescence *in situ* hybridization

### Author contributions

Jason Cournoyer— conceptualization, methodology, validation, formal analysis, investigation, data curation, visualization, writing

Tianyou Yao—methodology, validation, formal analysis, investigation, data curation

Ido Golding— conceptualization, methodology, formal analysis, resources, data curation, supervision, project administration, funding acquisition

Angad Mehta— conceptualization, methodology, formal analysis, resources, data curation, supervision, project administration, funding acquisition

### Introduction

According to the endosymbiotic theory, membrane-bound organelles such as chloroplasts evolved through the combining of two distinct phylogenetic lineages into a single organism[3,5]. It is hypothesized that approximately 1-2 Gya, a β-cyanobacterium was engulfed by a heterotrophic eukaryotic cell and retained[13–15]. Over billions of years of evolution, endosymbionts descended from this branch of cyanobacteria lost most of their genomes through endosymbiotic gene transfer (EGT) and became fully-integrated organelles incapable of conducting most metabolic functions on their own[2,17]. Instead, most of those functions are encoded by the nucleus of the host cell. Recently, we engineered an artificial, photosynthetic endosymbiotic platform for studying molecular-level drivers of primary endosymbiosis, whereby mutant *S. elongatus* PCC 7942 (Syn7942) acts as a source of ATP from within the cytosol of respiration-deficient *S. cerevisiae*[75]. A wider goal of this artificial endosymbiosis project is to recapitulate evolutionary events in the initial stages of primary plastid endosymbiosis in a laboratory as mediated by metabolic selection pressures (directed endosymbiosis). Directed and long-term evolution experiments have previously been used to find molecules and microorganisms with emergent traits[181]. This is not the case only for plasmid libraries and mono-culture microorganisms, but for artificial microbial communities (consortia) as well[182]; in principle, similar emergent traits should be observed through directed endosymbiosis studies using artificial, chimeric organisms, which can be conceptualized as a form of microbial community.

Metabolic interconnectivity emerges as a necessary outcome of sustained endosymbiosis[183–185]. This trait has been studied in modern-day plastids and photosynthetic endosymbionts[183,184], but it is unclear how these networks were initially established. In this study, we aimed to use smFISH methodology to test whether adaptations could be observed at the transcriptional level between endosymbiotic Syn7942 cells within yeast. Designed to monitor stochastic gene expression on a cell-to-cell basis, smFISH probes single mRNA molecules with dozens of singly-labeled, fluorescent oligonucleotides[186]. In this way, individual mRNAs produce sufficient fluorescent signal to be discretely counted within the outlines of cells viewed by a fluorescent microscope[186]. Although we have isolated a series of auxotrophic Syn7942 mutants as a step towards engineering metabolic interdependency between hosts and endosymbionts (see **Chapter 3**), Syn7942 and *S. cerevisiae* are inherently asynchronous at the metabolic level. Thus, their forced cooperation under photosynthetic selection conditions may induce adaptations at the genetic or transcriptional level which cause some chimeric cells to gain a selective advantage over others. As a model target, we

designed probes to hybridize with transcripts for core photosystem reaction center proteins (psbAI and psaA). Previously, it has been shown that the levels of these transcripts are altered drastically upon exposure to high photosynthetically active radiation (PAR)[187–189], and that those changes can be detected by microscopy[190]. For the purposes of this study, this property was critical for two reasons: 1) automated mRNA-counting algorithms reported by Skinner and co-workers[186] require targets with varying expression levels as inputs, and 2) the regulation of photosynthesis is expected to be a critical factor in the fitness of chimeric organisms which rely on photosynthesis to generate ATP[75]. Insights gained from probing transcripts involved in photosynthesis could serve as guiding principles for optimizing PAR conditions under which to generate and propagate artificial photosynthetic chimeras. Other potential targets include transcripts of amino acid biosynthesis genes: in the case of auxotrophic endosymbionts fused with host cells, adaptations could occur at the transcriptional level which serve to complement nutrient deficiencies.

**Results**

**Designing sensitive probes for Syn7942**

*In situ* hybridization probes targeting bacterial rRNA have been utilized previously in whole-cell fluorescence assays with Syn7942 and related cyanobacteria[191,192]. These studies demonstrated that fluorescently-tagged nucleic acid probes could produce sufficient signal to detect binding above the noise of cyanobacterial auto-fluorescence. As shown in **Figure 4.4**, Syn7942 broadly absorbs blue and red light due to their enrichment of chlorophylls and phycobilins[64]. Accordingly, fluorescent probes are designed to absorb light in the green-yellow "trough" in the PAR range; in particular, 5-carboxytetramethylrhodamine (TAMRA) and BODIPY-FL fluorophores have been shown to function reliably[192,193]. As a first step, we acquired a universal ssDNA probe, EUB338[192,194] (5'-GCTGCCTCCCGTAGGAGT-3'[3' mdC(TEG-amino)]), which binds 16s rRNA (**Table 4.1**). The unlabeled probe was labeled with BODIPY-FL NHS ester (**Figure 4.5**) and hybridized with wild-type Syn7942 cells and *E. coli* cells as a control using established protocols[186,190,195]. Using confocal microscopy, we detected increased whole cell fluorescence after excitation with 488 nm laser scanning (**Figure 4.6**). Next, we designed oligonucleotide probes to detect discrete copy numbers of specific mRNAs. Single molecule fluorescence *in situ* hybridization (smFISH) is a method developed, whereby dozens of singly-labeled, ~20 bp probes hybridize to a single mRNA molecule, producing sufficient signal to detect the molecule as a spot localized within the cell outline by microscopy[196]. Analysis of spot numbers and intensities within populations of cells allow mRNA copy numbers to be estimated[186]. Using this method, researchers can determine the factors influencing stochastic gene expression (i.e., intrinsic fluctuations in expression levels on the level of individual cells)[197]. We first designed 32 probes to hybridize with an antibiotic resistance cassette (kanR transcript) integrated into the genome of mutant *S. elongatus* cells (*SynJEC12, ΔnadA::kan*) (**Table 4.1**). In the experiment, we coupled both TAMRA and BODIPY-FL NHS esters with the kanR probes (**Figure 4.5**). Inherently, low levels of off-target binding occur in smFISH experiments; the intensity of the resulting "false positive" spots serves as a threshold for detecting true spots in automated algorithms[186]. By probing wild-type and SynJEC12 cells, the kanR transcript served as a simple "on/off" to determine a false positive threshold. As expected, we observed fluorescent spots in Syn7942 cells probes with BODIPY-FL-tagged kanR probes (**Figure 4.7**). However, the total intensity of fluorescent signal (measured as pixel values in single-channel images of cells) was substantially greater in SynJEC12 cells expressing *kanR* gene (**Figure 4.8**). Unlike with the BODIPY-FL-tagged probes, no spots at all

were detected in SynJEC12 cells hybridized with kanR[TAMRA] probes (**Figure 4.9**). Possibly, this discrepancy can be attributed to some physicochemical property such as quantum yield ($\Phi_{TAMRA} = 0.1$[198]; $\Phi_{BODIPY\text{-}FL} = 0.9$[199]) affecting the ease with which FISH signal surpasses the innate fluorescence of Syn7942 cells. Consequently, no false positive threshold for mRNA detection could be determined using TAMRA-labeled probes, and subsequent experiments in probing Syn7942 exclusively used BODIPY-FL-labeled probes.

**Fluctuation of Syn7942 transcripts in response to stimuli**

Having shown that mRNAs could be detected in an on/off binary, we next sought to alter expression levels of a physiologically-relevant transcript in order to quantify fluctuations of gene expression over time in response to a stimulus. The automated spot-calculation algorithm reported by Skinner and co-workers[186] uses negative (false-positive), low expression and high expression images as inputs; accordingly, we identified genes encoding photosystem proteins (*psbAI* and *psaA*, encoding PSII D1 protein[187] and PSI apoprotein AI[200], respectively) as ideal targets. Previous studies have shown the abundance of these transcripts to decline up to 30-fold when cultures are shifted from low-light (50-125 μmol photons $m^{-2}s^{-1}$) to high-light conditions (500-750 μmol photons $m^{-2}s^{-1}$)[187–189]. Levels of psbAI and psbAIII transcripts increase in response to the same stimulus[188,189]. Recently, Mahbub and co-workers used 40-48 singly-labeled FISH probes targeting *psbAI* and *psaA* transcripts to determine the subcellular localization of mRNAs encoding these core photosystem proteins in Syn7942 and *Synechocystis sp.* PCC 6803[190]. Inspired by this study, we acquired unlabeled ssDNA probes targeting these same transcripts (the sequences match a subset of the probes previously reported by Mahbub et al., **Table 4.1**) and labeled them with BODIPY-FL NHS ester (**Figure 4.5**). We hybridized Syn7942 cells with these probes following pre-treatment with high light and imaged them with simultaneous phase contrast and fluorescence microscopy in multiple z-sections or focal lengths (see **Methods** section). This experiment resulted in an mRNA depletion profile broadly corroborating results previously reported by Mahbub et al. by FISH[190] and Kulkarni et al. by Northern blot[189] (**Figure 4.1**). Transcripts of *psbAI* were reduced drastically within the first 10 min of high light exposure, then recovered afterwards; transcripts of *psaA* depleted gradually throughout the 30 min time period (**Figure 4.2**). We next estimated the percent reduction of FISH signal in whole cells using image processing software (ImageJ 1.53c, **Methods**). We segmented multi-channel images into single, in-focus z-slices and detected the outlines of whole cells based on phycobilisome (PBS) fluorescence captured in the Cy5 channel. Our permeabilization/hybridization procedure utilizes multiple wash steps with EtOH, extracting a significant amount of chlorophyll *a* from the cells[190]. In contrast, the PBS content of the cells is mostly unaffected by treatment with EtOH. Consequently, fluorescent signal captured in the Cy5 channel is constant in all of the samples, given the exposure time is unchanged. After identifying >50 cell outlines per sample, we measured the mean fluorescence intensity within each outline in the GFP channel. Shown is **Table 4.2**, psaA transcript abundance was reduced by ~40% and psbAI transcript abundance was reduced by ~90%. Next, we took advantage of the constant PBS fluorescence detected in the images to determine a noise reduction threshold[190]. Briefly, we determined the ratio of total FISH signal to PBS signal in a population of cells (~12:1). The pixel values in each GFP channel image were divided by this ratio, and those values were subtracted from each raw GFP channel image (**Figure 4.10**). As shown in **Figure 4.11** and tabulated in **Table 4.2**, the background-corrected measurements were almost unchanged from measurements taken from the raw data.

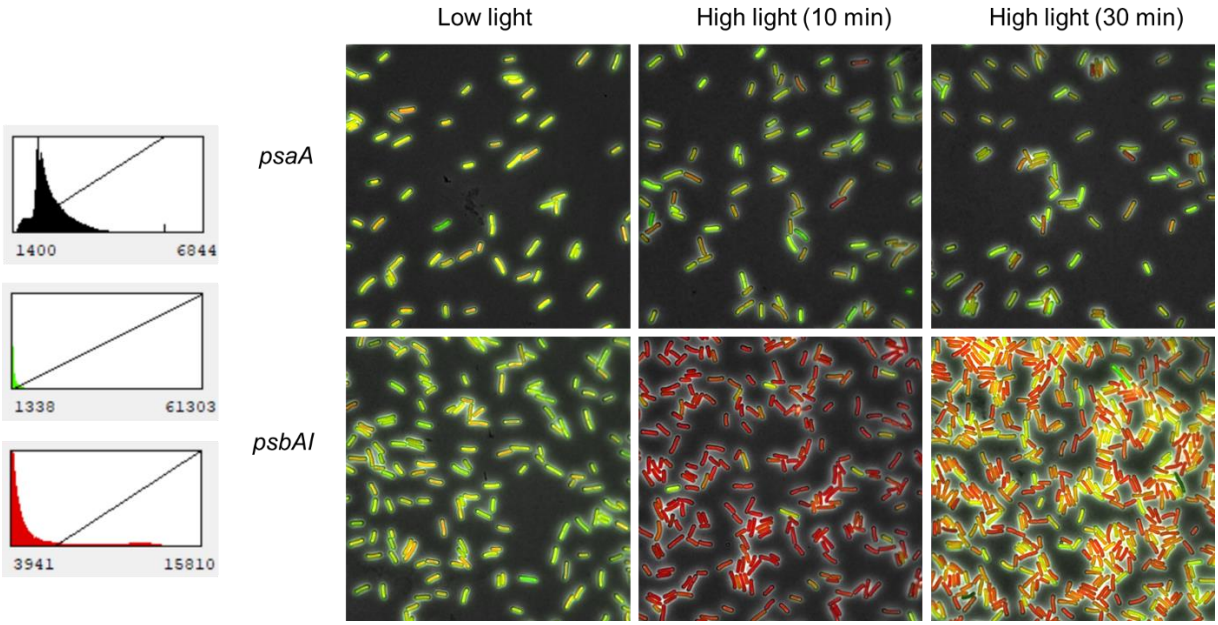



**Figure 4.1**. Effect of high light treatment on FISH signals in Syn7942. The microscopic images display an overlay of the phase contrast (B&W), FISH (BODIPY-FL, green), and phycobilisome fluorescence (PBS, red) channels. Overlap of the FISH and PBS channels appears yellow in color. Because PBS fluorescence is constant throughout the experiment, a shift in appearance of yellow/green to predominantly red-colored cells indicates a reduction (at the 10 min timepoint for *psbAI*) and subsequent recovery (30 min) of the corresponding RNAs. Left, the lookup tables (LUTs) are constant in all 6 images, meaning that the appearance of the images accurately reflects the pixel values of each channel.

Next, we used the same images to measure the mean intensity of mRNA foci (spots) within cell outlines, instead of whole cells. These intensities can be measured either in an automated fashion using the Spätzcells/Schnitzcells MATLAB programs[186] or by manually detecting spots in ImageJ through iterative rounds of edge detection. In this study, we used the latter approach: for each image, the threshold command was used to identify peaks in the isolated GFP channel within a narrow window of total pixel area and circularity. The window/level of the GFP channel image was continuously adjusted from highest to lowest LUT ramp position; thus, the brightest spots were detected first and added to a list of regions of interest (ROIs). By continuously adjusting the window/level, dim spots were revealed while bright spots exceeded the narrow size restriction and ignored (**Figure 4.12**). The ROIs in the list were filtered to remove doubly-counted spots, and the mean intensities within the bounds of those ROIs was measured. Shown in **Figure 4.12**, the intensities of these spots show a binomial distribution and can be fitted to 2D Gaussians.

**Adapting Syn7942 probe methodology to yeast-cyanobacteria chimeras**

Next, we sought to apply the methods above to an artificial, photosynthetic endosymbiotic system in order to monitor expression of genes involved in photosynthesis. In a previous study, we engineered yeast-cyanobacteria chimeras using respiration-deficient *Saccharomyces cerevisiae* as host cells and auxotrophic Syn7942 mutant cells as endosymbionts, providing the yeast cells with ATP[75]. Although this study showed that the endosymbiotic Syn7942 were capable of providing ATP generated by photosynthesis to the host yeast cells, the system is inherently

unpredictable given the asynchronous metabolisms of the partner cells. As directed endosymbiosis studies aim to engineer intricately interconnected metabolic networks between host cells and endosymbionts, smFISH-based methodology would be a powerful tool to monitor adaptations acquired by the chimeras to regulate gene expression. As a first step, we showed that hybridization protocols for yeast[196] and cyanobacteria could be combined into a single protocol without compromising the signal detection for yeast-specific or cyanobacteria-specific targets. As a yeast-specific probe, we used TAMRA-labeled ITS2 (**Table 4.1**) which binds a stem-loop region of 35S rRNA precursor[201] localized to the nucleolus[202] (**Figure 4.13**). We combined exponential-phase yeast and Syn7942 cultures shortly before fixation, permeabilization and hybridization with 1) 35S rRNA probe[TAMRA], and 2) 16S rRNA probe[BODIPY-FL]. Both yeast and cyanobacterial transcripts could be detected in tandem through the Cy5 and GFP channels, respectively (**Figure 4.14**). Next, we used an established protocol[75] to fuse yeast cells with SynJEC12 cells and hybridized the resulting endosymbiotic chimeras with 16S and kanR probes only, avoiding TAMRA-labeled probes in order to unambiguously attribute red fluorescent signals to the PBS (**Figure 4.3**, **Figure 4.15**). Here, we could detect clear instances of 16S- and kanR-specific FISH signals (GFP channel) co-localizing with innate PBS fluorescent signals. Because those mRNAs have half-lives lasting only minutes[189], this result suggested that the target genes were in a state of active transcription when the chimeras were fixed. Although the FISH signals were shown to necessarily co-localize with PBS signals, PBS signals did not always co-localize with FISH signals (**Figure 4.15**), suggesting the possibility that fragments of cyanobacterial membrane may fuse with yeast in the early stages of artificial endosymbiosis.

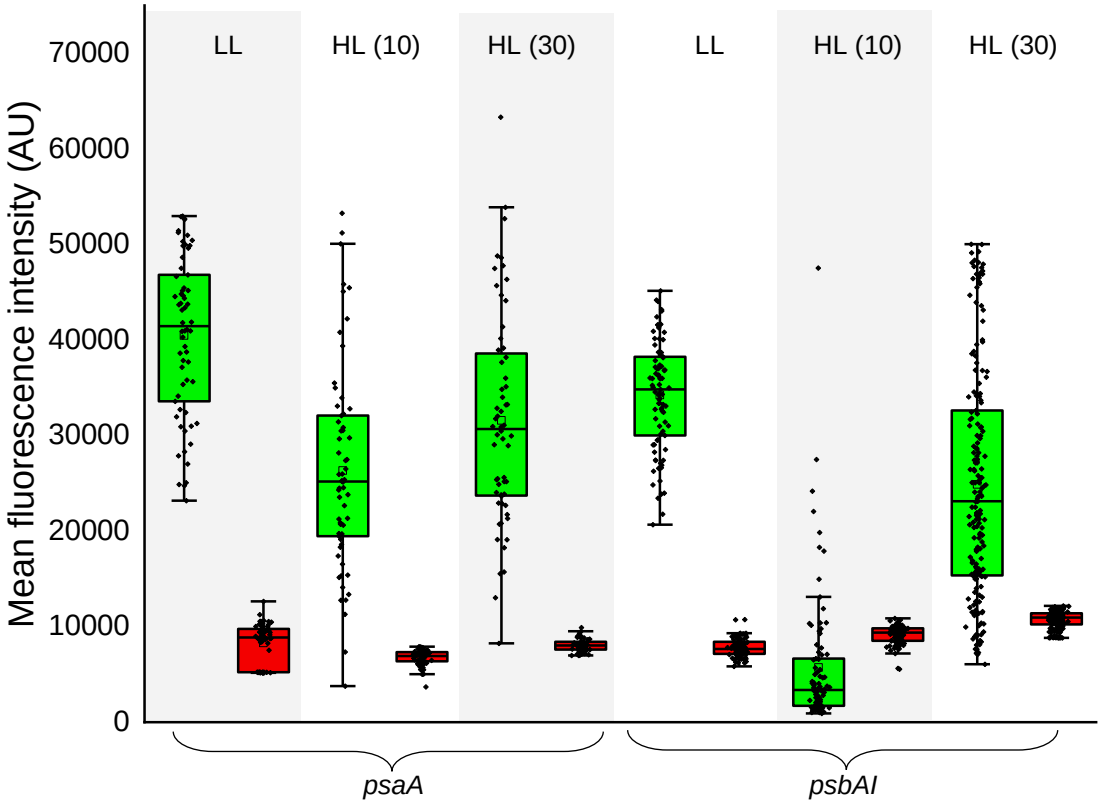



**Figure 4.2**. Comparison of mean whole-cell fluorescence values from the image sets shown in **Figure X-1** without background correction. Green boxes indicate the FISH channel and red boxes indicate the PBS channel. While PBS fluorescence is constant throughout the experiment, total FISH signal is altered in the full population of imaged cells. LL, low light; HL, high light (10 or 30 min).

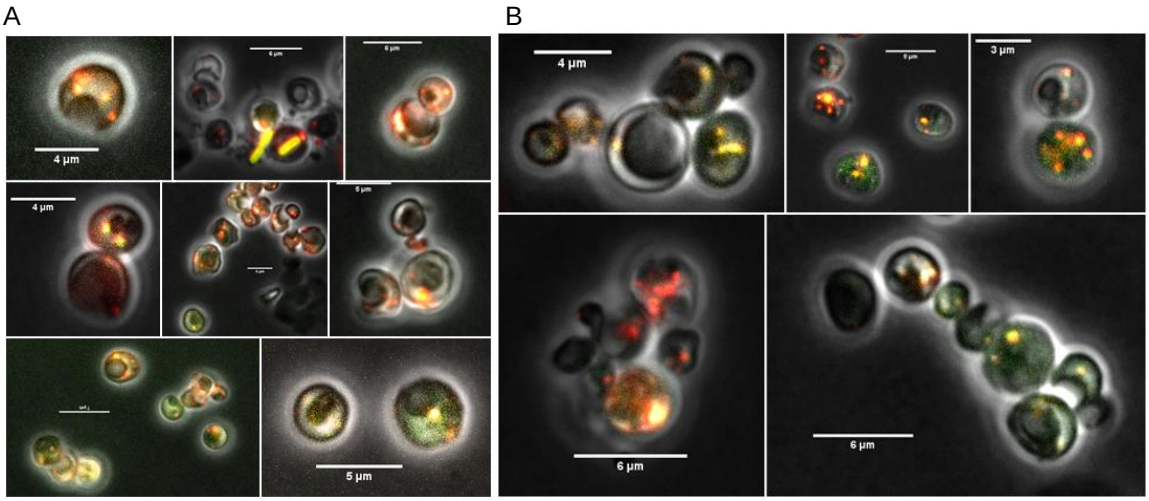



**Figure 4.3**. Probing yeast-cyanobacteria chimera. (A) Yeast-cyanobacteria chimera hybridized with kanR probes, (B) Yeast-cyanobacteria chimera hybridized with EUB338 (16S rRNA) probes. All images display merged phase contrast (B&W), FISH (GFP filter, green), and PBS (Cy5 filter, red) channels.

## Discussion

A necessary outcome of endosymbiosis is the evolution of metabolic interconnectivity between the endosymbiont and host cell[183–185]. While metabolic flux is well-studied in present-day endosymbiotic and organellar systems, comparatively little is known about how these networks emerge in the initial stages of endosymbiosis. We developed a synthetic endosymbiotic platform in order to recapitulate hypothesized adaptations which could have allowed primary endosymbionts to establish themselves as metabolically necessary components within larger cells[53,75,118]. In this platform, the core metabolic processes of the two organisms, *S. cerevisiae* and Syn7942, are inherently asynchronous; consequently, their forced metabolic coupling likely results in stochastic gene expression activity[203]. Cell-to-cell variations in the expression of genes relevant to processes such as stress response, cell cycle, and the circadian rhythm can be measured with smFISH[203]. In our synthetic endosymbiotic platform, the physiological response of chimeras to different light niches could provide insights into the role played by light sources in driving the acquisition of primary plastids.

In this study, we developed a protocol to simultaneously hybridize yeast and cyanobacterial cells with FISH probes and identified physiologically critical transcripts (psbAI and psaA) that can be counted inside single cells by microscopy. Currently, work is ongoing to show that psbAI and psaA transcript levels can be controlled and counted within endosymbiotic chimeras in the same manner as Syn7942 by itself. If this is shown in subsequent experiments, then those results could provide both evolutionary insights as well as a roadmap for optimizing culturing conditions of yeast-cyanobacteria chimeras. Another future application of this approach would be to probe the expression of host genes encoding amino acid biosynthesis enzymes. During the evolution of plastids, most of the genes encoding these enzymes were relinquished[204] by the endosymbiont; probing these genes encoded by Syn7942 or *S. cerevisiae* may show adaptations emerging at the transcriptional level which reflect this ceding of control by the endosymbiont and compensation by the host.

## Methods

### Strains and media

*S. elongatus* PCC 7942 was obtained from the lab of Prof. Susan Golden (UCSD). SynJEC12 cells were engineered by our group in a previous study (see **Chapter 3**). *S. cerevisiae* ρ + NB97 (MAT**a** *leu2-3,112 lys2 ura3-52 his3Δ HindIII arg8Δ::URA3* [*cox2-60*])[172] was obtained from the lab of Peter Schultz (Scripps Research). *E. coli* BL21(DE3) and 5-alpha strains were purchased from New England Biolabs. *E. coli* One Shot® *ccdB* Survival™ 2 T1R Chemically Competent Cells were purchased from Invitrogen. Cyanobacteria were cultured in BG-11 medium. Liquid cultures were shaken aerobically (250 rpm) in sterile Erlenmeyer flasks at 37 °C under 70-90 μmol photosynthetically active photons·$m^{-2}s^{-1}$ emitted by fluorescent bulbs. Agar plates were supplemented with $Na_2S_2O_3$ (100 μM) and grown at 30-34 °C under equivalent PAR emitted by an LED grow lamp. Yeast cells were shaken aerobically at 30 °C and 250 rpm in YPD medium (1% yeast extract, 2% peptone, 2% glucose) containing carbenicillin (50 mg/L). Selection medium I contained 1% yeast extract (Y), 2 % peptone (P), 3 % glycerol (G), 0.1 % glucose (D), 1 M sorbitol, 2 % agar, and 1X BG-11. Selection medium II contained 1% Y, 2 % P, 3 % G, 0.1 % D, 1 M sorbitol, 2 % agar, 1X BG-11 salts, and carbenicillin. Selection medium III contained 1% Y, 2 % P, 3 % G, no D, 1 M sorbitol,

2 % agar, 1X BG-11 salts, and carbenicillin. Plates containing Selection medium I, II and III were grown at 30 °C under 70-90 μmol photosynthetically active photons·$m^{-2}s^{-1}$ emitted by fluorescent bulbs.

**FISH probes**

Oligonucleotide probes corresponding to *metA, kanR, psbAI* and *psaA* transcripts were ordered from LGC Biosearch Technologies and designed to have a 3' mdC(TEG-amino) modification. They were synthesized on 50 nmol scale and purified by RPC. Upon receipt, the probes were centrifuged briefly and dissolved in TE buffer to a final concentration of 100 μM. The probes could be stably stored at -20 °C for >1 year. The ITS2 and EUB338 probes were ordered pre-labeled from the same vendor (3'-TAMRA for ITS2 and either 3'-TAMRA or C3-fluorescein for EUB338).

**Labeling FISH probes[186]**

Equal volumes of 100 μM oligo solutions corresponding to the *metA, kanR, psbAI*, or *psaA* transcripts were pooled to a final volume of 360 μL per respective target (i.e., 11.25 μL of metA-1, metA-2…metA-32 for each mRNA). To these oligo mixtures was added 40 μL of freshly-prepared 1 M $NaHCO_3$ dissolved in DEPC-treated water. TAMRA (Cayman 28509) or BODIPY-FL succinimide esters (Cayman 29508) (1 mg each) were dissolved in 2.5 μL DMSO and 25 μL 0.1 M $NaHCO_3$. The oligo solutions and fluorescent probe succinimide esters were mixed in an opaque black microcentrifuge tube and incubated in the dark overnight at 37 °C. Next, 3 M NaOAc (47 μL) was added to the labeled oligo solutions and 1180 μL of 100 % EtOH was added. The mixtures were stored at –80 °C for 3 h and then centrifuged at 15,000×g for 30 min (rt), leaving a precipitate. The supernatants were removed and the pellets were dissolved in DEPC-treated water (45 μL). 3 M NaOAc (5 μL) and 100 % EtOH (125 μL) were added, and the mixtures were stored again at –80 °C overnight. This centrifugation-precipitation step was repeated, and the mixtures were stored at -80 °C for another 3 h. The labeled probes were centrifuged at 15,000×g (rt), and the pellets were dissolved in 250 μL of 1× TE buffer. The labeled probes could be stored for > 1 year.

**Probe labeling efficiency**

Labeling efficiency (LE) of oligonucleotide probes is defined thus:

(2) $$LE = \frac{dye(\mu M)}{probe(\mu M)}$$

Single-stranded DNA concentration was measured in ng/uL using a Nanodrop UV-Vis spectrophotometer and converted to μM:

(3) $$Probe\ (\mu M)\ =\ (1{,}000\ /\ MWDNA)\ \times\ DNA\ (ng/\mu L)$$

Where $MW_{DNA}$ is estimated as the length of the probe (20 nt) × 303.7 g/mol per nucleotide. Concentration of the fluorescent dyes was measured using a Genesys 30 Visible Spectrophotometer with probe solutions diluted 20-fold in a 1 cm quartz cuvette ($A_{504}$ for BODIPY-FL and $A_{557}$ for TAMRA). Assuming εBODIPY = 70,000 $cm^{-1}M^{-1}$ and εTAMRA = 90,000 $cm^{-1}M^{-1}$, the concentration of the fluorophores was determined using Beer's law:

(4) $$A\ =\ \varepsilon \ell c$$

**Pre-treatment**

Syn7942 cultures were grown to log phase ($A_{750}$ = 0.5-0.7) and diluted in BG-11 medium (50 mL; $A_{750}$ = 0.3) to prevent occlusion of cyanobacterial cells during high light treatment. The diluted cultures were placed <1 cm from

a Feit Electric GLP24FS LED grow lamp (~750 μmol photons·$m^{-2}$·$s^{-1}$) and incubated at rt without shaking. The PAR intensity of the grow lamp was measured using either a Tekco Quantum PAR Meter (TK355PLUS) or the AMS AG TCS3701 ambient light sensor built into smartphones. The PAR intensity measurements were calibrated using the New Brunswick Innova 42 photosynthetic light bank[205].

**Probing of cyanobacterial cells[186,190]**

The following solutions were prepared prior to probing of yeast or cyanobacterial cells[186]:

1. Hybridization solution—40% (w/v) formamide, 2× SSC (Invitrogen AM9763), 10% dextran sulfate sodium salt, 0.1 mg/mL *E. coli* tRNA (Sigma R1753), and 10 mM of ribonucleoside vanadyl complex (VRC; NEB S1402S).
2. Wash solution—40% (w/v) formamide, 2× SSC

The hybridization solution could be stored for >1 year at -20 °C. The wash solution was prepared fresh immediately before probing and stored on ice.

To a log phase Syn7942 culture (50 mL; $A_{750}$ = 0.5-0.7) was added 1/9 volume of 37 % formaldehyde, to a final concentration of 3.7 %. The culture was transferred to a polypropylene culture tube and incubated for 30 min (rt) gentle mixing with a nutator. The cells were centrifuged at 400×g (rt) for 10 min and resuspended in 1× PBS (1 mL) in a microcentrifuge tube. The cells were washed twice with 1 mL 1× PBS using the same centrifugation conditions, then permeabilized by resuspending the cells in DEPC-treated water (300 μL) to which 100% EtOH (700 μL) was added. The cells were mixed for 1 h at room temperature, then centrifuged at 600×g (rt) for 10 min. The pellet was washed with wash solution with 5 min incubation with mixing. In separate, opaque black microcentrifuge tubes, oligonucleotide probe stock solutions were added to 50 μL aliquots of hybridization solution to a final concentration of 1 μM. The cells in wash solution were centrifuged under the same conditions as previous, resuspended in the hybridization solution aliquots with probes added, and incubated overnight at 30 °C. Portions of the samples (10 μL) were transferred to new opaque microcentrifuge tubes and washed three times, with 30 min incubation time (30 °C , no shaking) between each wash step, with wash solution (200 μL). The cells were centrifuged at 600×g (rt) for 10 min and resuspended in 10 μL of 2×SSC for imaging.

**Probing of yeast cells and yeast-cyanobacteria chimera**

Yeast cells and yeast-cyanobacteria chimeras were probed using an adapted protocol[206]. The following solutions were prepared prior:

1. Buffer B: (1.2 M sorbitol, 0.1M potassium phosphate dibasic)
2. Spheroplasting buffer: (10 mL Buffer B, 100 uL 200 mM vanadyl ribonucleoside complex (NEB))
3. Hybridization solution—40% (w/v) formamide, 2× SSC (Invitrogen AM9763), 10% dextran sulfate sodium salt, 0.1 mg/mL *E. coli* tRNA (Sigma R1753), and 10 mM of ribonucleoside vanadyl complex (VRC; NEB S1402S).
4. Wash solution—40% (w/v) formamide, 2× SSC

Yeast cells were cultured in 45 mL of synthetic defined (SD; minimal) medium to a density ($OD_{260}$) of 0.1-0.2. Formaldehyde (37%, 5 mL) was added directed to the growth medium and the cells were mixed for 45 min using a nutator. The cells were washed twice with 10 mL of ice-cold Buffer B, resuspended in Spheroplasting buffer, and

transferred to a microcentrifuge tube. Zymolyase 100T (Amsbio 120493-1, 1 mg) was added, and the cells were incubated for 15 min at 30 °C. The cells were centrifuged (500×g, 10 min, 4 °C) and washed twice with cold Buffer B. The cells were resuspended in DEPC-treated water (300 µL) and mixed with 100% EtOH (700 µL) and incubated overnight at 4 °C. The next day, the cells were washed twice with wash buffer and incubated overnight (30 °C, in the dark) in 25 µL of hybridization buffer containing 1 µM final concentration of probe(s), in a black opaque microcentrifuge tube. The next day, the yeast cells were washed 3 times with wash buffer (30 °C), with 30 min incubation time (30 °C, dark) between each wash step. For imaging, the cells were resuspended in 2×SSC.

Yeast-cyanobacteria chimeras were scraped from Selection-III medium plates, washed twice with 1×PBS medium and incubated overnight in 1×PBS to which formaldehyde had been added to a final concentration of 3.7 %. The chimeras were permeabilized with 70 % EtOH in two steps: 1) 1 h, room temperature, and 2) 2 h-overnight, 4 °C, in the dark. The washing, hybridization and imaging steps for the yeast-cyanobacteria chimeras proceeded in the same manner as the yeast cells following the extra permeabilization step.

**Microscopy**

For confocal microscopy, cyanobacterial samples were fixed with 3.7 % formaldehyde in PBS and wet-mounted on glass slides. The samples were with a Leica SP8 fluorescence confocal microscope equipped with 63X/1.40 HC PL APO Oil CS2 lens and excited with 488 nm and 561 nm lasers. Emission wavelengths in the 510/20 nm and >700 nm ranges were detected with photomultiplier tube (PMT) detectors.

SmFISH imaging was carried out as previously reported[207,208]. Briefly, 2 µL of each sample was resuspended in 2×SSC and spotted a circular coverslip. A pre-cast agarose pad (1.5 % agarose, 1×PBS) was placed on top of the spot and the sample was placed in an Attofluor Cell Chamber (Invitrogen A7816). The sample was viewed using an Eclipse Ti2E inverted microscope (Nikon) equipped with X-Cite XYLIS LED lamp, CMOS camera (Prime 95B, Photometrics) 100×, oil immersion phase contrast objective lens (CFI60 Plan Apo, Nikon; NA 1.45), and filter cubes corresponding to GFP (Nikon 96372), Cy3 (Nikon 96374), and Cy5 (Nikon 96376). Equivalent images were taken, for each sample, of the phase contrast and fluorescent channels, with 100 ms exposure time. For yeast and yeast-cyanobacteria chimeras, 21 sets of images were taken, in 300 nm increments of the. For bacterial cells, 5-7 images were taken in 300 nm increments of the z-axis. Each slide was imaged in multiple xy positions, as to see 100-200 distinct cell shapes per sample.

**Introduction of mutant cyanobacteria to S. cerevisiae cells:**

Syn7942 cells were introduced to *S. cerevisiae* spheroplasts as reported previously[75]. Briefly, mutant cyanobacteria were harvested in mid-log phase and resuspended in fresh BG-11 medium (500 µL). *S. cerevisiae cox2-60* cells were grown to a density ($OD_{600}$) of 0.8, harvested, and incubated for 30 min at 37 °C with Zymolyase 100 T (Amsbio 120493-1) in SCEM medium. The suspension was incubated for 30-60 min at 37 °C. The spheroplasts (750 µL) were washed with TSC buffer and mixed with dense Syn7942 culture (120 µL). The mixture was decanted into a round-bottom polypropylene tube containing PEG buffer (20% PEG 8000, 10 mM Tris-HCl, 2.5 mM $MgCl_2$, 10 mM $CaCl_2$, pH 8) (2 mL) and incubated statically (45 min, 30 °C). The cells were centrifuged (1500 × g, 10 min, rt), and YPDS was added on top without disturbing the pellet. The cells were incubated statically under light (5 h, 30 °C). The cells were resuspended in 1 M sorbitol (300 µL) and spread on Selection-I bottom agar medium. Selection-

I medium top agar was overlaid on top. Chimera were harvested after 4 d incubation in low light and propagated in spots on Selection-III medium.

