UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES

EXPLORING TWO STRATEGIES FOR THE HETEROLOGOUS BIOSYNTHESIS OF OLIVETOLIC ACID IN THE DIATOM PHAEODACTYLUM TRICORNUTUM

THÈSE PRÉSENTÉE

COMME EXIGENCE PARTIELLE DU

DOCTORAT EN BIOLOGIE CELLULAIRE ET MOLECULAIRE

PAR

NICOLAS LOUIS SENE

Juillet 2025

UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES

DOCTORAT EN BIOLOGIE CELLULAIRE ET MOLÉCULAIRE

Direction de recherche :

_____

Isabel Desgagné-Penix Directrice de recherche

_____

Hugo Germain Codirecteur de recherche

Jury d'évaluation de la thèse :

_____

Isabel Desgagné-Penix	Directrice de recherche
Hugo Germain	Codirecteur de recherche
Jörg Behnke	Évaluateur externe
Marina Cvetkovska	Évaluatrice externe
Geneviève Pépin	Présidente de jury

Thèse soutenue le 30/05/2025

To my family, although I know they will never read it. To my friends, those who battled for PhD and those who lived a happy life. To Alexandra Elbakyan who reminds the world that not everything should be monetized. And, to YOU reader, that has to skim through a thesis, by obligation, by curiosity, or by passion. Keep your flame lit!

Biiiisous!

ACKNOWLEDGEMENTS

PhD is a terribly long journey that is hard to bear on one's shoulder. The more people support us, the lighter the weight. I first of all want to nin utshetam muk utshin Professor Isabel Desgagné-Penix for giving me the opportunity to work in her laboratory and to Professor Hugo Germain for recruiting me and bringing me fishing for the first time of my life. You both have been an incredible support during this trip, and an admirable guidance. I want to thank Professor Geneviève Pépin for being my first and last examiner in Québec and for always being a benevolent and inspiring presence in our department, the Professor Marina Cvetkovska which ignited my scientific flame by just listening to her questions during the PhD defense of a friend and took the time to be a part of my jury, and the doctor Joerg Behnke for his very sympathetic chat in Montréal and his understanding of our struggle with our beloved diatom, and of course for being part of my defense jury (and for his amazing linkedin profile picture) vielen dank!

I want to thank my family for their unyielding support, and eternal love. To my mother who gave her life for us to become an eternal role model, even if she doesn't understand the nuance between pressure and motivation. And that we had to raise ourselves… Oχχχχuuuuu ! A mon père Nuiny qui sait toujours garder la tête froide et prodiguer des conseils sages et raisonnés loin du volcanisme familiale. A Boulinou mon frère, soutien moral et pilier de nos valeurs, me rappelant toujours l'essentiel et les bons moments même si nos échanges se limitent à un envoie de vidéo et image drôle une fois par mois et une réaction en émoji. A ma légendaire sœur et sa famille, qui ont aidé à rendre la distance plus passable en me faisant sentir un peu à leurs côtés. A ma partenaire de quasi 10 ans, Anne-Sophie qui a été ma partenaire de rires et de jeu, qui n'a besoin de rien faire de spéciale pour adoucir la vie.

I want to thank all my Greek family who wanted to help me do figures and reference my document. Ειστε οι πια αξιαγαπητη οικογεινια του κοσμου (και αστεια ρε βρε!!). Παρα πολα φιλακια αγαπητημενες μου !! Υιατι τα ελληνικα ειναι τοσο δυσκολο να γραφεις ρε βρε…

A mes amis en France qui n'ont pas arrêté de me demander « Tu rentres quand du coup » ? Merci à mon Seby pour être un pilier de la vie, un gentleman drôle, compréhensif et à la présence lumineuse. Mon seul ami toujours prêt à me raconter avec le plus grand enthousiasme du monde, des détails superflus sur des gens que je ne connais pas. A mon Thibouillou l'une des personnes que j'admire le plus au monde. Merci mon partenaire de voyage stellaire ! Merci à Morgane pour rester la drôle fille dont le meilleur public est elle-même, prête à se battre pour garder son trône de meilleure amie. Merci à Féyou pour être mon frère de philosophie créative et métaphysique du quotidien et des vidéos débiles. Merci à Alex, conteur d'aventures et mon frère de pèche nocturne. Merci à Bagnito et à mon Titan pour garder le temps de parler des choses essentielles de la vie comme le Seigneur des Anneaux et les figatellu. Nos conversations sont rares, mais nos cœurs battent en rythme. Merci à ma Thinhinou, une lumière vive qui continue de briller dans le brouillard de Paris.

Vielen dank to mein bro Alexander, a noble heart and a wise mind in the body of a hussler. Du fliegst über der Herde! 谢谢我的朋友们! Tiantian and Petitbout you are golden friends that teach me the best mandarin!

Et merci à tous ceux que je contacte une fois par an et qui me réponde comme si on ne s'était quitté que la veille. To my three cats Seliny, Snow, and Tshakapesh who despite being judgmental cats have the funniest and sweetest personalities (except Tshaki, you have to stop being so haughty!).

AND OF COURSE… To my Canadian crew!!!

Thank you to Fatma Meddeb, my first contact in Québec, and the lighthouse that I often had to guide so we would not BOTH end up lost, but always eventually brought me to shore. And… of course, to the heart of the galaxy of our lab, Natacha. Thank you for being such a bright, supportive, funny, smart and benevolent presence. To my PETI FRE, the man with the noblest heart, and the loudest laugh. You are a rare person that the world needs, to laugh and to see what abs can look like. To THeo, the druid of my heart, partner in trouble and eternally dissatisfied extraordinary chef. Thank you for always finding something incredibly stupid to make us laugh. I will both bear you as my brothers until death. To Claiwe the ambassador of French people in Québec and softest heart in the universe (I'm pretty sure I've checked). To Znehi, the VIP of my heart and the unnoticed diamond of this world. Your mind is bright and sharp, and your heart is almost as noble as mine. To BIBASTAAA…Bibel, the only one to fluently speak spanishos with me, and paladin of noble values. You are an admirable scientist and an OK friend. To the Latinas Ingrith, Baleria, Islé, and Gajos de Roble the sweetest poet, mi espagnole es mucho major a gracias de oustedes! Ingrith with your kind heart you really sooth the souls of people who talk to dyouu! Guay? Moudour! To Ellliiiiiiii and Fatima who have been guides and humorous presence in the office. To my chumdefilles and all my chumedegos. To Clouet the most respectable scientist in the universe and a sensitive heart for a brilliant mind, to Duru, a mentor in science always prone to joke around even on our demise, and to Clemuru, my sparring partner and a chef as sweet as his spirit. Thank you to all the lab members that made me laugh or touched me with your kind hearts (I'm not naming everyone, but you know that I'm talking about you!).

RÉSUMÉ

Les maladies neurodégénératives comme la maladie d'Alzheimer ou de Parkinson touchent aujourd'hui plusieurs millions d'individus. Ces maladies ont deux choses en commun : elles n'ont aucun remède, et sont reliées au système endocannabinoïde. Les cannabinoïdes sont des molécules connues pour leur interaction avec le système endocannabinoïdes et possèdent de nombreux effets sur la physiologie humaine. Provenant de *Cannabis sativa*, les phytocannabinoïdes les plus célèbres sont le Δ^9-tétrahydrocannabinol (THC) et le cannabidiol (CBD), pourtant plus d'une centaine de molécules ont été identifiées comme cannabinoïdes. L'un des obstacles majeurs à leur contribution au développement de traitements est la faible abondance de la plupart de ces molécules ainsi que leur forte similarité biochimique, rendant leur purification difficile.

Afin de simplifier l'accès aux cannabinoïdes, cette étude a donc tenté de transférer une partie de la voie de biosynthèse de cannabinoïdes de *Cannabis sativa* dans la diatomée *Phaeodactylum tricornutum* afin de produire des extraits plus purs. Nous avons cloné séparément dans *P. tricornutum* deux enzymes, la Tétrakétide synthase et l'acide olivetolique cyclase de *Cannabis sativa* (respectivement *Cs*TKS et *Cs*OAC) qui produisent en chaine respectivement le 3,5,7-trioxododecanoyl-CoA et l'acide olivetolique, le précurseur principal des cannabinoïdes. Nous avons aussi transféré l'enzyme SteelyA de l'amibe *Dictyostelium discoideum* capable, comme l'action combinée de *Cs*TKS et de *Cs*OAC à partir d'hexanoyl-CoA et de malonyl-CoA, de produire de l'olivetol.

Cette étude a confirmé la production des trois enzymes in vivo dans *Phaeodactylum tricornutum* et a testé l'activité de la *Cs*TKS par métabolomique non dirigée, montrant une différence de profile métabolique entre les clones et le contrôle négatif. L'expression de *SteelyA* a visiblement mené à la production d'une protéine insoluble inactive. Dans l'ensemble, cette étude montre que des connaissances supplémentaires sur le métabolisme de *P. tricornutum* sont requises pour comprendre les mécanismes sous-jacents entravant la biosynthèse d'acide olivetolique, et pour le futur de la bioproduction.

Mots clés : Diatomée, cannabinoïdes, tetrakétide synthase, acide olivetolique cyclase, expression hétérologue, métabolomique non dirigée, métabolisme.

ABSTRACT

Neurodegenerative diseases such as Alzheimer disease and Parkinson disease affect several millions of individuals nowadays. Those diseases have two things in common: they currently possess no cure, and they are connected to the endocannabinoid system. Cannabinoids are molecules known to interact with the endocannabinoid system and have numerous effects on human's physiology. Coming from *Cannabis sativa*, the most famous phytocannabinoids are the Δ^9-tetrahydrocannabinol (THC) and the cannabidiol CBD but over a hundred other molecules have been identified as cannabinoids. One major limitation for their contribution to the development of treatments is the low abundance of most of those molecules and their similar biochemical properties, rendering their purification difficult.

In order to simplify the access to cannabinoids, this study aimed at transferring a part of the *C. sativa* cannabinoid biosynthetic pathway in *Phaeodactylum tricornutum* to produce a smaller diversity of cannabinoids. We transferred separately the two *Cannabis sativa* enzymes, the tetraketide synthase and olivetolic acid cyclase (respectively *Cs*TKS and *Cs*OAC), responsible for the production of olivetolic acid, the main precursor of cannabinoids in *P. tricornutum*. In parallel, we also transferred the slime mold *Dictyostelium discoideum*'s SteelyA enzyme. This enzyme, similarly to the combined action of *Cs*TKS and *Cs*OAC producing olivetolic acid, starts from the precursors hexanoyl-CoA and malonyl-CoA to produce olivetol.

This study confirmed the production of the three enzymes in vivo in *P. tricornutum* and tested the activity of the *Cs*TKS in transconjugants by untargeted metabolomic, showing a difference in the metabolic profile between *Cs*TKS transconjugants and negative control. The expression of *SteelyA* seemingly led to the production of an

insoluble protein lacking activity. Overall, this work shows that further knowledge is required on *P. tricornutum*'s metabolism to understand the underlying mechanisms hindering the olivetolic acid biosynthesis, and for future industrial bioproduction.

Key words: Diatom, cannabinoids, tetraketide synthase, olivetolic acid cyclase, heterologous expression, untargeted metabolomic, metabolism.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS AND ACRONYMS

ALS Amyotrophic lateral sclerosis

APT Aromatic prenyltransferase

CBD Cannabidiol

CBGA Cannabigerolic acid

CBGAS Cannabigerolic acid synthase

CDS Coding sequence

DNA Deoxyribonucleic acid

EPA Ecosapentaenoic acid

EV Empty vector

FAS Fatty acid synthase

FTICR-MS Fourier-transform ion cyclotron-resonance mass spectrometer

GPP Geranyl pyrophosphate

HPLC High pressure liquid chromatography

HPLC-MS High pressure liquid chromatography coupled with mass spectrometer

LUCA Last common universal ancestor

MEP Methylerytriol 4-phosphate

OAC Olivetolic acid cyclase

OLS Olivetol synthase

ORF Open reading frame

PKS Polyketide synthaseTHC Δ^9-tetrahydrocannabinol

THCAS Δ^9-tetrahydrocannabinolic acid synthase

TKS Tetraketide synthase

CHAPTER I: Introduction

1.1 Humans and bioproduction

To place the necessity of this study in our society, let's start from the beginning. After Earth's formation 4.3 billion years ago, this celestial body could have remained an inert solid mass forever like all the other celestial bodies of our solar system (at least to our current knowledge). Fortunately for us, life appeared around 3.8 billion years ago, allowing us today to read molecular biology theses. In order to trace all the life emerging on our blue planet, we developed phylogeny. This discipline retraces all lifeforms and their affiliations to each other. It starts with our Last Universal Common Ancestor (LUCA) from which sprouted the tree of life separating Archea, Eubacteria, and Eukaryotes (Figure 1.1). All the organisms that rose through billions of years of natural selection (and most recently artificial selection) developed a tremendous arsenal of anatomical, cellular, and molecular, tools to survive. Ranging from photosynthesis harnessing the light from the sun, to saprophagy, symbiosis, or predation, lifeforms generated an incredible variety of means to obtain energy, attract partners, sabotage competition or deter threats. This sprouting of diversity dispersed a countless number of molecules precious to human society today. Born from this diversity, humanity quickly sought to find the elements that would contribute to its survival. Humans thus studied their ecosystem and drew everything they could to contribute and improve their survival and development.

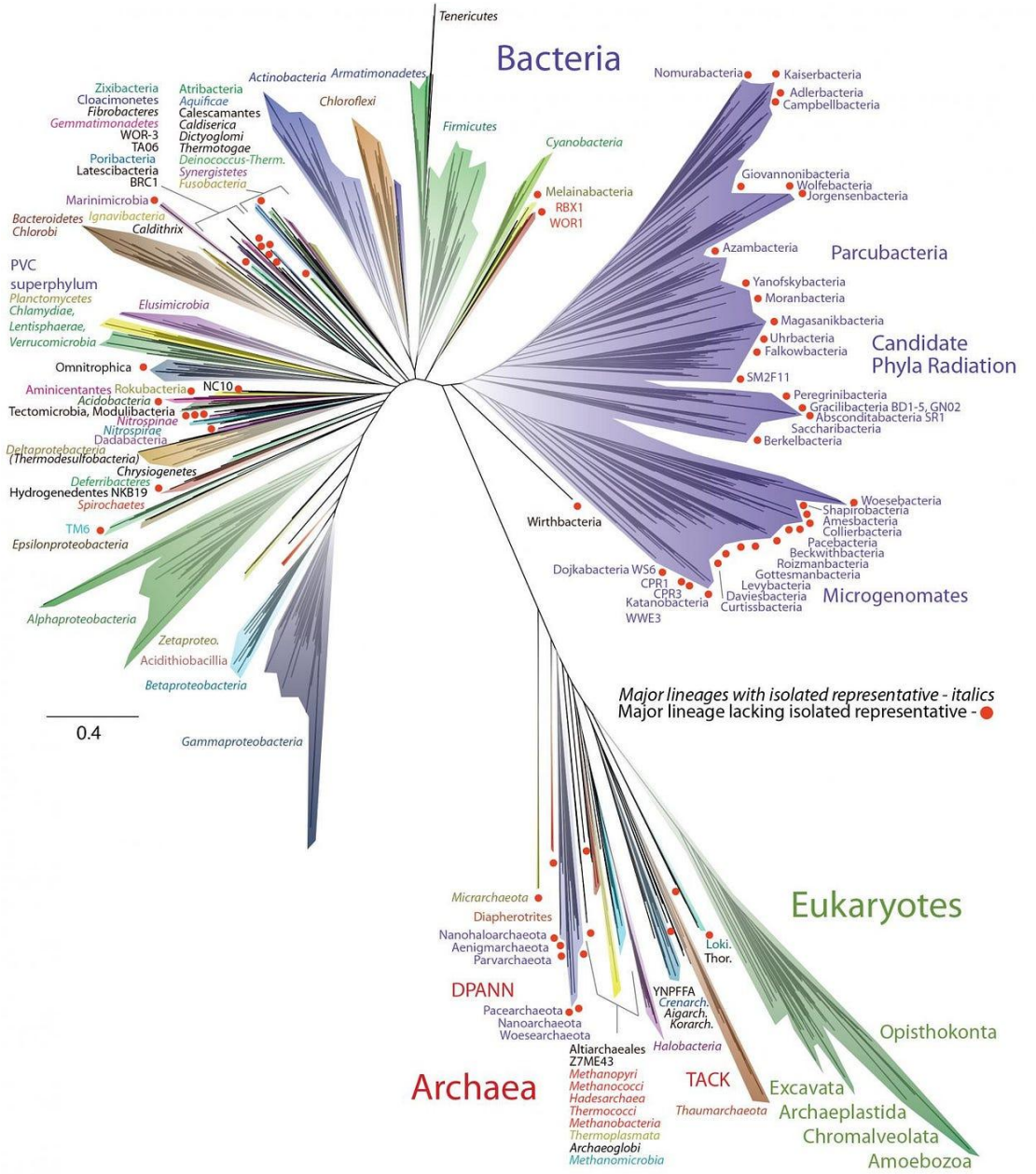


Figure 1.1 Phylogenetic tree of life

The tree includes 92 named bacterial phyla, 26 archaeal phyla and all five of the Eukaryotic supergroups. Major lineages are assigned arbitrary colours and named, with well-characterized lineage names, in italics. Lineages lacking an isolated representative are highlighted with non-italicized names and red dots. For details on taxon sampling and tree inference, see Methods. The names Tenericutes and Thermodesulfobacteria are bracketed to indicate that these lineages branch within the Firmicutes and the Deltaproteobacteria, respectively. Eukaryotic supergroups are noted, but not otherwise delineated due to the low resolution of these lineages. The CPR phyla are assigned a

single colour as they are composed entirely of organisms without isolated representatives and are still in the process of definition at lower taxonomic levels.

Source (Hug et al., 2016)

1.1.1 Survival to comfort, the need for extraction

To get a complete and healthy diet, humans extract organic matter from organisms in very diverse branches of the tree of life. To maintain homeostasis despite a diverse diet, the human microbiota includes Firmicutes and Bacteroidetes (depicted in Figure 1.1). Meanwhile, most human food sources originate from the Archaeplastida and Opisthokonta lineages (McCourt, 2016, Rinninella et al., 2019, Stojanov et al., 2020). Among the main bricks of life, the body needs to produce proteins composed of amino acids. It can however not produce all of them by itself and requires external intake (Teaford and Ungar, 2016, Mann, 2018). Those amino acids are called essential amino acids and can be found in milk, meat, and numerous diverse plants (Hou and Wu, 2018). The human body also needs lipids, sugars, vitamins, and minerals that are dispersed in different concentration in the world, to survive and avoid dysfunctioning. Fish and seafood bear omega-3 fatty acids that are good for the brain, selenium for the thyroid, and vitamins for homeostasis (Rayman, 2012, Lund, 2013), mushrooms contain minerals such as potassium, and copper, but also polysaccharides and vitamins[1] (Cheung, 2010, Roupas et al., 2012) and lipids can be found abundantly in animal fats, oils and nuts (Gorzynik-Debicka et al., 2018, Bhuyan et al., 2019, Alasalvar et al., 2020).

[1] Indigenous communities from the northern parts of the American continent thrived drawing vitamins from eating raw meat and raw fish and could not draw them from fruits and cereals as it did not grow in their region PARK, S., HONGU, N. & DAILY III, J. W. 2016. Native American foods: History, culture, and influence on modern diets. Journal of Ethnic Foods, 3, 171-177.

However, not all essential nutrients are equally accessible. Beyond survival, quality of life is also a key concern. To make them accessible to everyone and fit every person's need and deficiencies, society extracted nutrients from food and produced food supplements. Supplement market represented 152 billion USD in 2021 (Djaoudene et al., 2023). Protein production for protein shakes are heavily extracted from cow milk (Rodriguez-Lopez et al., 2022) and the protein shake market is evaluated at more than five billion USD on the global market in 2022 (Research, 2023a). The production of most supplements can be obtained by chemical synthesis, extractions from the organism naturally producing the compounds, or biosynthesis of engineered microorganisms not naturally producing those compounds (heterologous biosynthesis)(Lauterbach and Ober, 2000). All this development in society's needs and technology led us to rely tremendously on compounds extractions and chemical synthesis. Chemical synthesis bears the drawbacks of relying on possessing chemical data, reagents and physical parameters to obtain specific compounds thus rendering certain complex molecules inaccessible. On the other hand, extractions rely on obtaining compounds from a source naturally producing them. Extractions however bear different kind of burdens. When carried out from plants, consequent spaces and resources such as water, fertilizer and pesticides must be allocated to cultures to obtain enough quantities of compounds. Extractions of compounds from animals bring ethical questions. Finally, extractions can require the use of polluting solvents such as acetone, a volatile organic compound (one of majors air pollutant), and be highly energy consuming (Figure 1.2)(Kumar et al., 2021, Bitwell et al., 2023, Rakha et al., 2024). For example, drawing stevia out of *Stevia rebaudiana* Bertoni leaves requires reaching 100°C, while carbohydrates from barley's hull requires reaching temperatures as high as 160°C at 150 bars (Sarkar et al., 2014). Ultrasound assisted extraction even require temperatures to reach up to 5000°K to

extract compounds from plant matrix (Azmir et al., 2013).

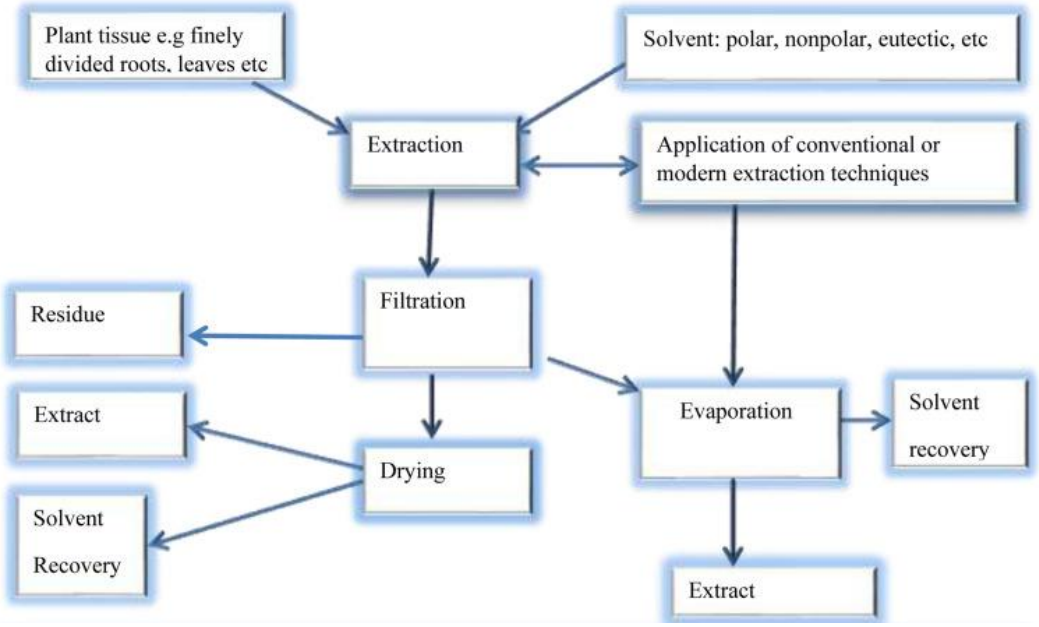


Figure 1.2 Scheme of steps involved in extraction of phytochemicals from plant sources.

Each box represents a step in the extraction process of molecules of interests. Arrows links the steps by incidence.

Source (Bitwell et al., 2023)

While nutrients are expected to ensure the proper functioning of body function, therapeutical molecules are defined by their ability to balance disorders. The distinction between a nutrient and a therapeutic molecule can sometimes be ambiguous. For example, palmitoleate, an omega-7 fatty acid found in macadamia nuts and avocados, has been shown to protect against Zika virus-induced trophoblast apoptosis (Muthuraj et al., 2022). Similarly, ginger exhibits antimutagenic, antioxidant, and anti-inflammatory properties (Puri et al., 2022). Foods with such medicinal properties are classified as nutraceuticals. While the integration of nutrition and healing has been a longstanding aspect of human culture, the diversity of ailments has driven efforts to

identify, isolate, and distribute the active molecules responsible for these therapeutic effects.

1.1.2 Humans and pharmaceuticals

The interactions of lifeforms with their ecosystem puts their homeostasis under pressure. Temperatures, infections, poisoning, hereditary ailment, wounds, the body is susceptible to many kinds of stress or damage. This need for cure brought human populations to study and develop remedies using the natural properties of the life surrounding them. To respond to the abundance of biotic and abiotic stresses, lifeforms produced a wide range of molecules and throughout history, societies have sought to harness these for therapeutic purposes (Hoffmann and Hercus, 2000). One of the most famous discoveries in this field is penicillin, a specialized metabolite produced by *Penicillium notatum* (an ascomycete). Alexander Fleming famously discovered its antibacterial properties by serendipity when he observed that it killed *Staphylococcus aureus* in a Petri dish (Fleming, 1944, Demain and Sanchez, 2009). The search for active molecules can be found in medicinal herbalism and be currently dated to 5000 years ago when populations used plants for healing and prophylaxis (Petrovska, 2012). *Asclepias*[2] genus is famous for bearing species widely used for their curing properties (expectorant, anti-inflammatory, cicatrization, anti-asthma, etc.) (Gaertner, 1979, Alonso-Castro et al., 2021, Gammatantrawet et al., 2024). The rod of Asclepios is still used as a medical symbol (Muhtaseb and Muhtaseb, 2022). This interest in medicinal plants is still currently high with an estimation of global market to surpass 100 billion

[2] Asclepios is the Greek God of medicine which gave the name to Asclepius plant, also known as Butterfly-weed. The Monarch butterflies lay their eggs on the Asclepias genus. A parallel can be found in mythology where similarly to Monarchs born on Asclepias, Macahon, name of a another butterfly, is one of the sons of Asclepios, FLOCKHART, D. T., PICHANCOURT, J. B., NORRIS, D. R. & MARTIN, T. G. 2015. Unravelling the annual cycle in a migratory animal: breeding-season habitat loss drives population declines of monarch butterflies. Journal of Animal Ecology, 84, 155-165.

USD per year and still more than 80% of the world's populations is using herbal drugs as main medicine (Vasisht et al., 2016). Willow bark was used as an analgesic and antipyretic for more than 3500 years, the source of its potency coming from the presence of salicin, now extracted abundantly for the production of aspirin (Desborough and Keeling, 2017).

As for salicin, the active molecule in Aspirin, society went from using medicinal plants to target and extract their active compounds of natural remedies to develop drugs. 25% of all medicines used worldwide come from plants (Rates, 2001) and more than 50% of drugs developed approved in pharmaceuticals are synthetic molecules derived from or inspired by natural molecules (Davis and Choisy, 2024). However, the abundance of molecules available today to cure the population covers about 22% of identified diseases (Espe, 2018, EveryCure, 2024). The increase in human longevity led to the spreading of diseases needing longer to take effect on the body such as cancer and neurodegenerative diseases (Crimmins, 2015). In the world, it is estimated that 53 M people are alive five years after being diagnosed with cancer (WHO, 2024), about 2.5 M people live with Huntington disease (Pringsheim et al., 2012), more than 50 M people live with Parkinson's disease (Gustavsson et al., 2023), about 350 000 people live with amyotrophic lateral sclerosis (ALS)(Xu et al., 2020), and more than 400 M live with Alzheimer's disease (Gustavsson et al., 2023). Significant efforts have been made to develop treatments for various disease; however, aside from cancer the neurodegenerative disorders mentioned above share two key characteristics: they currently have no identified cure and are all linked to the mammalian endocannabinoid system (Scotter et al., 2010, Cristino et al., 2020b, Vasincu et al., 2022). This system regulates numerous physiological functions through interactions of ligands with two cannabinoid receptors, cannabinoid type 1 receptor (CB1) and cannabinoid type 2

receptor (CB2) (Smith et al., 2010, Hilger et al., 2018). Consequently, a major research focus is aimed at finding and adapting molecules capable of modulating the endocannabinoid system. The discovery of the link between CB1- and CB2-binding molecules and their potential impact on neurodegenerative diseases has renewed interest in an already famous plant, the mother of cannabinoids: *Cannabis sativa*.

1.2 Cannabis

Cannabis is one of the oldest plants domesticated by humans (Hillig, 2005). It has the advantage of being easy to grow leading populations to cultivate it throughout the world (Halpin-McCormick et al., 2024). Originating in Central Asia, *Cannabis* has been transported and selectively bred across civilizations, making species differentiation increasingly complex (Yang et al., 2013, McPartland, 2017). An intensely debated theory gives two *Cannabis* species: *Cannabis sativa* (originating from India), and *Cannabis indica* (originating from central Asia)(Figure 1.3) (McPartland, 2017, McPartland and Small, 2020).

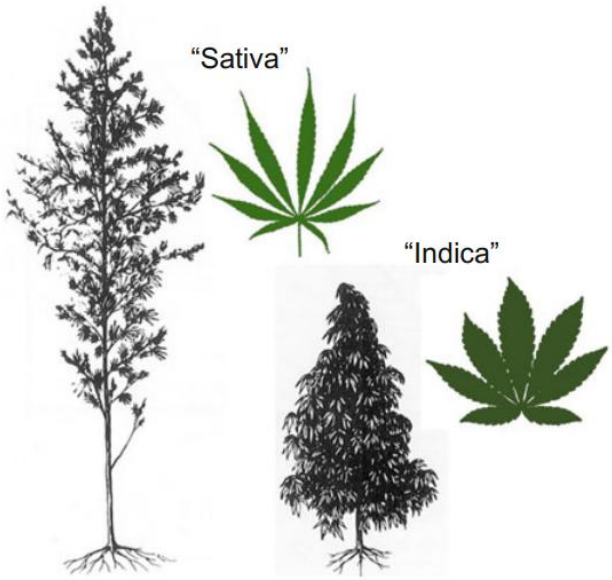


Figure 1.3 Cannabis vernacular taxonomy

Morphology of the plant and leaf of the two species of *Cannabis* defined as *sativa* (left) and *indica* (right).

Source (Anderson, 1980)

It is an annual plant (it takes a year to go from seed to seed) pollinated by wind and is dioecious, meaning that individuals have distinct sexual characters distinguishing male and female plants (Small, 2015, Govindarajan et al., 2023)(Figure 1.4). As this study is not trying to elucidate *Cannabis* phylogeny, it follows on this consensus herein.

Figure 1.4 Photography of *Cannabis sativa* in flowers.

Pistillate (female) plants at left, staminate (male) plants at right.

Source (Small, 2015)

Cannabis decline itself into various cultivars bearing different anatomical, physiological and biochemical properties. Civilizations therefore valued it for different reasons and selected and cultivated cultivars for purposes befitting those properties.

1.2.1 Cannabis plant in human society

Although *Cannabis* has always been famous for its recreational uses, numerous other uses made it valuable through history. Cultivars of *Cannabis* rich in fibers and low in psychoactive molecules are called 'industrial hemp', while the drug type 'medicinal cannabis', also known as marijuana, is coveted especially for its rich content in psychoactive molecules (Datwyler and Weiblen, 2006, Sawler et al., 2015). Grown mainly in temperate climate, hemp stalk contains the fibers that gives it its stiffness. Hemp fiber is water and rot resistant making it very valuable as a sail or recipient and its mechanical properties made it usable as rope. After treatment, the fiber could also be turned into fabric or paper[3] (Small, 2015, Musio et al., 2018). Beyond textiles and construction, hemp hurd, a byproduct of fiber production, has been used to make animal bedding. Varieties of hemp are nowadays cultivated to generate new composite materials with great mechanical properties as well as thermal and acoustic insulation (Shahzad, 2012, Musio et al., 2018, Crini et al., 2020).

Cannabis varieties grown for fibers were usually also bred for their achenes (dry fruit) also known as hempseeds. As they possess a high nutritional value with a rich content of polyunsaturated fatty acids, carbohydrates and proteins, hempseeds were consumed by prehistoric humans. Later in civilization's development hempseeds have been used to feed livestock (Callaway and Pate, 2009). An oil can be extracted from it and used in different applications encompassing nutrition, body care as soap and shampoo, painting for its property of polymerizing at ambient air, and more recently bio-plastics such from treated hemp paper (Callaway, 2004, Callaway and Pate, 2009, Beluns et al., 2023).

[3] Hemp has been used to make paper for one of the Gutenberg's bible printed in 15th century which is the first ever printed book using mass production methods, CRINI, G., LICHTFOUSE, E., CHANET, G. & MORIN-CRINI, N. 2020. Traditional and new applications of hemp. Sustainable agriculture reviews 42: hemp production and applications, 37-87.

The extraction of fiber and food from cannabis represented one of the main uses of *Cannabis* genus, however, it is not the one that made *Cannabis* notorious.