**Supplementary information**

**Table 4.1**. Oligonucleotide probes

| Target RNA | Sequence (5'→3') | Name |
|---|---|---|
| 35S rRNA precursor[209] | ATAGGCCAGCAATTTCAAGTTAACTCCAAAGAGTATCACTC | ITS2 |
| 16s rRNA | GCTGCCTCCCGTAGGAGT | EUB338 |
| kanR | AGACGTTTCCCGTTGAATAT | kanR-1 |
| | AATTTAATCGCGGCCTGGAG | kanR-2 |
| | CCATATAAATCAGCATCCAT | kanR-3 |
| | ATTATCGCGAGCCCATTTAT | kanR-4 |
| | TAGATTGTCGCACCTGATTG | kanR-5 |
| | ATCGGGCTTCCCATACAATC | kanR-6 |
| | GTTTCAGAAACAACTCTGGC | kanR-7 |
| | CATTGGCAACGCTACCTTTG | kanR-8 |
| | TCTGACCATCTCATCTGTAA | kanR-9 |
| | ATAAATTCCGTCAGCCAGTT | kanR-10 |
| | ATGCTTGATGGTCGGAAGAG | kanR-11 |
| | CATCATCAGGAGTACGGATA | kanR-12 |
| | GGATCGCAGTGGTGAGTAAC | kanR-13 |
| | ATACCTGGAATGCTGTTTTC | kanR-14 |
| | CACCTGAATCAGGATATTCT | kanR-15 |
| | TGCCAGCGCATCAACAATAT | kanR-16 |
| | CAGGAATCGAATGCAACCGG | kanR-17 |
| | CGATCGCTGTTAAAAGGACA | kanR-18 |
| | CGCCTGAGCGAGACGAAATA | kanR-19 |
| | CCAAACCGTTATTCATTCGT | kanR-20 |
| | CGTCATCAAAATCACTCGCA | kanR-21 |
| | TTCAACAGGCCAGCCATTAC | kanR-22 |
| | TATGCATTTCTTTCCAGACT | kanR-23 |
| | TCCGGTGAGAATGGCAAAAG | kanR-24 |
| | ATCACCATGAGTGACGACTG | kanR-25 |
| | CCCCTCGTCAAAAATAAGGT | kanR-26 |
| | GTCCAACATCAATACAACCT | kanR-27 |
| | TATCGGTCTGCGATTCCGAC | kanR-28 |
| | TTCCATAGGATGGCAAGATC | kanR-29 |
| | AGGAGAAAACTCACCGAGGC | kanR-30 |
| | TTTGAAAAAGCCGTTTCTGT | kanR-31 |
| | ACTCATCGAGCATCAAATGA | kanR-32 |

**Table 4.1 (cont.):** Oligonucleotide probes

| | | |
|---|---|---|
| psaA | GTCGGAACCGGATTTTTATC | psaA-1 |
| | CTAGGTTGATCGACAGTTGG | psaA-2 |
| | CTTACCCCACTTCTCAAAAG | psaA-3 |
| | ACGATGATGCTGATCGAACC | psaA-4 |
| | TCCAAATCCAAGTTGTGGTT | psaA-5 |
| | AAGTACGGATACGGAGGCAT | psaA-6 |
| | AAATCGTGAGCGTTAGCGTG | psaA-7 |
| | CAATCCAGATGTGGTGAGTA | psaA-8 |
| | TCAAGGTCACTGGTATGACT | psaA-9 |
| | TGAAAATGGCAGCGTGAGCA | psaA-10 |
| | ACCAAAGTGAGCGCTGAAGA | psaA-11 |
| | AGGAGGTTGTCGACATTCTT | psaA-12 |
| | ACCAGATAAAGATCACCGCA | psaA-13 |
| | GTGTCGTTGTGGATGTAGAG | psaA-14 |
| | CAGCCGCTGAAATTGGAGAA | psaA-15 |
| | CGAGAACATGTCTTGGGGAC | psaA-16 |
| | GAAGATCGGCCAAACGACTT | psaA-17 |
| | GATCTGAATGCCATGGAAGC | psaA-18 |
| | CCGTGACGTAGAGCTGAAAC | psaA-19 |
| | CTTTGTGGTAGTGGAACCAG | psaA-20 |
| | TCGACGTTTTGGAACCATTC | psaA-21 |
| | CAAGTGGTGGTTCAACATCG | psaA-22 |
| | CGACACGTGAATCTGGTGAC | psaA-23 |
| | TTTTGCCGTTGAGAACCAAC | psaA-24 |
| | CGGAATATCAGCCGCAGAAG | psaA-25 |
| | TCAGGCTGACATCCAAGAAC | psaA-26 |
| | AGCGTGAAGAAGGCTTTCAC | psaA-27 |
| | ACCTTTGAAGGTCAGGAAGT | psaA-28 |
| | GTGACCCGCAACAATGAAGA | psaA-29 |
| | CCAAGATTTCTTTGAGGCTG | psaA-30 |
| | GTGAAAGGACCTTTGTGAGC | psaA-31 |
| | TCTCATAGAGACCTTTGTGG | psaA-32 |

**Table 4.1 (cont):** Oligonucleotide probes

| | | |
|---|---|---|
| psbAI | TCGCGAAGAATGCTGGTCAT | psbA-1 |
| | ACATGAAGTTGAAGGTGCCG | psbA-2 |
| | GATCCCAAACGTTATCGCGG | psbA-3 |
| | TTGTGCTCTGCTTGGAACAC | psbA-4 |
| | CTGGTTACCCACTCACAAAA | psbA-5 |
| | TGTGGAAGGGGTGCATCAAA | psbA-6 |
| | CGAACCAACCCACGTAGATG | psbA-7 |
| | AGAACAGCGAACCACCGAAC | psbA-8 |
| | AGAGTGGGGATCATCAGCAC | psbA-9 |
| | CACCAACGAACCGTGCATTG | psbA-10 |
| | ATGAAGCAGATGGTGGCGGT | psbA-11 |
| | TAGTTTTGGCTCTCGGTCTC | psbA-12 |
| | AGGGGCTGCAATGAACGCAA | psbA-13 |
| | CTCTTGACCAAATTTGTAGC | psbA-14 |
| | ATACATGAGAGAGCCGGCAA | psbA-15 |
| | CCACGATGTTGTAGGTCTCT | psbA-16 |
| | CGCCGGAAATGATGTTGTTG | psbA-17 |
| | ATCAAGCGACCGAAGTAACC | psbA-18 |
| | GCGTTGCTGGAAGGAACAAC | psbA-19 |
| | GTTGAACGATGCGTATTGGA | psbA-20 |
| | GCTTCCCAAATCGGATAGAA | psbA-21 |
| | GGAAGAAGTGCAGCGAACGG | psbA-22 |
| | TTGTACAGCCACTCGTCGAG | psbA-23 |
| | CATGGAGGTAAACCAGATGC | psbA-24 |
| | CCACTAATTGGTAAGGACCA | psbA-25 |
| | ATACCCAGCAAGAAGTGGAA | psbA-26 |
| | ATTGACGACCCATGTAGCAG | psbA-27 |
| | ATACCGAGGCGGTACGACAG | psbA-28 |
| | ATGCAACACAGATCCAAGGG | psbA-29 |
| | AAAAGCAGCCGAGAGTGGAG | psbA-30 |
| | CCGATCGGGTAGATCAGAAA | psbA-31 |
| | CATGCCGTCCGAGAACGAAC | psbA-32 |

**Table 4.2**. Comparison of means from raw data

| Probe | Treatment | Channel | N (cells) | Mean | StDev | Median | Signal reduction |
|---|---|---|---|---|---|---|---|
| psaA | LL | FISH | 62 | 40469 | 8379 | 41451 | |
| psaA | HL (10) | FISH | 62 | 26332 | 10688 | 25185 | 39% |
| psaA | HL (30) | FISH | 60 | 31582 | 11100 | 30688 | 26% |
| psbAI | LL | FISH | 87 | 34238 | 5708 | 34821 | |
| psbAI | HL (10) | FISH | 91 | 5720 | 6980 | 3373 | 90% |
| psbAI | HL (30) | FISH | 187 | 24877 | 12127 | 23119 | 34% |
| psaA | LL | PBS | 62 | 8290 | 2131 | 8857 | |
| psaA | HL (10) | PBS | 62 | 6777 | 783 | 6919 | |
| psaA | HL (30) | PBS | 60 | 8052 | 669 | 8025 | |
| psbAI | LL | PBS | 87 | 7810 | 963 | 7652 | |
| psbAI | HL (10) | PBS | 91 | 9159 | 961 | 9369 | |
| psbAI | HL (30) | PBS | 187 | 10696 | 929 | 10939 | |

**Table 4.3**. Comparison of means with background correction

| Probe | Treatment | Channel | N (cells) | Mean | StDev | Median | Signal reduction |
|---|---|---|---|---|---|---|---|
| psaA | LL | FISH | 62 | 39778 | 8463 | 40789 | |
| psaA | HL (10) | FISH | 62 | 25767 | 10680 | 24634 | 40% |
| psaA | HL (30) | FISH | 45 | 25380 | 9575 | 23580 | 42% |
| psbAI | LL | FISH | 87 | 33587 | 5695 | 34086 | |
| psbAI | HL (10) | FISH | 91 | 4957 | 6995 | 2563 | 92% |
| psbAI | HL (30) | FISH | 187 | 25276 | 13460 | 22967 | 33% |

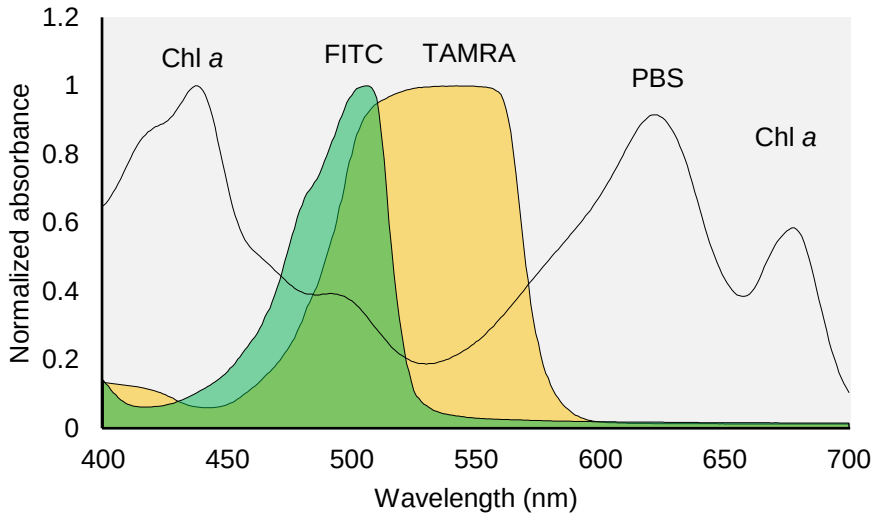
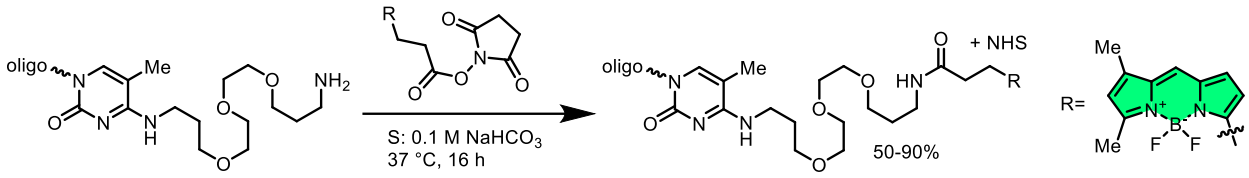


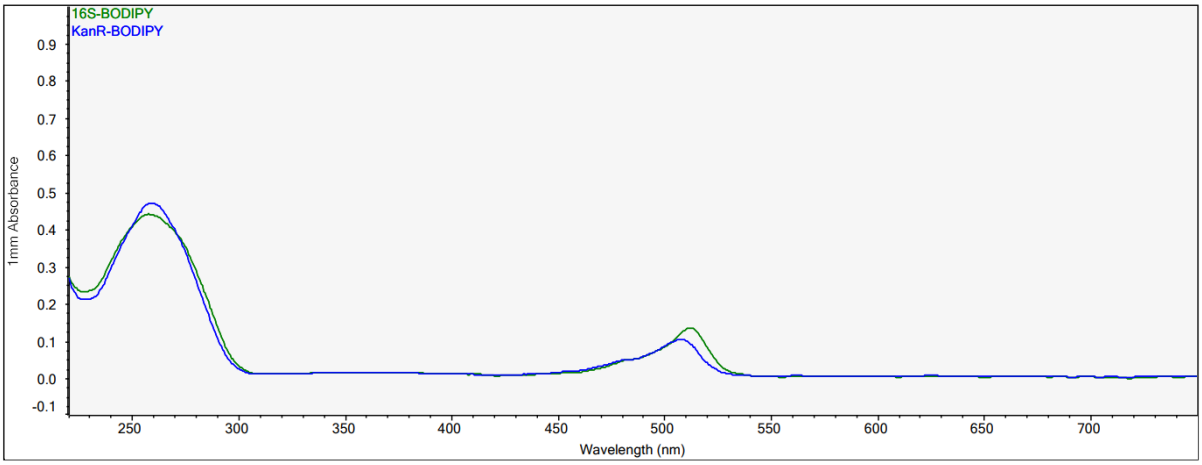

**Figure 4.4**. Spectrum of Syn7942 with probes. Green, FITC; Yellow, TAMRA; Chl *a*, chlorophyll *a;* PBS, phycobilisome.

A




B

| Sample ID | Nucleic Acid | Unit | A260 (Abs) | 260/280 | Factor | Sample Type | Dye 1 Abs (Abs) | Dye1 Name | Dye1 | Dye1 Unit |
|---|---|---|---|---|---|---|---|---|---|---|
| KanR-BODIPY | 150.5 | ng/µl | 0.467 | 1.79 | 33.00 | ssDNA | 0.083 | BODIPY-FL SE (pmol/µl) | 10.3 | pmol/µl |
| 16S-BODIPY | 140.4 | ng/µl | 0.436 | 1.51 | 33.00 | ssDNA | 0.082 | BODIPY-FL SE (pmol/µl) | 10.2 | pmol/µl |




C

| # | Sample ID | Nucleic Acid | Unit | A260 (Abs) | 260/280 | Factor | Sample Type | Dye 1 Abs (Abs) | Dye1 Name | Dye1 | Dye1 Unit |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | pS1 | 59.9 | ng/µl | 0.188 | 1.84 | 33.00 | ssDNA | 0.028 | BODIPY-FL SE (pmol/µl) | 3.4 | pmol/µl |
| 2 | pS2 | 65.9 | ng/µl | 0.208 | 1.88 | 33.00 | ssDNA | 0.035 | BODIPY-FL SE (pmol/µl) | 4.4 | pmol/µl |

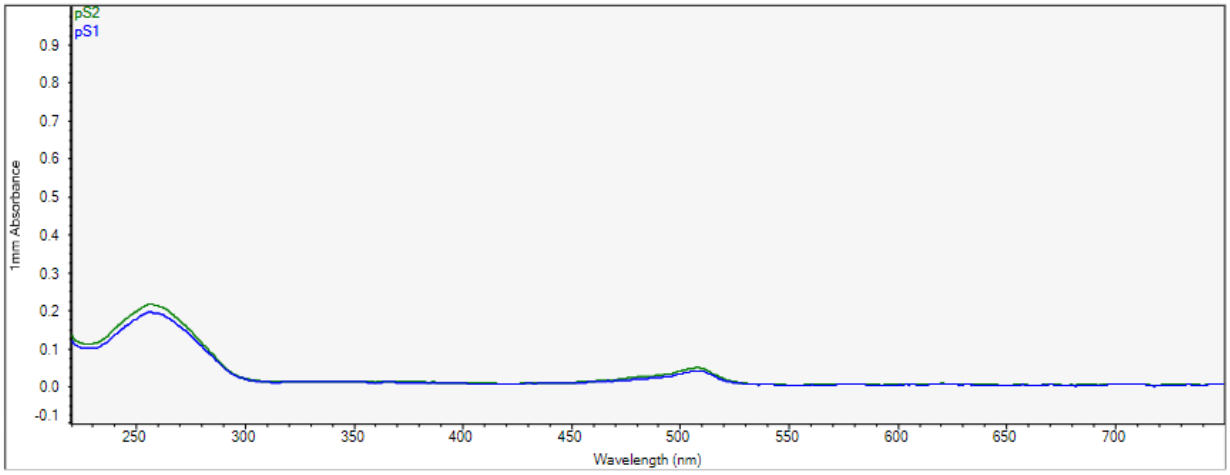



**Figure 4.5**. Probe labeling. (A) Synthetic scheme, labeling probes with 3'-TEG-amino group with BODIPY-FL NHS ester. (B) Nanodrop spectrophotometer traces of EUB338 and kanR probes labeled with BODIPY-FL. (C) Nanodrop spectrophotometer traces of psaA and psbAI probes labeled with BODIPY-FL.

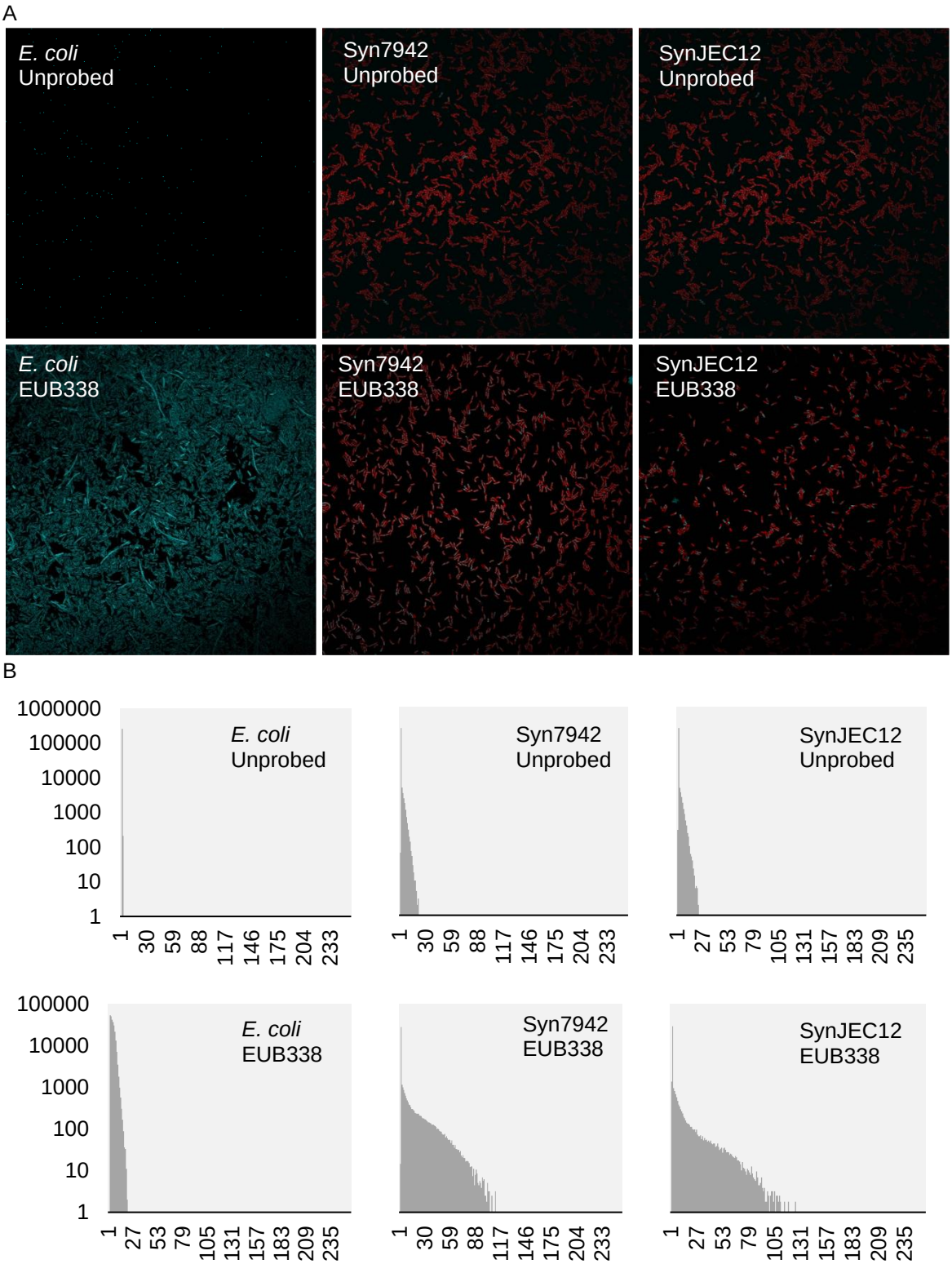



**Figure 4.6**. EUB338-BODIPY FL probes tested in *E. coli*, Syn7942, and SynJEC12 cells using confocal microscopy. (A) Representative microscopy images of cells excited with 488 nm laser. (B) Histograms of pixel values corresponding to each image in the green fluorescence channel only.

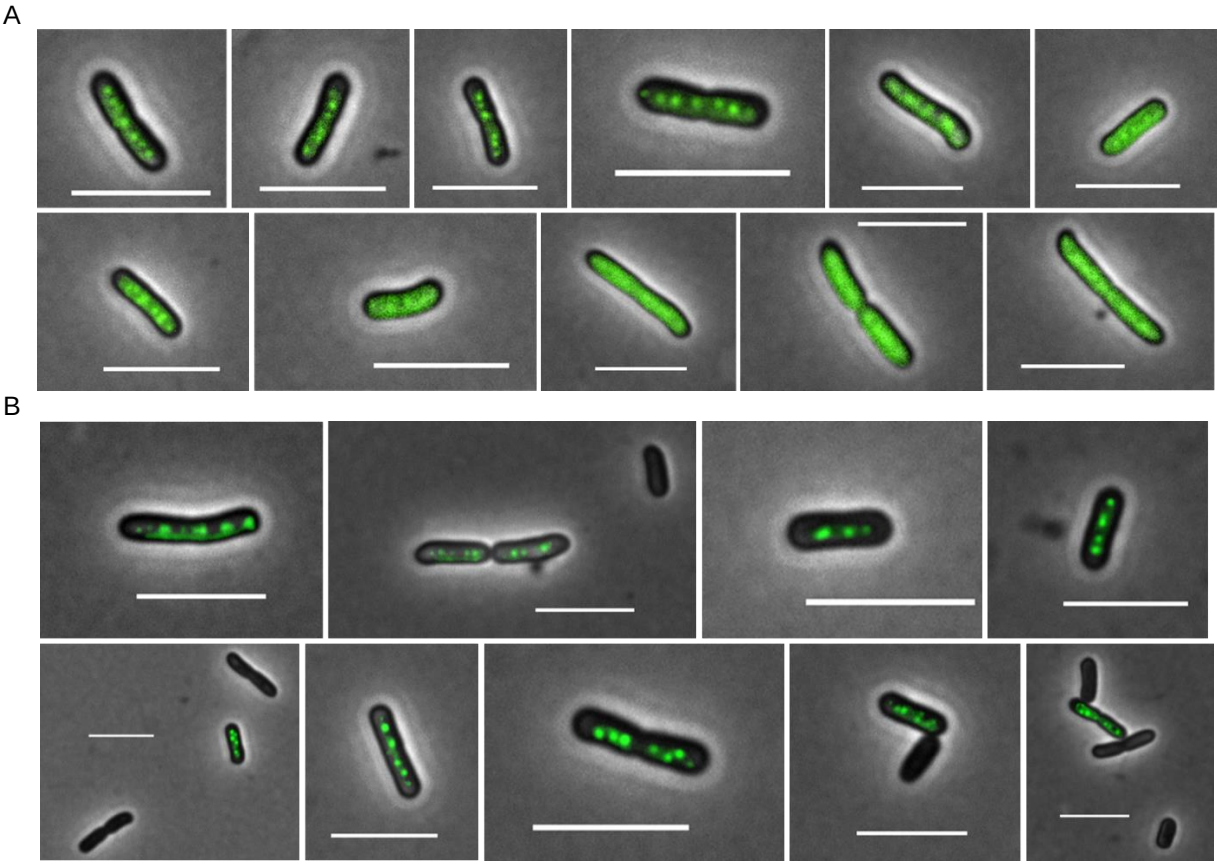



**Figure 4.7**. (A) Wild-type + kanR (BODIPY-FL)—Representative composite images (phase contrast + GFP cube). Green spots represent single mRNA molecules bound to probe (B) SynJEC12 + kanR (BODIPY-FL)—Representative composite images (phase contrast + GFP cube). Green spots represent single mRNA molecules bound to probe.

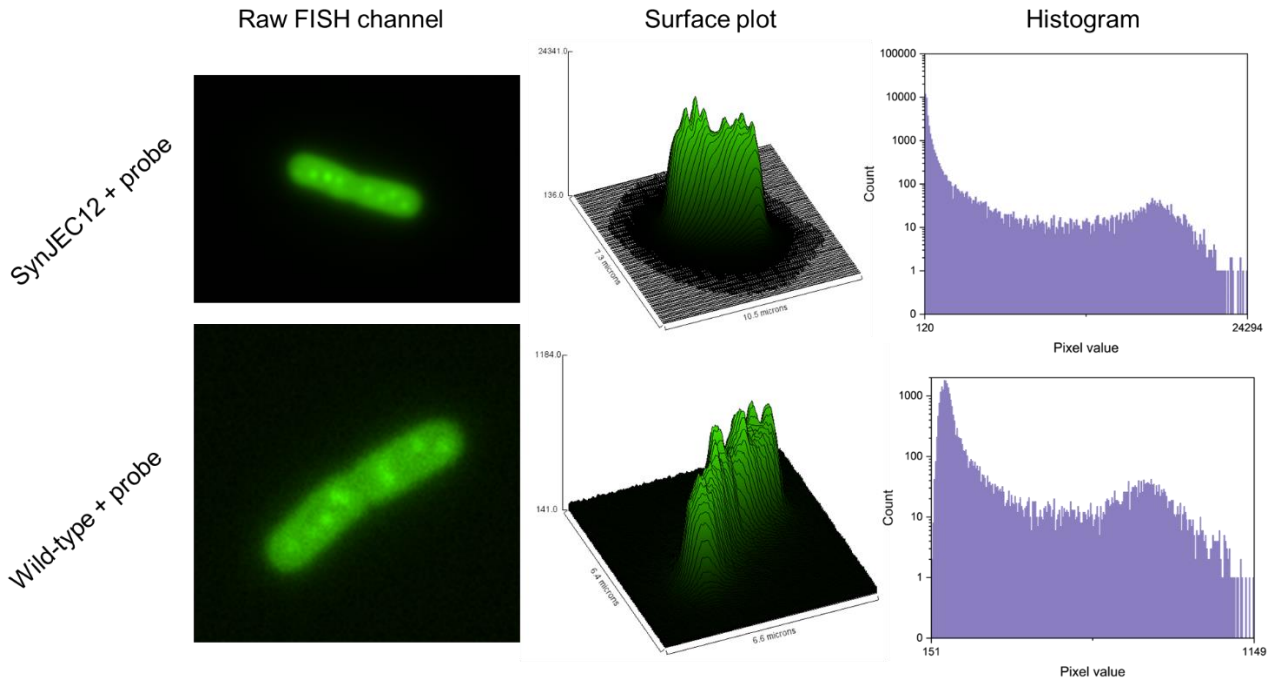



**Figure 4.8.** Comparison of off- and on-target binding in cells hybridized with kanR probes. Wild-type Syn7942 cells do not encode or express the *kanR* gene; spots represent off-target binding and false positive signal. SynJEC12 cells express the *kanR* gene; on-target binding is reflected in higher-intensity spots localized within the cell outlines.

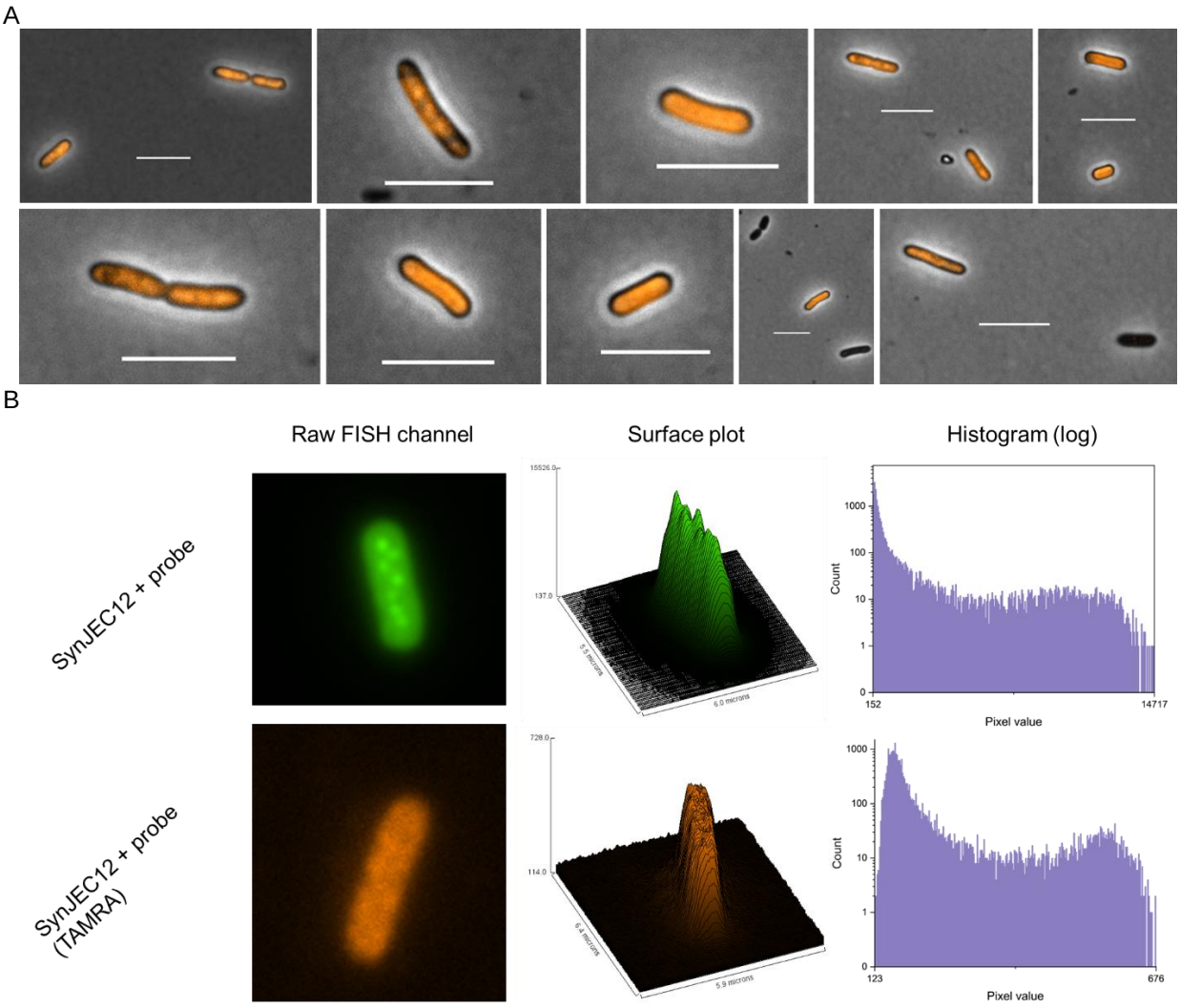



**Figure 4.9.** Comparison of kanR probes labeled with BODIPY-FL or TAMRA. (A) SynJEC12 + kanR (TAMRA)—Representative composite images (phase contrast + Cy3 cube). (B) Pixel values and surface plots of SynJEC12 cells hybridized with identical sets of kanR probes labeled with BODIPY-FL (green) or TAMRA (orange).

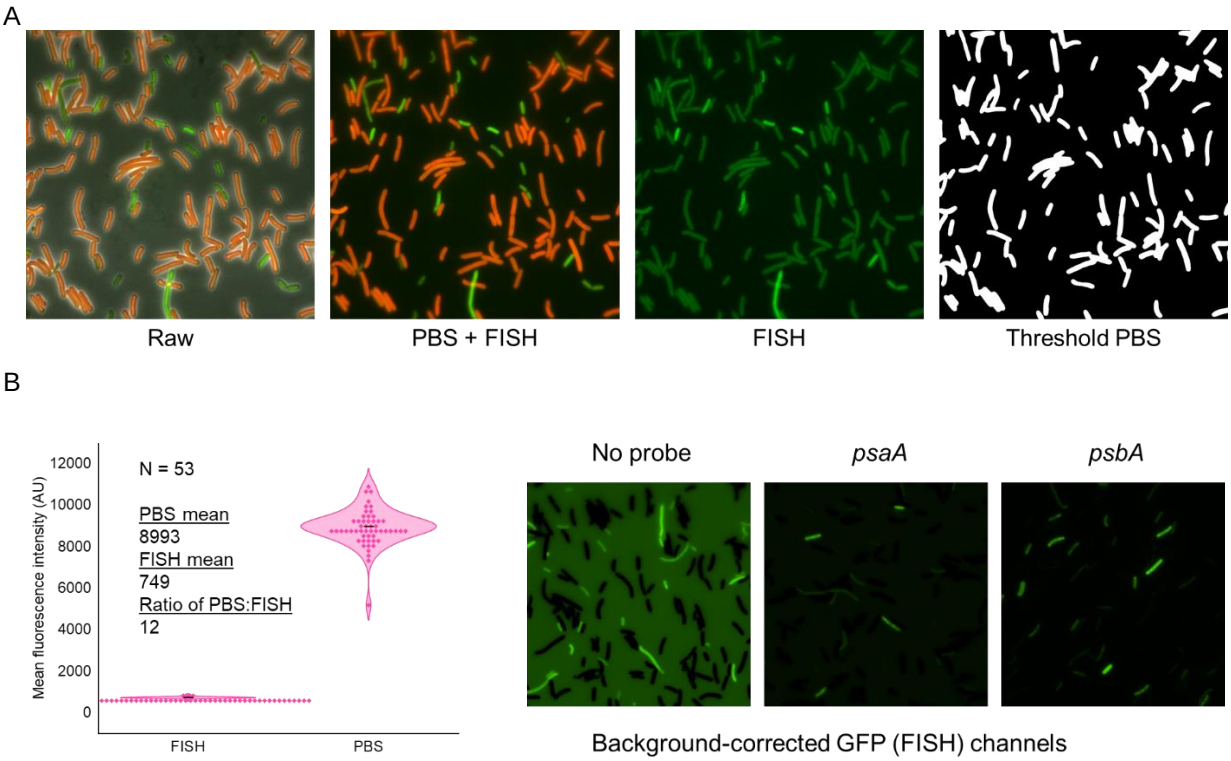



**Figure 4.10.** Data processing. (A) Comparison of raw data, PBS+FISH channels, FISH-only channel, and cell outlines detected in the PBS-only channel. (B) Noise reduction. Left, the FISH:PBS fluorescence ratio is calculated from a sample size of 53 images cells. Right, unprobed and probed cyanobacteria samples treated with noise reduction algorithm.

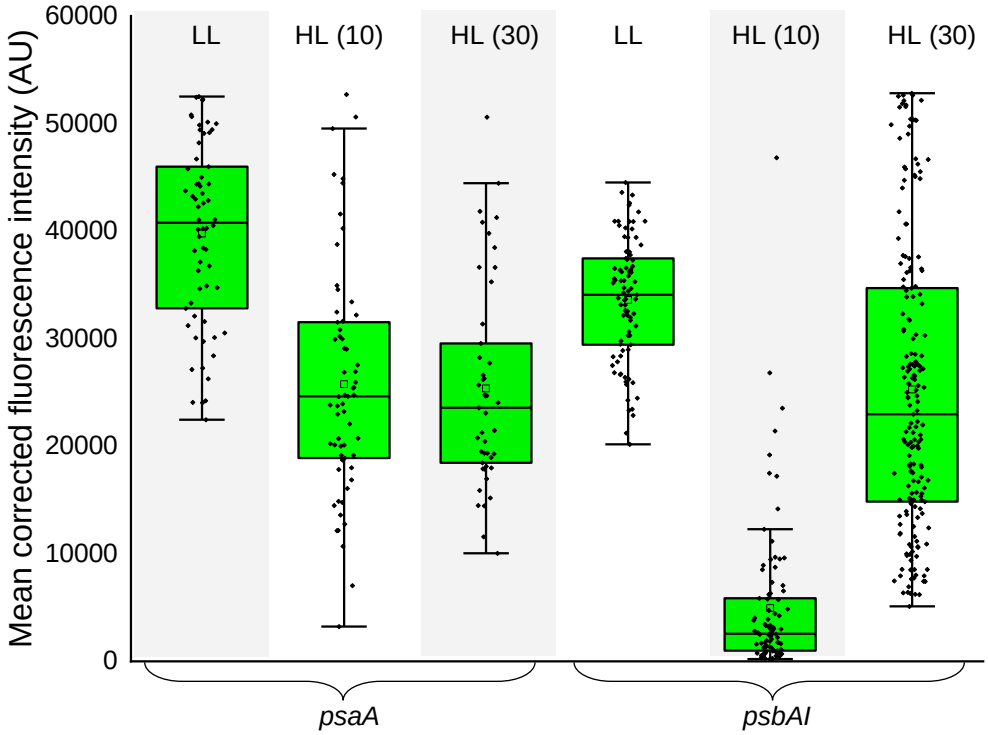



**Figure 4.11**. Comparison of whole-cell, mean fluorescence intensity of Syn7942 cells after high light treatment (after background correction shown in **Figure 4.10**).

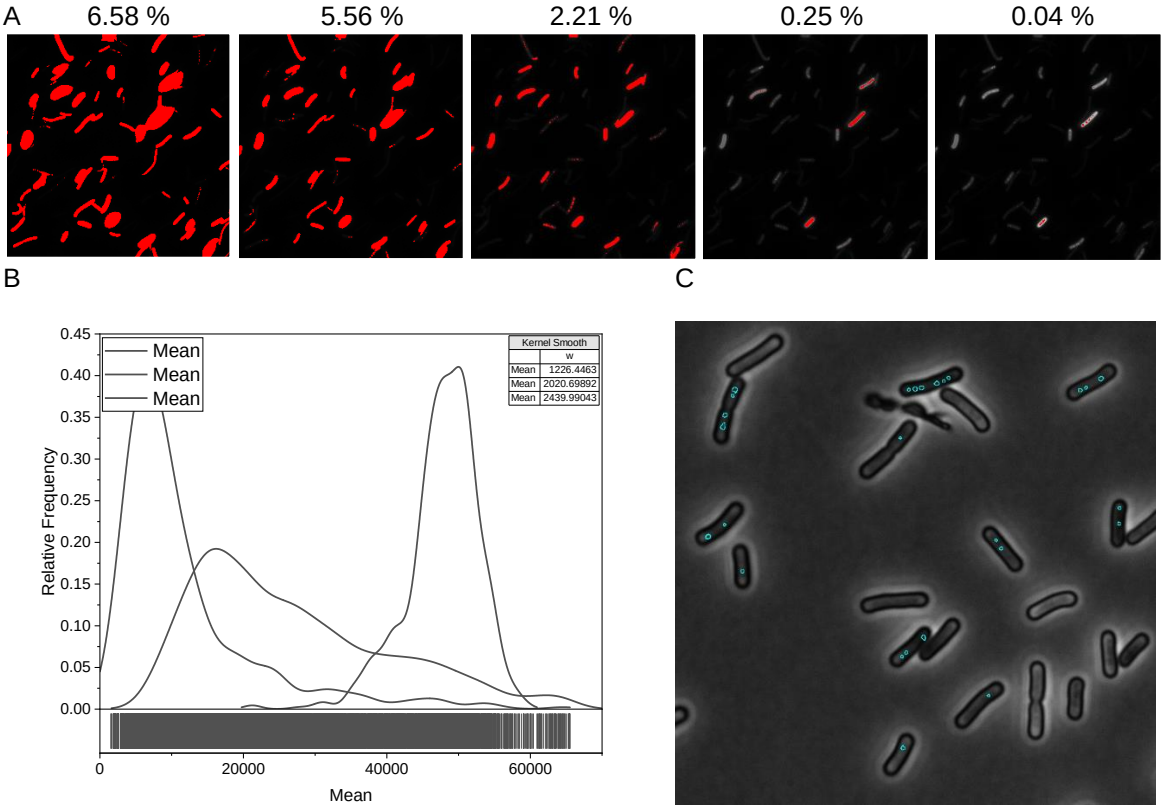



**Figure 4.12**. (A) Manual thresholding of the LUT ramp. At lower thresholds, new spots appear in the narrow area/circularity range, (B) Distribution of mean spot intensities in Syn7942 cells hybridized with *psbAI* probes, (C) representative microscope image of Syn7942 cell outlines (phase contrast) overlaid with ROIs found by thresholding.