1.2.2 Cannabis *and cannabinoids*

Cannabis can be found in pharmacopeia all around the world. Early traces of cannabis mention date from 4000 BC in China as part of herbalist ingredients as well as 2000 BC in Egypt (Pisanti and Bifulco, 2019). *Cannabis* leaves or seeds were grinded and mixed with other ingredients to obtain remedies for different kind of a diseases (Russo, 2007). They could observe that mixtures containing cannabis extracts could present anthelminthic, antipyretic, antiseptic, anticonvulsant and painkiller effects (Hollister, 1986). It has therefore been associated with healing in Egyptian culture with the goddess of wisdom Seshat being often represented with what seems to be a cannabis leaf on her head (Figure 1.5).

Figure 1.5 Picture of a bas-relief of the ancient Egyptian goddess Seshat (Goddess of wisdom, knowledge, writing, and architecture)

She is depicted with what seems to be a *Cannabis* leaf in her headdress.

Source (Bonini et al., 2018)

Traditional Chinese medicine used the oil to treat rheumatism, malaria, and fatigue and in some rural parts of Nepal, *Cannabis* plants are still used to treat human and livestock gastrointestinal diseases (Abbasi et al., 2013).

Among the active molecules contained in *C. sativa* and giving it its medicinal properties, some of them are psychoactive. Focusing on exploiting those molecules, around 500 years BC after burying their dead, Scythians horsemen in Siberia purified their bodies by throwing cannabis seeds in fires, inhaling the smoke, then "shrieked with delight at the fumes" (Russo, 2007), perhaps a tradition that should be reinstated after successful experiments... If other psychoactive mixtures were consumed like psilocybin and mescaline, cannabis was often used for ritualistic purpose as it could generate a feeling of spiritual elevation. This property led cannabis to be defined as entheogen[4] (Sayin, 2016, Ferrara, 2021). *Cannabis* is observed to have been used in shamanic trances for more than 10 000 years with traces being found originally in central Asia expanding to south America (Sayin, 2016). So, what makes this plant have those unique properties?

The impact of *Cannabis* on health and psyche is mainly due to its content in molecules named cannabinoids. In plants, they have shown antimicrobial effects, and are suspected to help defend against herbivores by being released from their stocking cells

[4] a psychoactive, hallucinogenic substance or preparation especially when derived from plants or fungi and used in religious, spiritual, or ritualistic contexts, "Entheogen." Merriam-Webster.com Dictionary, Merriam-Webster, https://www.merriam-webster.com/dictionary/entheogen. Accessed 20 Feb. 2025.

when mechanical pressure is applied to them and by their sticky texture and active properties hinder insect mandibles as well as disorienting bigger herbivores (Garrett and Hunt, 1974, Gülck and Møller, 2020, Sands et al., 2023). Besides their cyclic structure is expected to absorb and protect against UV and intense light (Figure 1.6)(Eichhorn Bilodeau et al., 2019).

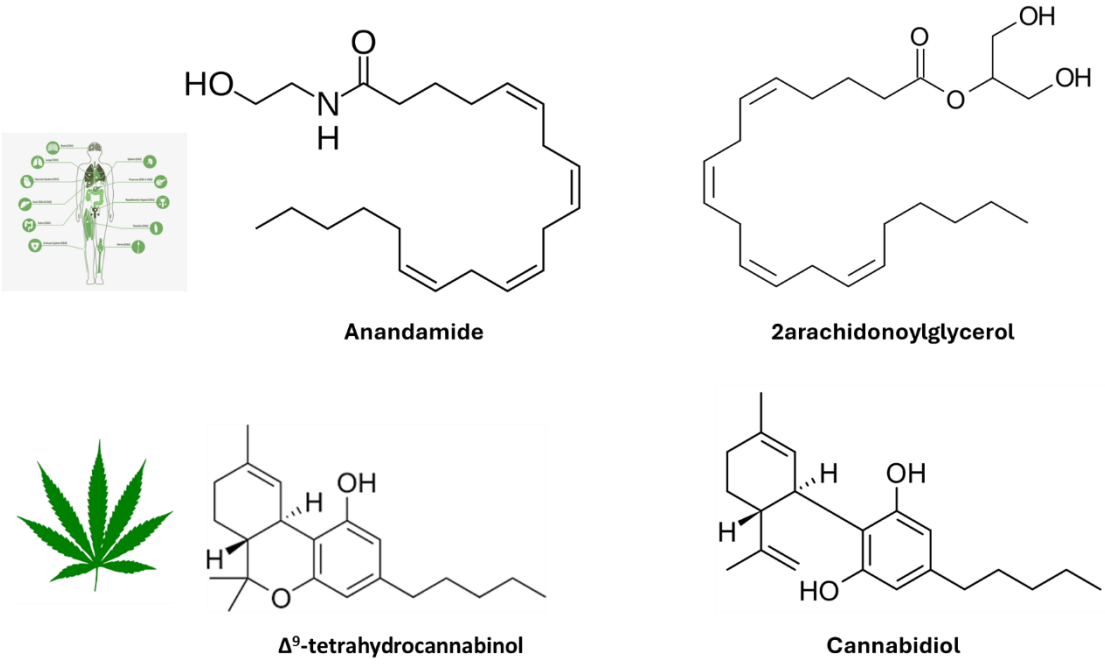


Figure 1.6 Structure of the most active endocannabinoids and phytocannabinoids

The top left and right molecules depict the structure of the two most active endocannabinoids detected so far. The two bottom molecules present the structure of the most abundant cannabinoids in *Cannabis sativa*.

This family of molecules is defined by their ability to interact with the mammalian endocannabinoid system (Ameri, 1999, Howlett and Abood, 2017). The most famous and abundant cannabinoids in the plant are Δ^9-tetrahydrocannabinol (THC) and cannabidiol (CBD) and through their direct interaction with the endocannabinoid system are responsible for a wide range of cannabis' effects. The mammalian endocannabinoid systems is regulated by the activation of two G-protein-coupled:

receptors CB1 and CB2 (Smith et al., 2010, Hilger et al., 2018). In humans, CB1 is mainly located in the central nervous system with higher concentrations in cerebral cortex, hippocampus and cerebellum, but also heart, lung and uterus (Howlett et al., 2002, Thakur et al., 2005), while CB2 is majorly located in the brainstem, spleen and immune cells (Van Sickle et al., 2005, Hashiesh et al., 2023) (Figure 1.7).

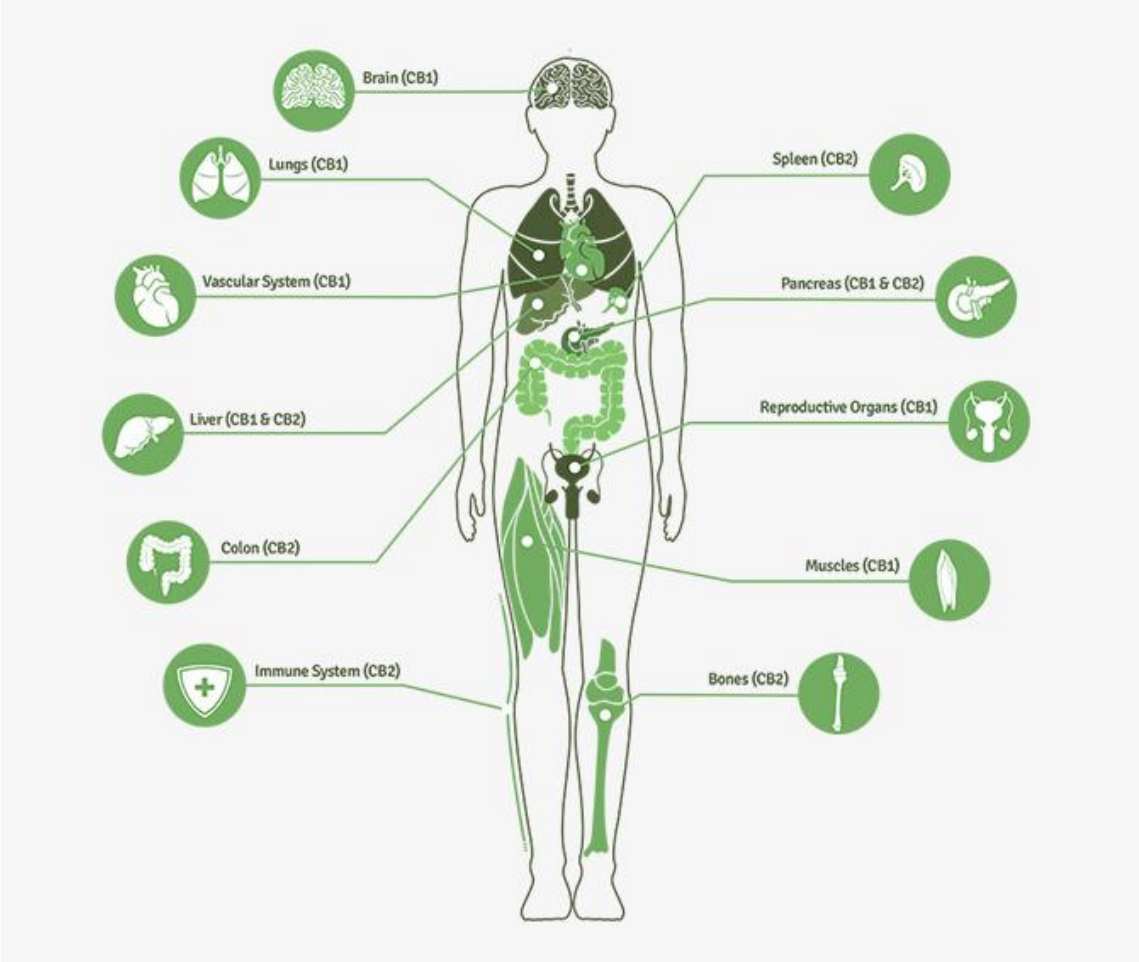


Figure 1.7 Endocannabinoid system in humans

Representation of the localization of CB1 and CB2 receptors in the human body.

Source (Ryan Zaklin, 2019)

Normally, those receptors are activated by molecules called endocannabinoids of which the most bioactive examples are anandamide[5] and 2-arachidonoylglycerol (figure 1.6) (Battista et al., 2012, Lu and Mackie, 2021). CB1 receptor was discovered when studying the effect of cannabis on humans which gave the name "cannabinoid receptor" and "endocannabinoids" for molecules naturally produced by the body capable of interacting with them (Howlett et al., 1986, Herkenham et al., 1990, Devane et al., 1992). This system regulates numerous functions in the human body such as appetite, learning, anxiety, depression, stroke, multiple sclerosis, neurodegeneration, sexual drive, epilepsy, etc. (Skaper and Di Marzo, 2012, Zou and Kumar, 2018, Śledziński et al., 2020)(Figure 1.8).

[5] Anandamide comes from the Sanskrit word "ananda" meaning "bliss", and can be found in chocolate. It has been observed in rat that low doses of anandamide increased sexual activity by increasing the amount of ejaculations and decreased the time between two ejaculations during coitus while high doses have adverse effects TUNICK, M. H. & NASSER, J. A. 2019. The chemistry of chocolate and pleasure. *Sex, smoke, and spirits: The role of chemistry.* ACS Publications, ANDROVICOVA, R., HORACEK, J., STARK, T., DRAGO, F. & MICALE, V. 2017. Endocannabinoid system in sexual motivational processes: Is it a novel therapeutic horizon? *Pharmacological Research,* 115**,** 200-208, ibid., ibid.

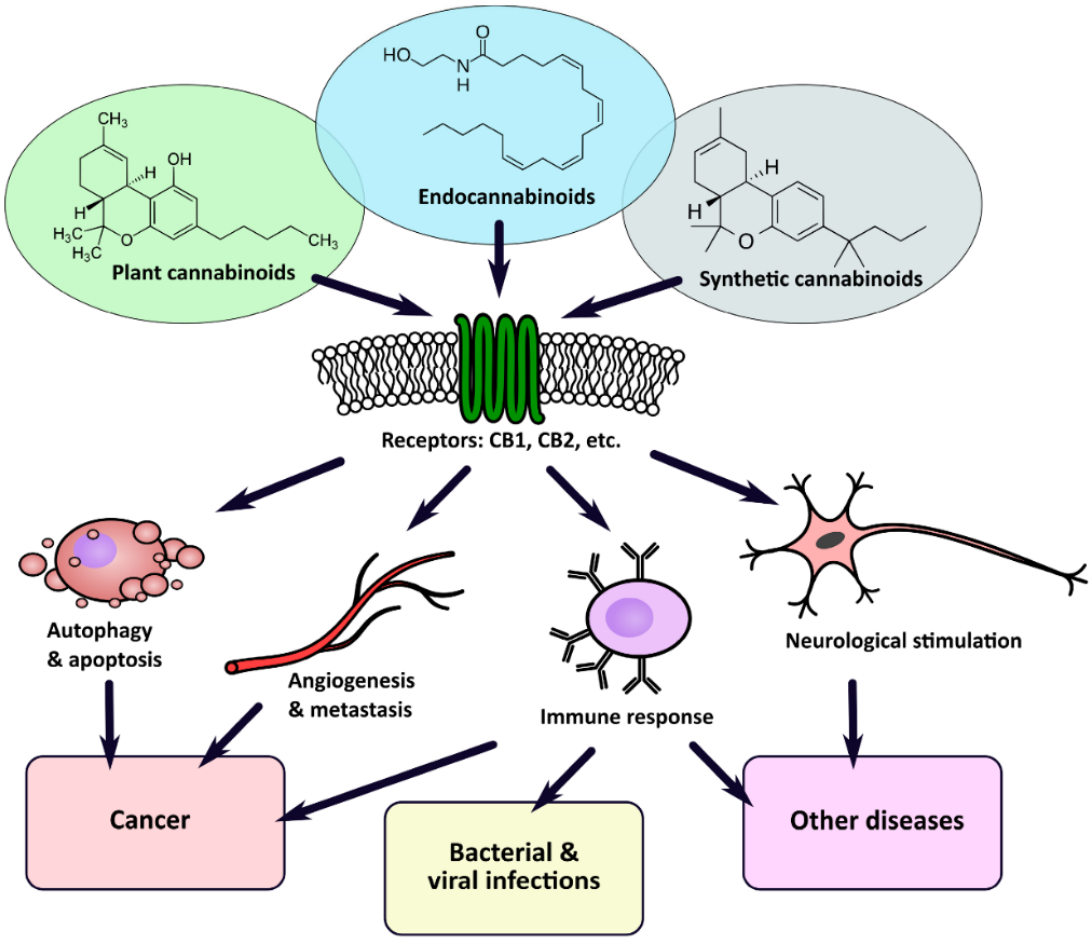


Figure 1.8 Endocannabinoid system functions

The action of cannabinoids on the cannabinoid receptors modulates numerous functions. The involvement of the endocannabinoid system in a variety of modulatory processes makes it a promising target in the therapy of many conditions.

Source (Śledziński et al., 2020)

The discovery of molecules capable of interacting with the endocannabinoid system and their capability of being administrated increased the potential of developing treatments for diseases so far uncurable. THC is an agonist (triggers a biological response) of CB1 and CB2 with a higher affinity for CB1, while CBD is a low affinity antagonist (blocks or dampens a biological response) of CB1 and agonist of CB2 (Malfitano et al., 2014, Bie et al., 2018, Pellati et al., 2018) and therefore mitigates the

effects of THC. Molecules produced by plants (especially *Cannabis*) capable of interacting with the cannabinoid receptors are called phytocannabinoids.

Henceforth, cannabinoids, particularly THC and CBD, are currently studied to develop treatments for diseases or alleviate symptoms. Their healing effects already domesticated by several cultures are now put to the test as analgesic, antiemetic during chemotherapy, aphrodisiac or even as cancer treatment (Pertwee, 2000, Pertwee, 2001, Porter and Felder, 2001, Howlett and Abood, 2017, Afrin et al., 2020, Gupta et al., 2020, Scheau et al., 2020). In the case of glioblastoma, the use of a mix containing Sativex (1:1 ratio of THC and CBD) yielded to increase the one year survival rate of patients by 39% (Afrin et al., 2020). Evidence also led to considering that agonists of CB2 leads to apoptosis of inflammatory tissues and immunoregulation (Herring et al., 1998, Afrin et al., 2020). Consequently, CBD binding to CB2 has been found to have anti-inflammatory effects in Crohn's disease and Parkinson disease (Pellati et al., 2018).

THC has been observed to enhance cognitive functions and could potentially enhance mitochondrial functions (Russo, 2007, Rao, 2024). Nabiximols, a mixture of THC or CBD and non-psychotic cannabinoids, is currently tested to ease neuropathic pain in multiple sclerosis and is tested for untreatable forms of pediatric epilepsy (Cristino et al., 2020b). Attempts are therefore actively being made using phytocannabinoids to regulate the endocannabinoid system in order to ease and treat complicated and long-lasting pathologies.