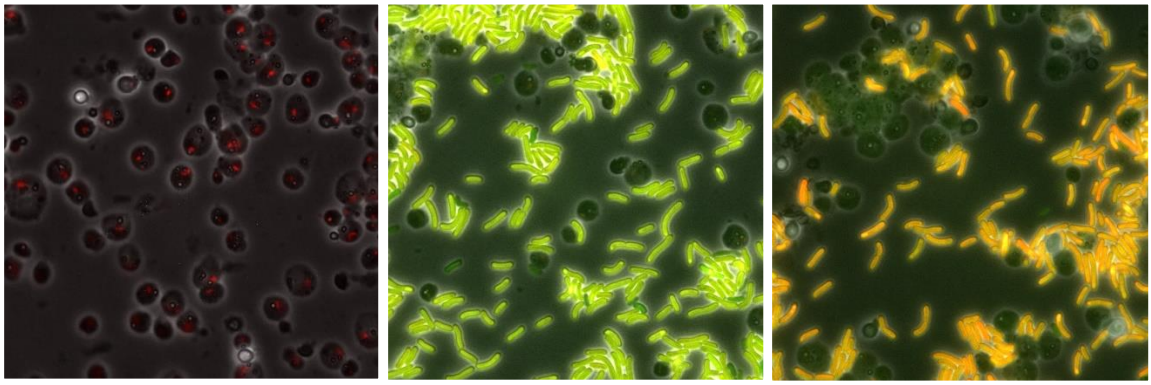


**Figure 4.13**. Yeast with ITS2. Co-culture with ITS2-TAMRA and 16s-BODIPY (merged phase, GFP, and Cy3 channels). Co-culture with ITS2-TAMRA and 16s-BODIPY (Merged phase, GFP, and Cy5 channels).

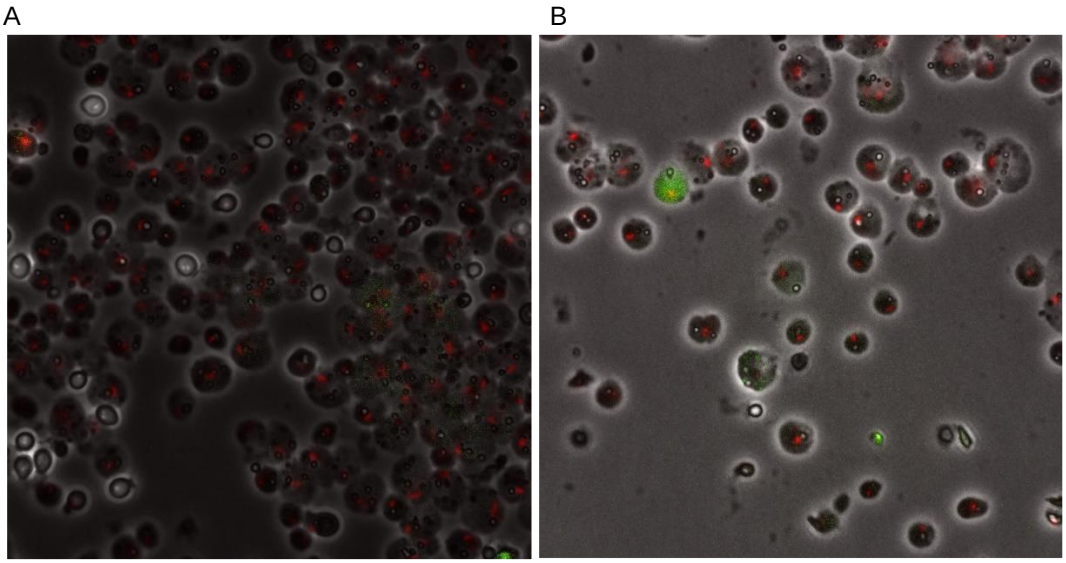



**Figure 4.14.** (A) *S. cereivisae cox2-60* cells hybridized with ITS2[TAMRA] and 16S-BODIPY; only TAMRA signal can be observed. (B) Yeast-cyanobacteria chimeras hybridized with ITS2[TAMRA] and 16S-BODIPY; nucleolar staining is visible, and disperse FISH signal is visible in some cells.

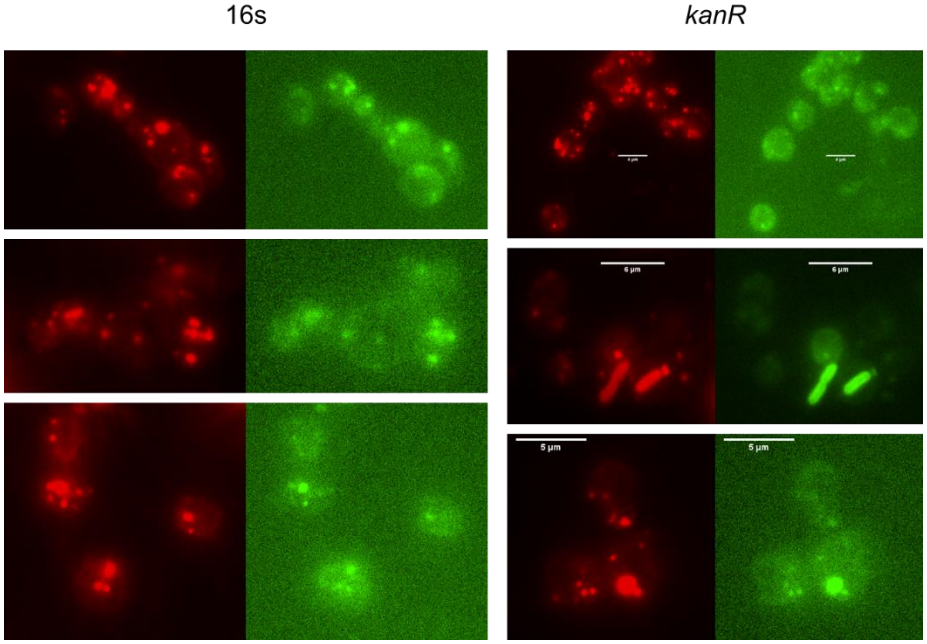



**Figure 4.15**. Segregated PBS (Cy5, red) and FISH (BODIPY-FL, green) channels in microscope images of yeast-cyanobacteria chimeras. FISH signal is not present in any loci lacking PBS signal.

**References**


(1) Margulis, L. Origin of Eukaryotic Cells; Evidence and Research Implications for a Theory of the Origin and Evolution of Microbial, Plant, and Animal Cells on the Precambrian Earth.; Yale University Press: New Haven, 1970.

(2) Zimorski, V.; Ku, C.; Martin, W. F.; Gould, S. B. Endosymbiotic Theory for Organelle Origins. Curr. Opin. Microbiol. **2014**, 22, 38–48. https://doi.org/10.1016/j.mib.2014.09.008.

(3) Sagan, L. On the Origin of Mitosing Cells. J. Theor. Biol. **1967**, 14 (3), 225-IN6. https://doi.org/10.1016/0022-5193(67)90079-3.

(4) Gray, M. W.; Doolittle, W. F. Has the Endosymbiont Hypothesis Been Proven? Microbiol. Rev. **1982**, 46 (1), 1–42.

(5) Mereschkowsky, C. Uber Natur Und Ursprung Der Chromatophoren Im Pflanzenreiche. Biologisches Centralblatt, 1905, 25, 293–604.

(6) Hohmann-Marriott, M. F.; Blankenship, R. E. Evolution of Photosynthesis. Annu. Rev. Plant Biol. **2011**, 62 (1), 515–548. https://doi.org/10.1146/annurev-arplant-042110-103811.

(7) Blankenship, R. E. Early Evolution of Photosynthesis. Plant Physiol. **2010**, 154 (2), 434–438. https://doi.org/10.1104/pp.110.161687.

(8) Martin, W. F.; Bryant, D. A.; Beatty, J. T. A Physiological Perspective on the Origin and Evolution of Photosynthesis. FEMS Microbiol. Rev. **2018**, 42 (2), 205–231. https://doi.org/10.1093/femsre/fux056.

(9) Xiong, J.; Fischer, W. M.; Inoue, K.; Nakahara, M.; Bauer, C. E. Molecular Evidence for the Early Evolution of Photosynthesis. Science **2000**, 289 (5485), 1724–1730. https://doi.org/10.1126/science.289.5485.1724.

(10) Fischer, W. W.; Hemp, J.; Johnson, J. E. Evolution of Oxygenic Photosynthesis. Annu. Rev. Earth Planet. Sci. **2016**, 44 (1), 647–683. https://doi.org/10.1146/annurev-earth-060313-054810.

(11) Flores Tinoco, V.; Herrera-Estrella, L.; Lopez-Arredondo, D. Back to Primary Endosymbiosis: From Plastids to Artificial Photosynthetic Life-Forms. Trends Plant Sci. **2023**. https://doi.org/10.1016/j.tplants.2023.03.026.

(12) McFadden, G. I. Primary and Secondary Endosymbiosis and the Origin of Plastids. J. Phycol. **2001**, 37 (6), 951–959. https://doi.org/10.1046/j.1529-8817.2001.01126.x.

(13) Raven, J. A.; Sánchez-Baracaldo, P. Gloeobacter and the Implications of a Freshwater Origin of Cyanobacteria. Phycologia **2021**, 60 (5), 402–418. https://doi.org/10.1080/00318884.2021.1881729.

(14) Shih, P. M.; Matzke, N. J. Primary Endosymbiosis Events Date to the Later Proterozoic with Cross-Calibrated Phylogenetic Dating of Duplicated ATPase Proteins. Proc. Natl. Acad. Sci. **2013**, 110 (30), 12355–12360. https://doi.org/10.1073/pnas.1305813110.

(15) Yoon, H. S.; Hackett, J. D.; Ciniglia, C.; Pinto, G.; Bhattacharya, D. A Molecular Timeline for the Origin of Photosynthetic Eukaryotes. Mol. Biol. Evol. **2004**, 21 (5), 809–818. https://doi.org/10.1093/molbev/msh075.

(16) Allen, J. F.; Raven, J. A.; Howe, C. J.; Barbrook, A. C.; Koumandou, V. L.; Nisbet, R. E. R.; Symington, H. A.; Wightman, T. F. Evolution of the Chloroplast Genome. Philos. Trans. R. Soc. Lond. B. Biol. Sci. **2003**, 358 (1429), 99–107. https://doi.org/10.1098/rstb.2002.1176.

(17) Allen, J. F. The Function of Genomes in Bioenergetic Organelles. Philos. Trans. R. Soc. B Biol. Sci. **2003**, 358 (1429), 19–38. https://doi.org/10.1098/rstb.2002.1191.

(18) Barbrook, A. C.; Howe, C. J.; Purton, S. Why Are Plastid Genomes Retained in Non-Photosynthetic Organisms? Trends Plant Sci. **2006**, 11 (2), 101–108. https://doi.org/10.1016/j.tplants.2005.12.004.

(19) Karpowicz, S. J.; Prochnik, S. E.; Grossman, A. R.; Merchant, S. S. The GreenCut2 Resource, a Phylogenomically Derived Inventory of Proteins Specific to the Plant Lineage. J. Biol. Chem. **2011**, 286 (24), 21427–21439. https://doi.org/10.1074/jbc.M111.233734.

(20) Martin, W.; Rujan, T.; Richly, E.; Hansen, A.; Cornelsen, S.; Lins, T.; Leister, D.; Stoebe, B.; Hasegawa, M.; Penny, D. Evolutionary Analysis of Arabidopsis, Cyanobacterial, and Chloroplast Genomes Reveals Plastid Phylogeny and Thousands of Cyanobacterial Genes in the Nucleus. Proc. Natl. Acad. Sci. **2002**, 99 (19), 12246–12251. https://doi.org/10.1073/pnas.182432999.

(21) Schleiff, E.; Becker, T. Common Ground for Protein Translocation: Access Control for Mitochondria and Chloroplasts. Nat. Rev. Mol. Cell Biol. **2011**, 12 (1), 48–59. https://doi.org/10.1038/nrm3027.

(22) Garrido, C.; Caspari, O. D.; Choquet, Y.; Wollman, F.-A.; Lafontaine, I. Evidence Supporting an Antimicrobial Origin of Targeting Peptides to Endosymbiotic Organelles. Cells **2020**, 9 (8), 1795. https://doi.org/10.3390/cells9081795.

(23) Caspari, O. D.; Lafontaine, I. The Role of Antimicrobial Peptides in the Evolution of Endosymbiotic Protein Import. PLOS Pathog. **2021**, 17 (4), e1009466. https://doi.org/10.1371/journal.ppat.1009466.

(24) Wollman, F.-A. An Antimicrobial Origin of Transit Peptides Accounts for Early Endosymbiotic Events. Traffic **2016**, 17 (12), 1322–1328. https://doi.org/10.1111/tra.12446.

(25) Caspari, O. D.; Garrido, C.; Law, C. O.; Choquet, Y.; Wollman, F.-A.; Lafontaine, I. Converting Antimicrobial into Targeting Peptides Reveals Key Features Governing Protein Import into Mitochondria and Chloroplasts. Plant Commun. **2023**, 4 (4), 100555. https://doi.org/10.1016/j.xplc.2023.100555.
(26) Singer, A.; Poschmann, G.; Mühlich, C.; Valadez-Cano, C.; Hänsch, S.; Hüren, V.; Rensing, S. A.; Stühler, K.; Nowack, E. C. M. Massive Protein Import into the Early-Evolutionary-Stage Photosynthetic Organelle of the Amoeba Paulinella Chromatophora. Curr. Biol. **2017**, 27 (18), 2763-2773.e5. https://doi.org/10.1016/j.cub.2017.08.010.
(27) Zhang, R.; Nowack, E. C. M.; Price, D. C.; Bhattacharya, D.; Grossman, A. R. Impact of Light Intensity and Quality on Chromatophore and Nuclear Gene Expression in Paulinella Chromatophora, an Amoeba with Nascent Photosynthetic Organelles. Plant J. **2017**, 90 (2), 221–234. https://doi.org/10.1111/tpj.13488.
(28) Karkar, S.; Facchinelli, F.; Price, D. C.; Weber, A. P. M.; Bhattacharya, D. Metabolic Connectivity as a Driver of Host and Endosymbiont Integration. Proc. Natl. Acad. Sci. **2015**, 112 (33), 10208–10215. https://doi.org/10.1073/pnas.1421375112.
(29) Tyra, H. M.; Linka, M.; Weber, A. P.; Bhattacharya, D. Host Origin of Plastid Solute Transporters in the First Photosynthetic Eukaryotes. Genome Biol. **2007**, 8 (10), R212. https://doi.org/10.1186/gb-2007-8-10-r212.
(30) Puri, K. M.; Butardo Jr, V.; Sumer, H. Evaluation of Natural Endosymbiosis for Progress towards Artificial Endosymbiosis. Symbiosis **2021**, 84 (1), 1–17.
(31) McCutcheon, J. P. The Genomics and Cell Biology of Host-Beneficial Intracellular Infections. Annu. Rev. Cell Dev. Biol. **2021**, 37, 115–142.
(32) McCutcheon, J. P.; Boyd, B. M.; Dale, C. The Life of an Insect Endosymbiont from the Cradle to the Grave. Curr. Biol. **2019**, 29 (11), R485–R495. https://doi.org/10.1016/j.cub.2019.03.032.
(33) Moran, N. A.; McCutcheon, J. P.; Nakabachi, A. Genomics and Evolution of Heritable Bacterial Symbionts. Annu. Rev. Genet. **2008**, 42 (1), 165–190. https://doi.org/10.1146/annurev.genet.41.110306.130119.
(34) Buchsbaum, R. Chick Tissue Cells and Chlorella in Mixed Cultures. Physiol. Zool. **1937**, 10 (4), 373–380. https://doi.org/10.1086/physzool.10.4.30151423.
(35) Anderson, J. C.; Clarke, E. J.; Arkin, A. P.; Voigt, C. A. Environmentally Controlled Invasion of Cancer Cells by Engineered Bacteria. J. Mol. Biol. **2006**, 355 (4), 619–627. https://doi.org/10.1016/j.jmb.2005.10.076.
(36) Narayanan, K.; Warburton, P. E. DNA Modification and Functional Delivery into Human Cells Using Escherichia Coli DH10B. Nucleic Acids Res. **2003**, 31 (9), e51. https://doi.org/10.1093/nar/gng051.
(37) Grillot-Courvalin, C.; Goussard, S.; Huetz, F.; Ojcius, D. M.; Courvalin, P. Functional Gene Transfer from Intracellular Bacteria to Mammalian Cells. Nat. Biotechnol. **1998**, 16 (9), 862–866. https://doi.org/10.1038/nbt0998-862.
(38) Critchley, R. J.; Jezzard, S.; Radford, K. J.; Goussard, S.; Lemoine, N. R.; Grillot-Courvalin, C.; Vassaux, G. Potential Therapeutic Applications of Recombinant, Invasive E. Coli. Gene Ther. **2004**, 11 (15), 1224–1233. https://doi.org/10.1038/sj.gt.3302281.
(39) Zhao, M.; Yang, M.; Li, X.-M.; Jiang, P.; Baranov, E.; Li, S.; Xu, M.; Penman, S.; Hoffman, R. M. Tumor-Targeting Bacterial Therapy with Amino Acid Auxotrophs of GFP-Expressing Salmonella Typhimurium. Proc. Natl. Acad. Sci. **2005**, 102 (3), 755–760. https://doi.org/10.1073/pnas.0408422102.
(40) Agapakis, C. M.; Niederholtmeyer, H.; Noche, R. R.; Lieberman, T. D.; Megason, S. G.; Way, J. C.; Silver, P. A. Towards a Synthetic Chloroplast. PLoS ONE **2011**, 6 (4), e18877. https://doi.org/10.1371/journal.pone.0018877.
(41) Alvarez, M.; Reynaert, N.; Chávez, M. N.; Aedo, G.; Araya, F.; Hopfner, U.; Fernández, J.; Allende, M. L.; Egaña, J. T. Generation of Viable Plant-Vertebrate Chimeras. PLOS ONE **2015**, 10 (6), e0130295. https://doi.org/10.1371/journal.pone.0130295.
(42) Matsunaga, T.; Hashimoto, K.; Nakamura, N.; Nakamura, K.; Hashimoto, S. Phagocytosis of Bacterial Magnetite by Leucocytes. Appl. Microbiol. Biotechnol. **1989**, 31 (4), 401–405. https://doi.org/10.1007/BF00257612.
(43) Bielecki, J.; Youngman, P.; Connelly, P.; Portnoy, D. A. Bacillus Subtilis Expressing a Haemolysin Gene from Listeria Monocytogenes Can Grow in Mammalian Cells. Nature **1990**, 345 (6271), 175–176. https://doi.org/10.1038/345175a0.
(44) Elani, Y.; Trantidou, T.; Wylie, D.; Dekker, L.; Polizzi, K.; Law, R. V.; Ces, O. Constructing Vesicle-Based Artificial Cells with Embedded Living Cells as Organelle-like Modules. Sci. Rep. **2018**, 8 (1), 4564. https://doi.org/10.1038/s41598-018-22263-3.
(45) Karas, B. J.; Moreau, N. G.; Deerinck, T. J.; Gibson, D. G.; Venter, J. C.; Smith, H. O.; Glass, J. I. Direct Transfer of a Mycoplasma Mycoides Genome to Yeast Is Enhanced by Removal of the Mycoides Glycerol Uptake Factor Gene glpF. ACS Synth. Biol. **2019**, 8 (2), 239–244. https://doi.org/10.1021/acssynbio.8b00449.

(46) Nass, M. M. k. Uptake of Isolated Chloroplasts by Mammalian Cells. Science **1969**, 165 (3898), 1128–1131. https://doi.org/10.1126/science.165.3898.1128.
(47) Du, Z.-Y.; Zienkiewicz, K.; Vande Pol, N.; Ostrom, N. E.; Benning, C.; Bonito, G. M. Algal-Fungal Symbiosis Leads to Photosynthetic Mycelium. eLife **2019**, 8, e47815. https://doi.org/10.7554/eLife.47815.
(48) Egede, E. J.; Jones, H.; Cook, B.; Purchase, D.; Mouradov, A. Application of Microalgae and Fungal-Microalgal Associations for Wastewater Treatment. In Fungal Applications in Sustainable Environmental Biotechnology; Purchase, D., Ed.; Springer International Publishing: Cham, 2016; pp 143–181. https://doi.org/10.1007/978-3-319-42852-9_7.
(49) Li, T.; Li, C.-T.; Butler, K.; Hays, S. G.; Guarnieri, M. T.; Oyler, G. A.; Betenbaugh, M. J. Mimicking Lichens: Incorporation of Yeast Strains Together with Sucrose-Secreting Cyanobacteria Improves Survival, Growth, ROS Removal, and Lipid Production in a Stable Mutualistic Co-Culture Production Platform. Biotechnol. Biofuels **2017**, 10 (1), 55. https://doi.org/10.1186/s13068-017-0736-x.
(50) Meaney, R. S.; Hamadache, S.; Soltysiak, M. P. M.; Karas, B. J. Designer Endosymbionts: Converting Free-Living Bacteria into Organelles. Curr. Opin. Syst. Biol. **2020**, 24, 41–50. https://doi.org/10.1016/j.coisb.2020.09.008.
(51) Shapiro, K. O. Plug and Play: Is "Directed Endosymbiosis" of Chloroplasts Possible? bioRxiv December 3, 2021, p 2021.12.03.471169. https://doi.org/10.1101/2021.12.03.471169.
(52) Mehta, A. P.; Supekova, L.; Chen, J.-H.; Pestonjamasp, K.; Webster, P.; Ko, Y.; Henderson, S. C.; McDermott, G.; Supek, F.; Schultz, P. G. Engineering Yeast Endosymbionts as a Step toward the Evolution of Mitochondria. Proc. Natl. Acad. Sci. **2018**, 115 (46), 11796–11801.
(53) Mehta, A. P.; Ko, Y.; Supekova, L.; Pestonjamasp, K.; Li, J.; Schultz, P. G. Toward a Synthetic Yeast Endosymbiont with a Minimal Genome. J. Am. Chem. Soc. **2019**, 141 (35), 13799–13802. https://doi.org/10.1021/jacs.9b08290.
(54) Göpfrich, K.; Platzman, I.; Spatz, J. P. Mastering Complexity: Towards Bottom-up Construction of Multifunctional Eukaryotic Synthetic Cells. Trends Biotechnol. **2018**, 36 (9), 938–951. https://doi.org/10.1016/j.tibtech.2018.03.008.
(55) Sulo, P.; Griač, P.; Klobučníková, V.; Kováč, L. A Method for the Efficient Transfer of Isolated Mitochondria into Yeast Protoplasts. Curr. Genet. **1989**, 15 (1), 1–6. https://doi.org/10.1007/BF00445745.
(56) Karas, B. J.; Jablanovic, J.; Sun, L.; Ma, L.; Goldgof, G. M.; Stam, J.; Ramon, A.; Manary, M. J.; Winzeler, E. A.; Venter, J. C.; Weyman, P. D.; Gibson, D. G.; Glass, J. I.; Hutchison, C. A.; Smith, H. O.; Suzuki, Y. Direct Transfer of Whole Genomes from Bacteria to Yeast. Nat. Methods **2013**, 10 (5), 410–412. https://doi.org/10.1038/nmeth.2433.
(57) M, H. Subsection I. (Formerly Chroococcales Wettstein 1924, Emends. Rippka, Deruelles, Waterbury, Herdman and Stanier 1979). Bergeys Man. Syst. Bacteriol. **2001**, 1, 493–514.
(58) Chen, Y.; Kay Holtman, C.; Magnuson, R. D.; Youderian, P. A.; Golden, S. S. The Complete Sequence and Functional Analysis of pANL, the Large Plasmid of the Unicellular Freshwater Cyanobacterium Synechococcus Elongatus PCC 7942. Plasmid **2008**, 59 (3), 176–192. https://doi.org/10.1016/j.plasmid.2008.01.005.
(59) van den Hondel, C. A.; Verbeek, S.; van der Ende, A.; Weisbeek, P. J.; Borrias, W. E.; van Arkel, G. A. Introduction of Transposon Tn901 into a Plasmid of Anacystis Nidulans: Preparation for Cloning in Cyanobacteria. Proc. Natl. Acad. Sci. **1980**, 77 (3), 1570–1574. https://doi.org/10.1073/pnas.77.3.1570.
(60) Shestakov, S. V.; Khyen, N. T. Evidence for Genetic Transformation in Blue-Green Alga Anacystis Nidulans. Mol. Gen. Genet. MGG **1970**, 107 (4), 372–375. https://doi.org/10.1007/BF00441199.
(61) Moore, L. R.; Coe, A.; Zinser, E. R.; Saito, M. A.; Sullivan, M. B.; Lindell, D.; Frois-Moniz, K.; Waterbury, J.; Chisholm, S. W. Culturing the Marine Cyanobacterium Prochlorococcus. Limnol. Oceanogr. Methods **2007**, 5 (10), 353–362. https://doi.org/10.4319/lom.2007.5.353.
(62) Bernstein, H. C.; Konopka, A.; Melnicki, M. R.; Hill, E. A.; Kucek, L. A.; Zhang, S.; Shen, G.; Bryant, D. A.; Beliaev, A. S. Effect of Mono- and Dichromatic Light Quality on Growth Rates and Photosynthetic Performance of Synechococcus Sp. PCC 7002. Front. Microbiol. **2014**, 5, 488. https://doi.org/10.3389/fmicb.2014.00488.
(63) Ditty, J. L.; Williams, S. B.; Golden, S. S. A Cyanobacterial Circadian Timing Mechanism. Annu. Rev. Genet. **2003**, 37 (1), 513–543. https://doi.org/10.1146/annurev.genet.37.110801.142716.
(64) Kaňa, R.; Prášil, O.; Komárek, O.; Papageorgiou, G. C.; Govindjee. Spectral Characteristic of Fluorescence Induction in a Model Cyanobacterium, Synechococcus Sp. (PCC 7942). Biochim. Biophys. Acta BBA - Bioenerg. **2009**, 1787 (10), 1170–1178. https://doi.org/10.1016/j.bbabio.2009.04.013.

(65) Yu, J.; Liberton, M.; Cliften, P. F.; Head, R. D.; Jacobs, J. M.; Smith, R. D.; Koppenaal, D. W.; Brand, J. J.; Pakrasi, H. B. Synechococcus Elongatus UTEX 2973, a Fast Growing Cyanobacterial Chassis for Biosynthesis Using Light and CO2. Sci. Rep. **2015**, 5 (1), 8132. https://doi.org/10.1038/srep08132.
(66) Mishra, S.; Kumari, N.; Singh, V.; Sinha, R. Cyanobacterial Biofuel: A Platform for Green Energy. Adv. Environ. Eng. Res. **2023**, 5 (3), 1–42. https://doi.org/10.21926/aeer.2303041.
(67) Hays, S. G.; Yan, L. L. W.; Silver, P. A.; Ducat, D. C. Synthetic Photosynthetic Consortia Define Interactions Leading to Robustness and Photoproduction. J. Biol. Eng. **2017**, 11 (1), 4. https://doi.org/10.1186/s13036-017-0048-5.
(68) Wang, J.; Song, X.; Chen, L.; Zhang, W. Application and Mechanism Analysis of Photosynthetic Microbial Coculture Systems for Bioproduction. Algal Biotechnol. **2023**, 32–53. https://doi.org/10.1079/9781800621954.0002.
(69) Whitmarsh, J.; Govindjee. The Photosynthetic Process. In Concepts in Photobiology: Photosynthesis and Photomorphogenesis; Singhal, G. S., Renger, G., Sopory, S. K., Irrgang, K.-D., Govindjee, Eds.; Springer Netherlands: Dordrecht, 1999; pp 11–51. https://doi.org/10.1007/978-94-011-4832-0_2.
(70) Govindjee. Bioenergetics of Photosynthesis,; Academic Press: New York, 1975.
(71) Sidler, W. A. Phycobilisome and Phycobiliprotein Structures. In The Molecular Biology of Cyanobacteria; Bryant, D. A., Ed.; Advances in Photosynthesis; Springer Netherlands: Dordrecht, 1994; pp 139–216. https://doi.org/10.1007/978-94-011-0227-8_7.
(72) Six, C.; Thomas, J.-C.; Garczarek, L.; Ostrowski, M.; Dufresne, A.; Blot, N.; Scanlan, D. J.; Partensky, F. Diversity and Evolution of Phycobilisomes in Marine Synechococcus Spp.: A Comparative Genomics Study. Genome Biol. **2007**, 8 (12), R259. https://doi.org/10.1186/gb-2007-8-12-r259.
(73) Grébert, T.; Garczarek, L.; Daubin, V.; Humily, F.; Marie, D.; Ratin, M.; Devailly, A.; Farrant, G. K.; Mary, I.; Mella-Flores, D.; Tanguy, G.; Labadie, K.; Wincker, P.; Kehoe, D. M.; Partensky, F. Diversity and Evolution of Pigment Types in Marine Synechococcus Cyanobacteria. Genome Biol. Evol. **2022**, evac035. https://doi.org/10.1093/gbe/evac035.
(74) Cohen, S. E.; Erb, M. L.; Pogliano, J.; Golden, S. S. Best Practices for Fluorescence Microscopy of the Cyanobacterial Circadian Clock. Methods Enzymol. **2015**, 551, 211–221. https://doi.org/10.1016/bs.mie.2014.10.014.
(75) Cournoyer, J. E.; Altman, S. D.; Gao, Y.; Wallace, C. L.; Zhang, D.; Lo, G.-H.; Haskin, N. T.; Mehta, A. P. Engineering Artificial Photosynthetic Life-Forms through Endosymbiosis. Nat. Commun. **2022**, 13 (1), 2254. https://doi.org/10.1038/s41467-022-29961-7.
(76) Mills, L. A.; Moreno-Cabezuelo, J. Á.; Włodarczyk, A.; Victoria, A. J.; Mejías, R.; Nenninger, A.; Moxon, S.; Bombelli, P.; Selão, T. T.; McCormick, A. J.; Lea-Smith, D. J. Development of a Biotechnology Platform for the Fast-Growing Cyanobacterium Synechococcus Sp. PCC 11901. Biomolecules **2022**, 12 (7), 872. https://doi.org/10.3390/biom12070872.
(77) Herdman, M.; Delaney, S. F.; Carr, N. G. Y. 1973. A New Medium for the Isolation and Growth of Auxotrophic Mutants of the Blue-Green Alga Anacystis Nidulans. Microbiology 79 (2), 233–237. https://doi.org/10.1099/00221287-79-2-233.
(78) Herdman, M.; Carr, N. G. Y. 1972. The Isolation and Characterization of Mutant Strains of the Blue-Green Alga Anacystis Nidulans. Microbiology 70 (2), 213–220. https://doi.org/10.1099/00221287-70-2-213.
(79) Herdman, M.; Delaney, S. F.; Carr, N. G. Mutation of the Cyanobacterium Anacystis Nidulans (Synechococcus PCC 6301): Improved Conditions for the Isolation of Auxotrophs. Arch. Microbiol. **1980**, 124 (2), 177–184. https://doi.org/10.1007/BF00427724.
(80) Baba, T.; Ara, T.; Hasegawa, M.; Takai, Y.; Okumura, Y.; Baba, M.; Datsenko, K. A.; Tomita, M.; Wanner, B. L.; Mori, H. Construction of Escherichia Coli K-12 in-Frame, Single-Gene Knockout Mutants: The Keio Collection. Mol. Syst. Biol. **2006**, 2, 2006.0008. https://doi.org/10.1038/msb4100050.
(81) Pronk, J. T. Auxotrophic Yeast Strains in Fundamental and Applied Research. Appl. Environ. Microbiol. **2002**, 68 (5), 2095–2100. https://doi.org/10.1128/AEM.68.5.2095-2100.2002.
(82) Montesinos, M. L.; Herrero, A.; Flores, E. Amino Acid Transport in Taxonomically Diverse Cyanobacteria and Identification of Two Genes Encoding Elements of a Neutral Amino Acid Permease Putatively Involved in Recapture of Leaked Hydrophobic Amino Acids. J. Bacteriol. **1997**, 179 (3), 853–862. https://doi.org/10.1128/jb.179.3.853-862.1997.
(83) Quintero, M. J.; Montesinos, M. L.; Herrero, A.; Flores, E. Identification of Genes Encoding Amino Acid Permeases by Inactivation of Selected ORFs from the Synechocystis Genomic Sequence. Genome Res. **2001**, 11 (12), 2034–2040. https://doi.org/10.1101/gr.196301.

(84) Elbourne, L. D. H.; Tetu, S. G.; Hassan, K. A.; Paulsen, I. T. TransportDB 2.0: A Database for Exploring Membrane Transporters in Sequenced Genomes from All Domains of Life. Nucleic Acids Res. **2017**, 45 (D1), D320–D324. https://doi.org/10.1093/nar/gkw1068.
(85) Elbourne, L. D. H.; Wilson-Mortier, B.; Ren, Q.; Hassan, K. A.; Tetu, S. G.; Paulsen, I. T. TransAAP: An Automated Annotation Pipeline for Membrane Transporter Prediction in Bacterial Genomes. Microb. Genomics **2023**, 9 (1). https://doi.org/10.1099/mgen.0.000927.
(86) Niederholtmeyer, H.; Wolfstädter, B. T.; Savage, D. F.; Silver, P. A.; Way, J. C. Engineering Cyanobacteria To Synthesize and Export Hydrophilic Products. Appl. Environ. Microbiol. **2010**, 76 (11), 3462–3466. https://doi.org/10.1128/AEM.00202-10.
(87) Mills, L. A.; McCormick, A. J.; Lea-Smith, D. J. Current Knowledge and Recent Advances in Understanding Metabolism of the Model Cyanobacterium Synechocystis Sp. PCC 6803. Biosci. Rep. **2020**, 40 (4), BSR20193325. https://doi.org/10.1042/BSR20193325.
(88) Bishé, B.; Taton, A.; Golden, J. W. Modification of RSF1010-Based Broad-Host-Range Plasmids for Improved Conjugation and Cyanobacterial Bioprospecting. iScience **2019**, 20, 216–228. https://doi.org/10.1016/j.isci.2019.09.002.
(89) Taton, A.; Unglaub, F.; Wright, N. E.; Zeng, W. Y.; Paz-Yepes, J.; Brahamsha, B.; Palenik, B.; Peterson, T. C.; Haerizadeh, F.; Golden, S. S.; Golden, J. W. Broad-Host-Range Vector System for Synthetic Biology and Biotechnology in Cyanobacteria. Nucleic Acids Res. **2014**, 42 (17), e136. https://doi.org/10.1093/nar/gku673.
(90) Ungerer, J.; Pakrasi, H. B. Cpf1 Is A Versatile Tool for CRISPR Genome Editing Across Diverse Species of Cyanobacteria. Sci. Rep. **2016**, 6 (1), 39681. https://doi.org/10.1038/srep39681.
(91) Wendt, K. E.; Ungerer, J.; Cobb, R. E.; Zhao, H.; Pakrasi, H. B. CRISPR/Cas9 Mediated Targeted Mutagenesis of the Fast Growing Cyanobacterium Synechococcus Elongatus UTEX 2973. Microb. Cell Factories **2016**, 15 (1), 115. https://doi.org/10.1186/s12934-016-0514-7.
(92) Taton, A.; Gilderman, T. S.; Ernst, D. C.; Omaga, C. A.; Cohen, L. A.; Rey-Bedon, C.; Golden, J. W.; Golden, S. S. Synechococcus Elongatus Argonaute Reduces Natural Transformation Efficiency and Provides Immunity against Exogenous Plasmids. mBio **2023**, 0 (0), e01843-23. https://doi.org/10.1128/mbio.01843-23.
(93) Hou, F.; Ke, Z.; Xu, Y.; Wang, Y.; Zhu, G.; Gao, H.; Ji, S.; Xu, X. Systematic Large Fragment Deletions in the Genome of Synechococcus Elongatus and the Consequent Changes in Transcriptomic Profiles. Genes **2023**, 14 (5), 1091. https://doi.org/10.3390/genes14051091.
(94) Rubin, B. E.; Wetmore, K. M.; Price, M. N.; Diamond, S.; Shultzaberger, R. K.; Lowe, L. C.; Curtin, G.; Arkin, A. P.; Deutschbauer, A.; Golden, S. S. The Essential Gene Set of a Photosynthetic Organism. Proc. Natl. Acad. Sci. **2015**, 112 (48), E6634–E6643. https://doi.org/10.1073/pnas.1519220112.
(95) Delaye, L.; González-Domenech, C. M.; Garcillán-Barcia, M. P.; Peretó, J.; de la Cruz, F.; Moya, A. Blueprint for a Minimal Photoautotrophic Cell: Conserved and Variable Genes in Synechococcus Elongatus PCC 7942. BMC Genomics **2011**, 12 (1), 25. https://doi.org/10.1186/1471-2164-12-25.
(96) Eleveld, T. F.; Bakali, C.; Eijk, P. P.; Stathi, P.; Vriend, L. E.; Poddighe, P. J.; Ylstra, B. Engineering Large-Scale Chromosomal Deletions by CRISPR-Cas9. Nucleic Acids Res. **2021**, 49 (21), 12007–12016. https://doi.org/10.1093/nar/gkab557.
(97) Adikusuma, F.; Piltz, S.; Corbett, M. A.; Turvey, M.; McColl, S. R.; Helbig, K. J.; Beard, M. R.; Hughes, J.; Pomerantz, R. T.; Thomas, P. Q. Large Deletions Induced by Cas9 Cleavage. Nature **2018**, 560 (7717), E8–E9. https://doi.org/10.1038/s41586-018-0380-z.
(98) Standage-Beier, K.; Zhang, Q.; Wang, X. Targeted Large-Scale Deletion of Bacterial Genomes Using CRISPR-Nickases. ACS Synth. Biol. **2015**, 4 (11), 1217–1225. https://doi.org/10.1021/acssynbio.5b00132.
(99) Yang, Y.; Wang, Y.; Li, Z.; Gong, Y.; Zhang, P.; Hu, W.; Sheng, D.; Li, Y. Increasing On-Target Cleavage Efficiency for CRISPR/Cas9-Induced Large Fragment Deletion in Myxococcus Xanthus. Microb. Cell Factories **2017**, 16 (1), 142. https://doi.org/10.1186/s12934-017-0758-x.
(100) So, Y.; Park, S.-Y.; Park, E.-H.; Park, S.-H.; Kim, E.-J.; Pan, J.-G.; Choi, S.-K. A Highly Efficient CRISPR-Cas9-Mediated Large Genomic Deletion in Bacillus Subtilis. Front. Microbiol. **2017**, 8.
(101) Clerico, E. M.; Ditty, J. L.; Golden, S. S. Specialized Techniques for Site-Directed Mutagenesis in Cyanobacteria. In Circadian Rhythms: Methods and Protocols; Rosato, E., Ed.; Methods in Molecular Biology™; Humana Press: Totowa, NJ, 2007; pp 155–171. https://doi.org/10.1007/978-1-59745-257-1_11.
(102) Hitchcock, A.; Hunter, C. N.; Canniffe, D. P. Progress and Challenges in Engineering Cyanobacteria as Chassis for Light-Driven Biotechnology. Microb. Biotechnol. **2020**, 13 (2), 363–367. https://doi.org/10.1111/1751-7915.13526.

(103) Huang, C.-H.; Shen, C. R.; Li, H.; Sung, L.-Y.; Wu, M.-Y.; Hu, Y.-C. CRISPR Interference (CRISPRi) for Gene Regulation and Succinate Production in Cyanobacterium S. Elongatus PCC 7942. Microb. Cell Factories **2016**, 15 (1), 196. https://doi.org/10.1186/s12934-016-0595-3.
(104) Ramey, C. J.; Barón-Sola, Á.; Aucoin, H. R.; Boyle, N. R. Genome Engineering in Cyanobacteria: Where We Are and Where We Need To Go. ACS Synth. Biol. **2015**, 4 (11), 1186–1196. https://doi.org/10.1021/acssynbio.5b00043.
(105) Riaz, S.; Jiang, Y.; Xiao, M.; You, D.; Klepacz-Smółka, A.; Rasul, F.; Daroch, M. Generation of Miniploid Cells and Improved Natural Transformation Procedure for a Model Cyanobacterium Synechococcus Elongatus PCC 7942. Front. Microbiol. **2022**, 13. https://doi.org/10.3389/fmicb.2022.959043.
(106) Pope, M. A.; Hodge, J. A.; Nixon, P. J. An Improved Natural Transformation Protocol for the Cyanobacterium Synechocystis Sp. PCC 6803. Front. Plant Sci. **2020**, 11, 372. https://doi.org/10.3389/fpls.2020.00372.
(107) Margulis, L. Origin of Eukaryotic Cells: Evidence and Research Implications for a Theory of the Origin and Evolution of Microbial, Plant and Animal Cells on the Precambrian Earth; Yale University Press, 1970.
(108) Zimorski, V.; Ku, C.; Martin, W. F.; Gould, S. B. Endosymbiotic Theory for Organelle Origins. Curr. Opin. Microbiol., 2014, 22, 38–48. https://doi.org/10.1016/j.mib.2014.09.008.
(109) Jensen, P. E.; Leister, D. Chloroplast Evolution, Structure and Functions. F1000prime reports, 2014, 6.
(110) Martin, W.; Kowallik, K. V. Annotated English Translation of Mereschkowsky's 1905 Paper 'Über Natur Und Ursprung Der Chromatophoren Im Pflanzenreiche.' Eur. J. Phycol. **1999**, 34 (3), 287–295.
(111) Bonen, L.; Cunningham, R.; Gray, M.; Doolittle, W. Wheat Embryo Mitochondrial 18S Ribosomal RNA: Evidence for Its Prokaryotic Nature. Nucleic Acids Res. **1977**, 4 (3), 663–671.
(112) McCutcheon, J. P.; Boyd, B. M.; Dale, C. The Life of an Insect Endosymbiont from the Cradle to the Grave. Curr. Biol. **2019**, 29 (11), R485–R495. https://doi.org/10.1016/j.cub.2019.03.032.
(113) Moran, N. A.; McCutcheon, J. P.; Nakabachi, A. Genomics and Evolution of Heritable Bacterial Symbionts. Annu. Rev. Genet. **2008**, 42 (1), 165–190. https://doi.org/10.1146/annurev.genet.41.110306.130119.
(114) Buchsbaum, R. Chick Tissue Cells and Chlorella in Mixed Cultures. Physiol. Zool. **1937**, 10 (4), 373–380. https://doi.org/10.1086/physzool.10.4.30151423.
(115) Buchsbaum, R.; Buchsbaum, M. An Artificial Symbiosis. Science **1934**, 80 (2079), 408–409. https://doi.org/10.1126/science.80.2079.408.
(116) Agapakis, C. M.; Niederholtmeyer, H.; Noche, R. R.; Lieberman, T. D.; Megason, S. G.; Way, J. C.; Silver, P. A. Towards a Synthetic Chloroplast. PLoS One **2011**, 6 (4), e18877.
(117) Nass, M. M. k. Uptake of Isolated Chloroplasts by Mammalian Cells. Science **1969**, 165 (3898), 1128–1131. https://doi.org/10.1126/science.165.3898.1128.
(118) Mehta, A. P.; Supekova, L.; Chen, J.-H.; Pestonjamasp, K.; Webster, P.; Ko, Y.; Henderson, S. C.; McDermott, G.; Supek, F.; Schultz, P. G. Engineering Yeast Endosymbionts as a Step toward the Evolution of Mitochondria. Proc. Natl. Acad. Sci. **2018**, 115 (46), 11796–11801.
(119) Mehta, A. P.; Ko, Y.; Supekova, L.; Pestonjamasp, K.; Li, J.; Schultz, P. G. Toward a Synthetic Yeast Endosymbiont with a Minimal Genome. J. Am. Chem. Soc. **2019**, 141 (35), 13799–13802. https://doi.org/10.1021/jacs.9b08290.
(120) Ponce-Toledo, R. I.; Deschamps, P.; López-García, P.; Zivanovic, Y.; Benzerara, K.; Moreira, D. An Early-Branching Freshwater Cyanobacterium at the Origin of Plastids. Curr. Biol. **2017**, 27 (3), 386–391. https://doi.org/10.1016/j.cub.2016.11.056.
(121) Sánchez-Baracaldo, P.; Raven, J. A.; Pisani, D.; Knoll, A. H. Early Photosynthetic Eukaryotes Inhabited Low-Salinity Habitats. Proc. Natl. Acad. Sci. **2017**, 114 (37), E7737–E7745. https://doi.org/10.1073/pnas.1620089114.
(122) Golden, S. S.; Brusslan, J.; Haselkorn, R. [12] Genetic Engineering of the Cyanobacterial Chromosome. Methods in enzymology, 1987, 153, 215–231.
(123) Taton, A.; Unglaub, F.; Wright, N. E.; Zeng, W. Y.; Paz-Yepes, J.; Brahamsha, B.; Palenik, B.; Peterson, T. C.; Haerizadeh, F.; Golden, S. S. Broad-Host-Range Vector System for Synthetic Biology and Biotechnology in Cyanobacteria. Nucleic Acids Res. **2014**, 42 (17), e136–e136.
(124) Zaremba-Niedzwiedzka, K.; Caceres, E. F.; Saw, J. H.; Bäckström, D.; Juzokaite, L.; Vancaester, E.; Seitz, K. W.; Anantharaman, K.; Starnawski, P.; Kjeldsen, K. U.; Stott, M. B.; Nunoura, T.; Banfield, J. F.; Schramm, A.; Baker, B. J.; Spang, A.; Ettema, T. J. G. Asgard Archaea Illuminate the Origin of Eukaryotic Cellular Complexity. Nature, 2017, 541, 353–358. https://doi.org/10.1038/nature21031 http://www.nature.com/nature/journal/v541/n7637/abs/nature21031.html#supplementary-information.

(125) Cavalier-Smith, T. Chloroplast Evolution: Secondary Symbiogenesis and Multiple Losses. Curr. Biol. **2002**, 12 (2), R62–R64. https://doi.org/10.1016/S0960-9822(01)00675-3.
(126) Raven, J. A.; Allen, J. F. Genomics and Chloroplast Evolution: What Did Cyanobacteria Do for Plants? Genome biology, 2003, 4, 209.
(127) Allen, J. F.; Raven, J. A.; Allen, J. F. The Function of Genomes in Bioenergetic Organelles. Philos. Trans. R. Soc. Lond. B. Biol. Sci. **2003**, 358 (1429), 19–38. https://doi.org/10.1098/rstb.2002.1191.
(128) Amiri, H.; Karlberg, O.; Andersson, S. G. E. Deep Origin of Plastid/Parasite ATP/ADP Translocases. J. Mol. Evol. **2003**, 56 (2), 137–150. https://doi.org/10.1007/s00239-002-2387-0.
(129) Major, P.; Embley, T. M.; Williams, T. A. Phylogenetic Diversity of NTT Nucleotide Transport Proteins in Free-Living and Parasitic Bacteria and Eukaryotes. Genome Biol. Evol. **2017**, 9 (2), 480–487. https://doi.org/10.1093/gbe/evx015.
(130) Paddon, C. J.; Keasling, J. D. Semi-Synthetic Artemisinin: A Model for the Use of Synthetic Biology in Pharmaceutical Development. Nat. Rev. Microbiol. **2014**, 12 (5), 355.
(131) Brochado, A. R.; Matos, C.; Møller, B. L.; Hansen, J.; Mortensen, U. H.; Patil, K. R. Improved Vanillin Production in Baker's Yeast through in Silico Design. Microb. Cell Factories **2010**, 9 (1), 84. https://doi.org/10.1186/1475-2859-9-84.
(132) Ostrov Nili; Jimenez Miguel; Billerbeck Sonja; Brisbois James; Matragrano Joseph; Ager Alastair; Cornish Virginia W. A Modular Yeast Biosensor for Low-Cost Point-of-Care Pathogen Detection. Sci. Adv. 3 (6), e1603221. https://doi.org/10.1126/sciadv.1603221.
(133) Schmitz-Esser, S.; Linka, N.; Collingro, A.; Beier, C. L.; Neuhaus, H. E.; Wagner, M.; Horn, M. ATP/ADP Translocases: A Common Feature of Obligate Intracellular Amoebal Symbionts Related to Chlamydiae and Rickettsiae. J. Bacteriol. **2004**, 186 (3), 683–691. https://doi.org/10.1128/JB.186.3.683-691.2004.
(134) Schmitz-Esser, S.; Haferkamp, I.; Knab, S.; Penz, T.; Ast, M.; Kohl, C.; Wagner, M.; Horn, M. Lawsonia Intracellularis Contains a Gene Encoding a Functional Rickettsia-like ATP/ADP Translocase for Host Exploitation. Journal of bacteriology, 2008, 190, 5746–5752.
(135) Bamber, L.; Harding, M.; Monné, M.; Slotboom, D.-J.; Kunji, E. R. S. The Yeast Mitochondrial ADP/ATP Carrier Functions as a Monomer in Mitochondrial Membranes. Proc. Natl. Acad. Sci. U. S. A. **2007**, 104 (26), 10830–10834. https://doi.org/10.1073/pnas.0703969104.
(136) Bonnefoy, N.; Bsat, N.; Fox, T. D. Mitochondrial Translation of Saccharomyces Cerevisiae COX2 mRNA Is Controlled by the Nucleotide Sequence Specifying the Pre-Cox2p Leader Peptide. Mol. Cell. Biol. **2001**, 21 (7), 2359–2372. https://doi.org/10.1128/MCB.21.7.2359-2372.2001.
(137) Poutre, C. G.; Fox, T. D. PET111, a Saccharomyces Cerevisiae Nuclear Gene Required for Translation of the Mitochondrial mRNA Encoding Cytochrome c Oxidase Subunit II. Genetics **1987**, 115 (4), 637–647.
(138) Steele, D. F.; Butler, C. A.; Fox, T. D. Expression of a Recoded Nuclear Gene Inserted into Yeast Mitochondrial DNA Is Limited by mRNA-Specific Translational Activation. Proc. Natl. Acad. Sci. **1996**, 93 (11), 5253–5257.
(139) Supekova, L.; Supek, F.; Greer, J. E.; Schultz, P. G. A Single Mutation in the First Transmembrane Domain of Yeast COX2 Enables Its Allotopic Expression. Proc. Natl. Acad. Sci. **2010**, 107 (11), 5047–5052. https://doi.org/10.1073/pnas.1000735107.
(140) Henze, K.; Martin, W.; Schnarrenberger, C. Chapter 27 - Endosymbiotic Gene Transfer: A Special Case of Horizontal Gene Transfer Germane to Endosymbiosis, the Origins of Organelles and the Origins of Eukaryotes. In Horizontal Gene Transfer (Second Edition); Syvanen, M., Kado, C. I., Eds.; Academic Press: London, 2002; pp 351–XII. https://doi.org/10.1016/B978-012680126-2/50034-7.
(141) Delevoye, C.; Nilges, M.; Dehoux, P.; Paumet, F.; Perrinet, S.; Dautry-Varsat, A.; Subtil, A. SNARE Protein Mimicry by an Intracellular Bacterium. PLoS Pathog. **2008**, 4 (3), e1000022.
(142) Wesolowski, J.; Paumet, F. SNARE Motif: A Common Motif Used by Pathogens to Manipulate Membrane Fusion. Virulence **2010**, 1 (4), 319–324.
(143) Casella, S.; Huang, F.; Mason, D.; Zhao, G.-Y.; Johnson, G. N.; Mullineaux, C. W.; Liu, L.-N. Dissecting the Native Architecture and Dynamics of Cyanobacterial Photosynthetic Machinery. Mol. Plant **2017**, 10 (11), 1434–1448. https://doi.org/10.1016/j.molp.2017.09.019.
(144) Six, C.; Thomas, J. C.; Brahamsha, B.; Lemoine, Y.; Partensky, F. Photophysiology of the Marine Cyanobacterium Synechococcus Sp. WH8102, a New Model Organism. Aquat. Microb. Ecol. **2004**, 35 (1), 17–29. https://doi.org/10.3354/ame035017.
(145) Karnovsky, M. A Formaldehyde-Glutaraldehyde Fixative of High Osmolality for Use in Electron Microscopy. J Cell Biol **1964**, 27.

(146) Facchinelli, F.; Weber, A. The Metabolite Transporters of the Plastid Envelope: An Update. Front. Plant Sci. **2011**, 2, 50. https://doi.org/10.3389/fpls.2011.00050.
(147) Doolittle, W. F. The Cyanobacterial Genome, Its Expression, and the Control of That Expression. In Advances in Microbial Physiology; Rose, A. H., Morris, J. G., Eds.; Academic Press, 1980; Vol. 20, pp 1–102. https://doi.org/10.1016/S0065-2911(08)60206-4.
(148) Hood, R. D.; Higgins, S. A.; Flamholz, A.; Nichols, R. J.; Savage, D. F. The Stringent Response Regulates Adaptation to Darkness in the Cyanobacterium Synechococcus Elongatus. Proc. Natl. Acad. Sci. **2016**, 113 (33), E4867–E4876. https://doi.org/10.1073/pnas.1524915113.
(149) Schmitz, O.; Tsinoremas, N. F.; Schaefer, M. R.; Anandan, S.; Golden, S. S. General Effect of Photosynthetic Electron Transport Inhibitors on Translation Precludes Their Use for Investigating Regulation of D1 Biosynthesis in Synechococcus Sp. Strain PCC 7942. Photosynth. Res. **1999**, 62 (2), 261–271. https://doi.org/10.1023/A:1006340423948.
(150) Gross, J.; Bhattacharya, D. Uniting Sex and Eukaryote Origins in an Emerging Oxygenic World. Biol. Direct **2010**, 5 (1), 53. https://doi.org/10.1186/1745-6150-5-53.
(151) Waldbauer, J. R.; Sherman, L. S.; Sumner, D. Y.; Summons, R. E. Late Archean Molecular Fossils from the Transvaal Supergroup Record the Antiquity of Microbial Diversity and Aerobiosis. Precambrian Res. **2009**, 169 (1), 28–47. https://doi.org/10.1016/j.precamres.2008.10.011.
(152) Lyons, T. W.; Reinhard, C. T.; Planavsky, N. J. The Rise of Oxygen in Earth's Early Ocean and Atmosphere. Nature **2014**, 506 (7488), 307–315. https://doi.org/10.1038/nature13068.
(153) Sessions, A. L.; Doughty, D. M.; Welander, P. V.; Summons, R. E.; Newman, D. K. The Continuing Puzzle of the Great Oxidation Event. Curr. Biol. **2009**, 19 (14), R567–R574. https://doi.org/10.1016/j.cub.2009.05.054.
(154) Burrows, C. J. Surviving an Oxygen Atmosphere: DNA Damage and Repair. ACS Symp. Ser. Am. Chem. Soc. **2009**, 2009, 147–156. https://doi.org/10.1021/bk-2009-1025.ch008.
(155) Hörandl, E.; Speijer, D. How Oxygen Gave Rise to Eukaryotic Sex. Proc. R. Soc. B Biol. Sci. **2018**, 285 (1872), 20172706. https://doi.org/10.1098/rspb.2017.2706.
(156) Koonin, E. V.; Aravind, L. Origin and Evolution of Eukaryotic Apoptosis: The Bacterial Connection. Cell Death Differ. **2002**, 9 (4), 394–404. https://doi.org/10.1038/sj.cdd.4400991.
(157) Delaye, L.; Valadez-Cano, C.; Pérez-Zamorano, B. How Really Ancient Is Paulinella Chromatophora? PLoS Curr. **2016**, 8, ecurrents.tol.e68a099364bb1a1e129a17b4e06b0c6b. https://doi.org/10.1371/currents.tol.e68a099364bb1a1e129a17b4e06b0c6b.
(158) Gavelis, G. S.; Gile, G. H. How Did Cyanobacteria First Embark on the Path to Becoming Plastids?: Lessons from Protist Symbioses. FEMS Microbiol. Lett. **2018**, 365 (19). https://doi.org/10.1093/femsle/fny209.
(159) Marin, B.; Nowack, E. C.; Glöckner, G.; Melkonian, M. The Ancestor of the Paulinella Chromatophore Obtained a Carboxysomal Operon by Horizontal Gene Transfer from a Nitrococcus-like γ-Proteobacterium. BMC Evol. Biol. **2007**, 7 (1), 85. https://doi.org/10.1186/1471-2148-7-85.
(160) Nowack, E. C. M.; Price, D. C.; Bhattacharya, D.; Singer, A.; Melkonian, M.; Grossman, A. R. Gene Transfers from Diverse Bacteria Compensate for Reductive Genome Evolution in the Chromatophore of Paulinella Chromatophora. Proc. Natl. Acad. Sci. **2016**, 113 (43), 12214–12219. https://doi.org/10.1073/pnas.1608016113.
(161) Nakayama, T.; Ishida, K. Another Acquisition of a Primary Photosynthetic Organelle Is Underway in Paulinella Chromatophora. Curr. Biol. **2009**, 19 (7), R284–R285. https://doi.org/10.1016/j.cub.2009.02.043.
(162) Wang, Z.; Wu, M. Phylogenomic Reconstruction Indicates Mitochondrial Ancestor Was an Energy Parasite. PLOS ONE **2014**, 9 (10), e110685. https://doi.org/10.1371/journal.pone.0110685.
(163) Whatley, J. M.; John, P.; Whatley, F. R.; Richmond, M. H.; Smith, D. C. From Extracellular to Intracellular: The Establishment of Mitochondria and Chloroplasts. Proc. R. Soc. Lond. B Biol. Sci. **1979**, 204 (1155), 165–187. https://doi.org/10.1098/rspb.1979.0020.
(164) Gardeström, P. Interactions between Mitochondria and Chloroplasts. Biochim. Biophys. Acta BBA - Bioenerg. **1996**, 1275 (1), 38–40. https://doi.org/10.1016/0005-2728(96)00047-3.
(165) Glass, J. I.; Merryman, C.; Wise, K. S.; Hutchison, C. A.; Smith, H. O. Minimal Cells—Real and Imagined. Cold Spring Harb. Perspect. Biol. **2017**, 9 (12), a023861. https://doi.org/10.1101/cshperspect.a023861.
(166) Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, J. C.; Hutchison, C. A.; Smith, H. O. Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases. Nat. Methods **2009**, 6 (5), 343–345. https://doi.org/10.1038/nmeth.1318.

(167) Clerico, E. M.; Ditty, J. L.; Golden, S. S. Specialized Techniques for Site-Directed Mutagenesis in Cyanobacteria. In Circadian Rhythms: Methods and Protocols; Rosato, E., Ed.; Methods in Molecular Biology™; Humana Press: Totowa, NJ, 2007; pp 155–171. https://doi.org/10.1007/978-1-59745-257-1_11.
(168) You, H.; Lattmann, S.; Rhodes, D.; Yan, J. RHAU Helicase Stabilizes G4 in Its Nucleotide-Free State and Destabilizes G4 upon ATP Hydrolysis. Nucleic Acids Res. **2017**, 45 (1), 206–214. https://doi.org/10.1093/nar/gkw881.
(169) Reynolds, E. S. THE USE OF LEAD CITRATE AT HIGH pH AS AN ELECTRON-OPAQUE STAIN IN ELECTRON MICROSCOPY. J. Cell Biol. **1963**, 17 (1), 208–212. https://doi.org/10.1083/jcb.17.1.208.
(170) Chen, Y.; Taton, A.; Go, M.; London, R. E.; Pieper, L. M.; Golden, S. S.; Golden, J. W. Y. 2016. Self-Replicating Shuttle Vectors Based on pANS, a Small Endogenous Plasmid of the Unicellular Cyanobacterium Synechococcus Elongatus PCC 7942. Microbiology 162 (12), 2029–2041. https://doi.org/10.1099/mic.0.000377.
(171) Bockholt, R.; Scholten-Beck, G.; Pistorius, E. K. Construction and Partial Characterization of an L-Amino Acid Oxidase-Free Synechococcus PCC 7942 Mutant and Localization of the l-Amino Acid Oxidase in the Corresponding Wild Type. Biochim. Biophys. Acta BBA - Gene Struct. Expr. **1996**, 1307 (1), 111–121. https://doi.org/10.1016/0167-4781(96)00029-2.
(172) Bonnefoy, N.; Bsat, N.; Fox, T. D. Mitochondrial Translation of Saccharomyces Cerevisiae COX2 mRNA Is Controlled by the Nucleotide Sequence Specifying the Pre-Cox2p Leader Peptide. Mol. Cell. Biol. **2001**, 21 (7), 2359–2372. https://doi.org/10.1128/MCB.21.7.2359-2372.2001.
(173) Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, J. C.; Hutchison, C. A.; Smith, H. O. Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases. Nat. Methods **2009**, 6 (5), 343–345. https://doi.org/10.1038/nmeth.1318.
(174) You, H.; Lattmann, S.; Rhodes, D.; Yan, J. RHAU Helicase Stabilizes G4 in Its Nucleotide-Free State and Destabilizes G4 upon ATP Hydrolysis. Nucleic Acids Res. **2017**, 45 (1), 206–214. https://doi.org/10.1093/nar/gkw881.
(175) Shimoi, H.; Kitagaki, H.; Ohmori, H.; Iimura, Y.; Ito, K. Sed1p Is a Major Cell Wall Protein of Saccharomyces Cerevisiae in the Stationary Phase and Is Involved in Lytic Enzyme Resistance. J. Bacteriol. **1998**, 180 (13), 3381–3387.
(176) Elliott, B.; Futcher, B. Stress Resistance of Yeast Cells Is Largely Independent of Cell Cycle Phase. Yeast **1993**, 9 (1), 33–42. https://doi.org/10.1002/yea.320090105.
(177) Petsalaki, E. I.; Bagos, P. G.; Litou, Z. I.; Hamodrakas, S. J. PredSL: A Tool for the N-Terminal Sequence-Based Prediction of Protein Subcellular Localization. Genomics Proteomics Bioinformatics **2006**, 4 (1), 48–55. https://doi.org/10.1016/S1672-0229(06)60016-8.
(178) Small, I.; Peeters, N.; Legeai, F.; Lurin, C. Predotar: A Tool for Rapidly Screening Proteomes for N-Terminal Targeting Sequences. PROTEOMICS **2004**, 4 (6), 1581–1590. https://doi.org/10.1002/pmic.200300776.
(179) Armenteros, J. J. A.; Salvatore, M.; Emanuelsson, O.; Winther, O.; Heijne, G. von; Elofsson, A.; Nielsen, H. Detecting Sequence Signals in Targeting Peptides Using Deep Learning. Life Sci. Alliance **2019**, 2 (5). https://doi.org/10.26508/lsa.201900429.
(180) Dubendorf, J. W.; Studier, F. W. Controlling Basal Expression in an Inducible T7 Expression System by Blocking the Target T7 Promoter with Lac Repressor. J. Mol. Biol. **1991**, 219 (1), 45–59. https://doi.org/10.1016/0022-2836(91)90856-2.
(181) Sánchez, Á.; Vila, J. C. C.; Chang, C.-Y.; Diaz-Colunga, J.; Estrela, S.; Rebolleda-Gomez, M. Directed Evolution of Microbial Communities. Annu. Rev. Biophys. **2021**, 50 (Volume 50, 2021), 323–341. https://doi.org/10.1146/annurev-biophys-101220-072829.
(182) Xie, L.; Yuan, A. E.; Shou, W. Simulations Reveal Challenges to Artificial Community Selection and Possible Strategies for Success. PLOS Biol. **2019**, 17 (6), e3000295. https://doi.org/10.1371/journal.pbio.3000295.
(183) Endosymbionts in Paramecium; Fujishima, M., Ed.; Microbiology Monographs; Springer Berlin Heidelberg: Berlin, Heidelberg, 2009; Vol. 12. https://doi.org/10.1007/978-3-540-92677-1.
(184) He, M.; Wang, J.; Fan, X.; Liu, X.; Shi, W.; Huang, N.; Zhao, F.; Miao, M. Genetic Basis for the Establishment of Endosymbiosis in Paramecium. ISME J. **2019**, 13 (5), 1360–1369. https://doi.org/10.1038/s41396-018-0341-4.
(185) McCutcheon, J. P.; McDonald, B. R.; Moran, N. A. Convergent Evolution of Metabolic Roles in Bacterial Co-Symbionts of Insects. Proc. Natl. Acad. Sci. **2009**, 106 (36), 15394–15399. https://doi.org/10.1073/pnas.0906424106.

(186) Skinner, S. O.; Sepúlveda, L. A.; Xu, H.; Golding, I. Measuring mRNA Copy-Number in Individual Escherichia Coli Cells Using Single-Molecule Fluorescent in Situ Hybridization (smFISH). Nat. Protoc. **2013**, 8 (6), 1100–1113. https://doi.org/10.1038/nprot.2013.066.
(187) Nair, U.; Thomas, C.; Golden, S. S. Functional Elements of the Strong psbAI Promoter of Synechococcus Elongatus PCC 7942. J. Bacteriol. **2001**, 183 (5), 1740–1747. https://doi.org/10.1128/JB.183.5.1740-1747.2001.
(188) Hihara, Y.; Kamei, A.; Kanehisa, M.; Kaplan, A.; Ikeuchi, M. DNA Microarray Analysis of Cyanobacterial Gene Expression during Acclimation to High Light. Plant Cell **2001**, 13 (4), 793–806. https://doi.org/10.1105/tpc.13.4.793.
(189) Kulkarni, R. D.; Schaefer, M. R.; Golden, S. S. Transcriptional and Posttranscriptional Components of psbA Response to High Light Intensity in Synechococcus Sp. Strain PCC 7942. J. Bacteriol. **1992**, 174 (11), 3775–3781. https://doi.org/10.1128/jb.174.11.3775-3781.1992.
(190) Mahbub, M.; Hemm, L.; Yang, Y.; Kaur, R.; Carmen, H.; Engl, C.; Huokko, T.; Riediger, M.; Watanabe, S.; Liu, L.-N.; Wilde, A.; Hess, W. R.; Mullineaux, C. W. mRNA Localization, Reaction Centre Biogenesis and Thylakoid Membrane Targeting in Cyanobacteria. Nat. Plants **2020**, 6 (9), 1179–1191. https://doi.org/10.1038/s41477-020-00764-2.
(191) Binder, B. J.; Liu, Y. C. Growth Rate Regulation of rRNA Content of a MarineSynechococcus (Cyanobacterium) Strain. Appl. Environ. Microbiol. **1998**, 64 (9), 3346–3351. https://doi.org/10.1128/AEM.64.9.3346-3351.1998.
(192) Worden, A. Z.; Chisholm, S. W.; Binder, B. J. In Situ Hybridization of Prochlorococcus and Synechococcus (Marine Cyanobacteria) Spp. with rRNA-Targeted Peptide Nucleic Acid Probes. Appl. Environ. Microbiol. **2000**, 66 (1), 284–289.
(193) Paerl, R. W.; Johnson, K. S.; Welsh, R. M.; Worden, A. Z.; Chavez, F. P.; Zehr, J. P. Differential Distributions of Synechococcus Subgroups Across the California Current System. Front. Microbiol. **2011**, 2, 59. https://doi.org/10.3389/fmicb.2011.00059.
(194) Daims, H.; Brühl, A.; Amann, R.; Schleifer, K.-H.; Wagner, M. The Domain-Specific Probe EUB338 Is Insufficient for the Detection of All Bacteria: Development and Evaluation of a More Comprehensive Probe Set. Syst. Appl. Microbiol. **1999**, 22 (3), 434–444. https://doi.org/10.1016/S0723-2020(99)80053-8.
(195) Mahbub, M.; Mullineaux, C. W. Locations of Membrane Protein Production in a Cyanobacterium. J. Bacteriol. **2023**, e00209-23. https://doi.org/10.1128/jb.00209-23.
(196) Raj, A.; van den Bogaard, P.; Rifkin, S. A.; van Oudenaarden, A.; Tyagi, S. Imaging Individual mRNA Molecules Using Multiple Singly Labeled Probes. Nat. Methods **2008**, 5 (10), 877–879. https://doi.org/10.1038/nmeth.1253.
(197) Shahrezaei, V.; Swain, P. S. Analytical Distributions for Stochastic Gene Expression. Proc. Natl. Acad. Sci. U. S. A. **2008**, 105 (45), 17256–17261. https://doi.org/10.1073/pnas.0803850105.
(198) Unruh, J. R.; Gokulrangan, G.; Wilson, G. S.; Johnson, C. K. Fluorescence Properties of Fluorescein, Tetramethylrhodamine and Texas Red Linked to a DNA Aptamer¶. Photochem. Photobiol. **2005**, 81 (3), 682–690. https://doi.org/10.1111/j.1751-1097.2005.tb00244.x.
(199) Makrigiorgos, G. M. Detection of Lipid Peroxidation on Erythrocytes Using the Excimer-Forming Property of a Lipophilic BODIPY Fluorescent Dye. J. Biochem. Biophys. Methods **1997**, 35 (1), 23–35. https://doi.org/10.1016/S0165-022X(97)00020-1.
(200) Azarin, K.; Usatov, A.; Makarenko, M.; Kozel, N.; Kovalevich, A.; Dremuk, I.; Yemelyanova, A.; Logacheva, M.; Fedorenko, A.; Averina, N. A Point Mutation in the Photosystem I P700 Chlorophyll a Apoprotein A1 Gene Confers Variegation in Helianthus Annuus L. Plant Mol. Biol. **2020**, 103 (4), 373–389. https://doi.org/10.1007/s11103-020-00997-x.
(201) Venema, J.; Tollervey, D. Ribosome Synthesis in Saccharomyces Cerevisiae. Annu. Rev. Genet. **1999**, 33 (1), 261–311. https://doi.org/10.1146/annurev.genet.33.1.261.
(202) Dubois, M.-L.; Boisvert, F.-M. The Nucleolus: Structure and Function. Funct. Nucl. **2016**, 29–49. https://doi.org/10.1007/978-3-319-38882-3_2.
(203) Raj, A.; van Oudenaarden, A. Stochastic Gene Expression and Its Consequences. Cell **2008**, 135 (2), 216–226. https://doi.org/10.1016/j.cell.2008.09.050.
(204) Henze, K.; Martin, W.; Schnarrenberger, C. Chapter 27 - Endosymbiotic Gene Transfer: A Special Case of Horizontal Gene Transfer Germane to Endosymbiosis, the Origins of Organelles and the Origins of Eukaryotes. In Horizontal Gene Transfer (Second Edition); Syvanen, M., Kado, C. I., Eds.; Academic Press: London, 2002; pp 351–XII. https://doi.org/10.1016/B978-012680126-2/50034-7.

(205) Jarvis, J.; Sha, M. The Lighting Intensity of New Brunswick™ Innova® 42R Incubator Shaker Photosynthetic Light Bank. 4.
(206) Raj, A.; Tyagi, S. Chapter 17 - Detection of Individual Endogenous RNA Transcripts In Situ Using Multiple Singly Labeled Probes. In Methods in Enzymology; Walter, N. G., Ed.; Single Molecule Tools: Fluorescence Based Approaches, Part A; Academic Press, 2010; Vol. 472, pp 365–386. https://doi.org/10.1016/S0076-6879(10)72004-8.
(207) Pountain, A. W.; Jiang, P.; Yao, T.; Homaee, E.; Guan, Y.; McDonald, K. J. C.; Podkowik, M.; Shopsin, B.; Torres, V. J.; Golding, I.; Yanai, I. Transcription–Replication Interactions Reveal Bacterial Genome Regulation. Nature **2024**, 626 (7999), 661–669. https://doi.org/10.1038/s41586-023-06974-w.
(208) Yao, T.; Coleman, S.; Nguyen, T. V. P.; Golding, I.; Igoshin, O. A. Bacteriophage Self-Counting in the Presence of Viral Replication. Proc. Natl. Acad. Sci. **2021**, 118 (51), e2104163118. https://doi.org/10.1073/pnas.2104163118.
(209) Oeffinger, M.; Zenklusen, D.; Ferguson, A.; Wei, K. E.; El Hage, A.; Tollervey, D.; Chait, B. T.; Singer, R. H.; Rout, M. P. Rrp17p Is a Eukaryotic Exonuclease Required for 5′ End Processing of Pre-60S Ribosomal RNA. Mol. Cell **2009**, 36 (5), 768–781. https://doi.org/10.1016/j.molcel.2009.11.011.