Among all phytocannabinoids detected in *Cannabis* plants, THC and CBD are high in concentration the most famous as they also are the most abundant. Some other phytocannabinoids are abundant enough to be tested such as cannabichromene and tetrahydrocannabivarin and are still considered major cannabinoids, both presenting

controversial effects linked to reducing inflammation (Turner and Elsohly, 1981, Pollastro et al., 2018, Walsh et al., 2021b). However, for numerous of them called minor cannabinoids, their low quantity and similar molecular properties to THC and CBD make them harder to separate and thus analyze.

1.2.3 Biosynthetic routes of cannabinoids

The complete biosynthetic pathway of cannabinoids in *C. sativa* was elucidated in the late 20th and early 21st centuries (Fellermeier and Zenk, 1998, Fellermeier et al., 2001). Although trace amounts of phytocannabinoids can be found in seeds, leaves, and even roots of certain cultivars, their primary site of production is a structure called the glandular trichome (Brousseau et al., 2021, Tanney et al., 2021, Govindarajan et al., 2023). In*Cannabis*, glandular trichomes are multicellular structures scattered in highest abundance on the surface of female flowers. Cannabinoid biosynthesis occurs within the secretory cells of these trichomes, with metabolites accumulating and getting stored in the secretory cavity (Figure 1.9)(Tahir et al., 2021).

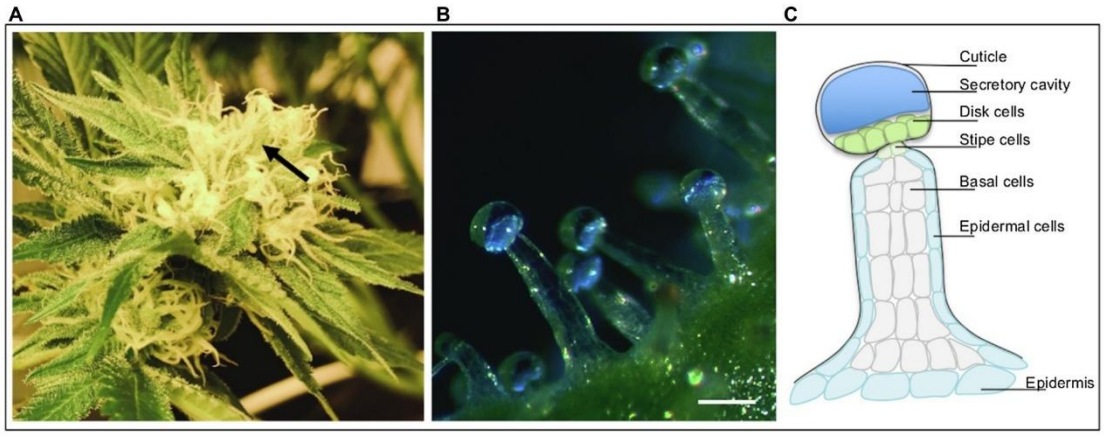


Figure 1.9 *Cannabis sativa* female flower inflorescence and trichomes.

A. An individual inflorescence, with majority of the organs covered in stalked glandular trichomes. Arrow indicates cluster of calyces and bracts covered with trichomes. B. Dark field micrograph of stalked glandular trichomes. The metabolites

are stored in the clear subcuticular cavity above the secretory disk cells; this cavity will turn milky white to dark brown over the course of flower maturity. C. Graphic illustration of stalked glandular trichome structure. Biosynthesis of secondary metabolites occurs in the secretory disk cells lining the base of the globular trichome head.

Source (Tanney et al., 2021).

The biosynthesis of cannabinoids is initiated in the cytosol of the disc cells by the condensation of hexanoyl-CoA, and malonyl-CoA by the tetraketide synthase (TKS), formely known as olivetol synthase (OLS), yielding 3,5,7-trioxododecanoyl-CoA tetraketide (Figure 1.10). Olivetolic acid cyclase (OAC) then cyclizes this intermediate to produce olivetolic acid (OA). OA is transported to the chloroplast where aromatic prenyltransferase (APT) adds a geranyl pyrophosphate (GPP), derived from the methylerytriol 4-phosphate pathway (MEP), resulting in cannabigerolic acid (CBGA). Finally, CBGA is transported to the apoplast (an extracellular space between plant's cell walls facilitating fluid movement) where it undergoes enzymatic conversion by either Δ^9-tetrahydrocannabinolic acid synthase (Δ^9-THCAS simplified to THCAS) or cannabidiolic acid synthase (CBDAS), leading to the production of Δ^9-tetrahydrocannabinolic acid (THCA) or cannabidiolic acid (CBDA), respectively (Figure 1.10) (Degenhardt et al., 2017, Gülck and Møller, 2020, Govindarajan et al., 2023, Sands et al., 2023).

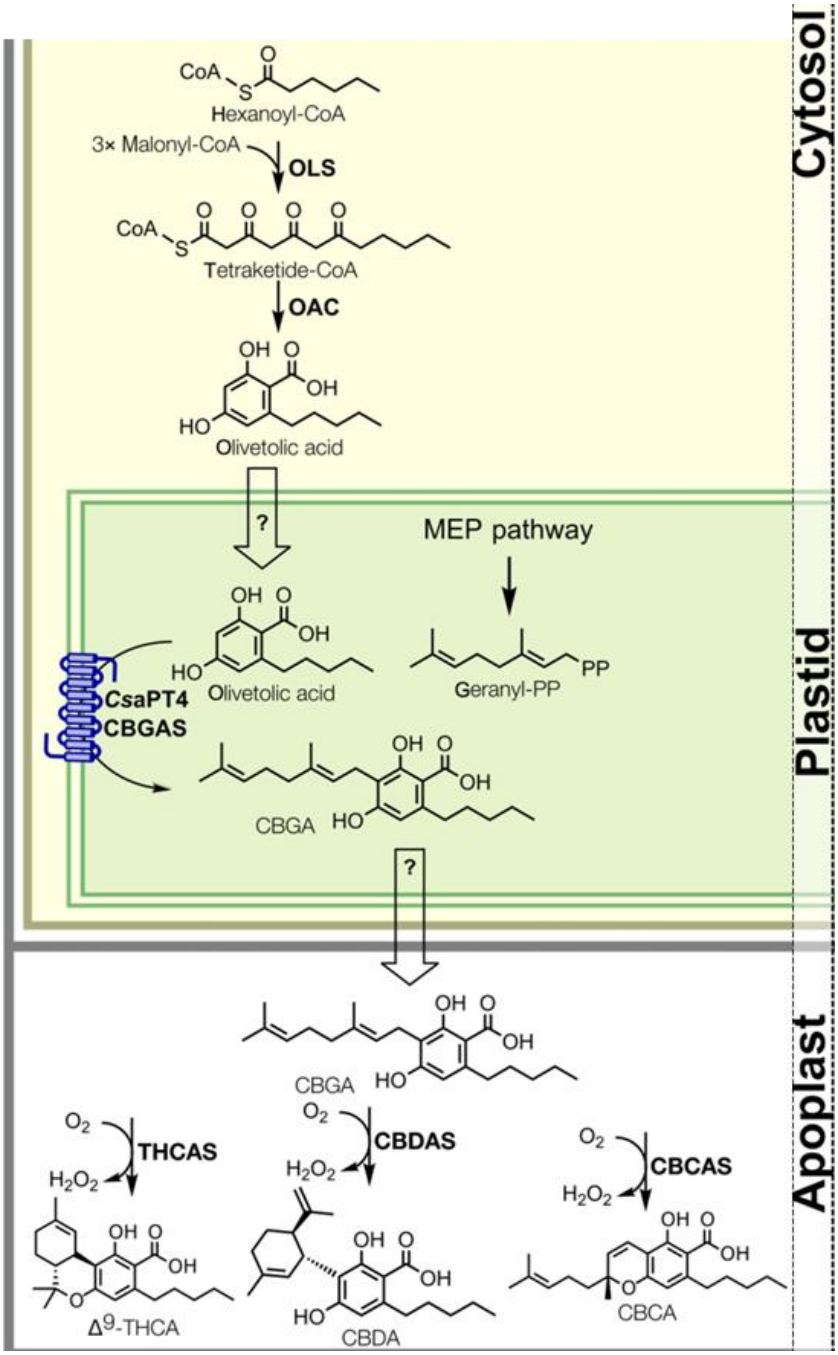


Figure 1.10 Main reactions catalyzing phytocannabinoid biosynthesis in *Cannabis sativa* and the subcellular distribution of enzymes

OLS: Olivetol synthase or Tetraketide synthase; OAC: Olivetolic acid cyclase, *Cs*apt4/CBGAS: *Cannabis sativa* aromatic prenytransferase/cannabigerolic acid synthase; THCAS: Δ^9-tetrahydrocannabinolic acid synthase; CBDAS: Cannabidiolic

acid synthase; CBCAS cannabichromenic acid synthase. The bold arrows represent known reactions while empty arrows represent unknown mecanisms.

Source (Gülck and Møller, 2020)

The acid form of cannabinoids is then secreted and accumulated in the secretory cavity (Figure 1.11)(Conneely et al., 2021, Xie et al., 2023).

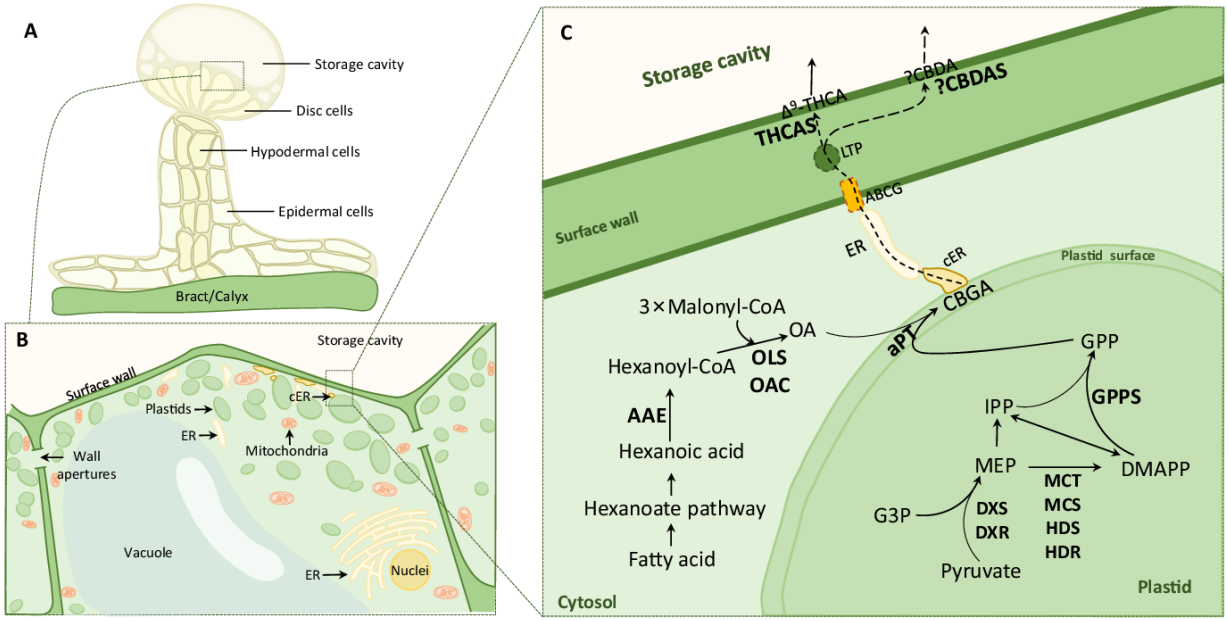


Figure 1.11 Location of the phytocannabinoids biosynthesis reactions in the *Cannabis sativa* glandular trichomes

A. Overall glandular trichome structure on *Cannabis sativa* female flowers. B. Disc cell intracellular structure. C. Intracellular localization and interactions of enzymes and substrates necessary for the biosynthesis of cannabinoids.

Source (Xie et al., 2023)

It is the action of heat that eventually entails the decarboxylation of CBDA and THCA and forms THC and CBD.

To summarize, starting from hexanoyl-CoA and malonyl-Co, the action of two enzymes (TKS and OAC) lead to the production of olivetolic acid the main phytocannabinoid precursor, then two more enzymes (CBGAS and THCAS or CBDAS) are required to

produce THCA and CBDA. Interestingly, the four major enzymes CBGAS, THCAS, CBDAS, and cannabichromenic acid synthase (CBCAS) are hypothesized to be responsible for the biosynthesis of most the hundred identified cannabinoids (Degenhardt et al., 2017).

As shown in figure 1.2, extraction of compounds from plants can be costly and complicated. In the case of cannabinoids, the discrepancy in abundance between female and male flowers forces to apply genetic selection to crops or to discard male plants. Besides, the abundance ratio of THCA and CBDA, and the minor cannabinoids and the similarity of their structures render them difficult to separate and purify (Luca et al., 2021). Finally, the main extractions techniques in use are currently generating waste and pollution, therefore a search for alternative source of cannabinoids is being actively carried out (Lazarjani et al., 2021).

1.2.4 Alternative synthesis routes of cannabinoids

C. sativa has for long been thought to be the only organism naturally producing cannabinoids. Its close history to human societies, its easy cultivation and the ability to cross it to match needs made it the main source of phytocannabinoids until now. Recently, more organisms have been discovered producing naturally fewer minor cannabinoids (Figure 1.12)(Gülck and Møller, 2020). Indeed, some *Rhododendron* species produce cannabichromene-like and cannabicyclol-like (another minor cannabinoid) molecules in their flowers that inhibit histamine release in humans and could thus act as anti-allergic agents (Yang et al., 2018, Gülck and Møller, 2020). The liverwort *Radula marginata* uses stilbene acid or dihydrostilbene acid to produce bibenzyl CBGA analogs: perrottetinenic acid. We can also find production of phytocannabinoids in other kind of organism than plants as the parasitic mushroom C*ylindrocarpon olidum* has shown production of cannabiorcichromenic and 8-

chlorocannabiorcichromenic acid which are homologs of cannabichromenic acid (Quaghebeur et al., 1994). One issue remains, it is the low/variable abundance of those molecules in their host.

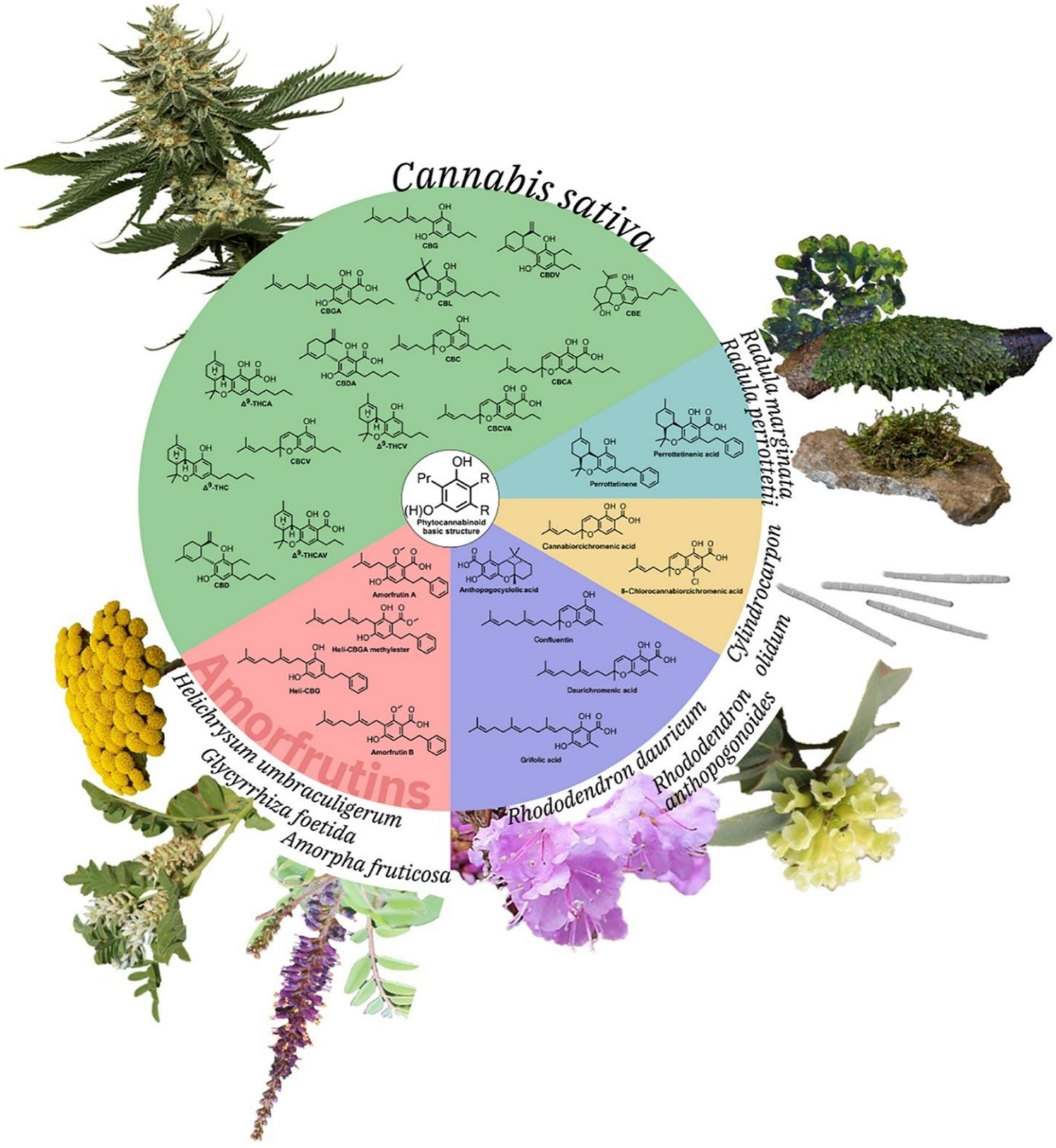


Figure 1.12 Illustration of the Structural Characteristics of Cannabinoids Found in Different Plant Species

One way of attributing molecules to the cannabinoids family is through their structure. A molecule can be defined as a cannabinoid if it possesses a resorcinol ring, two carbon chains and a propyl group between the two hydroxyl groups. Abbreviations: CBC, cannabichromene; CBCA, cannabichromenic acid; CBCV, cannabichromevarine; CBCVA, cannabichromevarinic acid; CBD, cannabidiol; CBDA, cannabidiolic acid; CBDV, cannabidivarine; CBE, cannabielsoin; CBG, cannabigerol; CBGA, cannabigerolic acid; CBL, cannabicyclol; Δ^9-THC, Δ^9-tetrahydrocannabinol; Δ^9-THCA, Δ^9-tetrahydrocannabinolic acid; Δ^9-THCAV, Δ^9-tetrahydrocannabivarinic acid; Δ^9-THCV, Δ^9-tetrahydrocannabivarine.

Source (Gülck and Møller, 2020)

To date, if no other organism has been found producing THCA or CBDA (Gertsch et al., 2010), other organisms produce olivetolic acid, or its decarboxylated form, olivetol. Although no cannabinoids have been detected yet, the lichen *Cetrelia monachorum* produces olivetol and olivetolic acid (Oettl et al., 2013). Surprisingly, we also find olivetol in the animal kingdom with the ant *Crematogaster deformis* producing 5-pentylresorcinol (olivetol) as a defense mechanism (Attygalle et al., 1989).

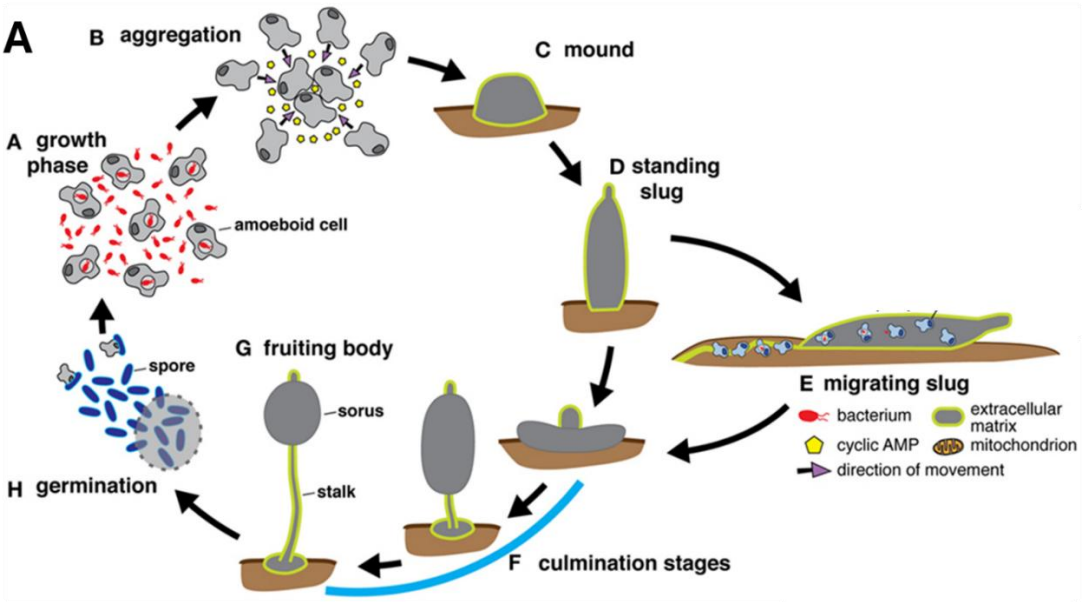
Among all the organisms discovered to be capable of producing olivetol, one presented a particular potential.

1.2.5 Dictyostelium discoideum

The slime mold *Dictyostelium discoideum* is a protozoon with a life cycle that is defying the definition of individuality. Indeed, it can spend its whole life cycle as an autonomous unicellular organism (like any conventional protozoon), however, when the cells are exposed to famine, they agglutinate into a slug superorganism[6] and search for nutrients before forming a mushrooms-like structure (Figure 1.13)(Aubry and Firtel, 1999,

[6] A group of organisms which function together in a highly integrated way to accomplish tasks at the group level such that the whole can be considered collectively as an individual, Cronin, A.L. (2022). Superorganism. In: Vonk, J., Shackelford, T.K. (eds) Encyclopedia of Animal Cognition and Behavior. Springer, Cham. https://doi.org/10.1007/978-3-319-55065-7_383

Fisher, 2001, Dunn et al., 2018). Cells previously autonomous then differentiate into stalk cells, and sorus cells (thus abandoning individuality) where spores are being produced before being scattered and form new individuals. Conversly to an ant colony (also considered as a super organism) where ants are separated by their functions at birth with soldiers, breeders, and workers, *D. discoideum* is an autonomous cell that can differenciate itself to becomes a part of a bigger organism. To do so, it has to abandon its other feeding and dividing functions and fit the goal attributed to it during differenciation. It can be compared to if humans fused with all the people around them to become part of a super snake and started looking for food before differenciating into a foot or a nipple for big scale reproduction...



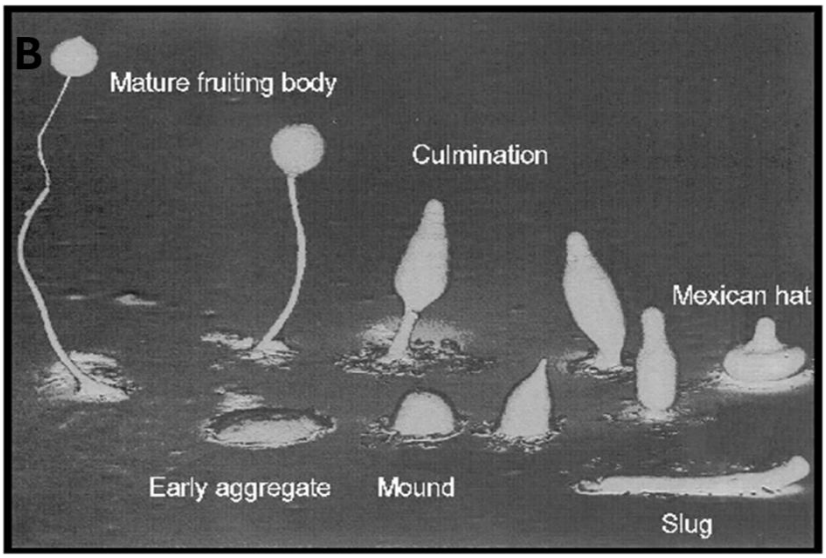


Figure 1.13 Dictyostelium *discoideum* differentiation cycle.

A. schematic representation of the cell differentiations steps. B. Scanning electron micrograph of *Dictyostelium discoideum* differentiation cycle.

Source (Fisher, 2001, Dunn et al., 2018)

During this differentiation, similarly to *C. sativa*, the slime mold produces methyl-olivetol from malonyl-CoA and hexanoyl-CoA under the action of a sole enzyme called SteelyA (Figure 1.13)(Reimer et al., 2022). This massive enzyme of 352 kDa is a hybrid of a multidomain type I fatty acid synthase (FAS) in N-terminal, and a type III polyketide synthase (PKS) in C-terminal. Besides the production of methyl-olivetol that plays a role as a differentiation factor, it also produces 4-methyl-5-pentylbenzene-1, 3-diol (MPBD) inducing spore differentiation (Narita et al., 2011, Reimer et al., 2022). SteelyA enzyme is therefore capable of cumulating the action of the *C. sativa* TKS (*Cs*TKS) and OAC (*Cs*OAC) on its own. Even if no cannabinoids have been detected so far, the interest in this organism has grown for its natural production of cannabinoid precursors, and its easy culture in laboratory conditions (Annesley and Fisher, 2009).

To circumvent the drawbacks of their complicated extraction, an alternative to obtain cannabinoids is their chemical synthesis. Called synthetic cannabinoids, those molecules were originally produced to study the endocannabinoid system, but their use

is currently split between therapeutical use, and illegal recreational use. They are analogs of phyto or endocannabinoids and tend to present a higher toxicity than natural phytocannabinoids which majorly impairs their use in clinical conditions (Mills et al., 2015, Le Boisselier et al., 2017). Therefore, new leads turned towards rerouting organisms as factories for the production of biological compounds.

1.3 Biofactories

To survive, every organism biosynthesizes a wide array of molecules. Biosynthesis, through the power of enzymes, achieves transformations that physical or chemical processes could accomplish only at great energetic cost or time scale. By guiding substrates through low-energy intermediate steps or enhancing their reactivity, enzymes carry out reactions with a dramatically increased efficiency and precision (Chaplin and Bucke, 1990).

1.3.1 Production of valuable products through heterologous biosynthesis

In order to use biosynthesis as a tool for the acquisition of rare, costly or hardly accessible molecules, it was first required to find and tame the appropriate organisms. *Escherichia coli* was discovered by the pediatrician and microbiologist Theodore Escherich in 1885 in children lower digestive tractus (Escherich, 1886, Hacker and Blum-Oehler, 2007). It has since been the most studied organism thanks to its numerous advantages: it is very easy to obtain, grows at a very fast pace, is cost effective and easy to cultivate, is conventionally harmless although it is an opportunistic organism, and is easy to mutate. That is why through time, it became the most tamed organism in the world: the first model organism (Zimmer, 2009, Taj et al., 2014, Blount, 2015). Slowly, the amount of data amassed on this organism and others became sufficient to give birth to a new discipline: systems biology. This field establishes models to better understand

the complex and integrated interactions of molecules, pathways, and signals in an organism, usually leading to applied purposes (Bruggeman and Westerhoff, 2007). It has since been used as a base for the building of chassis strains (strains that are modified for industrial production purposes)(Calero and Nikel, 2019).

This knowledge allowed the modifications of organisms using metabolic engineering methods leading to the development of synthetic biology. This discipline aims at applying more or less extensive genetic modifications to an organism in order to increase the intensity of a native metabolic pathway, or introduce a heterologous biosynthetic pathway, usually for bioproduction or bioremediation (Benner and Sismour, 2005, Xu et al., 2012b).

Enhancing or transferring pathways to *E.coli* allowed for the cost effective and clean production of many compounds such as solvents (ethanol, methanol)(Clomburg and Gonzalez, 2010, Reiter et al., 2024), biofuels (Koppolu and Vasigala, 2016), biomaterials (Webb et al., 2018) and pharmaceuticals (vitamins, proteins, specialized metabolites, etc.) (Baeshen et al., 2015, Fang et al., 2018, Yang et al., 2020).

This gram-negative bacterium still bears the drawback of being a prokaryote and therefore processes a smaller diversity of post-translational modifications. As more than 650 protein modifications have been identified today (among the most famous are glycosylation, phosphorylation, methylation, etc.), using prokaryotes as sole model for production restrains dramatically the possibilities of obtention of valuable proteinic compounds (Zhong et al., 2023). The budding yeast *Saccharomyces cerevisiae* thus came to the rescue. This eukaryotic organism has a closer relationship to humanity than *E. coli* as it has been used for millennia to produce the famous and coveted: beer. Louis Pasteur discovered its implication in alcoholic fermentation in 1858 (Pasteur, 1858).

Since then it has been used in laboratory to understand eukaryotic metabolism and turned into a biofactory for products that would be requiring post translational modifications, that would be too complex, or require extensive modifications for *E.coli,* (Huang et al., 2008, Grewal et al., 2018, Yang et al., 2024a).

With the increasing knowledge on organisms and biosynthetic pathways, especially of molecules of interest, the range of biofactories has widened to find the most appropriate hosts for bioproduction of the variety of coveted compounds. There is now a wide array of organisms used for biosynthesis presenting each their own advantages for the production of valuable molecules (figure 1.14).

Organism	Advantages	Disadvantage	Main industrial uses	References
Escherichia coli	✚ Fast growth ✚ Simple culture procedures ✚ Cost-effective ✚ Adaptable ✚ Widest genetic and metabolic knowledge ✚ Widest molecular toolbox available	✚ Fewer post-translational modifications ✚ Risk of translational errors because of a large number of rare codons ✚ Expensive and often challenging purification process ✚ Easily contaminated	✚ Pharmaceuticals ✚ Nutraceuticals ✚ Food industry (sweeteners, pigments, etc.) ✚ Cosmetics ✚ Biofuels ✚ Chemicals	(Avesani et al., 2014, Zhang et al., 2016, Mendez-Perez et al., 2017, Pham et al., 2019, Yang et al., 2020)
Bacillus sp	✚ Fast growth ✚ Simple culture procedure ✚ Gram positive ✚ Outstanding fermentation properties and protein production yield ✚ Adaptable	✚ Fewer post-translational modifications ✚ Primarily used in enzyme production. ✚ Plasmid instability ✚ Proteases hinder the production of recombinant proteins.	✚ Pharmaceuticals ✚ Nutraceuticals ✚ Chemicals	(Pham et al., 2019, Lajis, 2020, Yang et al., 2024b)

		• CRISP-Cas9 methods are often toxic		
Saccharomyces cerevisiae	• Fast growth rate • Cost-effective • Advanced fermentation knowledge • Wide genetic and metabolic knowledge • Wide molecular toolbox available • Crabtree positive • G.R.A.S.	• Crabtree positive • Glycosylation of protein different from humans • Codon usage is biased toward A and T	• Pharmaceuticals • Nutraceuticals • Food industry • Chemicals • Cosmetics	(Nielsen, 2013, Rasala and Mayfield, 2015, Pham et al., 2019, Parapouli et al., 2020)
Chinese hamster ovarian cell lines	• Abundance of culture media available and easily adaptable • High productivity • Human profiled glycosylation • Close model to human physiology	• Cell-lines can be very specialized • Costly • Complex proteins synthesis can have low yields • Vulnerable to contamination • Vulnerable to shear forces of agitations • Accumulates waste easily • Easily contaminated • Time consuming cell-line development	• Pharmaceuticals	(Walker et al., 2005, Omasa et al., 2010, Rasala and Mayfield, 2015, Stolfa et al., 2018, Kelly et al., 2018, Henderson, 2020, Arias et al., 2022)
Cell free system	• Easy manipulation and monitoring • Long term storage • Practical adjustments of reactions conditions • Fast reaction rates • Tolerance to toxic products • High yield • No cell maintenance	• Costly • Complicated purification	• Pharmaceuticals • Chemicals	(Li et al., 2018, Bogart et al., 2021, Buntru et al., 2022, Claassens et al., 2019)
Plants (*Nicotiana benthamiana* and *Arabidopsis thaliana*)	• Photosynthetic • Economical • Easy scale up • Consumable extracts	• Hard to contain ecologically • Slow growth	• Pharmaceuticals • Nutraceuticals • Food industry • Cosmetics • Chemicals	(Twyman et al., 2003, Walker et al., 2005, Rigano et al., 2009, Sharma and Sharma,

	• Stable transformants without selective pressure • High rates achievable • Nucleus and plastid transformable • Adaptable	• Vulnerable to pests • Contamination of product by mycotoxins, pesticides and herbicides possible • Glycosylation different from humans		2009, Obembe et al., 2011, Xu et al., 2012a)
Chlamydomonas reinhardtii	• Photosynthetic • Fast growth • Can accumulate fatty acids • Low maintenance cost • Green product extraction • G.R.A.S/ organism • Nucleus and plastid transformable	• Low recombinant protein accumulation • Inconsistant nuclear transformation • Variable strain stability	• Pharmaceuticals • Nutraceuticals • Food industry • Cosmetics • Biofuels	(Walker et al., 2005, Rasala and Mayfield, 2015, Abate et al., 2015, Scranton et al., 2015, Fernandes and Cordeiro, 2021, Arias et al., 2022)
Phaedactylum tricornutum)	• Photosynthetic • Fast growth • Rich in fatty acids • Green product extraction • Can be grown outdoors • Outcompete other microalgae • Can grow in low light • Capable of mixotrophy • Low maintenance cost • Can grow on waste waters • Nucleus and plastid transformable	• Low metabolic knowledge • Limited molecular tools	• Pharmaceuticals • Nutraceuticals • Food industry • Cosmetics • Biofuels	(Hamilton et al., 2014, Pérez-López et al., 2014, Abate et al., 2015, Butler et al., 2020, Fernandes and Cordeiro, 2021)

Table 1.1 Non-exhaustive table of the main organic systems used for bioproduction

GRAS = Generally recognized as safe

Crabtree positive organism produce ethanol under oxygenated conditions.