**Chapter 5—Appendix**

**BLAST method**

```
Blast4-request ::= {
  body queue-search {
    program "blastp",
    service "plain",
    queries bioseq-set {
      seq-set {
        seq {
          id {
            local str "Query_4104"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">At2g33260 15225823 HAAAP putative
 tyrosine-specific transport protein [Arabidopsis thaliana ]"
                }
              }
            },
            title "At2g33260 15225823 HAAAP putative tyrosine-specific
 transport protein [Arabidopsis thaliana ]"
          },
          inst {
            repr raw,
            mol aa,
            length 436,
            seq-data ncbieaa
"MDDLEITHETKKGKSFWAAVSLIIGTAVGPGMLGLPAATIRSGSIPST
IALLCSWVYVISSILLVAELSFAAMEEDNAAEVSFTGLATKSFGNKFGVFVAFVYASLSFSLMVACVSGIGSIVSQW
F
PSMNPFLANAIFPLVSGILIGFFPFNAIDFTNRGLCFLMLFSITSLVAIGLSVARSNVLASFGQSCWKVSMVLPAVP
V
MVLTLGFHVITPFICNLAGDSVSDARRAILVGGVVPLAMVLSWNLIVLGLARITVPAAPSSTIDPISLLLSVNPSAL
S
AVQGFAFSALATSLIGYAVSFPKQLLDTWKLVSKQSNGNGRLGSVSFSSKERDRRTNGRASYNEPARARDGFEAVVM
L
FVLGVPALIATFFPSTFSRALDFAGVYANCFLFGVLPPAMAYIQQSRKKLRPWVLPGGNFTLLILFAIAIILGIWH"
          }
        },
        seq {
          id {
            local str "Query_4105"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">BA3306 30263218 HAAAP serine/threonine
 transporter family protein [Bacillus anthracis Ames]"
                }
```

```
              }
            },
            title "BA3306 30263218 HAAAP serine/threonine transporter family
 protein [Bacillus anthracis Ames]"
          },
          inst {
            repr raw,
            mol aa,
            length 439,
            seq-data ncbieaa
"MNGNTAKNIEFQVDNTTVKNEKYLDPKKWHKQDTTWALSLFGTAIGAG
VLFLPINAGSGGLLSLLLITILAYPVMYYSHRALAKMIYASNSADEGITGTIREYFGNKASIIFNIVYFVSIYTIVL
M
YSVALTNTASSFIVHQLHMPEPPRAILSLVLVIGLITILNFGQDITVKIMSMLVYPFIASLLFIAISLIPQWNTSML
S
FSSVSTASTGTGYFGTIWMILPIIVFSFNHSPMISSFVMKQRATYGIDATDAKCAQIQKVCYIMTFAVVMFFVWSST
L
SLTPDDLKVAKEQNLSILSYLANELNSPVITIAAPIIAFVAITKSFLGHYIGAYEVMRDMIIKSGKKRGKDIGEKTV
K
TMILTFVVLTCWYVAYTNPSILGIIDALSGPLVAAILCLLPMYAIRKVPVLAKYRGKMSNVFVIIIGILTVLASIKS
LF"
          }
        },
        seq {
          id {
            local str "Query_4106"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">BA3451 30263357 HAAAP serine/threonine
 transporter family protein [Bacillus anthracis Ames]"
                }
              }
            },
            title "BA3451 30263357 HAAAP serine/threonine transporter family
 protein [Bacillus anthracis Ames]"
          },
          inst {
            repr raw,
            mol aa,
            length 439,
            seq-data ncbieaa
"MNGNTAKKIEVQAENTALNNEQYADPKKWHKQDTTWALSLFGTAIGAG
VLFLPINAGSGGLLSLLLITLLAYPVMYYSHRALAKMIYASNSADEGITGTIREYFGNKASIIFNIVYFGSIYTIVL
M
YSVALTNTASSFIVHQLHMPEPPRAILSLVLVLGLIAILNFGQDITVKVMSMLVYPFIVSLLFIAISLIPQWNTSML
S
FSAVSTASTGTGYFGTIVMILPIIVFSFNHSPMISSFVVKQRATYGIEATDAKCAQIQKVCYIMTFAVVMFFVWSSA
L
SLTPDDIKMAKEQNLLILSYLANELNSPVITIAAPIIAFVAITKSFLGHYIGAFEVMRDMIIKFGKSRGKDIEEKTI
K
TIILTFVVLSCWFVAYTNPSILGLIDSLSGPLVAAILCLLPMYAIQKVPVLAKYKGKMSNVFVIIVGVLTVLASIKS
LF"
```

```
          }
        },
        seq {
          id {
            local str "Query_4107"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">BC3189 30021299 HAAAP Serine transporter
 [Bacillus cereus ATCC14579]"
                }
              }
            },
            title "BC3189 30021299 HAAAP Serine transporter [Bacillus cereus
 ATCC14579]"
          },
          inst {
            repr raw,
            mol aa,
            length 439,
            seq-data ncbieaa
"MNGNTAKKIELQSENTVITNEKYLDPKKWHKQDTTWALSLFGTAIGAG
VLFLPINAGSGGLLSLLLITLLAYPVMYYSHRALAKMIYASNSADEGITGTIREYFGNTASIIFNIVYFVSIYTIVL
M
YSVALTNTASSFIVNQLHMKEPSRAILSLVLVLGLIAILNFGQDITVKIMSLLVYPFIASLLFIAISLIPQWNTSML
S
FSDVSTASTGTGYLGTIWMILPIIVFSFNHSPMISSFVIKQRSTYGIEATDAKCAQIQKICYIMTFVVVMFFVWSSA
L
SLTPDDLKVAKEQNLSILSYLANELNSPVITIAAPIIAFMAITKSFLGHYIGSYEVMRDMIIKFGKTRGKDIEEKTV
K
TVILTFVVLTCWYVAYANPSILGLIDALSGPLVAAILCLLPMYAIRKVPVLAKYRGKMSNVFVIIVGVLTILASIKS
LF"
          }
        },
        seq {
          id {
            local str "Query_4108"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">BC3398 30021502 HAAAP Serine transporter
 [Bacillus cereus ATCC14579]"
                }
              }
            },
            title "BC3398 30021502 HAAAP Serine transporter [Bacillus cereus
 ATCC14579]"
          },
          inst {
```

```
            repr raw,
            mol aa,
            length 439,
            seq-data ncbieaa
"MNGNTAKKIEVQAENTALKNEQYADPKKWHKQDTTWALSLFGTAIGAG
VLFLPINAGSGGLLSLLLITLLAYPVMYYSHRALAKMIYASNSADEGITGTIREYFGNKASIIFNIVYFGSIYTIVL
M
YSVALTNTASSFIVHQLHMPEPPRAILSLVLVLGLIAILNFGQDITVKVMSMLVYPFIVSLLFIAISLIPQWNTSML
S
FSAVSTASTGTGYFGTILMILPIIVFSFNHSPMISSFVVKQRATYGIEATDAKCAQIQKVCYIMTFAVVMFFVWSSA
L
SLTPDDIKMAKEQNLSILSYLANELNSPVITIAAPIIAFVAITKSFLGHYIGAFEVMRDMIIKFGKSRGKDIEEKTI
K
TIILTFVVLSCWFVAYTNPSILGLIDSLSGPLVAAILCLLPMYAIQKVPVLAKYKGKMSNVFVIIVGVLTVLASIKS
LF"
          }
        },
        seq {
          id {
            local str "Query_4109"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">CBU1539 29654830 HAAAP tryptophan/tyrosine
 permease family protein [Coxiella burnetii RSA493]"
                }
              }
            },
            title "CBU1539 29654830 HAAAP tryptophan/tyrosine permease family
 protein [Coxiella burnetii RSA493]"
          },
          inst {
            repr raw,
            mol aa,
            length 426,
            seq-data ncbieaa
"MIMEPPNILFDESQDVTFVNEIESPKAETTRGKVFGAILLIIGTSVGG
GMLALPMAIAAGGYYHSIFLFFGAWLITVLAAFYILEANLWLPENTNLISMARITLGKAGQLITWISYLLLLYTLLA
A
YMSGGTDLIHNLLSLINIPTPTWLDSILFVVILGAVLFYGVRAVDWVNRGLMSTKLAVYFLLVLFISPHIDVGKLSG
G
RLALLSGAVMVIITSFGYATIVPTLRSYLKSQVNALRLTIAIGSLMPLILYLLWTFVVNGTLGSHGANGLIHMAAST
D
AVSELSNTLSAHVNGELVKGLIHFFTSICIATSFLGVSLCLSDFMADGLKIKKEKNGRWLLVALTLAPPLLVILFYP
G
AFIASLNYAGVLCVILLIFLPALMVWSGRYVKKIAIGYEVIGGKTFIILEILVALALLIFALMHLG"
          }
        },
        seq {
          id {
            local str "Query_4110"
          },
          descr {
```

```
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">CCA00787 29840544 HAAAP aromatic amino acid
 transport protein [Chlamydophila caviae GPIC]"
                }
              }
            },
            title "CCA00787 29840544 HAAAP aromatic amino acid transport
 protein [Chlamydophila caviae GPIC]"
          },
          inst {
            repr raw,
            mol aa,
            length 395,
            seq-data ncbieaa
"MSNKVLGGSLIVTGTAIGAGVLAVPVITAYAGFLPTTLLYVLSWLFSV
ASGLCYLEIMTWFKEKQQVNLLSMAQYTLGDIGKIFMWLLYLFLFYSLLIAYFCEGGNILFRIFGCQGLDIPWIRHM
A
PLAFAVLICPALMMGAKVVDYCNRFFVLGLAIAFAVFCILGVFSLQPQLLLRASWARSTEGLSVLFLSFGFHNVVPS
L
YYYMDKNVKDVKKAIFIGSLIPLILYVIWEALVLGVVPLDFLMKAKEHGYTAVEAMKTSLQCSMFYLAGEFFGFFAL
V
SSFLGVALGVMDFLADALQWNKKKRSFSIFFLTVIIPLAWSMCYPEIVLKCLSYAGGFGAALIIGVCPVLMIWKGRY
G
KKHYQAKHLVPGGKIVLVLMLVVIVINLASFYYKF"
          }
        },
        seq {
          id {
            local str "Query_4111"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">NCgl0464 19551724 HAAAP amino acid permease
 [Corynebacterium glutamicum ATCC13032]"
                }
              }
            },
            title "NCgl0464 19551724 HAAAP amino acid permease
 [Corynebacterium glutamicum ATCC13032]"
          },
          inst {
            repr raw,
            mol aa,
            length 415,
            seq-data ncbieaa
"MTTESIVAHNAAGTAPQNVSSAKKKYLSVAQGVALIYGTNIGAGVLSL
PYAARNGGFLALVVALLIAGTLTTISMLYIAEVSLRTKKPLQLSGLAEKYLGQWGRWLVFIAIVVNSVGALIAYASG
S
```

```
GILIGNLTGLPPIVGTLGFFVLGTLIMWKGLHTASFVEALITTGMATIIIVLCGWTVLGPGISADNLIVFHPFFIVP
I
MNLAVFTFLAQYVVPEIARGVNPATPKAVPRAIIIGMVATGVTLAAVPFAALGLLGTGVSEVVTISWGEALAPVAYY
M
ANAFALLAMFTSFIAIGFTAMRNVLDIGHWPQHGWQRSVAVGLTVLPPLAISLAGLGGFVAALSYAGGFAGAIMSII
P
VLLLRNSRKSGDQEPVWKATWQAHPIFQILLIVVYSLAFVYSVLAIVGLMPAGWA"
          }
        },
        seq {
          id {
            local str "Query_4112"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">Cj1625c 15792930 HAAAP serine transporter
 [Campylobacter jejuni NCTC11168]"
                }
              }
            },
            title "Cj1625c 15792930 HAAAP serine transporter [Campylobacter
 jejuni NCTC11168]"
          },
          inst {
            repr raw,
            mol aa,
            length 416,
            seq-data ncbieaa
"MNTPKWTSHDTRWVLSLFGTAIGAGVLLLPISAGLGGLIPLLVILVLA
FPMTYLAHRNLCRFVLSSSNPKDDITFVAESYFGKGGGFLITLLYFFAILPILLVYSANLTTTLLEFLINQFNFNAD
L
THAARWWVSFLIVGVLVLISILGENVVTKAMSFLVFPFIIFLFIFSLLLIPQWNLSLFANVDFSVISTSNFWVTLWL
V
IPVMVFSFNHSPIISSLACYCKKEYGGYAEPRARKIISLAIILMVFVVMFFVFSCALTFTPEDFASAKDQNINILTF
I
ANKFPEVSLLAYVGPIVALVAISKSFLGHYLGSQEGLNGILYKASNGRIQGKFAQTLTAIITFAIAWLVAYKNPSVI
G
IIEAIGGPVLAILLFLMPLYCIYRFDILARFRNKFLDLFVLVMGIVAISAAIHDLL"
          }
        },
        seq {
          id {
            local str "Query_4113"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">TC0204 15834824 HAAAP Mtr/TnaB/TyrO permease
 family protein [Chlamydia muridarum Nigg]"
                }
```

```
              }
            },
            title "TC0204 15834824 HAAAP Mtr/TnaB/TyrO permease family
protein
 [Chlamydia muridarum Nigg]"
          },
          inst {
            repr raw,
            mol aa,
            length 398,
            seq-data ncbieaa
"MINKMLGGALIVAGTTIGAGVLAVPIATSEGGFLPTTLLYVVSWLIAV
ASGYCFLEVLTWLHARKNVNMVSMAECTLGYKSKVIMWLVYLLLFYSLLVAYFCDGGNILMRVMGCRDWDTPWIRHA
M
PIVFFALFSPLLMAKTSIVDQCNRVFVLGLGISFAMFCYFGFPLMKTELLVRSSWGGTLKGFPILFLAFGFQNVVPT
L
YHYMDKNVRDVKKAIVIGSFIPLVLYVIWEAIVLGAVPVSFLEQAKMEGWTAIGALQGALKCSAFYIAGEFFGFFAL
I
SSFIGVALGLKDFFIDAFQWDEKKRKLEIFLLVFVFPLVWAVFYPGIVLKCLECTGALGETIVLGVFPVLMVWKGRY
G
KKRYYGQRILPGGKGALLVMSGLVLVNLVLIVQKFLGY"
          }
        },
        seq {
          id {
            local str "Query_4114"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">TC0205 15834825 HAAAP Mtr/TnaB/TyrO permease
 family protein [Chlamydia muridarum Nigg]"
                }
              }
            },
            title "TC0205 15834825 HAAAP Mtr/TnaB/TyrO permease family
protein
 [Chlamydia muridarum Nigg]"
          },
          inst {
            repr raw,
            mol aa,
            length 395,
            seq-data ncbieaa
"MRNKCIGGVLIVAGTVIGAGVLAVPVLTALDGFLPAALLYMLAWLVSL
ASGYGYLEVLTWCKGNKQANLCSMAEETLGKVGRIVLCLVYLFLFYSLLVAYFCDGGNILSRMLGDGVLENPWARHV
M
PILFFCIFAPLLMAKTSIIDYCNRVFVFGLILVFGLFCILGAPRVQGDLLLRASWFSSLNSLPIFFLAFGFQNVVPS
L
YHYLDGDVREVKRAIFIGSLIPLVLYVIWEALVLGTVPLVYLLKAKELGWTAAGALQGALKNSAFHIAGELFGFFAL
V
TSFIGTALALKDFYIDIFKWDARKQRLNLFLLVLVFPLVWAVSYPEIVLSCLRYAGGIGGACIIVLFPVAMLWNGRY
G
KRHCSGKQILPGGKTVLLILLGYTVLNLAPLYYTF"
```

```
          }
        },
        seq {
          id {
            local str "Query_4115"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">CP0889 16752061 HAAAP Mtr/TnaB/TyrO permease
 family protein [Chlamydia pneumoniae AR39]"
                }
              }
            },
            title "CP0889 16752061 HAAAP Mtr/TnaB/TyrO permease family
protein
 [Chlamydia pneumoniae AR39]"
          },
          inst {
            repr raw,
            mol aa,
            length 396,
            seq-data ncbieaa
"MSNKVLGGSLLIAGSAIGAGVLAVPVLTAKGGFFPATFLYIVSWLFSM
ASGLCLLEVMTWMKESKNPVNMLSMAESILGHVGKISICLVYLFLFYSLLIAYFCEGGNILCRVFNCQNLGISWIRH
L
GPLGFAILMGPIIMAGTKVIDYCNRFFMFGLTVAFGIFCALGFLKIQPSFLVRSSWLTTINAFPVFFLAFGFQSIIP
T
LYYYMDKKVGDVKKAILIGTLIPLVLYVLWEVVVLGAVSLPILSQAKIGGYTAVEALKQAHRSWAFYIAGELFGFFA
L
VSSFVGVALGVMDFLADGLKWNKKSHPFSIFFLTFIIPLAWAVCYPEIVLTCLKYAGGFGAAVIIGVFPTLIVWKGR
Y
GKQHHREKQLVPGGKFALFLMFLLIVINVVSIYHEL"
          }
        },
        seq {
          id {
            local str "Query_4116"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">CP0891 16752063 HAAAP Mtr/TnaB/TyrO permease
 family protein [Chlamydia pneumoniae AR39]"
                }
              }
            },
            title "CP0891 16752063 HAAAP Mtr/TnaB/TyrO permease family
protein
 [Chlamydia pneumoniae AR39]"
          },
```

```
          inst {
            repr raw,
            mol aa,
            length 396,
            seq-data ncbieaa
"MSNKVLGGSLLIAGSAIGAGVLAVPVLTAKGGFFPATFLYIVSWLFSM
ASGLCLLEVMTWMKESKNPVNMLSMAESILGHVGKISICLVYLFLFYSLLIAYFCEGGNILCRVFNCQNLGISWIRH
L
GPLGFAILMGPIIMAGTKVIDYCNRFFMFGLTVAFGIFCALGFLKIQPSFLVRSSWLTTINAFPVFFLAFGFQSIIP
T
LYYYMDKKVGDVKKAILIGTLIPLVLYVLWEVVVLGAVSLPILSQAKIGGYTAVEALKQAHRSWAFYIAGELFGFFA
L
VSSFVGVALGVMDFLADGLKWNKKSHPFSIFFLTFIIPLAWAVCYPEIVLTCLKYAGGFGAAVIIGVFPTLIVWKGR
Y
GKQHHREKQLVPGGKFALFLMFLLIVINVVSIYHEL"
          }
        },
        seq {
          id {
            local str "Query_4117"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">CpB1006 33242337 HAAAP tyrosine-specific
 transport protein [Chlamydophila pneumoniae TW-183]"
                }
              }
            },
            title "CpB1006 33242337 HAAAP tyrosine-specific transport protein
 [Chlamydophila pneumoniae TW-183]"
          },
          inst {
            repr raw,
            mol aa,
            length 396,
            seq-data ncbieaa
"MSNKVLGGSLLIAGSAIGAGVLAVPVLTAKGGFFPATFLYIVSWLFSM
ASGLCLLEVMTWMKESKNPVNMLSMAESILGHVGKISICLVYLFLFYSLLIAYFCEGGNILCRVFNCQNLGISWIRH
L
GPLGFAILMGPIIMAGTKVIDYCNRFFMFGLTVAFGIFCALGFLKIQPSFLVRSSWLTTINAFPVFFLAFGFQSIIP
T
LYYYMDKKVGDVKKAILIGTLIPLVLYVLWEVVVLGAVSLPILSQAKIGGYTAVEALKQAHRSWAFYIAGELFGFFA
L
VSSFVGVALGVMDFLADGLKWNKKSHPFSIFFLTFIIPLAWAVCYPEIVLTCLKYAGGFGAAVIIGVFPTLIVWKGR
Y
GKQHHREKQLVPGGKFALFLMFLLIVINVVSIYHEL"
          }
        },
        seq {
          id {
            local str "Query_4118"
          },
          descr {
```

```
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">CT818 3329287 HAAAP Tyrosine Transport
[Chlamydia
 trachomatis serovarD]"
                }
              }
            },
            title "CT818 3329287 HAAAP Tyrosine Transport [Chlamydia
 trachomatis serovarD]"
          },
          inst {
            repr raw,
            mol aa,
            length 397,
            seq-data ncbieaa
"MCMRNKCVGGILIVAGTVIGAGVLAVPILTAVEGFFPAVVLYVLAWLV
SLASGYGYLEVLTWCKGNRQANLCSMAEETLGRVGRIVLCLVYLFLFYSLLVAYFCDGGNILSRVIGESFFSYPWMR
H
VMPLLFFSLFAPLLMANTSVIDYCNRGFVFGLIFVFGLLCVLGVPRIQGELLLRASWFSSLNSLPIFFLAFGFQNVV
P
SLYHYLDGNIREVKRAILIGSLIPLILYIAWEALVLGTVPLVDLLKAKDLGWTAAGALQGSLKNSAFYIAGELFGFF
A
LVTSFIGTALALKDFYIDIFKWDARKKRVSLFFLVQVFPLVWAIFYPEIVLSCLRYAGGIGGACIIVLFPVAMLWNG
R
YGKRRCFGKRILPGGKTVLLILTGYTVLNLATLYYTF"
          }
        },
        seq {
          id {
            local str "Query_4119"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">CT817 6578114 HAAAP Tyrosine Transport
[Chlamydia
 trachomatis serovarD]"
                }
              }
            },
            title "CT817 6578114 HAAAP Tyrosine Transport [Chlamydia
 trachomatis serovarD]"
          },
          inst {
            repr raw,
            mol aa,
            length 398,
            seq-data ncbieaa
"MINKMLGGALLVAGTTIGAGVLAVPVSTSEGGFLPTTLLYIVSWFIAV
```

```
ASGYCFLEVLTWTHSRKNVNMVSMAEYTLGHKSKIIMWLVYLLLFYSLLVAYFCDGGNILMRVMGCRSWDTPWIRHA
M
PVVFFALFSPLLMAKTSIIDQCNRVFVFGLGIAFAMFCYFGFPLMKTDLLVRSAWGATLKGFPILFLAFGFQNVVPT
L
YHYMDKNVKDVKKAIVIGSSIPLVLYIIWEAIVLGAVPISFLEQAKVEGWTAIGALQTALKCSAFYVAGEFFGFFAL
I
SSFIGVSLGLKDFFIDAFQWDEKKRKVEIFFLVFVFPLVWAVFYPGIVLKCLECTGALGETIVLGVFPVLMVWKGRY
G
KKRYYGKRILPGGKGTLLVMSGLVLLNLVLIAQKFLGY"
          }
        },
        seq {
          id {
            local str "Query_4120"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">TyrP 16129857 HAAAP tyrosine-specific transport
 system [Escherichia coli K12-MG1655]"
                }
              }
            },
            title "TyrP 16129857 HAAAP tyrosine-specific transport system
 [Escherichia coli K12-MG1655]"
          },
          inst {
            repr raw,
            mol aa,
            length 403,
            seq-data ncbieaa
"MKNRTLGSVFIVAGTTIGAGMLAMPLAAAGVGFSVTLILLIGLWALMC
YTALLLLEVYQHVPADTGLGTLAKRYLGRYGQWLTGFSMMFLMYALTAAYISGAGELLASSISDWTGISMSATAGVL
L
FTFVAGGVVCVGTSLVDLFNRFLFSAKIIFLVVMLVLLLPHIHKVNLLTLPLQQGLALSAIPVIFTSFGFHGSVPSI
V
SYMDGNIRKLRWVFIIGSAIPLVAYIFWQVATLGSIDSTTFMGLLANHAGLNGLLQALREMVASPHVELAVHLFADL
A
LATSFLGVALGLFDYLADLFQRSNTVGGRLQTGAITFLPPLAFALFYPRGFVMALGYAGVALAVLALIIPSLLTWQS
R
KHNPQAGYRVKGGRPALVVVFLCGIAVIGVQFLIAAGLLPEVG"
          }
        },
        seq {
          id {
            local str "Query_4121"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">SdaC 16130703 HAAAP probable serine transporter
```

```
 [Escherichia coli K12-MG1655]"
                }
              }
            },
            title "SdaC 16130703 HAAAP probable serine transporter
 [Escherichia coli K12-MG1655]"
          },
          inst {
            repr raw,
            mol aa,
            length 429,
            seq-data ncbieaa
"METTQTSTIASKDSRSAWRKTDTMWMLGLYGTAIGAGVLFLPINAGVG
GMIPLIIMAILAFPMTFFAHRGLTRFVLSGKNPGEDITEVVEEHFGIGAGKLITLLYFFAIYPILLVYSVAITNTVE
S
FMSHQLGMTPPPRAILSLILIVGMMTIVRFGEQMIVKAMSILVFPFVGVLMLLALYLIPQWNGAALETLSLDTASAT
G
NGLWMTLWLAIPVMVFSFNHSPIISSFAVAKREEYGDMAEQKCSKILAFAHIMMVLTVMFFVFSCVLSLTPADLAAA
K
EQNISILSYLANHFNAPVIAWMAPIIAIIAITKSFLGHYLGAREGFNGMVIKSLRGKGKSIEINKLNRITALFMLVT
T
WIVATLNPSILGMIETLGGPIIAMILFLMPMYAIQKVPAMRKYSGHISNVFVVVMGLIAISAIFYSLFS"
          }
        },
        seq {
          id {
            local str "Query_4122"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">b2845 16130749 HAAAP putative transporter
protein
 [Escherichia coli K12-MG1655]"
                }
              }
            },
            title "b2845 16130749 HAAAP putative transporter protein
 [Escherichia coli K12-MG1655]"
          },
          inst {
            repr raw,
            mol aa,
            length 409,
            seq-data ncbieaa
"MSNIWSKEETLWSFALYGTAVGAGTLFLPIQLGSAGAVVLFITALVAW
PLTYWPHKALCQFILSSKTSAGEGITGAVTHYYGKKIGNLITTLYFIAFFVVVLIYAVAITNSLTEQLAKHMVIDLR
I
RMLVSLGVVLILNLIFLMGRHATIRVMGFLVFPLIAYFLFLSIYLVGSWQPDLLTTQVEFNQNTLHQIWISIPVMVF
A
FSHTPIISTFAIDRREKYGEHAMDKCKKIMKVAYLIICISVLFFVFSCLLSIPPSYIEAAKEEGVTILSALSMLPNA
P
AWLSISGIIVAVVAMSKSFLGTYFGVIEGATEVVKTTLQQVGVKKSRAFNRALSIMLVSLITFIVCCINPNAISMIY
A
```

```
ISGPLIAMILFIMPTLSTYLIPALKPWRSIGNLITLIVGILCVSVMFFS"
          }
        },
        seq {
          id {
            local str "Query_4123"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">YhaO 16131003 HAAAP putative transport system
 permease protein [Escherichia coli K12-MG1655]"
                }
              }
            },
            title "YhaO 16131003 HAAAP putative transport system permease
 protein [Escherichia coli K12-MG1655]"
          },
          inst {
            repr raw,
            mol aa,
            length 425,
            seq-data ncbieaa
"MPGMSESEWREAIKFDSTDTGWVIMSIGMAIGAGIVFLPVQVGLMGLW
VFLLSSVIGYPAMYLFQRLFINTLAESPECKDYPSVISGYLGKNWGILLGALYFVMLVIWMFVYSTAITNDSASYLH
T
FGVTEGLLSDSPFYGLVLICILVAISSRGEKLLFKISTGMVLTKLLVVAALGVSMVGMWHLYNVGSLPPLGLLVKNA
I
ITLPFTLTSILFIQTLSPMVISYRSREKSIEVARHKALRAMNIAFGILFVTVFFYAVSFTLAMGHDEAVKAYEQNIS
A
LAIAAQFISGDGAAWVKVVSVILNIFAVMTAFFGVYLGFREATQGIVMNILRRKMPAEKINENLVQRGIMIFAILLA
W
SAIVLNAPVLSFTSICSPIFGMVGCLIPAWLVYKVPALHKYKGMSLYLIIVTGLLLCVSPFLAFS"
          }
        },
        seq {
          id {
            local str "Query_4124"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">TdcC 16131009 HAAAP anaerobically inducible
 L-threonine, L-serine permease [Escherichia coli K12-MG1655]"
                }
              }
            },
            title "TdcC 16131009 HAAAP anaerobically inducible L-threonine,
 L-serine permease [Escherichia coli K12-MG1655]"
          },
          inst {
```

```
            repr raw,
            mol aa,
            length 443,
            seq-data ncbieaa
"MSTSDSIVSSQTKQSSWRKSDTTWTLGLFGTAIGAGVLFFPIRAGFGG
LIPILLMLVLAYPIAFYCHRALARLCLSGSNPSGNITETVEEHFGKTGGVVITFLYFFAICPLLWIYGVTITNTFMT
F
WENQLGFAPLNRGFVALFLLLLMAFVIWFGKDLMVKVMSYLVWPFIASLVLISLSLIPYWNSAVIDQVDLGSLSLTG
H
DGILITVWLGISIMVFSFNFSPIVSSFVVSKREEYEKDFGRDFTERKCSQIISRASMLMVAVVMFFAFSCLFTLSPA
N
MAEAKAQNIPVLSYLANHFASMTGTKTTFAITLEYAASIIALVAIFKSFFGHYLGTLEGLNGLVLKFGYKGDKTKVS
L
GKLNTISMIFIMGSTWVVAYANPNILDLIEAMGAPIIASLLCLLPMYAIRKAPSLAKYRGRLDNVFVTVIGLLTILN
I
VYKLF"
          }
        },
        seq {
          id {
            local str "Query_4125"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">Mtr 16131053 HAAAP tryptophan-specific transport
 protein [Escherichia coli K12-MG1655]"
                }
              }
            },
            title "Mtr 16131053 HAAAP tryptophan-specific transport protein
 [Escherichia coli K12-MG1655]"
          },
          inst {
            repr raw,
            mol aa,
            length 414,
            seq-data ncbieaa
"MATLTTTQTSPSLLGGVVIIGGTIIGAGMFSLPVVMSGAWFFWSMAAL
IFTWFCMLHSGLMILEANLNYRIGSSFDTITKDLLGKGWNVVNGISIAFVLYILTYAYISASGSILHHTFAEMSLNV
P
ARAAGFGFALLVAFVVWLSTKAVSRMTAIVLGAKVITFFLTFGSLLGHVQPATLFNVAESNASYAPYLLMTLPFCLA
S
FGYHGNVPSLMKYYGKDPKTIVKCLVYGTLMALALYTIWLLATMGNIPRPEFIGIAEKGGNIDVLVQALSGVLNSRS
L
DLLLVVFSNFAVASSFLGVTLGLFDYLADLFGFDDSAVGRLKTALLTFAPPVVGGLLFPNGFLYAIGYAGLAATIWA
A
IVPALLARASRKRFGSPKFRVWGGKPMIALILVFGVGNALVHILSSFNLLPVYQ"
          }
        },
        seq {
          id {
            local str "Query_4126"
          },
```

```
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">YhjV 16131411 HAAAP putative transporter protein
 [Escherichia coli K12-MG1655]"
                }
              }
            },
            title "YhjV 16131411 HAAAP putative transporter protein
 [Escherichia coli K12-MG1655]"
          },
          inst {
            repr raw,
            mol aa,
            length 423,
            seq-data ncbieaa
"MQHNTLSKHNQKLPFTRYDFGWVLLCIGMAIGAGTVLMPVQIGLKGIW
VFITAAIIAYPATWVVQDIYLKTLSESDSCNDYTDIISHYLGKNWGIFLGVIYFLMIIHGIFIYSLSVVFDSASYLK
T
FGLTDADLSQSLLYKVAIFAVLVAIASGGERLLFKISGPMVVVKVGIIVVFGFAMIPHWNFANITAFPQASVFFRDV
L
LTIPFCFFSAVFIQVLNPMNIAYRKREADKVLATRLALRTHRISYITLIAVILFFAFSFTFSISHEEAVSAFEQNIS
A
LALAAQVIPGHIIHITSTVLNIFAVLTAFFGIYLGFHEAIKGIILNLLSRIIDTKKINSRVLTLAICAFIVITLTIW
V
SFRVSVLVFFQLGSPLYGIVSCLIPFFLIYKVAQLEKLRGFKAWLILLYGILLCLSPLLKLIE"
          }
        },
        seq {
          id {
            local str "Query_4127"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">TnaB 16131577 HAAAP low affinity tryptophan
 permease [Escherichia coli K12-MG1655]"
                }
              }
            },
            title "TnaB 16131577 HAAAP low affinity tryptophan permease
 [Escherichia coli K12-MG1655]"
          },
          inst {
            repr raw,
            mol aa,
            length 415,
            seq-data ncbieaa
"MTDQAEKKHSAFWGVMVIAGTVIGGGMFALPVDLAGAWFFWGAFILII
AWFSMLHSGLLLLEANLNYPVGSSFNTITKDLIGNTWNIISGITVAFVLYILTYAYISANGAIISETISMNLGYHAN
P
```

```
RIVGICTAIFVASVLWLSSLAASRITSLFLGLKIISFVIVFGSFFFQVDYSILRDATSSTAGTSYFPYIFMALPVCL
A
SFGFHGNIPSLIICYGKRKDKLIKSVVFGSLLALVIYLFWLYCTMGNIPRESFKAIISSGGNVDSLVKSFLGTKQHG
I
IEFCLLVFSNLAVASSFFGVTLGLFDYLADLFKIDNSHGGRFKTVLLTFLPPALLYLIFPNGFIYGIGGAGLCATIW
A
VIIPAVLAIKARKKFPNQMFTVWGGNLIPAIVILFGITVILCWFGNVFNVLPKFG"
          }
        },
        seq {
          id {
            local str "Query_4128"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">Z2963 15802316 HAAAP tyrosine-specific transport
 system [Escherichia coli O157:H7 EDL933]"
                }
              }
            },
            title "Z2963 15802316 HAAAP tyrosine-specific transport system
 [Escherichia coli O157:H7 EDL933]"
          },
          inst {
            repr raw,
            mol aa,
            length 403,
            seq-data ncbieaa
"MKNRTLGSVFIVAGTTIGAGMLAMPLAAAGVGFSVTLILLIGLWALMC
YTALLLLEVYQHVPADTGLGTLAKRYLGRYGQWLTGFSMMFLMYALTAAYISGAGELLASSISDWTGISMSATAGVL
L
FTFVAGGVVCVGTSLVDLFNRFLFSAKIIFLVVMLVLLLPHIHKVNLLTLPLQQGLALSAIPVIFTSFGFHGSVPSI
V
SYMDGNIRKLRWVFITGSAIPLVAYIFWQVATLGSIDSTTFMGLLANHAGLNGLLQALREMVASPHVELAVHLFADL
A
LATSFLGVALGLFDYLADLFQRSNTVGGRLQTGAITFLPPLAFALFYPRGFVMALGYAGVALAVLALIIPSLLTWQS
R
KHNPQAGYRVKGGRPALVVVFLCGIAVIGVQFLIAAGLLPEVG"
          }
        },
        seq {
          id {
            local str "Query_4129"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">Z4113 15803318 HAAAP probable serine transporter
 [Escherichia coli O157:H7 EDL933]"
                }
```

```
            }
          },
          title "Z4113 15803318 HAAAP probable serine transporter
 [Escherichia coli O157:H7 EDL933]"
        },
        inst {
          repr raw,
          mol aa,
          length 429,
          seq-data ncbieaa
"METTQTSTIASKDSRSAWRKTDTMWMLGLYGTAIGAGVLFLPINAGVG
GMIPLIIMAILAFPMTFFAHRGLTRFVLSGKNPGEDITEVVEEHFGIGAGKLITLLYFFAIYPILLVYSVAITNTVE
S
FMSHQLGMTPPPRAILSLILIVGMMTIVRFGEQMIVKAMSILVFPFVGVLMLLALYLIPQWNGAALETLSLDTASAT
G
NGLWMTLWLAIPVMVFSFNHSPIISSFAVAKREEYGDMAEQKCSKILAFAHIMMVLTVMFFVFSCVLSLTPADLAAA
K
EQNISILSYLANHFNAPVIAWMAPIIAIIAITKSFLGHYLGAREGFNGMVIKSLRGKGKSIEINKLNRITALFMLVT
T
WIVATLNPSILGMIETLGGPIIAMILFLMPMYAIQKVPAMRKYSGHISNVFVVVMGLIAISAIFYSLFS"
        }
      },
      seq {
        id {
          local str "Query_4130"
        },
        descr {
          user {
            type str "CFastaReader",
            data {
              {
                label str "DefLine",
                data str ">Z4165 15803365 HAAAP putative transporter
protein
 [Escherichia coli O157:H7 EDL933]"
              }
            }
          },
          title "Z4165 15803365 HAAAP putative transporter protein
 [Escherichia coli O157:H7 EDL933]"
        },
        inst {
          repr raw,
          mol aa,
          length 409,
          seq-data ncbieaa
"MSNIWSKEETLWSFALYGTAVGAGTLFLPIQLGSAGAVVLFITALVAW
PLTYWPHKALCQFILSSKTSAGEGITGAVTHYYGKKIGNLITTLYFIAFFVVVLIYAVAITNSLTEQLAKHMVIDLR
I
RMLVSLGVVLILNLIFLMGRHATIRVMGFLVFPLIAYFLFLSIYLVGSWQPDLLTTQVEFNQNTLHQIWISIPVMVF
A
FSHTPIISTFAIDRREKYGEHAMDKCKKIMKVAYLIICISVLFFVFSCLLSIPPSYIEAAKEEGVTILSALSMLPNA
P
AWLSISGIIVAVVAMSKSFLGTYFGVIEGATEVVKTTLQQVGVKKSRAFNRALSIMLVSLITFIVCCINPNAISMIY
A
ISGPLIAMILFIMPTLSTYLIPALKPWRSIGNLITLIVGILCVSVMFFS"
        }
```

```
        },
        seq {
          id {
            local str "Query_4131"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">Z4463 15803650 HAAAP putative transport system
 permease protein [Escherichia coli O157:H7 EDL933]"
                }
              }
            },
            title "Z4463 15803650 HAAAP putative transport system permease
 protein [Escherichia coli O157:H7 EDL933]"
          },
          inst {
            repr raw,
            mol aa,
            length 443,
            seq-data ncbieaa
"MEIASNKGVIADASTPAGRAGMSESEWREAIKFDSTDTGWVIMSIGMA
IGAGIVFLPVQVGLMGLWVFLLSSVIGYPAMYLFQRLFINTLAESPECKDYPSVISGYLGKNWGILLGALYFVMLVI
W
MFVYSTAITNDSASYLHTFGVTEGLLSDSPFYGLVLICILVAISSRGEKLLFKISTGMVLTKLLVVAALGVSMVGMW
H
LYNVGSLPPLGLLVKNAIITLPFTLTSILFIQTLSPMVISYRSREKSIEVARHKALRAMNIAFGILFIIVFFYAVSF
T
LAMGHDEAVKAYEQNISALAIAAQFISGDGAAWVKVVSVILNIFAVMTAFFGVYLGFREATQGIVMNILRRKMPAEK
I
NENLVQRGIMIFAILLAWSAIVLNAPVLSFTSICSPIFGLVGCLIPAWLVYKVPALHKYKGMSLYLIIVTGLLLCVS
P
FLAFS"
          }
        },
        seq {
          id {
            local str "Query_4132"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">Z4468 15803655 HAAAP anaerobically inducible
 L-threonine, L-serine permease [Escherichia coli O157:H7 EDL933]"
                }
              }
            },
            title "Z4468 15803655 HAAAP anaerobically inducible L-threonine,
 L-serine permease [Escherichia coli O157:H7 EDL933]"
          },
          inst {
```

```
            repr raw,
            mol aa,
            length 443,
            seq-data ncbieaa
"MSTSDSIVSSQTKQSSWRKSDTTWTLGLFGTAIGAGVLFFPIRAGFGG
LIPILLMLVLAYPIAFYCHRALARLCLSGSNPSGNITETVEEHFGKTGGVVITFLYFFAICPLLWIYGVTITNTFMT
F
WENQLGFAPLNRGFVALFLLLLMAFVIWFGKDLMVKVMSYLVWPFIASLVLISLSLIPYWNSAVIDQVDLGSLSLTG
H
DGILITVWLGISIMVFSFNFSPIVSSFVVSKREEYEKDFGRDFTERKCSQIISRASMLMVAVVMFFAFSCLFTLSPA
N
MAEAKAQNIPVLSYLANHFASMTGTKTTFAITLEYAASIIALVAIFKSFFGHYLGTLEGLNGLVLKFGYKGDKTKVS
L
GKLNTISMIFIMGSTWVVAYANPNILDLIEAMGAPIIASLLCLLPMYAIRKAPSLAKYRGRLDNVFVTVIGLLTILN
I
VYKLF"
          }
        },
        seq {
          id {
            local str "Query_4133"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">Z4522 15803703 HAAAP tryptophan-specific
 transport protein [Escherichia coli O157:H7 EDL933]"
                }
              }
            },
            title "Z4522 15803703 HAAAP tryptophan-specific transport protein
 [Escherichia coli O157:H7 EDL933]"
          },
          inst {
            repr raw,
            mol aa,
            length 414,
            seq-data ncbieaa
"MATLTTTQTSPSLLGGVVIIGGTIIGAGMFSLPVVMSGAWFFWSMAAL
IFTWFCMLHSGLMILEANLNYRIGSSFDTITKDLLGKGWNVVNGISIAFVLYILTYAYISASGSILHHTFAEMSLNV
P
ARAAGFGFALLVAFVVWLSTKAVSRMTAIVLGAKVITFFLTFGSLLGHVQPATLFNVAESNASYAPYLLMTLPFCLA
S
FGYHGNVPSLMKYYGKDPKTIVKCLVYGTLMALALYTIWLLATMGNIPRPEFIDIAEKGGNIDVLVQALSGVLNSRS
L
DLLLVVFSNFAVASSFLGVTLGLFDYLADLFGFDDSAVGRLKTALLTFAPPVVGGLLFPNGFLYAIGYAGLAATIWA
A
IVPALLARASRKRFGSPKFRVWGGKPMITLILVFGVGNALVHILSSFNLLPVYQ"
          }
        },
        seq {
          id {
            local str "Query_4134"
          },
```

```
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">Z4956 15804082 HAAAP putative transporter
protein
 [Escherichia coli O157:H7 EDL933]"
                }
              }
            },
            title "Z4956 15804082 HAAAP putative transporter protein
 [Escherichia coli O157:H7 EDL933]"
          },
          inst {
            repr raw,
            mol aa,
            length 423,
            seq-data ncbieaa
"MQHNTLPKHDQKLPFTRYDFGWVLLCIGMAIGAGTVLMPVQIGLKGIW
VFITAAIIAYPATWVVQDIYLKTLSESDSCNDYTDIISHYLGKNWGXFLGVIYFLMIIHGIFIYSLSVVFDSASYLK
T
FGLTDADLSQSLLYKVAIFAVLVAIASGGERLLFKISGPMVVVKVGIIVVFGFTMIPHWNFANITAFPQASVFFRDV
L
LTIPFCFFSAVFIQVLNPMNIAYRKREADKVLATRLALRTHRISYITLIAVILFFAFSFTFSISHEEAVSAFEQNIS
A
LALAAQVIPGHIIHITSTVLNIFAVLTAFFGIYLGFHEAIKGIILNLLSRIIDTKKINSRVLTLAICAFIVITLTIW
V
SFRVSVLVFFQLGSPLYGIVSCLIPFFLIYKVAQLEKLRGFKTWLILLYGILLCLSPLLKLIE"
          }
        },
        seq {
          id {
            local str "Query_4135"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">Z5204 15804306 HAAAP low affinity tryptophan
 permease [Escherichia coli O157:H7 EDL933]"
                }
              }
            },
            title "Z5204 15804306 HAAAP low affinity tryptophan permease
 [Escherichia coli O157:H7 EDL933]"
          },
          inst {
            repr raw,
            mol aa,
            length 415,
            seq-data ncbieaa
"MTDQAEKKHSAFWGVMVIAGTVIGGGMFALPVDLAGAWFFWGAFILII
```

```
AWFSMLHSGLLLLEANLNYPVGSSFNTITKDLIGNTWNIISGITVAFVLYILTYAYISANGAIISETISMNLGYHAN
P
RIVGICTAIFVASVLWISSLAASRITSLFLGLKIISFVIVFGSFFFLVDYSILRDATSSTAGTSYFPYIFMALPVCL
A
SFGFHGNIPSLIICYGKRKDKLIKSVVFGSLLALVIYLFWLYCTMGNIPRESFKAIISSGGNVDSLVKSFLGTKQHG
I
IEFCLLVFSNLAVASSFFGVTLGLFDYLADLFKIDNSHGGRFKTVLLTFLPPALLYLIFPNGFIYGIGGAGLCATIW
A
VIIPAVLAIKARKKFPNQMFTVWGGNLIPAIVILFGITVILCWFGNVFNVLPKFG"
          }
        },
        seq {
          id {
            local str "Query_4136"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">HD0648 33151831 HAAAP tryptophan-specific
 transport protein [Haemophilus ducreyi 35000HP]"
                }
              }
            },
            title "HD0648 33151831 HAAAP tryptophan-specific transport
protein
 [Haemophilus ducreyi 35000HP]"
          },
          inst {
            repr raw,
            mol aa,
            length 399,
            seq-data ncbieaa
"MQKQPSIFGGACIIAGVCVGAGMLGLPTSGAGAWTIWSIISLTLTMIV
MTLSGWLLLNVYQNYDARASFNTVTKDLLGNNININIVNNLAVYFVGGILLYAYTTASGGILSNLTNSTFELGEYSNRI
W
STIFVFIFSFFVWHSTRLVDRISVLLILFMAFSFAFSISGLVINIDLPTLFDQQNQQGEYAKYALAMFPIALTSFGY
H
HSVSSMRAYYGEEQKAKYAIASGTFIALILYLLWIISILGNLPRQQFIPIIASNGDLEILLNTVGKVVESPYVRQAI
N
AFSMAAILSSFIGVGLGTFDFLADFFKFDNSKMGRMKSWAATFLPPLVFSLIAPLGFLKAIGYAGAVATLWTCIIPA
L
LAYKARKGHLLVITLVLLFGVCTAIFHFLAMFEILPVFN"
          }
        },
        seq {
          id {
            local str "Query_4137"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
```

```
                  data str ">HD1143 33152262 HAAAP serine transporter
 [Haemophilus ducreyi 35000HP]"
                }
              }
            },
            title "HD1143 33152262 HAAAP serine transporter [Haemophilus
 ducreyi 35000HP]"
          },
          inst {
            repr raw,
            mol aa,
            length 408,
            seq-data ncbieaa
"MVRLSNRFSKTWVLNLFGTAVGAGVLFLPINAGMSGFYPLIIMTLIVG
PMTYLAHRGLTRFVLSSQKPGSDITNVVREHFGEQAGKFITLLYFFAIFPILLIYGVGITNTVSSFIENQLHMTSPP
R
VLLSGALIAVLISVMLLNEQAMLRITTYLVYPLVLILFGLSIYLIPEWNSAALEQMPSTGDFITTLWLTIPVLVFAF
N
HSPAISSFALSQQKHHQDPEKNDVESGKVLRSTATILVLFVMFFVFSCVLTLTPVELAEAKVQNISILSYLANKFDN
P
IISYLGPFIAFLAIGSSFFGHYLGAREGLEGLVNQLRDKPIASHKIHKITAVVFFLILWIVATINPSILGFIESLGG
P
IIAMILFIMPVYAVYTVPALAKYKGQFSNIFVAIMGSIAISAILYGLL"
          }
        },
        seq {
          id {
            local str "Query_4138"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">HD1499 33152552 HAAAP tyrosine-specific
transport
 protein [Haemophilus ducreyi 35000HP]"
                }
              }
            },
            title "HD1499 33152552 HAAAP tyrosine-specific transport protein
 [Haemophilus ducreyi 35000HP]"
          },
          inst {
            repr raw,
            mol aa,
            length 405,
            seq-data ncbieaa
"MKNKILGSALMIAGTTIGAGMLAMPLTSAGMGFGATLFLLISLWGLLA
YTGLLFMEVYQTAPKRDVGVASLAEQYFGIVGRVLATATLLVLLYALLAAYITGGGSLLSGILPEIMDSQLTNKIAI
L
LFTTLLGTFVVVGIKGVDGLTRVLFMGKIIAFLLVLAMMLPKAKIENLFAMPLDNLLVISVVPIFFTSFGFHVIMAS
I
NNYLEADIGKIRIAIYIGTAIPLIAYLLWQLATHGVLTQAEFVEILKSDPSLNGLVKATSQITGSTFLGEMLRLFSA
L
```

```
ALITSFLGVAMGIFECIGDLLKRINLPTNRLVLTLVTFIPPLLFALFYPNGFIAALGYAGLLFAFYGMLLPIGLAWR
A
RKLHPNLPYRVSGGNSALLIALLMGIIIMIIPFLIQFGYLPQVMG"
          }
        },
        seq {
          id {
            local str "Query_4139"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">HD1875 33152864 HAAAP tyrosine-specific
transport
 protein [Haemophilus ducreyi 35000HP]"
                }
              }
            },
            title "HD1875 33152864 HAAAP tyrosine-specific transport protein
 [Haemophilus ducreyi 35000HP]"
          },
          inst {
            repr raw,
            mol aa,
            length 408,
            seq-data ncbieaa
"MIMNKTLGSTLLVSGTMIGGGMLAMPLTSAGIGFCLTLVLLVVLWILL
TYSALLFVEVYQTAEHDVGIGTLAARYFGKTGRVIATLVLMVFLYALLSAYVTGGGNILASMFPAAPDHLMTEKEKI
A
IVMFTIIFGAFIVFGTRSVDGINRFLFFIMLAGLALVLFALMPFIQIDNLLAMPVDNLLLVSASPIFFTAFGFHGSI
P
SLNNYLNGNVKALRFAIIVGSLITLMVYILWQMATHGLLSQTAFLDILQKDPTLNGLVVATAQTTGSNIIGYIVKFF
S
AFALITSFLGVALGLFECINDLLKQSFNFASNRLNVGLLTFSPPLLFALFYPEGFVLALRYAGQMFAFYAVVLPVLL
V
WKVRQLHPNLPYRVWGGNVLLWLVALLGIVITCIPFMVKTGCLPTVVG"
          }
        },
        seq {
          id {
            local str "Query_4140"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">HH1100 32262650 HAAAP serine transporter
 [Helicobacter hepaticus ATCC51449]"
                }
              }
            },
            title "HH1100 32262650 HAAAP serine transporter [Helicobacter
```

```
 hepaticus ATCC51449]"
          },
          inst {
            repr raw,
            mol aa,
            length 459,
            seq-data ncbieaa
"MNKWNNFDTSWIISLFGTAVGAGILYLPIKAGGGGIWPVIAMCFIIFP
MVYLSHRALSRFVCQANGNDKDITYAAEEYFGRNVSIFISILYFFAIFPICLAYCVGITNTFESFIYHQILPLVNEN
H
PAVAPLGTLAGFIQDIYYDPYSFLHQSESGALLKLKSENYATLHPFWRAGLVFILVSAFMLIMLFSEKLIIKVCQWL
V
YPLCAILFIFSLYLIPQWNLESITFVPDIEDFLIIVWLTLPVLVFSFNHSPAISTFALSAKREYGADAIIKSDSILL
R
TSIMLLVFVMFFVISCVLSLTPQELIEARAQNIPVLSYFANKLDNPLISYGGPLIAFLAISSSFFGHYFGAREGAYG
I
VRKCCKIAGNKEPNLKLIAAICTFVMYVIMLITAYINPSVLGFIEDLGGPIIAAILFLMPIIAIYSVSKMKQFKNPA
L
DAFVFITGLLTIFTITYKLIL"
          }
        },
        seq {
          id {
            local str "Query_4141"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">HI0287 16272242 HAAAP tryptophan-specific
 transport protein (mtr) [Haemophilus influenzae KW20]"
                }
              }
            },
            title "HI0287 16272242 HAAAP tryptophan-specific transport
protein
 (mtr) [Haemophilus influenzae KW20]"
          },
          inst {
            repr raw,
            mol aa,
            length 418,
            seq-data ncbieaa
"MIQQKSPSLLGGAMIIAGTAIGAGMLANPTSTAGVWFIGSILALIYTW
FCMTTSGLMILEANLHYPTGSSFDTIVKDLLGKSWNTINGLSVAFVLYILTYAYITSGGGITQNLLNQAFSSAESAV
D
IGRTSGSLIFCLILAAFVWLSTKAVDRFTTVLIVGMVVAFFLSTTGLLSSVKTAVLFNTVAESEQTYLPYLLTALPV
C
LVSFGFHGNVPSLVKYYDRDGRRVMKSIFIGTGLALVIYILWQLAVQGNLPRTEFAPVIEKGGDVSALLEALHKYIE
V
EYLSVALNFFAYMAISTSFLGVTLGLFDYIADLFKFDDSLLGRTKTTLVTFLPPLLLSLQFPYGFVIAIGYAGLAAT
I
WAAIVPALLAKASRQKFPQASYKVYGGNFMIGFVILFGILNIVAQIGANLGWFASFTG"
          }
        },
```

```
        seq {
          id {
            local str "Query_4142"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">HI0289 16272244 HAAAP serine transporter (sdaC)
 [Haemophilus influenzae KW20]"
                }
              }
            },
            title "HI0289 16272244 HAAAP serine transporter (sdaC)
 [Haemophilus influenzae KW20]"
          },
          inst {
            repr raw,
            mol aa,
            length 412,
            seq-data ncbieaa
"MKSTEKLKWNKFDATWMLNLFGTAVGAGVLFLPINAGMGGFWPLVLMA
IIVGPMTYFAHRGLAYFVLSSKNPGSDITEVVEEHFGKTAGKLITLLYFFAIFPILLIYGNGITNTVDSFIVNQLGM
A
SPNRVILSFVLIAVLISVMLFNEKVMLKITEWLVYPLVLILFVLSIYLIPEWNSAVLYEFPTVGGFLTTLWLTIPVL
V
FSFNHSPAISSFTCSQFREYKTFDGAERHISHTEKGASTILLFFVMFFVFSCVLTLTPEELVAAKEQNISILSFLAN
K
FDNPYISYFGPLVAFLAITSSFFGHYMGAREGLEGLYLKMKGEAVNRKKLNYATALFFLLTLWGVAIINPSILGLIE
S
LGGPIIAMILFIMPMYAIRKIPAMKRYSGRFSNVFVTIMGLIAISAVVYGLL"
          }
        },
        seq {
          id {
            local str "Query_4143"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">HI0477 16272424 HAAAP tyrosine-specific
transport
 protein (tyrP) [Haemophilus influenzae KW20]"
                }
              }
            },
            title "HI0477 16272424 HAAAP tyrosine-specific transport protein
 (tyrP) [Haemophilus influenzae KW20]"
          },
          inst {
            repr raw,
            mol aa,
```

```
            length 400,
            seq-data ncbieaa
"MNKTVGSTLLVAGTMIGAGMLAMPLTSAGIGFGFTLVLLLGLWALLTF
SALLFVELYQTAESDAGIGTLAEQYFGKTGRIIATAVLIIFLYALIAAYISGGGSLLKDLLPESFGDKVSVLLFTVI
F
GSFIVIGTHSVDKINRVLFFVMLAAFAVVLSLMLPEIKFDNLMATPIDKALIISASPVFFTAFGFHGSIPSLNKYLD
G
NVKALRFSILVGSAITLCAYILWQLSTHGLLTQNEFLQILKEDATLNGLVKATFAITGSNVIASAVKLFSTLALITS
F
LGVGLGLLECIEDLLKRSFNVTAGRISLGLLTFIPPLVFALFYPEGFILALGYAGQMFAFYAVVLPVSLVWKARRAH
A
NLPYKVWGGNLTLIIVLVLGVLITSIPFAIRAGYLPFVVG"
          }
        },
        seq {
          id {
            local str "Query_4144"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">HI0528 16272472 HAAAP tyrosine-specific
transport
 protein (tyrP) [Haemophilus influenzae KW20]"
                }
              }
            },
            title "HI0528 16272472 HAAAP tyrosine-specific transport protein
 (tyrP) [Haemophilus influenzae KW20]"
          },
          inst {
            repr raw,
            mol aa,
            length 406,
            seq-data ncbieaa
"MLKNKTFGSALIIAGTTIGAGMLAMPLTSAGMGFGYTLLLLVGLWALL
VYSGLLFVEVYQTADQLDDGVATLAEKYFGVPGRIFATLSLLVLLYALSAAYITGGGSLLSGLPTAFGMEAMSLKTA
I
IIFTVVLGSFVVVGTKGVDGLTRVLFIGKLIAFAFVLFMMLPKVATDNLMALPLDYAFVVSAAPIFLTSFGFHVIMA
S
VNSYLGGSVDKFRRAILIGTAIPLAAYLVWQLATHGVLSQSEFVRILQADPTLNGLVNATREITGSHFMGEVVRVFS
S
LALITSFLGVMLGVFEGLGDLFKRYHLPNNRFVLTIAAFLPPLVFALFYPEGFITALSYAGLLCAFYCLILPISLAW
R
TRIENPTLPYRVAGGNFALVLALLIGVVIMLIPFLIQWGYLPVVAG"
          }
        },
        seq {
          id {
            local str "Query_4145"
          },
          descr {
            user {
              type str "CFastaReader",
```

```
              data {
                {
                  label str "DefLine",
                  data str ">HP0133 15644763 HAAAP serine transporter (sdaC)
 [Helicobacter pylori 26695]"
                }
              }
            },
            title "HP0133 15644763 HAAAP serine transporter (sdaC)
 [Helicobacter pylori 26695]"
          },
          inst {
            repr raw,
            mol aa,
            length 413,
            seq-data ncbieaa
"MAQEKAVPRDPKKLNAFDLRWMVSLFGTAVGAGILFLPIRAGGHGVWA
IVVMSAIIFPLTYLGHRALAYFIGSKDKEDITMVVRSHFGAQWGFLITLLYFLAIYPICLVYGVGITNVFDHFFTNQ
L
HLAPFHRGLLAVALVSLMMLVMVFNATIVTRICNALVYPLCLILLLFSLYLIPYWQGANLFVVPSFKEFVLAIWLTL
P
VLVFAFDHSPIISTFTQNVGKEYGVFKEYKLNQIELGTSLMLLGFVMFFVFSCVMCLNADDFVKAREQNIPILSYLA
N
TLNNPLINYAGPVVAFLAIFSSFFGHYYGAKEGLEGIIIQSLKLKKASKPLSVSVTIFLWLTITLVAYINPNILDFI
E
NLGGPIIALILFVMPMIAFYSVSSLKRFRNFKVDIFVFVFGSLTALSVFLGLF"
          }
        },
        seq {
          id {
            local str "Query_4146"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">lp_0502 28377393 HAAAP serine transporter
 [Lactobacillus plantarum WCFS1]"
                }
              }
            },
            title "lp_0502 28377393 HAAAP serine transporter [Lactobacillus
 plantarum WCFS1]"
          },
          inst {
            repr raw,
            mol aa,
            length 426,
            seq-data ncbieaa
"MTNIFKGWRKNDTFWMLSLFGTAIGAGVLFLPIGVGTAGILGIIMILI
LALPTTFFAHRGLSRFVLSAKNDGDDITDVVEQHFGFKIGLLFTIIYFFSIYPILLVYSVSITNTVSKFITDQMHMQ
T
PPRWLLSLVLVGILIGIARFGTSLITKVMSGLAFPFIIVIVLFSFYLIPHWNGAILSTFSSSVSGGHLVGTLGNLWL
L
```

```
IPVMIFSFNHSPIISSFSVAERKEYSGEGKDKVDKKISTILLSAETLMVIVAMFFVISCTLALTPDQILEAKNENVS
I
LDYVATAFNNPIIKYVSPVIAFIAIAKSFLGHYLGTSEGLRGIIRKMEEKSNKTLSSRTVSTIVDLVLLLSAWIVAW
V
NPSIMGMIETIIGPIIAFILFLMPMYAIHKSPKLQQYAGKHSNVFVVVIGLIAVSGILFNIIKLFI"
          }
        },
        seq {
          id {
            local str "Query_4147"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">NMB2031 15677855 HAAAP tryptophan transporter
 [Neisseria meningitidis MC58]"
                }
              }
            },
            title "NMB2031 15677855 HAAAP tryptophan transporter [Neisseria
 meningitidis MC58]"
          },
          inst {
            repr raw,
            mol aa,
            length 413,
            seq-data ncbieaa
"MPNKTPSLFGGAMIIAGTVIGAGMLANPTATSGVWFTGSLAVLLYTWF
SMLSSGLMILEVNTHYPHGASFDTMVKDLLGRGWNIINGIAVAFVLYLLTYAYIFVGGDLTAKGLGSAAGGDVSLTV
G
QLVFFGILAFCVWASARLVDRFTGVLIGGMVLTFIWAAGGLIADAKPSVLFDTQAPAGTNYWIYAATALPVCLASFG
F
HGNVSSLLKYFKGDAPKVAKSIWTGTLIALVIYVLWQTAIQGNLPRNEFAPVIAAEGQVSVLIETLSKFAQTGNMDK
I
LSLFSYMAIATSFLGVTLGLFDYIADIFKWNDSISGRTKTAALTFLPPLISCLLFPTGFVTAIGYVGLAATVWTGII
P
AMLLYRSRKKFGAGKTYKVYGGLWLMVWVFLFGIVNIAAQVLSQMELVPVFKG"
          }
        },
        seq {
          id {
            local str "Query_4148"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">PA1916 15597112 HAAAP probable amino acid
 permease [Pseudomonas aeruginosa PAO1]"
                }
              }
            },
```

```
            title "PA1916 15597112 HAAAP probable amino acid permease
 [Pseudomonas aeruginosa PAO1]"
          },
          inst {
            repr raw,
            mol aa,
            length 410,
            seq-data ncbieaa
"MGMREANDNASGLKAHPLGSLETIAIIIGTNIGAGVLSMAYAARKVGY
VPLLVCLALTCLFCIVTMLYVTEACLRTRGNNQLSGLSRRYLGPLGGWLIFIAVAANSYGALVAYMTGSGNIMLEFF
G
QYGLTRQVGSLIFFVPSALVLYLGLKALGVGEKLISAGMVAIVCILIGATLMHDDARLAHLWQSQWQYVVPVFNLAV
F
VFGAQFLVPELVRGNLATPGRVPRLIVVGMLLTFLLVAAIPASVIALVGPDNLSEVATLSWGRSLGQWAYYVANTFA
L
LAMLTSYWGLGGCLFTNIFDHFRLGSETNRRKRLAVLAAVSVPPFLLAYAGLGGFVNALYFAGTFGGVLMGIIPILL
L
NAARRHGDQLPAFSCGWYAHPLIQGLIVLVFASSGVYAIASALGMLPASW"
          }
        },
        seq {
          id {
            local str "Query_4149"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">PA3766 15598961 HAAAP probable aromatic amino
 acid transporter [Pseudomonas aeruginosa PAO1]"
                }
              }
            },
            title "PA3766 15598961 HAAAP probable aromatic amino acid
 transporter [Pseudomonas aeruginosa PAO1]"
          },
          inst {
            repr raw,
            mol aa,
            length 418,
            seq-data ncbieaa
"MSDTRVQQYQDHAGAEAVDSSGLEVKRLTFLEAVAMIVGTNIGAGVLS
MAYASRKAGFMPLLLWLAVAGLFTTISMLYVSETALRTRTHNQLSGLAQRYVGSFGAWAIFLSVAVNSIGALIAYMS
G
SGKILSAFFGISPALGSVLFFIPAAGVLYLGLSAIGKGEKFISIGMVSMILILVAATLLNDNTEFARLLDGDWIYMV
P
VFNIAVFCFSAQYIVPEMARGFSHAPERLPKAVITGMLTTFVLLSIVPLSVIALTGLENQSEVATLAWGKALGEWAF
F
TANTFALCAMLTSYWGLGGSFLTNIFDKFHKLGPENRPKTRLLVLAIVALPPFVLAYSGLVGFVNALYFAGAFSGVI
L
SIMPILMLRSARRNGDFEPAWKCGWIAHPAIQAVIVTIYLASAAYAICSAMNLLPAGW"