1.3.2 Heterologous biosynthesis of cannabinoids

As extraction of cannabinoids bears numerous hurdles, synthetic biology has been considered to obtain higher yields of easily accessible cannabinoids, precursors, or cannabinoid derived molecules. A search for the most suitable host for fast, high yield, pure, controllable and sustainable production of cannabinoids has been initiated (Blatt-Janmaat and Qu, 2021a, Kearsey et al., 2023). As expected, attempts have been made with *E. coli* and first led to the production of insoluble proteins that generated inclusion bodies (Lange et al., 2015), however, CBG in vivo production has been achieved in low amount (Kearsey et al., 2023). The TKS, OAC from *C. sativa* were inserted in *E. coli* to produce low amounts of OA, then the *At*APT from *Aspergillus terreus* leading to titers below g. L^{-1}, necessary for industrial production. The cannabinoid was mostly secreted in the medium, but the low amount produced make it difficult to use at an industrial scale. Eventually, the highest production of CBGA in *E. coli* after supplementation of olivetolic acid and GPP reached 1.2 mg. L^{-1} (Ayakar et al., 2024). The complete pathway has so far never been functional *E. coli*.

The yeast *S. cerevisiae* has also been used to produce cannabinoids yielding low levels of olivetolic acid (Blatt-Janmaat and Qu, 2021a). Most of the studies inserted one enzyme and fed its substrate to obtain the product. One study has achieved a complete biosynthesis from fatty acids precursor inserting the mevalonate pathway in the yeast to make it produce GPP eventually leading to the production of cannabinoids analogue such as THCA, THCVA, and CBGA (Luo et al., 2019a). Although the achievement is remarkable, the extensive modifications of the yeast and the cultivation costs make the host currently not used for industrial bioproduction of cannabinoids. A patent from Hyasynth Bio Inc presents the production of olivetol and methyl olivetol by inserting the *Dyctiostelium discoideum* SteelyA enzyme in *Saccharomyces cerevisiae*

reaching 76 mg. L^{-1} olivetol and 66 mg. L^{-1} (Mookerjee et al., 2018, Mookerjee et al., 2022).

THCA was produced at low levels in the yeast *Komagataella phaffii* limited particularly by the enzymes expression levels (Zirpel et al., 2017a, Luo et al., 2019a). A strain has successfully produced 96 mg. L^{-1} of CBGA after olivetolic acid supplementation (Szamecz et al., 2024). CBGA and cannabigerovarinic acid have been produced to levels of 1.25 g. L^{-1} using cell-free system reaching levels two magnitudes higher than those reached in yeasts (Luo et al., 2019a, Valliere et al., 2020).

Trying to harness the metabolic proximity of *Nicotiana benthamiana* with *Cannabis sativa*'s, the heterologous expression of cannabis enzyme THCAS in hairy roots of the tobacco plant in liquid cultures produced THCA from CBGA feeding reaching 2.7 mg. L^{-1} (Sirikantaramas et al., 2004b). Finally, attempts to produce cannabinoids harnessing the natural advantage of microalgae *Chlamydomonas reihardtii* producing high levels of fatty acids, a patent managed to produce de novo detectable levels of CBGA but too low to be used at an industrial scale (Laban, 2021). The overall heterologous production of cannabinoids has been troublesome, requiring numerous genetic and environmental conditions to achieve production. As displayed in table 1.1, every bioproduction platforms bears drawbacks, the search for the best fitting host is therefore still ongoing. Combining the advantages of a unicellular being, a prokaryote, and a photosynthetic organism, fortunately, a hope still remains.

Overall, the strategy in all organisms tested have been very similar: inserting *Cannabis sativa'* cannabinoids biosynthetic pathway in foreign organisms, trying to modulate the feeding of substrates. Each strategy aims at harnessing the advantages of the host organisms to improve the yields of productions of cannabinoids even though it seems the complete biosynthesis of cannabinoids in heterologous organisms require

comprehensive understanding of the host's metabolism and sometime extensive modifications.

1.4 Phaeodactylum tricornutum

The diatom *Phaeodactylum tricornutum* has been discovered in 1987 by Knut Harald Bohlin in the waters of Plymouth, UK (Bohlin, 1897). It is a unicellular, eukaryotic organism that belongs to the group of Bacillariophyta, a group containing more than 10 000 species of diatoms (Fernandes and Cordeiro, 2021). This phylogenic family has been observed to contribute to about 20 % of the global primary production as well as being a major oxygen generator (Field et al., 1998).

The phylogeny of *P. tricornutum* remains nowadays unclear, however it is hypothesized that it originates from two endosymbiosis, the first one of a cyanobacteria and a heterotrophic eukaryote, then a photoautotrophic eukaryotic red algae and a heterotrophic Stramenopiles (McFadden, 2001, Gruber and Oborník, 2024). *P. tricornutum* bears several unique properties as it is the only diatom know to be polymorphic. Indeed, it can grow having an oval, a fusiform, or a triradiate morphotypes (Figure 1.14).

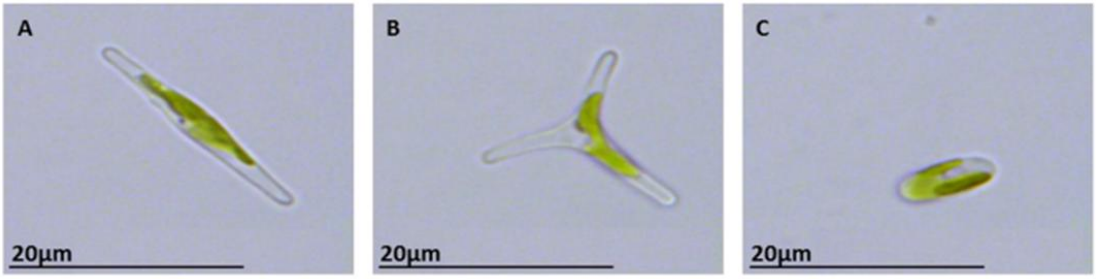


Figure 1.14 *Phaeodactylum tricornutum* observed morphotypes under a lightmicroscope.

A. Fusiform morphotype. B. Triradiate morphotype. C. Oval morphotype.

Source (Ovide et al., 2018)

Those morphotypes appear to be determined by the genetic properties and growth conditions (Tesson et al., 2009a, Galas et al., 2021, Bao et al., 2025). It has been observed that the oval form is efficient for accumulating proteins and pigments while the fusiform and triradiate morphotype accumulate higher lipid content (Song et al., 2020).

Diatoms are characterized by the presence of a silicate wall made of two parts called theca, fitting each other like the two parts of a petri dish and thus form the frustule. This is where their name comes from διατομος (diatomos) meaning "cut in half", or "in two parts" (Armbrust, 2009). However, as the diatom *P. tricornutum*'s cell wall is composed of weakly polymerized silica, it has been observed to grow well in silica depleted medium, keeping a cell wall composed of polysaccharides, lipids and proteins making it more convenient for laboratory cultivation (Tesson et al., 2009b, Katsaros and Heimann, 2012). This lack of silicate in the cell wall also seems to contribute to its transformability. Besides, its ability to grow in suspension rather than clamped, its reproducible genetic transformation, its genetic stability through generations, and its ability to grow in autotrophic and mixotrophic[7] conditions made it a valuable organism for laboratory work (Sabir et al., 2018, Butler et al., 2020, Russo et al., 2023). There has also been through time an accumulation of knowledge and the development of variety of tools to allow genetic manipulation and laboratory cultivation of *P. tricornutum*. It is for instance possible to insert heterologous DNA through various methods and apply antibiotic selections as well as protein and metabolite extraction

[7] Mixotrophy represents the ability of an organism to grow through autotrophy or heterotrophy, PATEL, A. K., CHOI, Y. Y. & SIM, S. J. 2020. Emerging prospects of mixotrophic microalgae: Way forward to sustainable bioprocess for environmental remediation and cost-effective biofuels. *Bioresource technology,* 300**,** 122741.

processes making this organism convenient for research and bioproduction (Karas et al., 2015b, Russo et al., 2023).

Major methods allow for the insertion of DNA in *P. tricornutum*:

1) Microparticle bombardment, also called biolistic or gene gun, is a method relying on delivering a tungsten or gold nanoparticle coated in nucleic acids or protein inside cells (Sodeinde and Kindle, 1993). The particles are shot at high pressure into the cells but their dimensions close to cell size often damage them leading to a low efficiency of transformation. It however allows for nuclear and plastid transformation and has been confirmed to function with plasmids reaching 10 kb (Stewart et al., 2018, Li et al., 2025). It also requires high DNA concentration for successful transformations (Sharma et al., 2018). DNA transformed in the diatom this way is also inserted randomly in the genome making the DNA stable in the genome but its expression inconsistent between clones (Sharma et al., 2018, Li et al., 2025).
2) Interkingdom episomal conjugation has been developed by Karas et al, 2015 allowing for the delivery of DNA from *E. coli* to *P. tricornutum* (Karas et al., 2015b). A strain of the bacteria contains a plasmid bearing all the required machinery for the delivery of the cargo plasmid. The cargo must be designed to contain an origin of transfer sequence (oriT) besides episome maintenance sequence ARS. This method also functions with low DNA concentrations and high transformants quantities. It permits the stable maintenance of an episome when bearing a selection marker and reduces the variability of expression levels between clones (Li et al., 2025).
3) Electroporation is a method that applies electric pulses on cells to open pores on cell membrane (Weaver, 1995). The transformation efficiency can be as

much as ten time higher than biolistic. It allows for the transformation of linear and circular DNA and permits cotransformation (Li et al., 2025). It is also functioning in many organisms but requires adaptation of parameters to each of them, adding a tuning step to enhance the transformation efficiency. It has the advantage of working with very low amounts of DNA to generate successful transformations (Walker et al., 2024, Li et al., 2025).

4) A fourth method is still being tuned to transform diatoms, the polyethylene glycol (PEG)-mediated transformation. The precise mechanism underlying the permeation of the cells to DNA is not elucidated yet (Karas et al., 2015b). This method is currently being improved with higher transformants generation by Walker et al.

An array of promoters are also available in *P. tricornutum*, among which the Fcp (fucoxanthin chlorophyll-a/c-binding protein) and is an endogenous, light induced promoter active especially during exponential phase thus allowing for a proper accumulation of proteins (Siaut et al., 2007). Another promoter involved in the production of a highly expressed secreted protein has been identified: Hasp1pro (highly abundantly secreted protein 1 promoter)(Erdene-Ochir et al., 2019b). As it is also highly active in stationary phase of growth, and is thus convenient for the production and accumulation of proteins throughout growth, or the production of slowly synthesized proteins.

1.4.1 Phaeodactylum tricornutum for bioproduction

The expansion of the genetic and molecular toolbox available for *P. tricornutum* turned it into an increasingly interesting model for the industrial field.

Indeed, aside of its lab practical properties, it also possesses characteristics that make it valuable for bioproduction. It has a naturally high content in omega-3 fatty acids which are, with omega-6 fatty acids, essential for the human health[8]. The most abundant omega-3 present in *P. tricornutum* are ecosapentaenoic acid (EPA) and docosahexaenoic acid (DHA) (Pudney et al., 2019, Cui et al., 2021). This high content of fatty acids is not only valuable for the production of food supplements but, despite bearing currently several limitations, is also highly promising for the production biofuels (Hildebrand et al., 2012, Lam and Lee, 2012, Caporgno et al., 2016, Sanghamitra et al., 2020, Zhang et al., 2022b). Using algae for the production of biofuels not only leads to the synthesis of non-toxic biodegradable fuels, produces less greenhouse gas when burned, but can also use waste waters as a medium for their cultivation (Demirbas, 2009). As mentioned before, diatoms also are major carbon fixators and produce a higher mass-volume ratio of biomass than plant crops and can be grown outdoors in ponds (Field et al., 1998, Lam and Lee, 2012, Branco-Vieira et al., 2020, Butler et al., 2020).

Thanks to all its assets, *P. tricornutum* is already used for the bioproduction of compounds such as the pigment fucoxanthin[9] (Pocha et al., 2022, Pang et al., 2024a, Elshobary et al., 2025), of fatty acids (Gu et al., 2022b, Gu et al., 2022a) but are for now only being tested for the production of biofuels (Sanghamitra et al., 2020, Zhang et al., 2022b). Its metabolism made it suitable for the production of terpenoids such as

[8] Omega-3 and omega-6 fatty acids reduce the risks of heart diseases and contribute to brain development, SIMOPOULOS, A. P. 2010. The omega-6/omega-3 fatty acid ratio: health implications. *Oléagineux, Corps gras, Lipides,* 17**,** 267-275.

[9] Fucoxanthin extracts given to young human adults seemed to improve cognitive functions such as focus, memory and vigilance, YOO, C., MAURY, J., GONZALEZ, D. E., KO, J., XING, D., JENKINS, V., DICKERSON, B., LEONARD, M., ESTES, L. & JOHNSON, S. 2024. Effects of Supplementation with a Microalgae Extract from Phaeodactylum tricornutum Containing Fucoxanthin on Cognition and Markers of Health in Older Individuals with Perceptions of Cognitive Decline. *Nutrients,* 16**,** 2999.

geraniol for the cosmetic and agricultural industry (Fabris et al., 2020b) and the production of betulin for the pharmaceutical industry (D'Adamo et al., 2019).

P. tricornutum is therefore already an active biofactory and being slowly optimized for the production of more compounds at greater yields.

1.5 Research objectives

The number of tools available for metabolic engineering, its natural production of malonyl-CoA and GPP, its expected production of hexanoyl-CoA, its production of fatty acids, its cultivation advantages, and its current use in bioproductions, made *P. tricornutum* a very promising host for the production of cannabinoids. That is why the company Algae-C contacted Isabel Desgagné-Penix's laboratory and developed a project to transfer the *Cannabis* pathway into *P. tricornutum*. As the metabolic pathway starts from hexanoyl-CoA and malonyl-CoA, the transfer of the *Cs*TKS and *Cs*OAC was expected to yield olivetolic acid. In 2023, (Awwad et al., 2023a) assembled a series of episomal constructs containing expression cassettes of *Cs*TKS and *Cs*OAC with both enzyme in a cistron where protein were linked by a self-cleavable T2A sequence. It appeared that olivetolic acid has been detected temporarily but its production declined through time and could not be reobtained. In parallel, (Fantino et al., 2024a) produced CBGA using *P. tricornutum* crude extract for enzymatic assay after inserting the *Streptomyces sp nphB* aromatic prenyltransferase in *P. tricornutum*. By securing the production of olivetolic acid, it is therefore possible that *P. tricornutum* produces cannabinoids de novo.

The first objective of this thesis work was therefore to insert separately the two *Cs*TKS and *Cs*OAC into *P. tricornutum* and assess the production of olivetolic acid or olivetol. This research describes in chapter II the article of the insertion of *Cs*TKS and *Cs*OAC

in two separate episomes to transform *P. tricornutum* and generate a *Pt*TKS strain (pTKS) and a *Pt*OAC strain (pOAC). Total protein extracts were pooled to test the combined activity of the two enzymes, with and without addition of malonyl-Coa and hexanoyl-CoA. The impact of the *Cs*TKS alone on the diatom's metabolome was then analyzed.

The second objective of this thesis work was the utilization of *Dictyostelium discoideum*'s hybrid enzyme SteelyA, capable of producing olivetolic acid and olivetol from malonyl-coA and hexanoyl-CoA, as potential alternative to the insertion of *Cs*TKS and *Cs*OAC. The chapter III presents the draft of the article on the insertion of an episome expression the SteelyA enzyme in *P. tricornutum*. In this study, we verify the protein's production and its activity.

In addition, an article in which I am a coauthor, "Extrachromosomal expression of functional *Cannabis sativa* cannabidiolic acid synthase in *Phaedodactylum tricornutum*" is presented in the Annex C of this thesis.

CHAPTER II: Impact of heterologous expression of *Cannabis sativa* tetraketide synthase on *Phaeodactylum tricornutum* metabolic profile

Nicolas Sene[1], Karen Cristine Gonçalves dos Santos[1], Natacha Merindol[1], Sarah-Eve Gélinas[1], Alexandre Custeau[1], Fatima Awwad[1], Elisa Fantino[1], Fatma Meddeb-Mouelhi[1,2], Hugo Germain[1,2], and Isabel Desgagné-Penix[1,2*]

[1]Department of biochemistry, chemistry, physics, and forensic science, Université du Québec à Trois-Rivières, 3351 Boulevard des Forges, Trois-Rivières, QC G9A 5H7, Canada

[2]Groupe de Recherche en Biologie Végétale, Université du Québec à Trois-Rivières, Trois-Rivières, Québec, Canada

***Correspondence:**

Isabel Desgagné-Penix

Isabel.Desgagne-Penix@uqtr.ca

Accepted for publication in Biotechnology for Biofuels and Bioproducts.

Author contributions

NS: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. KCGS: Data curation, Formal analysis, Writing – review & editing. NM: Writing – review & editing. SEG: Formal analysis, Writing – review & editing. AC: Formal analysis, Validation. FA: Conceptualization, Writing – review & editing. EF: Formal analysis, Investigation, Methodology. FM: Formal analysis, Investigation, Methodology, Writing – review & editing. HG: Funding acquisition, Resources, Supervision, Writing – review & editing. IDP: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing.

Résumé :

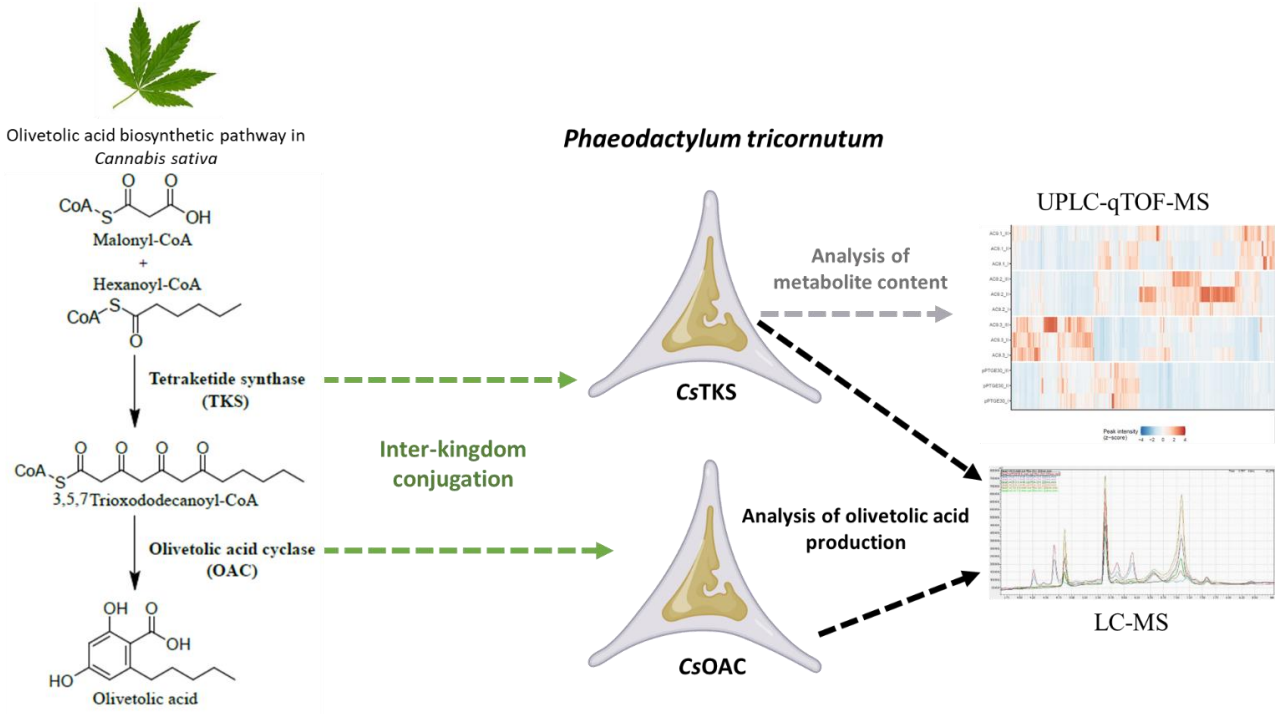
La sécurité pharmaceutique est une priorité mondiale croissante, particulièrement en tenant en compte que la demande pour des composés thérapeutiques augmente avec la population. Les phytocannabinoides, une classe de molécules bioactives issue des plantes, ont gagné une attention significative à cause de leur interaction avec le système endocannabinoïde humain, offrant de potentielles options pour contrôler et moduler une multitude de symptômes et de maladies. Le modèle traditionnel d'extraction depuis les plantes de cannabis présente des contraintes réglementaires, environnementales, et de rendement. En conséquence, la biosynthèse microbienne est devenue une alternative prometteuse pour produire des cannabinoïdes dans un cadre contrôlé, adaptable à différentes échelles, et de façon durable. Le développement de bio-usines à partir de diatomées représente une étape cruciale dans l'avancement des biotechnologies permettant ainsi la production efficace de composés de haute valeur telles que les cannabinoïdes.

Dans cette étude, nous avons modifié la diatomée *Phaeodactylum tricornutum* un organisme photosynthétique unicellulaire prisée pour son haut contenu en lipides pour produire de l'acide olivetolique (OA), un précurseur clé dans la synthèse de cannabinoïdes. Les gènes codant pour la synthèse de l'enzyme tétracétide synthase et l'acide olivetolique cyclase de *Cannabis sativa* ont été clonés dans un épisome chacun et introduits dans la diatomée par conjugaison bactérienne pour évaluer leur activité enzymatique et la production d'OA in vivo. Les deux gènes ont été exprimés avec succès et leurs enzymes ont été accumulées dans les lignées transconjugantes. Cependant, malgré les tests sur des extraits individuels ou combinés, aucune accumulations d'OA n'a pu être détectée suggérant une potentielle conversion ou l'utilisation d'OA par des voies métaboliques inhérentes à la diatomée. Afin

d’approfondir les recherches, nous avons analysé l’impact de l’expression de la *Cs*TKS sur le métabolome, révélant une altération significative indiquant une redirection du flux métabolique.

Mots clés : Diatomée ; Cannabinoïdes ; Ingénierie métabolique ; Polycétides ; Bio-usine photosynthétique ; Métabolomique non dirigée ; Voie métabolique du shikimate.

Graphical Abstract



Background

Cannabinoids are a class of molecules notorious for their recreational properties. They are naturally produced in highest abundance in the glandular trichomes of the female flowers of the *Cannabis* genus. The plant of cannabis was used for millennia for its anti-inflammatory, antiemetic, and as a pain relief. Those properties are due to the ability of cannabinoids to interact with the mammalian endocannabinoid system involved in several metabolic functions (Śledziński et al., 2020, Zou and Kumar, 2018, Crocq, 2020a). Today, more than a hundred terpenophenolic compounds in the glandular trichomes of *Cannabis sativa* (*C. sativa*) have been identified as cannabinoids. Among them, Δ^9-tetrahydrocannabinol (THC) and cannabidiol (CBD) are the most well-known and abundant (Punja et al., 2023, Conneely et al., 2021). Recently, the interest in cannabinoids focused on their potential in fighting neurodegenerative disease such as Alzheimer's and Parkinson's diseases (Micale et al.,

2007, Śledziński et al., 2020). The pathway to biosynthesize THC and CBD starts from hexanoyl-CoA and malonyl-CoA moieties, the tetraketide synthase (TKS) condensates hexanoyl-CoA with three malonyl-CoA to produce the enzyme-bound 3,5,7-trioxododecanoyl-CoA (TdCoA) compound which is then cyclized in olivetolic acid (OA) by the olivetolic acid cyclase (OAC). An aromatic prenyltransferase (APT) fuses OA with the prenyl group of geranyl pyrophosphate (GPP) to produce cannabigerolic acid (CBGA). The action of cannabidiol acid synthase (CBDAS) or tetrahydrocannabinol acid synthase (THCAS) leads respectively to the production of cannabidiolic acid (CBDA), or Δ^9-tetrahydrocannabinolic acid (THCA). Cannabinoids therefore belong to the terpenophenols family as their final structure present a cyclic monoterpene, brought by GPP, and a diphenol alkylated chain, brought by OA (Farag and Kayser, 2017). Combustion triggers the decarboxylation of the THCA and CBDA, resulting in the formation of THC and CBD (Fig. 2.1) (Gagne et al., 2012, Luo et al., 2019a, Tahir et al., 2021). To study and harness their potential, demand for cannabis cultivation has increased, with the global legal market projected to exceed US$400 billion by 2030, driven by rising medical and economic demands (Xie et al., 2023, Hammond et al., 2022, Insight, 2024). However, several hurdles make cannabis cultivation complicated. Among them, a restrictive regulation, low yield of cannabinoids due to restricted production in female flowers only, and difficult purification led to the search for alternatives (Lewis et al., 2017, Backer et al., 2019).

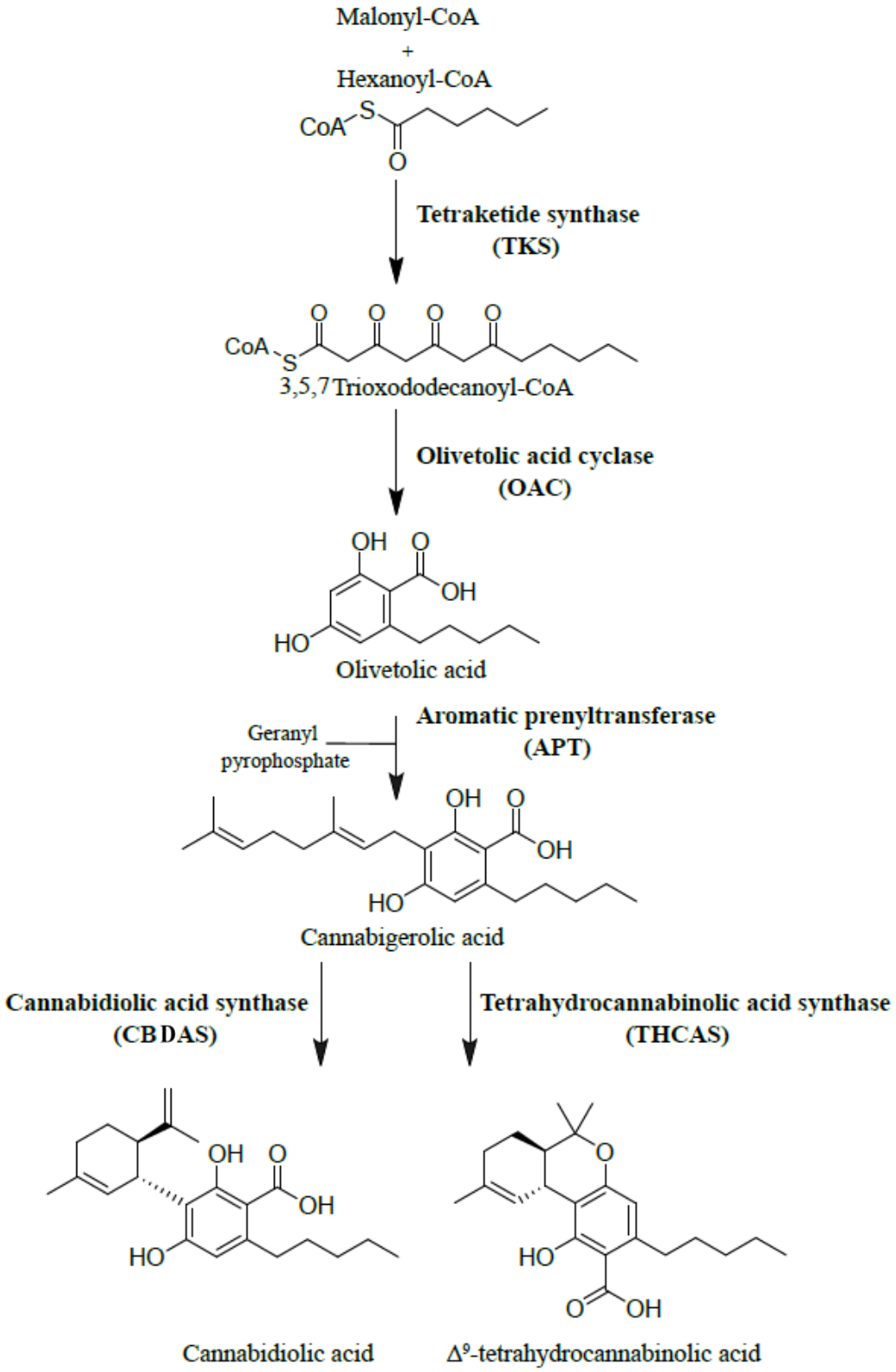


Figure 2.1. Scheme of the proposed cannabinoids synthesis in *Cannabis sativa* from hexanoyl-CoA and malonyl-CoA. Arrows represent enzymatic reactions. Enzymes' names are in bold. D

Bioengineering of model organisms has increasingly advanced, both to confirm biological knowledge, and as an alternative way to produce compounds of interest. It does not only allow to circumvent all the constraints associated with cultivation but also has a higher organism mass volume ratio. Attempts of producing cannabinoids in *E. coli* have been performed on purified TKS using synthetic *N*-acetylcysteamine as a substrate and leading to the production of olivetol. However, attempts of producing THCA never succeeded due to the complex folding of THCAS therefore considering the bacteria unsuitable for the production of cannabinoids synthase (Wiles et al., 2022). To overcome the limits of bacteria, the production of cannabinoids has been achieved in yeast with the limit of reaching poor yields despite supplementation (Luo et al., 2019a, Zirpel et al., 2017a). Algae are promising heterologous hosts, as the number of available tools and data gathered have dramatically increased over recent years (Kumar et al., 2020, Gao et al., 2021, Nouemssi et al., 2020). Compared to conventional platforms, algae offer additional advantages, such as the ability to fix CO_2 through photosynthesis. In 2019, a patent reported that cannabinoids can be heterologously produced and harvested in microalgae *Chlamydomonas reinhardtii* and *Synechococcus elongatus* (Laban, 2021). Compared to other microalgae, a unique advantage of diatoms is their estimated contribution to around 20% of Earth's oxygen production (Kassaw et al., 2022, George et al., 2020b). Among diatoms, *Phaeodactylum tricornutum* has emerged as valuable platform to produce fatty acid-derived molecules, due to its fast-growing rate, high lipid content, and natural production of omega-3 fatty acids like eicosapentaenoic acid (Pudney et al., 2019, Xue et al., 2015). Efficient protocols therefore have been developed for the extraction of apolar compounds such as cannabinoids (Lazarjani et al., 2021, Fajardo et al., 2007). *P. tricornutum* also naturally produces malonyl-CoA (Cui et al., 2019), GPP (Fabris et al., 2020b), and is predicted

to produce hexanoyl-CoA (BioCyC, 2024), all essential precursors for cannabinoid biosynthesis. Together, these features suggest that *P. tricornutum* is a suitable model organism for the heterologous biosynthesis of cannabinoids. Indeed, the introduction of S*treptomyces sp.* aromatic prenyltransferase NphB (Fantino et al., 2024a), and CBDAS (Fantino et al., 2024b) into *P. tricornutum*, has led to the *in vitro* production of CBGA and CBDA, respectively. Additionally, co-expression of a fused construction of the first two enzymes of the pathway — TKS and OAC from *C. sativa* — in *P. tricornutum* lead to the transient production of trace amounts of OA (Awwad et al., 2023a). The reason for the loss of OA accumulation over time remains unclear. It is possible that the coexpression of the two genes in a cistron might hinder the stability of the construct over time and entail silencing. We hypothesized that expressing TKS and OAC separately in *P. tricornutum* strains, rather than as a fused construct, might enhance enzyme functionality and stability, thereby improving OA synthesis. Therefore, in this study, we aimed to produce olivetolic acid in *P. tricornutum* by inserting TKS and OAC enzymes in different transconjugants and test their activity individually or pulled together. In order to observe the extend of *Cs*TKS impact on the metabolite content and deepen our understanding on the potential compatibility of *P. tricornutum* to produce cannabinoids, using UPLC-qTOF-MS we assessed the impact of *Cs*TKS accumulation on transconjugants metabolite content.