          }
        },
        seq {
          id {
```

```
            local str "Query_4150"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">PA5434 15600627 HAAAP tryptophan permease
 [Pseudomonas aeruginosa PAO1]"
                }
              }
            },
            title "PA5434 15600627 HAAAP tryptophan permease [Pseudomonas
 aeruginosa PAO1]"
          },
          inst {
            repr raw,
            mol aa,
            length 417,
            seq-data ncbieaa
"MSSSPAQTPSRRPSLLGGSMIIAGTAVGAGMFSLPIAMSGIWFGWSVA
VFLLTWFCMLLSGMMILEANLNYPVGSSFSTITRDLLGQGWNVVNGLSIAFVLYILTYAYISGGGSIIGYTLSSGLG
V
TLPEKLAGLLFALAVALVVWWSTRAVDRITTLMLGGMIITFGLSISGLLGRIQPAILFNSGEPDAVYWPYLLATLPF
C
LTSFGYHGNVPSLMKYYGKDPQRISRSLWIGTLIALAIYLLWQASTLGTIPREQFKGIIAGGSNVGTLVEYLHRITA
S
DSLKALLTTFSNLAVASSFLGVTLGLFDYLADLCRFDDSHFGRFKTALLTFVPPTIGGLLFPNGFIYAIGFAGLAAA
F
WAVIVPALMARASRKRFGSPLFRAWGGTPAIVLVLLFGVANAVAHILASLHWLPEYR"
          }
        },
        seq {
          id {
            local str "Query_4151"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">PM0037 15601902 HAAAP SdaC [Pasteurella
multocida
 Pm70]"
                }
              }
            },
            title "PM0037 15601902 HAAAP SdaC [Pasteurella multocida Pm70]"
          },
          inst {
            repr raw,
            mol aa,
            length 413,
            seq-data ncbieaa
"MATKSCVSWSKFDLVWALNLFGTAVGAGVLFLPINAGKGGFWPLVCMT
```

```
LLVGPMTYLAHRGLTRFVLSSAKSGSDITEVVEEHFGQTAGKFITLLYFFAIFPILLIYGNGITNTVDSFIVHQLHM
N
SPNRALLAFILIAGLISVMLMNEKIMLKITEFLVYPLVCILLALSLYLIPEWNTSMLTEFPSTQALFSTLWITIPVL
V
FSFNHSPAISSFAQSQQREYGDLTNTETHTNRTLKITSSILLVFVMFFVFSCVLSLTPAELAQAKTQNISILSFLAN
K
FDNPYISYLGPFVAFLAITSSFFGHYLGAKEGLEGLIIKMSGKKEIKCQKLNYFTALFFLLTLWIVAIINPSILGLI
E
SFGGPIIAAILFIMPMYAIRKIPAMQRYQGRFSNLFVTIMGMFAISAVVYGLF"
          }
        },
        seq {
          id {
            local str "Query_4152"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">PM0732 15602597 HAAAP TyrP [Pasteurella
multocida
 Pm70]"
                }
              }
            },
            title "PM0732 15602597 HAAAP TyrP [Pasteurella multocida Pm70]"
          },
          inst {
            repr raw,
            mol aa,
            length 403,
            seq-data ncbieaa
"MKNKTFGSALLVAGTTIGAGMLAMPLTSAEMGFTYTLILLFILWGLLS
YSALLFVEVYQKAETKNAGIATLAEQYFGLPGRILATLSLVVFMYAILSAYVTGGGSLLAGVLPFLGEHATPISIVS
F
TIVLGVFIVISTGAVDMLTRFLFMIKLVAFVFVLLMMLPLVNGDNLLAMPLKDFLIISASPVFFTSFGFHVIIPSIN
S
YLDGNIRRLRIAIITGTAIPLVAYIIWQLATHGVFPQAQFVQILNTDPTLNGLITATYQATESAIISHAMRLFFTLA
L
ITSFLGVALSLFDCLYDLLKRVKIKTNRVSLGLLTFLPPLIFALFYPEGFVMALGYAGQMFTFYGLVLPVGMAWRAR
K
RYPDLPYRVMGGNLTLFGALILGILIMNVPFLIKAGYLPAVIG"
          }
        },
        seq {
          id {
            local str "Query_4153"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
```

```
                  data str ">PM0810 15602675 HAAAP TyrP [Pasteurella
multocida
 Pm70]"
                }
              }
            },
            title "PM0810 15602675 HAAAP TyrP [Pasteurella multocida Pm70]"
          },
          inst {
            repr raw,
            mol aa,
            length 404,
            seq-data ncbieaa
"MNKTLGSTLLVSGTMIGAGMLAMPLTSAGIGFTFTVVLLLVLWFLLTY
SALLFVEAYQTVESDAGIGSLSAVYFGGFGRFISTAALLIFLYALLAAYVTGGGGLLSSILPTIQSAETTSHISIVI
F
TTVFGAFIIIGTQTVDGLNRLLFFTMLVALGAVLFLLVPEVKMDNLMAMPIDKALLISASPVFFTSFGFHGSIPSLN
K
YLGGNIKALRISILVGSFITLCGYILWQFGTHGVLTQSVFLQILQDEPTLNGLVAATKQVTGSSVISGAVKLFSALA
L
ITSFLGVALGLFECIEDLLSRVFKFKAGRITLGLLTFIPPLLFALFYPKGFILALGYAGQMFAFYAVVLPAALVWKV
R
KLHPNLPYRVSGGSALLVISAVLGVIITSIPFAIKAGYLPAVVG"
          }
        },
        seq {
          id {
            local str "Query_4154"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">PM1192 15603057 HAAAP Mtr [Pasteurella multocida
 Pm70]"
                }
              }
            },
            title "PM1192 15603057 HAAAP Mtr [Pasteurella multocida Pm70]"
          },
          inst {
            repr raw,
            mol aa,
            length 418,
            seq-data ncbieaa
"MTTKQLPSLLGGAMIIAGTAIGAGMLANPTATAGVWFIGSLLLLAYTW
FCMTTSGLMILEANLHYPTGASFDTIVKDLLGKRWNLINGLSIAFVLYILTYAYITSGGGITQHLLNQVLSRPDNTV
E
IGRGLGSLLFCCVLAGFVWLSTKAVDRFTTVLIGGMIIAFLLSTSGLMASVRSDVLFNNLAQGETHYLPYFLTALPV
C
LVSFGFHGNVPSLVKYYDRHAQSVVKAIFLGTGIAFVIYAFWQLAVQGNLPRHQFAPVIEKGGDIAALLKALSQYIQ
T
DYMGVVLNFFAYMAIASSFLGVSLGLFDYLADLLKFDDSSLGRTKTTLVTFLPPLLLSLQFPYGFVVAIGYAGLAAT
I
WAAIVPALLAKATRQRFTESHYTVYGGNFMIYFVILFGLLNILAQIAAQFGVLPTFLG"
```

```
          }
        },
        seq {
          id {
            local str "Query_4155"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">PM1419 15603284 HAAAP TnaB [Pasteurella
multocida
 Pm70]"
                }
              }
            },
            title "PM1419 15603284 HAAAP TnaB [Pasteurella multocida Pm70]"
          },
          inst {
            repr raw,
            mol aa,
            length 408,
            seq-data ncbieaa
"MKKTPSIFGGACIIASVCVGAGMLGLPSSGAGAWTIWSILTICFTMFT
MTISGWLLLEAYKHYDLRASFNTVTKSMLGNSVNTINNLAVYFVGGILLYAYTTASGGILSGLAAPYFALDSRIWSV
I
FVFVFSFFVWHSTRIVDRVSVLLIIFMALSFLFSVFGLTVNIDLSTLFDTKGTESHYAMYAMGLLPVALTSFGYHHS
V
ATMRAYYGDERKAKYAILGGTGIALTLYLLWVVSIFGNLPRDQFAPVLASDGNLDVLLGALGNVIESASVKQAINAF
S
IAAILSSFIGVGLGVFDFLADFFKFDNSKEGRAKSWAVTFLPPLILSVAYPLGFLKAIGYAGAVATIWTCIIPALLA
R
KSRSLPNGQTGFMVGGGNITIVVVMVFGIVTALFHFLAMFGMLPAFKG"
          }
        },
        seq {
          id {
            local str "Query_4156"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">PP3589 26990300 HAAAP serine transporter SdaC
 [Pseudomonas putida KT2440]"
                }
              }
            },
            title "PP3589 26990300 HAAAP serine transporter SdaC [Pseudomonas
 putida KT2440]"
          },
          inst {
            repr raw,
```

```
            mol aa,
            length 426,
            seq-data ncbieaa
"MNEQAPSVEQRFAESTPAALGTWARHDTTWMLGLFGTAIGAGTLFLPI
NAGLGGFWPLLILAVLAFPMTYFAHRGLTRFVLSGRDGSDITEVVKEHFGITAGASITVLYFFAIFPILLIYSVALT
N
TVSSFMEHQLHMQPPPRAILSFVLILGLLAIVRCGEQATVKVMSLLVYPFIVALALLGLYLVPHWTGGILDSATEVP
P
ASAFLHTLWLAIPVMVFSFNHSPIISAFAVDQKRRYGAHADERSGQILRRAHLLMVVMVLFFVFSCVLTLSSAQLAE
A
KAQNLSILSYLANHFSNPTIAFAAPLIAFVAIAKSFLGHYIGASEGLKGIIAKTGARPGAKALDRVVAALMLVVCWI
V
ATLNPSILGMIESLGGPILAVLLFLMPMYAIRRVPSMRKYSGAASNVFVVVVGLVALTSVVYGLVS"
          }
        },
        seq {
          id {
            local str "Query_4157"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">PSPTO1760 28868966 HAAAP threonine/serine
 transporter [Pseudomonas syringae pv. tomato DC3000]"
                }
              }
            },
            title "PSPTO1760 28868966 HAAAP threonine/serine transporter
 [Pseudomonas syringae pv. tomato DC3000]"
          },
          inst {
            repr raw,
            mol aa,
            length 425,
            seq-data ncbieaa
"MNDQANGVVERLDVAPESIASWNRNDTTWMLGLFGTAIGAGTLFLPIN
AGIGGFWPLLALALLAFPMTFYAHRGLTRFVLSGREGADITEVVEQHFGKSAGAMITLLYFFAIFPILLIYSVALTN
T
VGSFLEHQLHIAPPPRAVLAFLLIMGLLALVRCGERFIVKAMSVMVYPFIVALLFLAMFLIPHWTGGILSTATTLPE
P
SAFLSTLWLAIPVMVFSFNHAPIISAFAVDQKRQYGENAEVRSSQILARAHVLMVVMVLFFVFSCVLTLSPAELAEA
K
AQNISILSYLANHFNNPTIAFVAPLIAFVAISKSFLGHYIGASEGLKGLVLKSGRRPAPKVLDRMTAAFMLVVCWIV
A
TLDPSILGMIENLGGPVISVLLFLMPMYAIYKVPSMRKYAGAWSNYFVIAAGLVAISALIFSLTR"
          }
        },
        seq {
          id {
            local str "Query_4158"
          },
          descr {
            user {
              type str "CFastaReader",
```

```
              data {
                {
                  label str "DefLine",
                  data str ">SF1953 24113287 HAAAP tyrosine-specific
transport
 system [Shigella flexneri 2a 301]"
                }
              }
            },
            title "SF1953 24113287 HAAAP tyrosine-specific transport system
 [Shigella flexneri 2a 301]"
          },
          inst {
            repr raw,
            mol aa,
            length 403,
            seq-data ncbieaa
"MKNRTLGSVFIVAGTTIGAGMLAMPLAAAGVGFSVTLILLIGLWALMC
YTALLLLEVYQHVPADTGLGTLAKRYLGRYGQWLTGFSMMFLMYALTAAYISGAGELLASSISDWTGISMSATAGVL
L
FTFVAGGVVCVGTSLVDLFNRFLFSAKIIFLVVMLVLLLPHIHKVNLLTLPLQQGLALSAIPVIFTSFGFHGSVPSI
V
SYMDGNIRKLRWVFIIGSAIPLVAYIFWQVATLGSIDSTTFMGLLANHAGLNGLLQALREMVASPHVELAVHLFADL
A
LATSFLGVALGLFDYLADLFQRSNTVGGRLQTGAITFLPPLAFALFYPRGFVMALGYAGVALAVLALIIPSLLTWQS
R
KHNPQAGYRVKGGRPALVVVFLCGIAVIGVQFLIAAGLLPEVG"
          }
        },
        seq {
          id {
            local str "Query_4159"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">SF2810 24114081 HAAAP probable serine
transporter
 [Shigella flexneri 2a 301]"
                }
              }
            },
            title "SF2810 24114081 HAAAP probable serine transporter
[Shigella
 flexneri 2a 301]"
          },
          inst {
            repr raw,
            mol aa,
            length 429,
            seq-data ncbieaa
"METTQTSTIASKDSRSAWRKTDTMWMLGLYGTAIGAGVLFLPINAGVG
GMIPLIIMAILAFPMTFFAHRGLTRFVLSGKNPGEDITEVVEEHFGIGAGKLITLLYFFAIYPILLVYSVAITNTVE
S
```

```
FMSHQLGMTPPPRAILSLILIVGMMTIVRFGEQMIVKAMSILVFPFVGVLMLLALYLIPQWNGAALETLSLDTASAT
G
NGLWMTLWLAIPVMVFSFNHSPIISSFAVANREEYGDMAEQKCSKILAFAHIMMVLTEMIFVFICVLSLTPADLAAA
K
EHNISILSYLANHFNAPVIAWMAPIIAIIAITKSFLGHYLGAREGFNGMVIKSLRGKGKSIEINKLNRITALFMLVT
T
WIVATLNPSILGMIETLGGPIIAMILFLMPMYAIQKVPAMRKYSGHISNVFVVVMGLIAISAIFYSLFS"
          }
        },
        seq {
          id {
            local str "Query_4160"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">SF2855 24114124 HAAAP putative transporter
 protein [Shigella flexneri 2a 301]"
                }
              }
            },
            title "SF2855 24114124 HAAAP putative transporter protein
 [Shigella flexneri 2a 301]"
          },
          inst {
            repr raw,
            mol aa,
            length 372,
            seq-data ncbieaa
"MVLFITALVAWPLTYWPHKALCQFILSSKTSAGEGITGAVTHYYGKKI
GNLITTLYFIAFFVVVLIYAVAITNSLTEQLAKHMVIDLRIRMLVSLGVVLILNLIFLMGRHATIRVMGFLVFPLIA
Y
FLFLSIYLVGSWQPDLLTTQVEFNQNTLHQIWISIPVMVFAFSHTPIISTFAIDRREKYGEHAMDKCKKIMKVAYLI
I
CISVLFFVFSCLLSIPSSYIEAAKEEGVTILSALSMLPNAPAWLSISGIIVAVVAMSKSFLGTYFGVIEGATEVVKT
T
LQQVGVKKSRAFNHALSIMLVSLITFIVCCINPNAISMIYAISGPLIAMILFIMPTLSTYLIPALKPWRSIGNLITL
I
VGILCVSVMFFS"
          }
        },
        seq {
          id {
            local str "Query_4161"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">SF3151 24114405 HAAAP putative transport system
 permease protein [Shigella flexneri 2a 301]"
                }
```

```
              }
            },
            title "SF3151 24114405 HAAAP putative transport system permease
 protein [Shigella flexneri 2a 301]"
          },
          inst {
            repr raw,
            mol aa,
            length 443,
            seq-data ncbieaa
"MEIASNKGVIADASTPAGRAGMSESEWREAIKFDSTDTGWVIMSIGMA
IGAGIVFLPVQVGLMGLWVFLLSSVIGYPAMYLFQRLFINTLAESPECKDYPSVISGYLGKNWGILLGALYFVMLVI
W
MFVYSTAITNDSASYLHTFGVTEGLLSDSPFYGLVLICILVAISSRGEKLLFKISTDMVLTKLLVVAALGVSMVGMW
H
LYNVGSLPPLGLLVKNAIITLPFTLTSILFIQTLSPMVISYRSREKSIEVARHKALRAMNIAFGILFVTVFFYAVSF
T
LAMGHDEAVKAYEQNISALAIAAQFISGDGAAWVKVVSVILNIFAVMTAFFGVYLGFREATQGIVMNILRRKMPAEK
I
NENLVQRGIMIFAILLAWSSIVLNAPVLSFTSICSPIFGMVGCLIPAWLVYKVPALHKYKGMSLYLIIVTGLLLCVS
L
FLAFS"
          }
        },
        seq {
          id {
            local str "Query_4162"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">SF3156 24114410 HAAAP anaerobically inducible
 L-threonine, L-serine permease [Shigella flexneri 2a 301]"
                }
              }
            },
            title "SF3156 24114410 HAAAP anaerobically inducible L-threonine,
 L-serine permease [Shigella flexneri 2a 301]"
          },
          inst {
            repr raw,
            mol aa,
            length 443,
            seq-data ncbieaa
"MSTSDSIVSSQTKQSSWRKSDTTWTLGLFGTAIGAGVLFFPIRAGFGG
LIPILLMLVLAYPIAFYCHRALARLCLSGSNPSGNITETVEEHFGKTGGVVITFLYFFAICPLLWIYGVTITNTFMT
F
WENQLGFAPLNRGFVALFLLLLMAFVIWFGKDLMVKVMSYLVWPFIASLVLISLSLIPYWNSAVIDQVDLGSLSLTG
H
DGILITVWLGISIMVFSFNFSPIVSSFVVSKREEYEKDFGRDFTERKCSQIISRASMLMVAVVMFFAFSCLFTLSPA
N
MAEAKAQNIPVLSYLANHFASMTGTKTTFAITLEYAASIIALVAIFKSFFGHYLGTLEGLNGLILKFGYKGDKTKVS
L
```

```
GKLNTLSMIFIMGSTWVVAYANPNILDLIEAMGAPIIASLLCLLPMYAIRKAPSLAKYRGRLDNVFVTVIGLLTILN
I
VYKLF"
          }
        },
        seq {
          id {
            local str "Query_4163"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">SF3202 24114452 HAAAP tryptophan-specific
 transport protein [Shigella flexneri 2a 301]"
                }
              }
            },
            title "SF3202 24114452 HAAAP tryptophan-specific transport
protein
 [Shigella flexneri 2a 301]"
          },
          inst {
            repr raw,
            mol aa,
            length 414,
            seq-data ncbieaa
"MATLTTTQTSPSLLGGVVIIGGTIIGAGMFSLPVVMSGAWFFWSMAAL
IFTWFCMLHSGLMILEANLNYRIGSSFDTITKDLLGKGWNVVNGISIAFVLYILTYAYISASGSILHHTFAEMSLNV
P
ARAAGFGFALLVAFVVWLSTKAVSRMTAIVLGAKVITFFLTFGSLLGHVQPATLFNVAESNASYAPYLLMTLPFCLA
S
FGYHGNVPSLMKYYGKDPKTIVKCLVYGTLMALALYTIWLLATMGNIPRPEFIGIAEKGGNIDVLVQALSGVLNSRS
L
DLLLVVFSNFAVASSFLGVTLGLFDYLADLFGFDDSAVGRLKTALLTFAPPVVGGLLFPNGFLYAIGYAGLAATIWA
A
IVPALLARASRKRFGSPKFRVWGGKPMIALILVFGVGNALVHILSSFNLLPVYQ"
          }
        },
        seq {
          id {
            local str "Query_4164"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">SF3574 24114808 HAAAP putative transporter
 protein [Shigella flexneri 2a 301]"
                }
              }
            },
            title "SF3574 24114808 HAAAP putative transporter protein
```

```
 [Shigella flexneri 2a 301]"
          },
          inst {
            repr raw,
            mol aa,
            length 411,
            seq-data ncbieaa
"MPFTRYDFGWVLLCIGMAIGAGTVLMPVQIGLKGIWVFITAAIIAYPV
TWVVQDIYLKTLSESDSCNDYTDIISHYLGKNWGIFLGVIYFLMIIHGIFIYSLSVVFDSASYLKTFGLTDADLSQS
L
LYKVAIFAVLIAIASGGERLLFKISGPMVVVKVGIIVVFGFAMIPHWNFANITAFPQASVFFRDVLLTIPFCFFSAV
F
IQVLNPMNIAYRKREADKVLATRLALRTHRISYITLIAVILFFAFSFTFSISHEEAVSAFEQNISALALAAQVIPGH
I
IHITSTVLNIFAVLTAFFGIYLGFHEAIKGIILNLLSRIIDTKKINSRMLTLAICAFIVITLTIWVSFRVSVLVFFQ
L
GSPLYGIVSCLIPFFLIYKVAQLEKLRGFKAWLILLYGILLCLSPLLKLIE"
          }
        },
        seq {
          id {
            local str "Query_4165"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">SF3753 24114979 HAAAP low affinity tryptophan
 permease [Shigella flexneri 2a 301]"
                }
              }
            },
            title "SF3753 24114979 HAAAP low affinity tryptophan permease
 [Shigella flexneri 2a 301]"
          },
          inst {
            repr raw,
            mol aa,
            length 415,
            seq-data ncbieaa
"MTDQAEKKNSAFWGVMVIAGTVIGGGMFALPVDLAGAWFFWGAFILII
AWFSMLHSGLLLLEANLNYPVGSSFNTITKDLIGNTWNIISGITVAFVLYILTYAYISANGAIISETISMNLGYHAN
P
RIVGICTAIFVASVLWISSLAASRITSLFLGLKIISFVIVFGSFFFQVDYSILRDATSTTAGTSYFPYIFMALPVCL
A
SFGFHGNIPSLIICYGKRKDKLIKSVVFGSLLALVIYLFWLYCTMGNIPRESFKAIISSGGNIDSLVKSFLGTKQHG
I
IEFCLLVFSNLAVASSFFGVTLGLFDYLADLFKIDNSHGGRFKTVLLTFLPPALLYLIFPNGFIYGIGGAGLCATIW
A
VIIPAVLAIKARKKFPNQMFTVWGGNLIPAIVILFGITVILCWFGNVFNVLPKFG"
          }
        },
        seq {
          id {
            local str "Query_4166"
```

```
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">SO0919 24372506 HAAAP serine transporter,
 putative [Shewanella oneidensis MR-1]"
                }
              }
            },
            title "SO0919 24372506 HAAAP serine transporter, putative
 [Shewanella oneidensis MR-1]"
          },
          inst {
            repr raw,
            mol aa,
            length 430,
            seq-data ncbieaa
"MQKDAISVLETPQSTQESTTRRLPWTRQDTTWMLSLFGTAVGAGILFL
PINAGMGGFWPLVLMAIIIGPMTYLAHRGLSRFVCSSSIPGSDITQVVEEHFGIGAGKAITLLYFLAIYPIVLIYGV
G
ITNTVDSFIVNQLGMASPPRFLLSGILIFGMMAVMVAGEQFMLKVTQLLVYPLVGILAFMSVYLIPEWKMDALQVVP
E
TGAFLGTVWLTIPVLVFAFNHSPAISQFSVSLKRDHGANAARKADVILRNTSMMLVGFVMLFVFSCVLSLSPEQLAE
A
KAKNLPILSYLANVHDSGFVSYFGPFIAFIAIVSSFFGHYMGATEGMKGIIVKQLRSSKKQVSEDKLNKFILVFMFA
T
IWGVAIKNPSILGMIEALGGPVIAAILYLMPMYAVYKVPALKAYRHRISNVFVVIAGLLAMTAILFGLLS"
          }
        },
        seq {
          id {
            local str "Query_4167"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">SO1074 24372659 HAAAP tyrosine-specific
transport
 protein, putative [Shewanella oneidensis MR-1]"
                }
              }
            },
            title "SO1074 24372659 HAAAP tyrosine-specific transport protein,
 putative [Shewanella oneidensis MR-1]"
          },
          inst {
            repr raw,
            mol aa,
            length 395,
            seq-data ncbieaa
"MNSKMLGSIAIVAGTAIGAGMLALPLATAALGMVPAILLMVVIWGLSA
```

```
YTSLLMLEINLRSGVGDNVHAITGKLLGKKGQIVQGASFLSLLVALTAAYLTGGSSLLVLKAQNMFDIVLDNQLAVV
L
FTIVLGGFAALGVAWVDKVSRFLFSLMILLLIVVVLFLLPEVSISTIATSAVAQSLTSSWMAAIPVVFTSFGFHVCI
A
TLVRYLDGDTVSLRKVLLIGSTIPLACYIFWLLVTLGTVGGNEINGFNGSLPALISALQEIAHTPLISKCISLFADL
A
LITSFLGVTLSLYDFVAELTRAKKTFVGRAQTWLLTFVPPLLCALYVPEGFVAVLGFAAVPLVVMIIFLPIAMALRQ
R
QIQPQGYQVVGGTFALGLAGILGAIIIGAQLFVAL"
          }
        },
        seq {
          id {
            local str "Query_4168"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">SO2065 24373625 HAAAP tyrosine-specific
transport
 protein [Shewanella oneidensis MR-1]"
                }
              }
            },
            title "SO2065 24373625 HAAAP tyrosine-specific transport protein
 [Shewanella oneidensis MR-1]"
          },
          inst {
            repr raw,
            mol aa,
            length 395,
            seq-data ncbieaa
"MTQNKFFGSLLLIAGTTIGAGMLALPIASAGLGFGMSSMIMLIFWALM
AYTALLMVEIHQFAPSDASLNQLARHLLGTKGQVVACIALMFLLYALCAAYIAGGGEQVHQKLTSWLGLALPPQAGA
I
LFTLLIGTIVGLGTHCVDLINRVLFSLKIIALILMLALLLPQVEGTHLLELPLEQGLIVSAIPVIFTSFGFHGSIPS
V
VRYLGIEVKALRKIMLLGSALPLLIYLLWQLGSQGVLSQSQLMTNQSLSGFISQLASVLHSQYLSSAISVFADLALA
T
SFLGVSLGLFDFIAANLRQEDNATGRSVTAAITFVPPLGFALFYPQGFITALGYAAIALVILAIFLPVSMVWVQRKQ
R
DKAKLPQGYRVAGGTLGLVLALLCGLAVISAQLLG"
          }
        },
        seq {
          id {
            local str "Query_4169"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
```

```
                  data str ">SO4601 24376074 HAAAP tryptophan-specific
 transport protein [Shewanella oneidensis MR-1]"
                }
              }
            },
            title "SO4601 24376074 HAAAP tryptophan-specific transport
protein
 [Shewanella oneidensis MR-1]"
          },
          inst {
            repr raw,
            mol aa,
            length 414,
            seq-data ncbieaa
"MNAVKNKPVGKSLLGGAMIIAGTTVGAGMFSLPVVGSGMWFGYSILML
LGIWFCMLMSGLLLLETNLHFEPGASFDTLTKETLGQFWRIVNGVSIAFVLYILTYAYISGGGSIVNHSLQGMGIEL
P
QSVAGLVFTVVLACIVLISTKAVDRITTIMLGGMIITFFLAIGNLLIEIDVTKLLEPDGNQRFIPYLWVALPFGLAS
F
GYHGNVPSLVKYYGKDSSTIIKAIFVGTFIALIIYVCWLVATMGNIPRSQFSEIIAQGGNMGVLVGALSKVMESSWL
N
SMLTLFANLAVASSFLGVTLGLFDYLADLFGFDDSRSGRMKTAIVTFVPPTVFGLLFPDGFLIAIGFAALAATVWAV
I
VPALMAYKSRQLFPNHQGFRVFGGTPLIILVVLFGVVTGACHLLAMANLLPQFS"
          }
        },
        seq {
          id {
            local str "Query_4170"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">STY2145 16760884 HAAAP tyrosine-specific
 transport protein [Salmonella typhi CT18]"
                }
              }
            },
            title "STY2145 16760884 HAAAP tyrosine-specific transport protein
 [Salmonella typhi CT18]"
          },
          inst {
            repr raw,
            mol aa,
            length 403,
            seq-data ncbieaa
"MKNRTLGSIFIVAGTTIGAGMLAMPLAAAGVGFSVTLGLLIGLWALMC
YTALLLLEVYQHVPADTGLGSLAKRYLGRYGQWLTGFSMMFLMYALTAAYISGAGELLASSINNWLGATLSPAAGVL
L
FTFVAGGVVCVGTSLVDLFNRFLFSAKIIFLVIMLALLTPHIHKVNLLTLPLQQGLALSAIPVIFTSFGFHGSVPSI
V
SYMNGNIRRLRWVFMTGSAIPLVAYIFWQLATLGSIDSPTFRGLLASHAGLNGLLQALREVVASPHVELAVHLFADL
A
```

```
LATSFLGVALGLFDYLADLFQRRSTVSGRLQTGLITFLPPLAFALFYPRGFVMALGYAGVALAVLALLIPAMLVWQC
R
KQSPQAGYRVAGGTPALALVFICGIVVIGVQFSIALGFLPDPG"
          }
        },
        seq {
          id {
            local str "Query_4171"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">STY3109 16761748 HAAAP putative serine
 transporter [Salmonella typhi CT18]"
                }
              }
            },
            title "STY3109 16761748 HAAAP putative serine transporter
 [Salmonella typhi CT18]"
          },
          inst {
            repr raw,
            mol aa,
            length 429,
            seq-data ncbieaa
"MKTTQTSTIASIDSRSAWRKTDTMWMLGLYGTAIGAGVLFLPINAGVG
GMIPLIIMAILAFPMTFFAHRGLTRFVLSGKNPGEDITEVVEEHFGIGAGKLITLLYFFAIYPILLVYSVAITNTVE
S
FLTHQLAINPPPRAILSLILIVGMMTIVRFGEQMIVKAMSILVFPFVAALMLLALYLIPQWNGAALETLSFDSAAST
G
NGLWMTLWLAIPVMVFSFNHSPIISSFAVAKREEYGEGAEKKCSKILAFAHIMMVLTVMFFVFSCVLSLTPADLAAA
K
EQNISILSYLANHFNAPIIAWMAPIIAMIAITKSFLGHYLGAREGFNGMVIKSLRGKGKSIEINKLNKITALFMLVT
T
WIVATLNPSILGMIETLGGPIIAMILFLMPMYAIQKVPAMRKYSGHISNVFVVIMGLIAISAIFYSLFS"
          }
        },
        seq {
          id {
            local str "Query_4172"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">STY3167 16761800 HAAAP probable amino acid
 transport protein [Salmonella typhi CT18]"
                }
              }
            },
            title "STY3167 16761800 HAAAP probable amino acid transport
 protein [Salmonella typhi CT18]"
```

```
          },
          inst {
            repr raw,
            mol aa,
            length 409,
            seq-data ncbieaa
"MSKIWSKDETLWSFALYGTAVGAGTLFLPIQLGSAGAIVLFITALVAW
PLTYWPHKALCQFILSSKTSAGEGITGAVTHYYGKKIGSIITTLYFIAFFVVVLIYAVAITNSLTEQLAKHIQIDIR
I
RMAVSFGVVLILNMIFLMGRHATLRVMGFLVFPLIAYFLFLSLYLTGSWQPSLLTGQMSLDSHTLHQVWISIPVMVF
A
FSHTPIISTFAIDRRENFGDQAMDKCKKIMKVAYLIICLSVLFFVFSCLLSIPPSYIEDARNEGVTILSALSMMPNA
P
AWLSISGIIVAVVAMSKSFLGTYFGVIEGATEMVRTTLQQVGVKKSRAFNRALSIMLVSGITFIICCINPNAISMIY
A
ISGPLIAMILFIMPTLSTYLIPALKPYRSIGNFITLVVGLLCVSVMFFG"
          }
        },
        seq {
          id {
            local str "Query_4173"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">STY3460 16762044 HAAAP probable amino acid
 permease [Salmonella typhi CT18]"
                }
              }
            },
            title "STY3460 16762044 HAAAP probable amino acid permease
 [Salmonella typhi CT18]"
          },
          inst {
            repr raw,
            mol aa,
            length 414,
            seq-data ncbieaa
"MATLTTTQTSPSLLGGVVIIGGTIIGAGMFSLPVVMSGAWFFWSMAAL
VFTWFCMLHSGLMILEANLNYRIGSSFDTITKDLLGKGWNVVNGISIAFVLYILTYAYISASGSILHHTFAEMSLNV
P
ARAAGFAFALLVAFVVWLSTKAVSRMTAIVLGAKVITFFLTFGSLLGHVQPTTLFNVAESHASYTPYLLMTLPFCLA
S
FGYHGNVPSLMKYYGKDPRTIVKCLIYGTLLALALYSVWLLGTMGNIPRPEFIGIAQKGGNIDVLVQALSGVLNSRS
L
DLLLVVFSNFAVASSFLGVTLGLFDYLADLFGFDDSAMGRFKTALLTFLPPMIGGLLYPNGFLYAIGYAGLAATIWA
A
IVPALLARKSRERFGSPKFRVWGGKPMIALILVFGVGNAVIHILSSFNLLPVYQ"
          }
        },
        seq {
          id {
            local str "Query_4174"
          },
```

```
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">STY4173 16762677 HAAAP putative amino acid
 permease [Salmonella typhi CT18]"
                }
              }
            },
            title "STY4173 16762677 HAAAP putative amino acid permease
 [Salmonella typhi CT18]"
          },
          inst {
            repr raw,
            mol aa,
            length 432,
            seq-data ncbieaa
"MQDDTLPLNNSNATTTPLSTRLPFTKYDFGWVLLCIGMAIGAGTVLMP
VQIGLKGIWVFITAFIIAYPATYIVQDIYLKTLSESETCDDYTDIISHYLGKNWGVFLGVIYFLMIIHGVFIYSLSV
V
FDSASYIKTFGLTEADLSQSIIYKVAIFAVLVAIASGGEKLLFKISGPMVVVKVGIILIFGLAMIPHWNLDNISAFP
A
ASVFFRDVLLTIPFCFFSAVFIQVLNPMNIAYRKREPDRVLATRMAIRTHRISYITLIAIILFFSFSFTFSISHEEA
V
SAFEQNISALALAAQVIPGQIIHITSTILNIFAVLTAFFGIYLGFHEALKGIVLNVLSRIMDVKNINPLLLTSGICV
F
IVVTLVIWVSFRVSVLVFFQLGSPLYGIVACIIPFFLIYKVAQLEKLRGLKTWLILLYGILLCLSPLLKLIE"
          }
        },
        seq {
          id {
            local str "Query_4175"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">STM1937 16765278 HAAAP HAAAP family,
 tyrosine-specific transport protein [Salmonella typhimurium LT2]"
                }
              }
            },
            title "STM1937 16765278 HAAAP HAAAP family, tyrosine-specific
 transport protein [Salmonella typhimurium LT2]"
          },
          inst {
            repr raw,
            mol aa,
            length 403,
            seq-data ncbieaa
"MKNRTLGSIFIVAGTTIGAGMLAMPLAAAGVGFSVTLGLLIGLWALMC
YTALLLLEVYQHVPADTGLGSLAKRYLGRYGQWLTGFSMMFLMYALTAAYISGAGELLASSINNWLGATLSPAAGVL
L
```

```
FTFVAGGVVCVGTSLVDLFNRFLFSAKIIFLVIMLALLTPHIHKVNLLTLPLQQGLALSAIPVIFTSFGFHGSVPSI
V
SYMNGNIRRLRWVFMTGSAIPLVAYIFWQLATLGSIDSPTFRGLLASHAGLNGLLQALREVVASPHVELAVHLFADL
A
LATSFLGVALGLFDYLADLFQRRSTVSGRLQTGLITFLPPLAFALFYPRGFVMALGYAGVALAVLALLIPAMLVWQC
R
KQSPQAGYRVAGGTPALALVFICGIVVIGVQFSIALGFLPDPG"
          }
        },
        seq {
          id {
            local str "Query_4176"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">STM2970 16766275 HAAAP putative HAAAP family,
 serine transport protein [Salmonella typhimurium LT2]"
                }
              }
            },
            title "STM2970 16766275 HAAAP putative HAAAP family, serine
 transport protein [Salmonella typhimurium LT2]"
          },
          inst {
            repr raw,
            mol aa,
            length 429,
            seq-data ncbieaa
"METTQTSTIASIDSRSAWRKTDTMWMLGLYGTAIGAGVLFLPINAGVG
GMIPLIIMAILAFPMTFFAHRGLTRFVLSGKNPGEDITEVVEEHFGIGAGKLITLLYFFAIYPILLVYSVAITNTVE
S
FLTHQLAINPPPRAILSLILIVGMMTIVRFGEQMIVKAMSILVFPFVAALMLLALYLIPQWNGAALETLSFDSAAST
G
NGLWMTLWLAIPVMVFSFNHSPIISSFAVAKREEYGEGAEKKCSKILAFAHIMMVLTVMFFVFSCVLSLTPADLAAA
K
EQNISILSYLANHFNAPIIAWMAPIIAMIAITKSFLGHYLGAREGFNGMVIKSLRGKGKSIEINKLNKITALFMLVT
T
WIVATLNPSILGMIETLGGPIIAMILFLMPMYAIQKVPAMRKYSGHISNVFVVIMGLIAISAIFYSLFS"
          }
        },
        seq {
          id {
            local str "Query_4177"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">STM3022 16766324 HAAAP putative transport
protein
 [Salmonella typhimurium LT2]"
```

```
              }
            }
          },
          title "STM3022 16766324 HAAAP putative transport protein
 [Salmonella typhimurium LT2]"
        },
        inst {
          repr raw,
          mol aa,
          length 409,
          seq-data ncbieaa
"MSKIWSKDETLWSFALYGTAVGAGTLFLPIQLGSAGAIVLFITALVAW
PLTYWPHKALCQFILSSKTSAGEGITGAVTHYYGKKIGNIITTLYFIAFFVVVLIYAVAITNSLTEQLAKHIQIDIR
I
RMAVSFGVVLILNMIFLMGRHATLRVMGFLVFPLIAYFLFLSLYLTGSWQPSLLTGQMSLDSHTLHQVWISIPVMVF
A
FSHTPIISTFAIDRRENFGDQAMDKCKKIMKVAYLIICLSVLFFVFSCLLSIPPSYIEDARNEGVTILSALSMMPNA
P
AWLSISGIIVAVVAMSKSFLGTYFGVIEGATEMVRTTLQQVGVKKSRAFNRALSIMLVSGITFIICCINPNAISMIY
A
ISGPLIAMILFIMPTLSTYLIPALKPYRSVGNFITLVVGLLCVSVMFFG"
        }
      },
      seq {
        id {
          local str "Query_4178"
        },
        descr {
          user {
            type str "CFastaReader",
            data {
              {
                label str "DefLine",
                data str ">STM3239 16766538 HAAAP putative HAAAP family
 transport protein [Salmonella typhimurium LT2]"
              }
            }
          },
          title "STM3239 16766538 HAAAP putative HAAAP family transport
 protein [Salmonella typhimurium LT2]"
        },
        inst {
          repr raw,
          mol aa,
          length 443,
          seq-data ncbieaa
"MESASNTSVILDASAPARRAGMTESEWREAIKFDSTDTGWVIMSIGMA
IGAGIVFLPVQVGLMGLWVFLLSSIIGYPAMYLFQRLFINTLAESPECKDYPSVISGYLGKNWGILLGALYFVMLVI
W
MFVYSTAITNDSASYLHTFGVTEGLLSDSPFYGLVLICILVAISSRGEKLLFKISTGMVLTKLLVVAALGVSMVGMW
H
LYNIGALPPMALLIKNAIITLPFTLTSILFIQTLSPMVISYRSREKSIEVARHKALRAMNIAFGILFVTVFFYAVSF
T
LAMGHDEAVKAYEQNISALAIAAQFISGDGAGWVKVVSVILNIFAVMTAFFGVYLGFREATQGIVMNILRRKMPAEK
I
KENLVQRGIMIFAILLAWSAIVLNAPVLSFTSICSPIFGMVGCLIPAWLVYKVPALHKYKGASLYLIIITGLLLCVS
P
```

```
FLAFS"
          }
        },
        seq {
          id {
            local str "Query_4179"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">STM3243 16766542 HAAAP HAAAP family, L-
threonine/
 L-serine permease, anaerobically inducible [Salmonella typhimurium LT2]"
                }
              }
            },
            title "STM3243 16766542 HAAAP HAAAP family, L-threonine/ L-serine
 permease, anaerobically inducible [Salmonella typhimurium LT2]"
          },
          inst {
            repr raw,
            mol aa,
            length 443,
            seq-data ncbieaa
"MSTTDSIVSSQAKQSSWRKSDTTWTLGLFGTAIGAGVLFFPIRAGFGG
LIPILLMLVLAYPIAFYCHRALARLCLSGSNPSGNITETVEEHFGKTGGVVITFLYFFAICPLLWIYGVTITNTFMT
F
WENQLQMPALNRGFVALFLLLLMAFVIWFGKDLMVKVMSYLVWPFIASLVLISLSLIPYWNSAVIDQVDLSNIALTG
H
DGILVTVWLGISIMVFSFNFSPIVSSFVVSKREEYEKEFGREFTERKCSQIISRASMLMVAVVMFFAFSCLFTLSPQ
N
MADAKAQNIPVLSYLANHFASLSGTKSTFATVLEYGASIIALVAIFKSFFGHYLGTLEGLNGLVLKFGYKGDKTKVS
M
GKLNTISMIFIMGSTWVVAYANPNILDLIEAMGAPIIASLLCLLPMYAIRKAPSLAKYRGRLDNVFVTLIGLLTILN
I
VYKLF"
          }
        },
        seq {
          id {
            local str "Query_4180"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">STM3279 16766577 HAAAP HAAAP family,
 tryptophan-specific transport protein [Salmonella typhimurium LT2]"
                }
              }
            },
            title "STM3279 16766577 HAAAP HAAAP family, tryptophan-specific
```

```
 transport protein [Salmonella typhimurium LT2]"
          },
          inst {
            repr raw,
            mol aa,
            length 414,
            seq-data ncbieaa
"MATLTTTQTSPSLLGGVVIIGGTIIGAGMFSLPVVMSGAWFFWSMAAL
VFTWFCMLHSGLMILEANLNYRIGSSFDTITKDLLGKGWNVVNGISIAFVLYILTYAYISASGSILHHTFAEMSLNV
P
ARAAGFAFALLVAFVVWLSTKAVSRMTAIVLGAKVITFFLTFGSLLGHVQPATLFNVAESHASYTPYLLMTLPFCLA
S
FGYHGNVPSLMKYYGKDPRTIVKCLIYGTLLALALYSVWLLGTMGNIPRPEFIGIAQKGGNIDVLVQALSGVLNSRS
L
DLLLVVFSNFAVASSFLGVTLGLFDYLADLFGFDDSAMGRFKTALLTFLPPMIGGLLYPNGFLYAIGYAGLAATIWA
A
IVPALLARKSRERFGSPKFRVWGGKPMIVLILLFGVGNALVHILSSFNLLPVYQ"
          }
        },
        seq {
          id {
            local str "Query_4181"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">STM3625 16766912 HAAAP putative HAAAP family
 transport protein [Salmonella typhimurium LT2]"
                }
              }
            },
            title "STM3625 16766912 HAAAP putative HAAAP family transport
 protein [Salmonella typhimurium LT2]"
          },
          inst {
            repr raw,
            mol aa,
            length 432,
            seq-data ncbieaa
"MQDDTLPLNNSNATTTPLSTRLPFTKYDFGWVLLCIGMAIGAGTVLMP
VQIGLKGIWVFITAFIIAYPATYIVQDIYLKTLSESETCDDYTDIISHYLGKNWGIFLGVIYFLMIIHGVFIYSLSV
V
FDSASYIKTFGLTEADLSQSIIYKVAIFAVLVAIASGGEKLLFKISGPMVVVKVGIILIFGFAMIPHWNLDNISAFP
A
ASVFFRDVLLTIPFCFFSAVFIQVLNPMNIAYRKREPDRVLATRMAIRTHRISYITLIAIILFFSFSFTFSISHEEA
V
SAFEQNISALALAAQVIPGHIIHITSTILNIFAVLTAFFGIYLGFHEALKGIVLNVLSRIMDVKNVNPLLLTSGICV
F
IVVTLVIWVSFRVSVLVFFQLGSPLYGIVACIIPFFLIYKVAQLEKLRGLKTWLILLYGILLCLSPLLKLIE"
          }
        },
        seq {
          id {
            local str "Query_4182"
```

```
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">VCA0160 15600930 HAAAP tryptophan-specific
 transport protein [Vibrio cholerae El Tor N16961]"
                }
              }
            },
            title "VCA0160 15600930 HAAAP tryptophan-specific transport
 protein [Vibrio cholerae El Tor N16961]"
          },
          inst {
            repr raw,
            mol aa,
            length 411,
            seq-data ncbieaa
"MKPKDFIMSKQPSLFGGACIIASVCVGAGMLGLPSAGAGAWTLWSMLA
LALTMAVMTLSGWMLLEAFKHYDLRVSFNTVTKDMLGSHINRFNNLTVYFVGGILLYAYITSSGLILQDLLHINSKI
A
SILFVAVFSAFVWHSTRAVDRISVILIVFMVLSFIFGVSGLAINVKTSILFDTLNQSGEYAPYAMAMLPVALTSFGY
H
HSVSSMRAYYGEERKAKYAILGGTVIALSLYALWLFSIFGNLPRADFAPVIQQGGNVDVLLKALGSVVESEKVSQAI
N
AFSMAAILSSFIGVGLGVFDFLADLFQFSNCKQGRTKTWLVTFLPPLILSLLFPFGFIIAIGYAGAAATIWACIIPV
L
LARKSRTLANGAQGFVVPGGNLALGLVLIFGVLTAVFHMMAMANLLPAFKG"
          }
        },
        seq {
          id {
            local str "Query_4183"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">VCA0772 15601527 HAAAP tyrosine-specific
 transport protein [Vibrio cholerae El Tor N16961]"
                }
              }
            },
            title "VCA0772 15601527 HAAAP tyrosine-specific transport protein
 [Vibrio cholerae El Tor N16961]"
          },
          inst {
            repr raw,
            mol aa,
            length 400,
            seq-data ncbieaa
"MTQSKLLGSTLIIAGTTIGAGMLALPLASAGIGFSTSLMIMLGLWMLM
```

```
AFTALLMVEIHQYADKEATLHTLAKQILGDKGKWVATFAMLFLFYSLCAAYIAGGGAQFTQRITDFTGVNVESSSGT
L
LFTLIVALVVTVGTGTVDRVNRVLFAGKMIAMVAVLFFLAPNVSQSYLLSMPIQQGLIVAAIPVIFTSFGFHGSIPA
I
VNYLDGDTPALRKAILIGSAIPLVIYIFWQLVTLGVVSQSALLDNMGLTALIGVLSTTVHQSNLGNIIGVFADLALL
T
SFLGVSLGLFEFMGDSLRNQQGKMNRPLASVVTFLPPLIFALFYPQGFIMALGYAAIALAILAIFLPLVMVIKVRQQ
A
TEQHYQVTGGNGALLVTGLVGLLIIGAQLLITLGILPALG"
          }
        },
        seq {
          id {
            local str "Query_4184"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">VC1301 15641314 HAAAP serine transporter [Vibrio
 cholerae El Tor N16961]"
                }
              }
            },
            title "VC1301 15641314 HAAAP serine transporter [Vibrio cholerae
 El Tor N16961]"
          },
          inst {
            repr raw,
            mol aa,
            length 417,
            seq-data ncbieaa
"MNTTTAATSTARSASKWTYKDFTWALSLFGTAVGAGVLFLPIKAGAGG
FWPLVLLALIAAPMTWFAHKSLARFVLSAKNPDADITDTVEEHFGKAGANLITFAYFFAIYPIVLIYGVGITNTVDS
F
LVNQIGMESIPRWLLSGALITAMTAGVVFGKELMLKATSAMVYPLVFILLALSFYLIPDWNTSMMEVAPDWSAMPAI
V
WLAIPIIVFSFNHSPIISQFSKEQRQQFGDKAVQKTDMITGGAAMMLMGFVMFFVFSVVLSLSPAELAMAKEQNISV
L
SYLANEHASPIISYLGPIVAFAAITSSYFGHFLGAHEGLVGLVKSRSNMQVSKIEKISLGFIVITTWIVAIVNPSIL
G
MIETMGAPMIAAILFLLPVFAMHKVPAMAKFKTSAPVQIFTVICGLAAISSVIYGAL"
          }
        },
        seq {
          id {
            local str "Query_4185"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">VC1658 15641663 HAAAP serine transporter [Vibrio
```

```
 cholerae El Tor N16961]"
                }
              }
            },
            title "VC1658 15641663 HAAAP serine transporter [Vibrio cholerae
 El Tor N16961]"
          },
          inst {
            repr raw,
            mol aa,
            length 417,
            seq-data ncbieaa
"MKDSQEVTHFSAARSSKWTQHDTYWLLSLFGTAVGAGILFLPINIGLG
GFWPLVLMAVLAFPMTYLSHRGLARFVLSSRRPNADFTDVVEEHFGQNAGRLISLLYFLSIFPILLIYGVGLTNTVD
S
FLVNQADMVSPPRALLSGALVFGLIAIMLSGEKIMLRAFAIMVYPLAAILALLSLYLIPYWSMPSFDFPQTSDFLHT
V
WLAVPVVVFSFSHAAAISSFANIQRRHYGEQADQKSEQILRHTSVVLILFVLLFVFSCVLALSPQALQEAKAQNISV
L
TYLANVTDNPFIATLGPLVAFIAITSSFLGHFLGARESLSGLLTKNLGVSLRLAERLTIGFLFVTIWIAAILNPSIL
G
MMEALSGPVIAMILFIMPMIAVFKVPALREHQKRFSTFFVLLVGLLAVSALLYSLLR"
          }
        },
        seq {
          id {
            local str "Query_4186"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">VP1426 28898200 HAAAP tyrosine-specific
transport
 protein [Vibrio parahaemolyticus RIMD2210633]"
                }
              }
            },
            title "VP1426 28898200 HAAAP tyrosine-specific transport protein
 [Vibrio parahaemolyticus RIMD2210633]"
          },
          inst {
            repr raw,
            mol aa,
            length 401,
            seq-data ncbieaa
"MSRSKVFGSTLIIAGTTIGAGMLALPLASAGIGFTTSLIIMLSLWALM
AFTALLMLELHQYAESSATLHTLAKQILGQKGKWVASFAMLFLFYSLCAAYIAGGGAQFGDRLSQWFELDISGPTAT
V
IFTLIVTLVVTIGTGTVDKVNRVLFALKLVAMVAVLSFLAPNVTESYLLSMPIEQGLIVAAIPVIFTSFGFHGSIPA
I
VNYLDGDTSSLRKAVIVGSTIPLVIYIFWQIVTLGVVSQDALIENGGLSALIGQLSQTVHKSNLSSIVGVFADLALL
T
SFLGVSLGLFEFLGDTIKGKSEKPNRLLAAVITFTPPLGFALFYPQGFIMALGYAAIALAILAIFLPLVMVIKVRRS
D
```

```
DFSGEYRVAGGQGALVVTGIAGTGIVLAQVLITLGVLPALG"
          }
        },
        seq {
          id {
            local str "Query_4187"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">VP1879 28898653 HAAAP serine transporter [Vibrio
 parahaemolyticus RIMD2210633]"
                }
              }
            },
            title "VP1879 28898653 HAAAP serine transporter [Vibrio
 parahaemolyticus RIMD2210633]"
          },
          inst {
            repr raw,
            mol aa,
            length 417,
            seq-data ncbieaa
"MNTTTSAVSTAQSSSKFTYKDFTWCLSLFGTAVGAGVLFLPIKAGAGG
FWPLVILALIAAPMTWFAHKSLARFVLSAKNPEADITDTVEEHFGKTGANLITFAYFFAIYPIVLIYGVGITNTVDS
F
LVNQMGMESIPRWLLSGALIAAMTAGVVFGKELMLKATSAMVYPLVFILLALSFYLIPEWNTSMIEVAPDWAAMPTI
V
WLAIPIIVFSFNHSPIISQFSKEQRMQYGDEAYKKTDMITGGAAMMLMGFVMFFVFSVVLSLSPEQLASAKEQNISV
L
SYLANIHESPLISYMGPLVAFAAITSSYFGHFLGAHEGLVGLIKSRSQSPVSKIEKGSLLFIVITTWIVAIVNPSIL
G
MIETMGAPMIAAILFLMPVFAMQKVPAMAKYKTSAPVQIFTAICGLAAITSVIYGAL"
          }
        },
        seq {
          id {
            local str "Query_4188"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">VP2175 28898949 HAAAP putative HAAAP family
 transport protein [Vibrio parahaemolyticus RIMD2210633]"
                }
              }
            },
            title "VP2175 28898949 HAAAP putative HAAAP family transport
 protein [Vibrio parahaemolyticus RIMD2210633]"
          },
          inst {
```

```
            repr raw,
            mol aa,
            length 441,
            seq-data ncbieaa
"MSVTAPTQTLPIGKEPPHVKAGMALDEWKAATKFDSTDWGWVIMSIGM
AIGAGIVFLPVKVGVLGLWVFLASAVIGYPAMYLFQRLFINTLSSSPKCQDYAGVISGYLGNKWGALLGVLYFIMLV
I
WVFVYSTAINNDSASFLHSFGITEGLLSENPLYGLALVCALVAIASRGEKILFKVSTGLVLIKLCTVALLGVMMIEK
W
DVANVGSIPETGAGLKDAIELLPFTLTSILFIQSLSPMVISYRSKEKSIEVARFKAMRAMKIAFGILFVTVFFYAIS
F
TLAMSHEQAVKAAEENISALAMVAQGMDGTTLKLMSLMLNIFSVMTAFFGVYLGFRDSCQGLAMLALKKVMPEEKIN
K
DLVTKGIILFAILMAWGAIVLNLPVLSFTSVCSPIFGLIGCLIPAYLVYQVPSLHKYKGASLYLIIATGILLCVSPF
V
AFS"
          }
        },
        seq {
          id {
            local str "Query_4189"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">VPA0509 28900364 HAAAP putative tyrosine-
specific
 transport protein [Vibrio parahaemolyticus RIMD2210633]"
                }
              }
            },
            title "VPA0509 28900364 HAAAP putative tyrosine-specific
transport
 protein [Vibrio parahaemolyticus RIMD2210633]"
          },
          inst {
            repr raw,
            mol aa,
            length 385,
            seq-data ncbieaa
"MNLKLVGSSLIVAGTALGAGMLAIPMVLAQFGLLWGTLLMLFIWAGTT
YAALLLLEASCKVGGGVSMNAIARETLGKGGQLVTNGLLYALLVCLLMAYIIGAGDLVQKITASVGLSVSTVSSQVG
F
TILVGLIVSAGTGVVDKLNRGLFIGMIVALVLTLFALAPSVSFEGLNEVVSSDKMALIKTSSVLFTSFGFMVVIPSL
V
TYNKEASKTQLRNMIVVGSTIPLVCYLLWLFAVVGNLPPHELVQYSNVTELISVLGQQYNGLEFILSMFTGLALLTS
F
LGVAMALYDQNADLLKTSKPVVFVTTFILPLLGAVFAPEHFLAILSYAGIILVFLAVFVPLSMTMKVRRVPVEDNSV
Y
EAGGGVMGMSMIFLFGCFLLFAQAV"
          }
        },
        seq {
          id {
```

```
            local str "Query_4190"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">VPA1067 28900922 HAAAP serine transporter
[Vibrio
 parahaemolyticus RIMD2210633]"
                }
              }
            },
            title "VPA1067 28900922 HAAAP serine transporter [Vibrio
 parahaemolyticus RIMD2210633]"
          },
          inst {
            repr raw,
            mol aa,
            length 418,
            seq-data ncbieaa
"MKESRNTLNSSELNTTTSWSKHDTHWVLSLFGTAVGAGILFLPINLGI
GGFWPLVAMAFLAFPMTYLAHRGLARFVLSSKIKNADFTDVVEEHFGAKAGRSISLLYFLSIFPILLIYGVGITNTV
D
SFMVNQAGMEALPRELLSGVLVFALIAIMMAGEKVMLRAFAVMVYPLVAILAFLSFYLMPNWTMPVLDTPDMGAFAS
T
MWLAVPVVIFSFSHAAAISSFANVQRRHYGDDADAKAELILRCTSIMLIAFVLLFVFSCVLALSPEQLAQAKAQNVS
V
LSYLANATDNPFIATLGPLVAFVAITSSFLGHFLGARESLNGLITKHSNLSETRVDRISVVVLFLSIWAAAIMNPSI
L
GMMEALSGPVIAMILFIMPMLAVHKIESMKQYRGKLSTYFVLITGIVAVSALVFSLLS"
          }
        },
        seq {
          id {
            local str "Query_4191"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">VV12261 27365584 HAAAP Serine transporter
[Vibrio
 vulnificus CMCP6]"
                }
              }
            },
            title "VV12261 27365584 HAAAP Serine transporter [Vibrio
 vulnificus CMCP6]"
          },
          inst {
            repr raw,
            mol aa,
            length 417,
```

```
            seq-data ncbieaa
"MKTTSTAAPAVQSSSKFTYKDFTWCLSLFGTAVGAGVLFLPIKAGAGG
FWPLVMLALLAAPMTWFAHKSLARFVLSAKNPEADITDTVEEHFGKTGANLITFAYFFAIYPIVLIYGVGITNTVDS
F
LVNQMGMESIPRPLLSGALILAMTAGVVFGKELMLKATSAMVYPLVFVLLALSVYLIPDWNTSMMQVAPDWSSMPVV
V
WLAIPIIVFSFNHSPIISQFSKEQRLQHGDKAVQKTDAITGGAAMMLMGFVMFFVFSVVLSMSPEQLASAKEQNISV
L
SYLANVHESPLISYLGPLVAFAAITSSYFGHFLGAHEGLVGLIKSRSNSPISKIEKASLLFIVITTWIVAIVNPSIL
G
MIETMGAPMIAAILFLMPVFAMQKVPAMAKYKTSAAVQIFTVICGLASITSVIYGAL"
          }
        },
        seq {
          id {
            local str "Query_4192"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">VV12293 27365614 HAAAP Tyrosine-specific
 transport protein [Vibrio vulnificus CMCP6]"
                }
              }
            },
            title "VV12293 27365614 HAAAP Tyrosine-specific transport protein
 [Vibrio vulnificus CMCP6]"
          },
          inst {
            repr raw,
            mol aa,
            length 401,
            seq-data ncbieaa
"MSRSKIFGSTLIIAGTTIGAGMLALPLASAGIGFSTSVMIMIGFWVLM
SYTALLMVELHQEADRSATLHTLAKQILGTKGKWVASFAMLFLFYALCAAYIAGGGAQFGERVAAWFSIEMDASVAT
V
IFTVIVAAVVTSGTATVDKVNRVLFGLKMVAMVAVLFFLAPNVTESYLLSMPIQQGLIVAALPVIFTSFGFHSSIPA
I
VNYLDGDTKALRKAIVIGSSIPLVIYVFWQLVTLGVVNQEVLLNNSGLTALIAQLAAKVQQSYLSQLVGVFADLALL
T
SFLGVSLGLFEFLGDVIKKSGQKQNRVIPAVVTFTPPLMFALFYPQGFIMALGYAAIALAVLAIFLPVVMVNKVRKS
P
EKQGSYQVAGGQLALTLTGVTGVVIVAAQILITVGVLPALG"
          }
        },
        seq {
          id {
            local str "Query_4193"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
```

```
                label str "DefLine",
                data str ">VV20037 27366498 HAAAP Amino acid permease
 [Vibrio vulnificus CMCP6]"
              }
            }
          },
          title "VV20037 27366498 HAAAP Amino acid permease [Vibrio
 vulnificus CMCP6]"
        },
        inst {
          repr raw,
          mol aa,
          length 390,
          seq-data ncbieaa
"MSLFGTAVGAGILFLPINLGIGGFWPLVLLALLAFPMTFWGHRALARF
VLSSKQADADFTDVVEEHFGAKAGRLISLLYFLSIFPILLIYGVGITNTVDSFMVNQAGMESLPRVLLSGVLVFSLI
A
IMLGGEKLMLRAFAAMVYPLATILAFLSLYLIPSWQLPSLALPSAGDFLQTVWLAVPVVIFSFSHAAAISSFVGVQK
R
HYGEQANGKSEQILQRTSIMLIAFVLLFVFSCVLSLSPEQLAQAKAQNVSVLSYLANATENPFIATFGPLVAFIAIT
S
SFLGHFLGARESLNGLLTKHTSMPIQRADKMTIVSLFISIWIAAVINPSILGMMETLSGPVIAMILFIMPVYAIYRV
D
SLKIYRGKATTVFVLLTGLLAVSALLFAIV"
        }
      },
      seq {
        id {
          local str "Query_4194"
        },
        descr {
          user {
            type str "CFastaReader",
            data {
              {
                label str "DefLine",
                data str ">VV20855 27367261 HAAAP Amino acid permease
 [Vibrio vulnificus CMCP6]"
              }
            }
          },
          title "VV20855 27367261 HAAAP Amino acid permease [Vibrio
 vulnificus CMCP6]"
        },
        inst {
          repr raw,
          mol aa,
          length 404,
          seq-data ncbieaa
"MSKQPSLIGGACIIASVCVGAGMLGLPSAGAGAWTLWSLLALTLTMVV
MTLSGWLLLEAFKHYELRVSFNTVTKDMLGSKVNWFNNLMVYFVGGILLYAYITSSGLILQDVLQIDSKVASILFVA
V
FSAFVWHSTRAVDRISVILIVFMVLSFIFGVSGLAINVKTSVLFDTINENADYAPYAMAMLPVALTSFGYHHSVSSM
R
SYYGEERKAKYAILGGTIIALSLYFLWLLSIFGNLPRSEFGSVIEQGGNVDVLLKALGSVIESEKVSQAINAFSMAA
I
```

```
LSSFIGVGLGVFDFLADLFKFGNCKRGRSKTWLVTFLPPLILSLLFPFGFIIAIGYAGAAATVWACIIPVLLAYKSR
S
MENGYQGFVVPGGMSVLVLVLGFGVLTAIFHMLAMANMLPTFTG"
          }
        },
        seq {
          id {
            local str "Query_4195"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">Wbr0264 24323870 HAAAP Amino acid permeases
 [Wigglesworthia brevipalpis P-endosymbiont]"
                }
              }
            },
            title "Wbr0264 24323870 HAAAP Amino acid permeases
[Wigglesworthia
 brevipalpis P-endosymbiont]"
          },
          inst {
            repr raw,
            mol aa,
            length 429,
            seq-data ncbieaa
"MFYKKHKNICNICNNNYKKSNLIWILSLYGTAIGAGVLFLPINIGING
LFSLLIIMIIAFPMAFFSHRGLTRFVLSSKKPNKNLFNVIHENFNKQTSFIFILFYFFSIYPILLIYSTAITNTVMS
F
IIHQLNIAEQSRVFVSFILISVLMGIIILGEKVIIKFMSFLVIPFILSLIFFSFYIIQFWNYSLFYEIKNIFNLDIN
S
LLKIIWYIIPVTVFSFNHSPIISSFAVFKKRECSNCCEKTCSKILFYSNLLMIFTVIFFVFSCILSVPREAMQEAKD
K
NLSVLSYFSIYFEDNLMNFLGPIIAIIAIIKSFFGHYLGAKESINKILMTTFPKLKNKKRTYLTNFTTNSLIFTSSW
I
IAIFNPSILNIIEAIIGPIISIILFLIPAYSIYKVPNMKRYKKEKFTNLFLFVIGIFSFLAMMHKLIIF"
          }
        },
        seq {
          id {
            local str "Query_4196"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">YPO0277 16120616 HAAAP putative transmembrane
 transport protein [Yersinia pestis CO-92]"
                }
              }
            },
            title "YPO0277 16120616 HAAAP putative transmembrane transport
```

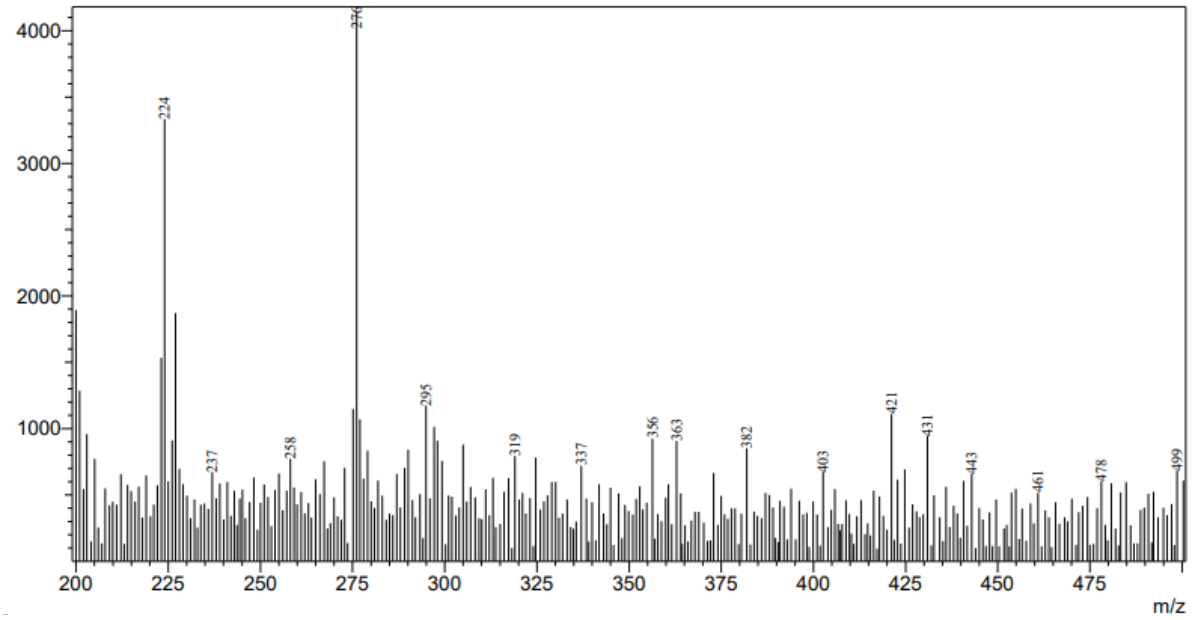
```
 protein [Yersinia pestis CO-92]"
          },
          inst {
            repr raw,
            mol aa,
            length 428,
            seq-data ncbieaa
"MSEISSSKSELKKLYKSVPFTQYDVGWVILCIGMAIGSGIVFMPVQVG
IKGIWVFIAAVIISYPAIYLLQNLYLRTLSESDNCTDYTSVITQYLGKNWGVGLGIAYFLMLLHGMFSYSLAVTFDS
A
SYIKTFGLTEGLLSDSIWYGLIILTVLVAIAAQGEKILFKVSGPMVCVKFGIIVVLGVVMVPYWDFNNISAFPELFS
F
LRDVLLTLPFTLFSILFVQILSPMNIAYRKVESDKRIATYRAIRANRVAYIILAVAVLFFAFSFTFSISHEQAVSAF
E
QNISALAIAAQVIPGSIVRIMTALLNIFAILTAFLGIYLGFQEAIKGIVVNIISRFIPEENINQKVLHIGVCVGVIL
T
LWLWVSTRFSILFFLQLGGPLFGVVSCLIPCYLVYKVPVLHKLKGPTIWFISFFGILLCLSPFFKFFE"
          }
        },
        seq {
          id {
            local str "Query_4197"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">YPO1209 16121500 HAAAP tyrosine-specific
 transport protein [Yersinia pestis CO-92]"
                }
              }
            },
            title "YPO1209 16121500 HAAAP tyrosine-specific transport protein
 [Yersinia pestis CO-92]"
          },
          inst {
            repr raw,
            mol aa,
            length 402,
            seq-data ncbieaa
"MKNRTLGSIFIVAGTTIGAGMLAMPLAAAGVGFGVTLALLVSLWLLMC
YTALLLVEVYQYEAADTGLGTLAKRYLGGHGQWLTGFSMMFLMYALTAAYISGAGELLATSISQWTQQSFPTYLGVL
L
FTLVAGGVVCIGTHSVDLFNRILFSAKIIFLIVMLALMMPHIEKTNLLTLPLEQGLALSAIPVIFTSFGFHGSVPSI
V
NYMGGNIRKLRWVFIIGSAIPLIAYIFWQLATLGAISSHTFVGILAQQAGLNGLLQAVRDVVASPHVELAVHLFADL
A
LATSFLGVALGLFDYLADLFKRRNSIRGRLQTGLITFTPPLLFALFYPRGFIMALGFAAVALSVLALILPAMLAWKA
R
KLHQGKYQVWGGRPALAIVFACGVTVIGIQIGIVTGALPAVG"
          }
        },
        seq {
          id {
            local str "Query_4198"
```

```
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">YPO1285 16121568 HAAAP putative
 tryptophan-specific transport protein [Yersinia pestis CO-92]"
                }
              }
            },
            title "YPO1285 16121568 HAAAP putative tryptophan-specific
 transport protein [Yersinia pestis CO-92]"
          },
          inst {
            repr raw,
            mol aa,
            length 414,
            seq-data ncbieaa
"MSMALSQPLSRPSVLGGAMIVAGTAVGAGMFSIPIVTAGVWFSGSVVL
LIYTWACMLVSGLMILEANLNYPSGASFHTMVQDLLGKVWSSINGLSITFVLYILTYAYISAGGSIIAHTLQNVAGI
G
QTSAGFIFALLVAFIVWLSTRAVDRLSTILIGGMVITFVMSVGDMFSHVESAVLFNQTDSNAHYLPYALAALPYLLT
S
FGYHGNVPGLVKYYNKDSKAVVRSLLYGTLIALVIYLLWQYAIQGNIARDAFKQVIADGGNIGNLLQQMDTVSASKA
T
SQLLNAFSYMALASSFLGVSLGLFDYISDLFDFTDDRSGRTKSALMTFVPPTLGALLFPDGFLYAIGFAGLAATIWA
V
IVPALMARASRKRFPQARYRAPGGRFMIGFIILFGLVNAVAHISALFGLLPVYQ"
          }
        },
        seq {
          id {
            local str "Query_4199"
          },
          descr {
            user {
              type str "CFastaReader",
              data {
                {
                  label str "DefLine",
                  data str ">YPO1321 16121603 HAAAP serine transporter
 [Yersinia pestis CO-92]"
                }
              }
            },
            title "YPO1321 16121603 HAAAP serine transporter [Yersinia pestis
 CO-92]"
          },
          inst {
            repr raw,
            mol aa,
            length 433,
            seq-data ncbieaa
"MDTTQTSTLASARKISSSTWRKSDTMWMLGLYGTAIGAGVLFLPINAG
```

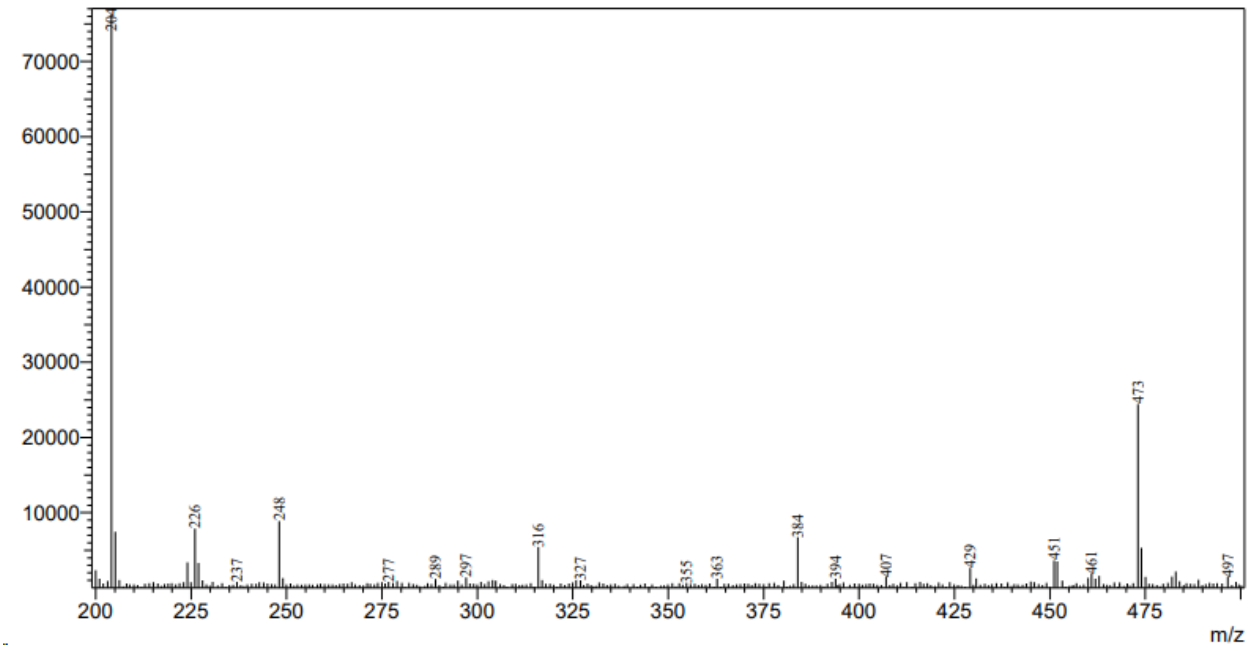
```
IGGLLPLIVMAIIAFPMTYYAHRGLCRFVLSGKNPGEDITEVVEEHFGVGAGKLITLLYFFAIYPILLVYSVAITNT
V
DSFITHQIHLPSPPRAILSLILIVGLMAIVRFGEQTIVKAMSILVFPFVAVLMMLAIYLIPNWSGAIFENVSMDGNG
T
GSGLWMTMWLIIPVMVFSFNHSPIISAFAVAKREEYGVDAEKKCSRILSFAHIMMVITVMFFVFSCVLSLSPADLMD
A
KNQNISILSYLANHFNTPVIAYMAPVIAFIAITKSFLGHYLGAREGFNGMVIKSLRSRGKTIEQNKLNRITALFMLV
T
TWAVATLNPSILGMIETLGGPIIAMLLFLMPMYAIHKVPAMRKYSGQISNVFVVLMGLIAISGILFNLSNLFN"
          }
        }
      }
    },
    subject database "nr",
    algorithm-options {
      {
        name "EvalueThreshold",
        value cutoff e-value { 5, 10, -2 }
      },
      {
        name "UngappedMode",
        value boolean FALSE
      },
      {
        name "PercentIdentity",
        value real { 0, 10, 0 }
      },
      {
        name "HitlistSize",
        value integer 1000
      },
      {
        name "EffectiveSearchSpace",
        value big-integer 0
      },
      {
        name "DbLength",
        value big-integer 0
      },
      {
        name "WindowSize",
        value integer 40
      },
      {
        name "SegFiltering",
        value boolean FALSE
      },
      {
        name "MaskAtHash",
        value boolean FALSE
      },
      {
        name "WordSize",
        value integer 3
      },
      {
        name "WordThreshold",
```

```
        value integer 11
      },
      {
        name "MatrixName",
        value string "BLOSUM62"
      },
      {
        name "GapOpeningCost",
        value integer 11
      },
      {
        name "GapExtensionCost",
        value integer 1
      },
      {
        name "CompositionBasedStats",
        value integer 2
      },
      {
        name "SmithWatermanMode",
        value boolean FALSE
      }
    },
    program-options {
      {
        name "EntrezQuery",
        value string "txid1117 [ORGN]"
      }
    },
    format-options {
      {
        name "Web_JobTitle",
        value string "96 sequences (At2g33260 15225823 HAAAP putative..."
      },
      {
        name "Web_RunPsiBlast",
        value boolean TRUE
      },
      {
        name "Web_StepNumber",
        value string "1"
      },
      {
        name "Web_OrganismName",
        value string "Cyanobacteriota (taxid:1117)"
      },
      {
        name "Web_ShortQueryAdjust",
        value boolean TRUE
      }
    }
  }
}
```

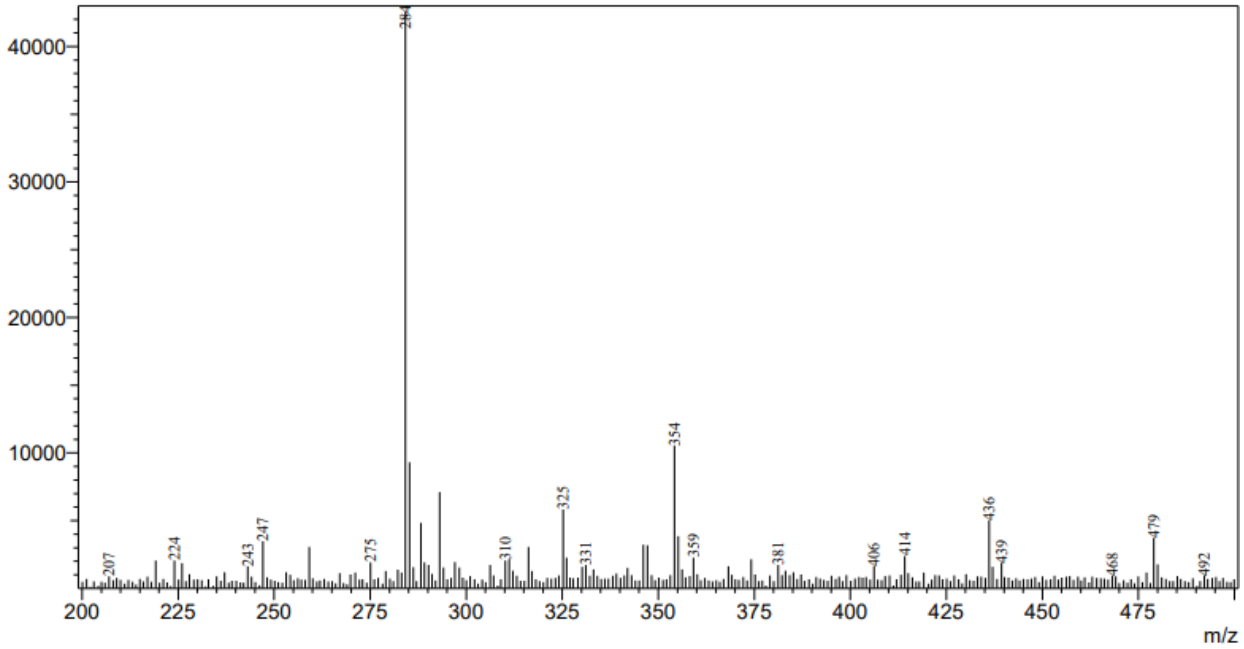
**MS spectra**



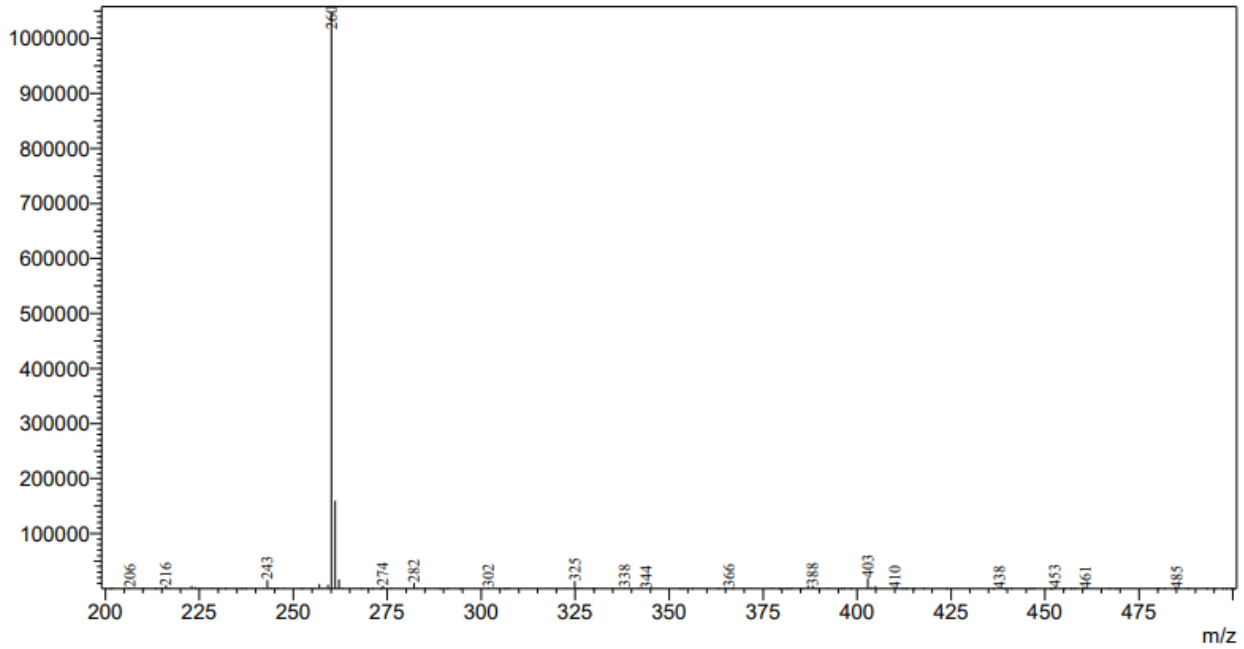

Lys-Glu



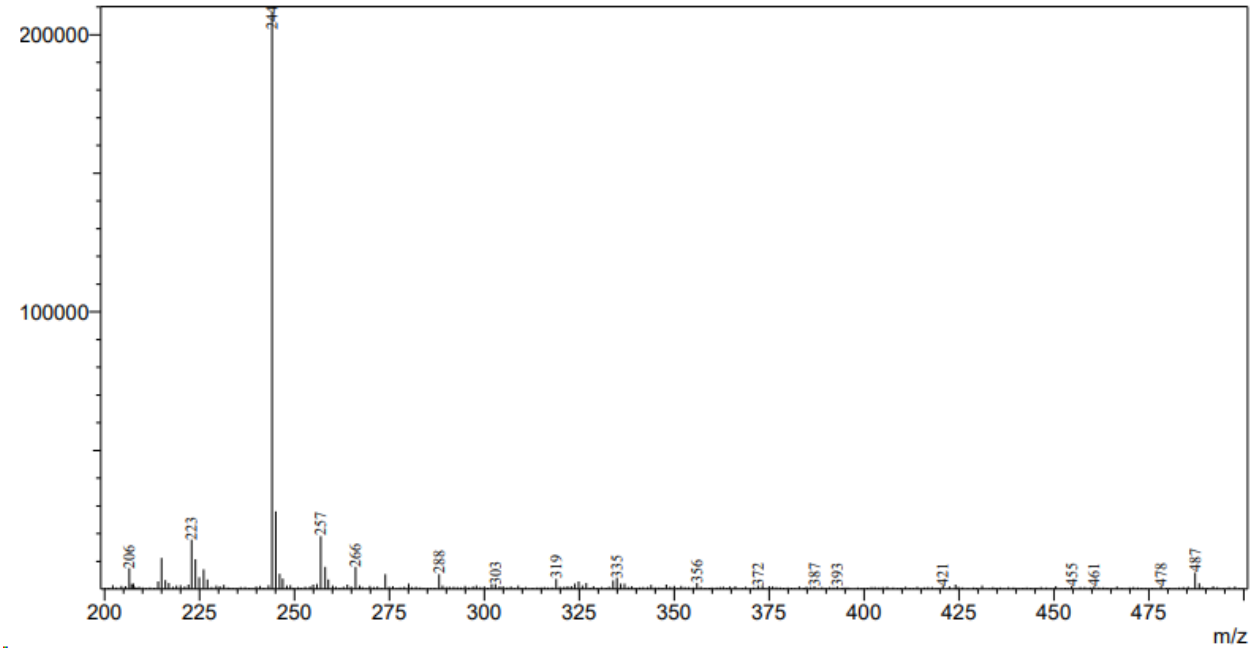

Lys-Gly



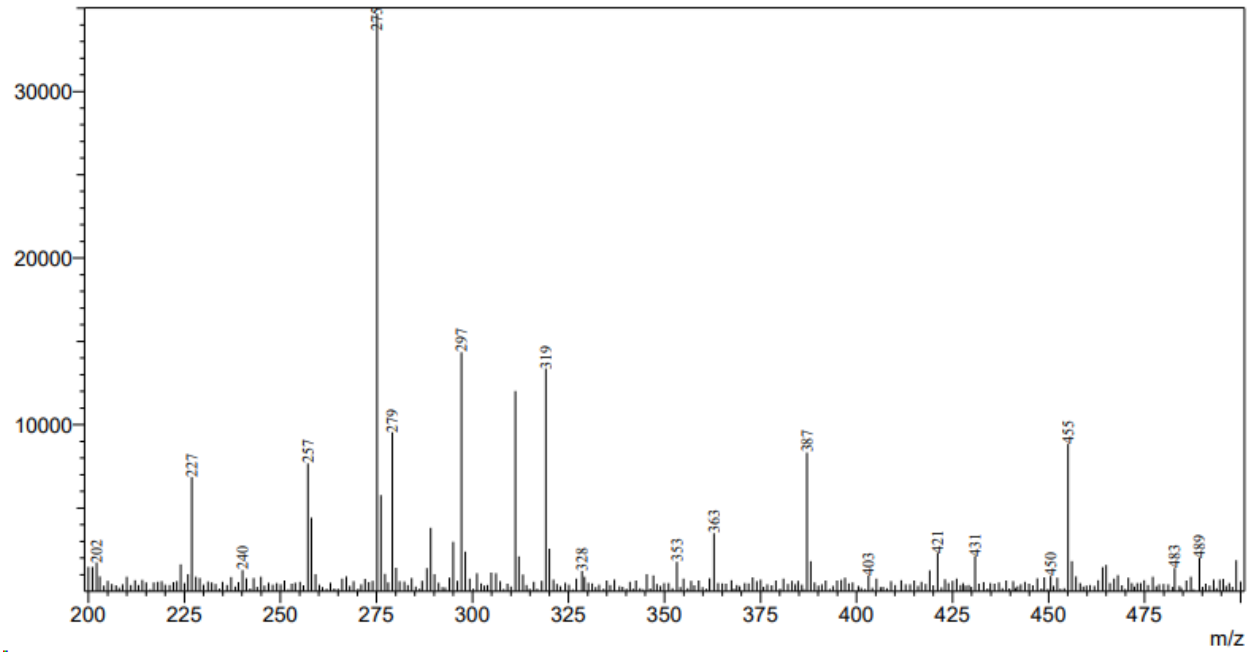

Lys-His



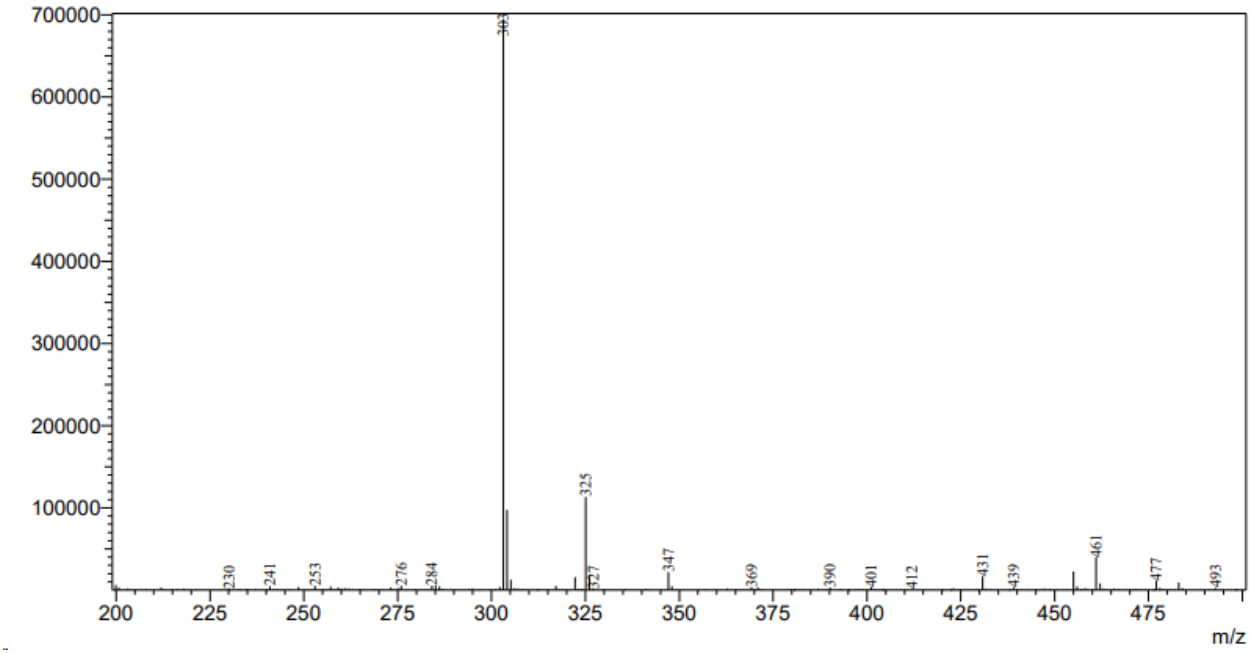

Lys-Leu



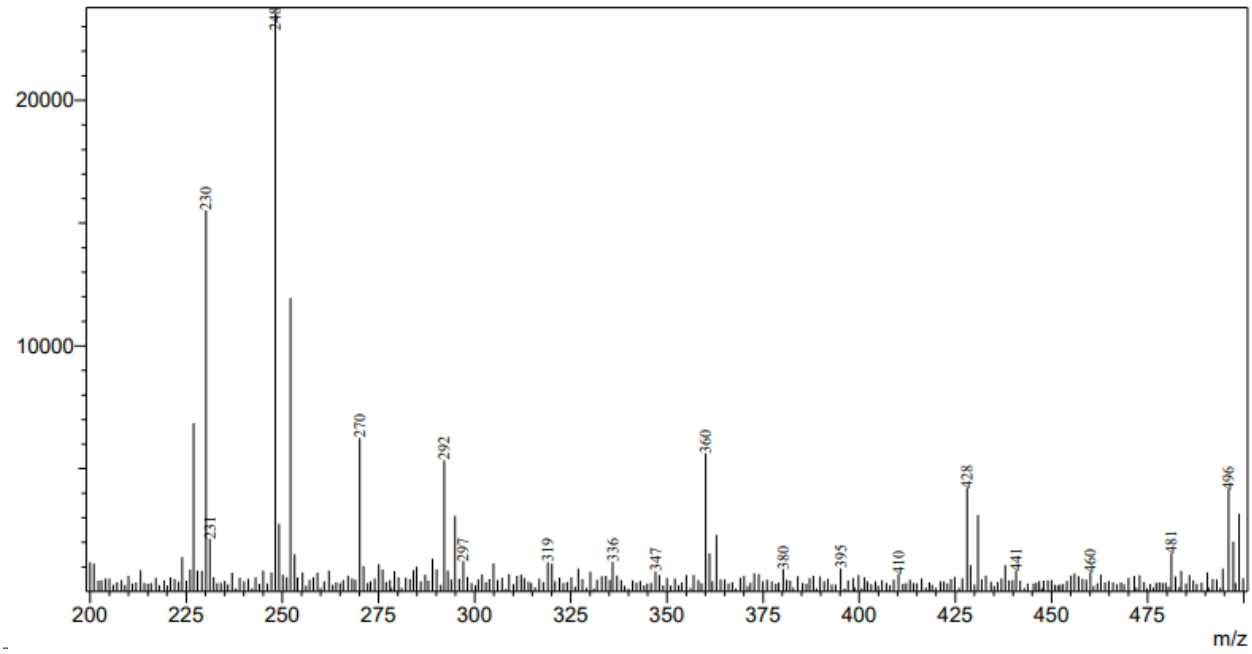

Lys-Pro



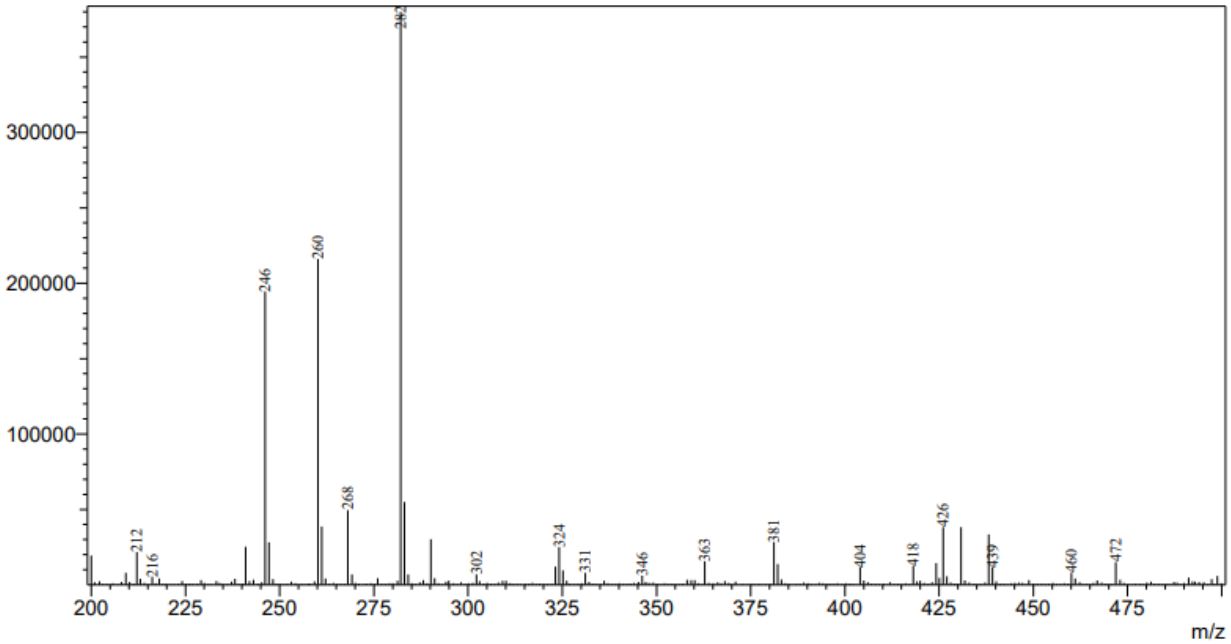

Lys-Gln




Lys-Arg




Lys-Thr


300000
200000
100000
246
260
268
212
216
302
324
331
346
363
381
404
418
426
439
460
472
200
225
250
275
300
325
350
375
400
425
450
475
m/z

Lys-Val