Material and methods

Microbial strains and growth conditions

Escherichia coli strains Epi300 (TransforMax Epi300 Biosearch™ Technologies, Novato, California, USA) and 10β (New England Biolab®, Ontario, Canada) were grown in Luria Bertani broth (LB) at 37°C with 220 rpm agitation when in liquid

medium, supplemented with chloramphenicol (35 μg.mL^{-1}) (Thermo Fisher Scientific, Ottawa, Ontario, Canada) and gentamicin (50 μg.mL^{-1}) (Thermo Fisher Scientific, Ottawa, Ontario, Canada) when containing pTA-MOB.

Saccharomyces cerevisiae VL6-48 strain was grown following (Karas et al., 2013, Awwad et al., 2023a). Cultures in rich medium were done in yeast extract peptone dextrose YEPD while cultures in minimum media were done in complete minimal (CM) medium lacking histidine and uracil or tryptophan and uracil (Teknova, Hollister, California, USA) at 30°C.

The *P. tricornutum* strain (CCAP 1055/1, Culture Collection of Algae and Protozoa, RRID:NCBITaxon_2850) was grown in L1 medium without silica at 18°C under white fluorescent lights (75 μE m^{-2} s^{-1}) with a photoperiod of 16 h of light and 8 h of darkness under the agitation of 130 rpm for liquid cultures with zeocin (50 μg.mL^{-1}) (Invitrogen, Burlington, Ontario, Canada) when cultivating transconjugant cells containing the *Sh Ble* resistance gene (Diamond et al., 2023b).

Plasmid constructions and insertion in *Phaeodactylum tricornutum*

Recombinant plasmids, named pTKS for *Cs*TKS and pOAC for *Cs*OAC, were assembled through yeast assembly using *S. cerevisiae* strain VL6-48 (ATCC MYA-3666: MATα his3-Δ200 trp1-Δ1 ura3-52 lys2 ade2-1 met14 cir^{0}) exactly as described in (Karas et al., 2013). The cassette containing the *CsTKS* under the control of *40SRPS8* promoter and *fcpA* terminator was inserted in the p*Pt*GE30 plasmid, named EV hereafter, composed of a pCC1BAC (RRID:Addgene_62862) backbone, containing a bleomycin resistance gene (*Sh Ble*) as a selection marker, and a *P. tricornutum* region contributing to plasmid stability.

Assembled pTKS and pOAC episomes were transformed into the *E. coli* Epi 300 strain containing pTA-MOB plasmid. The cargo plasmid was then transferred in the diatoms via conjugation as described in the literature (Karas et al., 2015d). Transconjugants of *P. tricornutum* were then selected following the protocol of (Fantino et al., 2024a) leading to a liquid culture of *P. tricornutum* clones in L1 supplemented with zeocin. Conjugations were verified by colony PCR. Positive clones were selected for plasmid extraction then sent for Next Generation Sequencing (NGS).

Growth kinetics

Optical density of the clones was measured at 730 nm (OD_{730nm}) every two days for 14 days. For each of them, 250 µL was added to a Microplate 96 Well, PS, F-Bottom Chimney Well µCLEAR®, black (Greiner Bio-One, Kremsmünster, Austria) and absorbance was recorded using Synergy H1 BioTek microplate reader (Agilent, Santa Clara, California, USA). Curves were obtained using GraphPad Prism software (RRID:SCR_002798). A t-test was performed using a Compare Groups of Growth Curves (CGGC) permutation test from Walter and Eliza Hall Institute of Medical Research with 10 000 permutations (Elso CM et al., 2004).

Plasmid rescue

For each transconjugant lines, 5 mL were centrifuged at 4 700 g for 10 min after seven days of culture. After supernatant removal, 235 µL of PDL1 Large Plasmid Extraction Kit (Geneaid™, New Taipei City, Taiwan), together with 5 µL Lysozymes (100 $mg.mL^{-1}$) and 5 µL hemicellulase (25 $mg.mL^{-1}$) were added to resuspend the pellet. Samples were incubated at 37°C for 30 min. 250 µL of PDL2 solution was added to the mix before mixing by inverting ten times, then 375 µL of PDL3 was added before mixing. The rest of the procedure was performed according to the manufacturer protocol until

elution in 50 µL of elution buffer preheated at 80°C. DNA concentration was measured with a N60/N50 Implen NanoPhotometer®. Purified plasmid DNA was then used to transform *E. coli* 10β strains from New England Biolabs (Whitby, Ontario, Canada) following their protocol. Bacteria were plated on LB with chloramphenicol (35 $mg.mL^{-1}$). After inoculating 5 mL LB with identical concentration of chloramphenicol, the liquid culture was incubated over night at 37°C under agitation. Plasmid DNA was extracted by miniprep (Geneaid™, New Taipei City, Taiwan). Purified plasmid DNA concentration was measured with the NanoPhotometer and the quality of the DNA sample was verified by digestion with restriction enzymes. The plasmid DNA sample was then used for NGS at Massachusetts General Hospital DNA Core.

Western blot

Approximately 160 mg of wet mass was resuspended in 1X Laemmli in PBS at a ratio of 1 $mL.750\ mg^{-1}$, heated at 95°C for 10 min, and centrifuged 5 min at 14 000 g at 4°C. At least 500 µg of protein was loaded on a 10% SDS-polyacrylamide gel and migrated at 80 V for 25 min, then at 120 V for 1 h. The gel was then transferred to a 0.2 µm polyvinylidene fluoride (PVDF) membrane, previously activated in methanol and then equilibrated in transfer buffer (25 mM Tris, 192 mM glycine and 20 % methanol). Transfer was carried out at 100 V, 400 mA, for 2 h in transfer buffer on ice. The membrane was blocked in 5% milk in Tris Buffer Saline (TBS) with 0.1% Tween 20 for 50 min at room temperature. Blots were washed three times for 10 min with TBST before being incubated with 6x-His monoclonal antibody (Invitrogen, Canada, RRID:AB_557403) (dilution 1:1 000) in 3% bovine serum albumin overnight at 4°C. After three washes in TBST, blots were incubated for 1 h with a 1:20 000 dilution, in 5 % milk of Immun- Star Goat Anti-Mouse (GAM)-HRP from Bio-Rad (Ontario, Canada, RRID:AB_11125753). After three more washes in TBST protein detection was

performed using Clarity Max, 10 ng of Multiple Tag (GenScript, Brockville, Ontario, Canada) were used as positive control. Red ponceau was applied for 2 min at room temperature and then rinsed with distilled water. Western ECL Substrate-Luminol solution from Bio-Rad. Images were acquired using ChemiDoc Imaging System (RRID:SCR_019037) with Image Lab™ Software (RRID:SCR_014210) (Bio-Rad, Hercules, California, USA).

Protein extraction

For each sample, 500 µL of cell buffer (0.75 mM SDS,10% Glycerol, 51.4 mM, Tris HCl pH8, 0.02 mM EDTA) was added to 125 mg of pellets. PMSF and protease inhibitor were added at a ratio of 1/10 and 1/100 then proteins were extracted by sonication with the following parameters: amplitude 35%, time 3 min, ON 30 s, OFF 25 s (Fisherbrand™ Model 505 Sonic Dismembrator). Extracts were then spun for 40 min at 4°C at 14 000 g and the supernatant was removed and spun a second time in the same conditions. The new supernatant was flash frozen and kept at -80°C until use.

Protein purification

The soluble fraction of protein extracts from about 300 mg of cell pellets were purified using Ni-NTA agarose resin (Invitrogen, Burlington, ON, Canada) targeting his-tagged protein. The protocol was followed until elution. Washes 1 and 2 as well as elutions 1 and 2 were tested by western blot.

To purify OAC, magnetic beads coated with anti-myc antibody were used following the protocol for low denaturation of enzymes (Cedarlane, Burlington, ON, Canada).

Enzymatic assay

Hexanoyl-CoA and malonyl-CoA were purchased from Millipore Sigma Canada Ltd (Oakville, Ontario, Canada) and the olivetol and olivetolic acid used were purchased from Santa Cruz biotechnologies (Dallas, Texas, United states). Protein concentration was measured by colorimetric RC DC™ Protein Assay (Bio-Rad, Hercules, California, USA). Every test was done in triplicate in 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) buffer 0.05 M pH 7, DTT 0.005 M, with 400 µg of TKS and/or OAC protein extract (800 µg total protein extract when pulled together), 0.2 mM hexanoyl-CoA, and 1 mM malonyl-CoA, completed to 100 µL with water. The reaction was performed at 25°C for 16 h and was stopped with trichloroacetic acid 2% v/v. 300 µL of methanol was added to the reaction mix then was filtered on PTFE 0.2 µm filter and transferred into new tubes. The samples were dried using a vacuum concentrator (SpeedVac Thermo Savant SPD 2010), then resuspended in 100 µL methanol before analyzing olivetolic acid production by high-performance liquid chromatography with diode-array detection (HPLC-DAD).

HPLC-DAD analysis

An InfinityLab Poroshell 120 EC-C18 (4.6 × 100 mm, 2.7 mm; Agilent Technologies, Mississauga, Ontario, Canada) column, maintained at 30°C, was used for separation. For each sample, 10 µL was injected into the HPLC device. The mobile phase used was composed at 30% of (A) formic acid 0.1 % v/v in milli-Q water and 70% of (B) formic acid 0.1 % v/v in methanol. Samples were processed at a flow rate of 1 $mL.min^{-1}$ following 0 min, 70 % B; 1.0 min, 70 % B; 6.0 min, 77 % B; 15.0 min, 90 % B; 15.1 min, 70 % B and 18.0 min, 70 % B. The total run time was 18.5 min allowing the reconditioning of the column for the next injection. D2 (deuterium) lamps acquired wavelengths from 190 to 400 nm. The analyses were done using UV wavelength at 200

nm. The maximum absorption wavelength and the retention times of each detected signal in samples were compared to standards to allow their identification.

Untargeted metabolomic

Metabolite extraction

Metabolites were extracted from three biological replicates for each clone. Each replicate was inoculated at OD_{730} 0.1 in 200 mL, homogenized, then separated into four 50 mL volumes in 250 mL flasks to optimize the growth rate. Cultures were grown for an additional week, then for each replicate, the 50 mL cultures were pooled together. Cultures were then centrifugated for 10 min at 1 500 g at 4°C and pellets were weighed. 1 mL of ethanol 95% (Fisher Chemicals, Pittsburgh, Pennsylvania, USA) per 100 mg of wet biomass was added to the pellet. Samples were mixed by vortexing and incubated overnight at -20°C. The following day, the samples were centrifuged at 4 000 g for 10 min at 4°C and supernatants were filtered using polytetrafluoroethylene (PTFE) 0.2 µm filter. All filtrates were dried using a vacuum concentrator (SpeedVac Thermo Savant SPD 2010) without heating, for 3h at maximum ramp.

Data Acquisition

Metabolomic extracts that were analyzed by reversed-phase chromatography (RP) were resuspended in 100 µL of pure water, while the ones analyzed using hydrophilic interaction liquid chromatography (HILIC) were reconstituted in 100 µL of mobile phase (ammonium formate 10 mM in acetonitrile/water (95:5), pH 3.8). Samples were vortexed and sonicated for 2 min, followed by a 5 minutes 16 000 g centrifugation. Supernatants were collected and diluted 1:100 in their initial solvents followed by a filtration using 0.45 µm AcroPrep filters in a 5 mm Hg vacuum. Acquisition was done with an Eksigent µU HPLC (Eksigent, Redwood City, California, USA) in conjunction

with an ABSciex TripleTOF 6600 (ABSciex, Foster City, California, USA) system with an electrospray interface. For data processing, acquisition and instrument control, the Analyst TF version 1.8 software was used. Data acquisition was performed in Data Dependent Acquisition (DDA) and SWATH mode for positive ionization mode. The source voltage was set to 5.5 kV, temperature at 220°C, curtain gas at 50 psi, ion source gas 1 at 40 psi, and ion source gas 2 at 50 psi. For all methods, the mobile phase was composed of: A (water with 0.1% formic acid and water with 10 mM ammonium fluoride) and solvent B (acetonitrile with 0.1% formic acid and acetonitrile: water (95:5) with 10 mM NH4F). Columns Luna Omega Polar 100x2.1 1.6 µm (RP positive) and BEH Z-HILIC 100x2.1 Particle 1.7 µm (HILIC) from Waters were kept at 50°C. The gradient began at 0% B for reversed-phase, and 100% for HILIC.

Bioinformatics Analysis

Raw mass spectra files in profile mode were aligned, deconvoluted and annotated with MS-Dial version 5.1.230912 (RRID:SCR_023076) (Lai et al., 2018, Tsugawa et al., 2015, Tsugawa et al., 2019). The parameters were set as follows for all three analyses: retention time range of 0-10 min; mass range of 50-1 250 Da; MS1 and MS2 tolerances, for deconvolution and alignment, of 0.01 Da and 0.025 Da, respectively; minimum peak height and width of 1000 amplitude and 5 Da, respectively; the smoothing method used was linear-weighted moving average, with smoothing level of 3 scans and mass slice of 0.1 Da; and the sigma window value was set to 0.5. For analyte identification, the following parameters were set: accurate mass tolerance 0.01 Da for MS1 and 0.05 Da for MS2; retention time tolerance was set to 0.1 minute; adduct ions were set to $[M+H]^+$ and $[M+NH_4]^+$ for ESI^+ and $[M-H]^-$ for ESI^-. Deconvoluted spectra were compared to MassBank ESI^+ and ESI^- (MoNa, 2025) and MS-Dial's "ESI^+-" and "ESI^--MS/MS from authentic standards" databases version 19 (all databases were obtained as MSP files

from MS-Dial's webpage on March 2024) for annotation. These comparisons were done with the default parameters, except for MS1 and MS2 mass tolerances, set to 0.01 Da and 0.001 Da, respectively. Annotations marked as "reference match" by MS-Dial were validated by checking that main peaks from reference were present in the current experiment's spectra.

Peak height tables exported from MS-Dial were then used for comparative analyses in R version 4.3.3. RP ESI^+ and ESI^- datasets included three blank samples, while HILIC ESI^+ dataset included two blank samples. For each analyte, the abundance estimate was defined as:

$$peak\_height_{replicate} - (3 * \max(peak\_height_{blank\_samples}))$$

If the abundance estimate was ≤ 0, the analyte was considered absent. For each strain and for both quality control replicates (pooled samples), only analytes present in all replicates were considered present. Finally, only analytes considered present in at least one strain were selected for further analysis. The abundance estimation table was prepared for MetaboAnalystR version 4.0 (RRID:SCR_016723) (Pang et al., 2024b), which was used for replacing missing values, filter analytes (parameters: $qc.filter =$ T, $rsd = 25, var.filter =$ "*none*", $var.cutoff = -1, int.filter =$ "*mean*", $int.cutoff = 0$) and normalize data (functions $PreparePreNormData$ and $Normalization("GroupPQN", transNorm = "LogNorm", scaleNorm =$ "$ParetoNorm", ref = "QC", ratio = F, ratioNum = 20$)). Next, MetaboAnalystR's filtered datasets were used to find significant differences in analyte abundances between pPtGE30 and AC9 strains using the function *sam* (parameters: $method =$ "*d.stat*", $B = 200, med = F, use.dm = T, var.equal = T, R.fold = 1$)

from the R package siggenes version 1.76.0 (Schwender, 2023). Fold change was calculated as follows:

$$log_2 \frac{mean(abundance_{Ac9\ clone})}{mean(abundance_{pPTGE30})}$$

If $p < 0.01$ and |fold change| > 1, the analyte was considered deregulated in the corresponding AC9 strain. Finally, principal component analysis (PCA) was performed using the combined normalized data from all three analyses (RP ESI$^+$, RP ESI$^-$ and HILIC ESI$^+$) with MetaboAnalystR function *PCA.Anal*.Results

Heterologous accumulation of CsTKS and CsOAC in *P. tricornutum*

Two expression cassettes were designed to independently characterize *CsTKS* and *CsOAC* expression in *P. tricornutum*. The plasmid referred to as pTKS, contained the *CsTKS* coding sequence (CDS) under the control of the endogenous *40SRPS8* promoter, tagged with a 6xHis marker, and terminated with the *FcpA* terminator. The plasmid referred to as pOAC, included *CsOAC* CDS under the control of the *FcpD* promoter, tagged with a myc marker, and terminated with the *FcpD* terminator (Figure 2.2). These plasmids along with the empty vector (EV), were introduced into *P. tricornutum* wild-type strains via interkingdom conjugation with *E. coli*. The plasmids harbored the *Sh ble* gene, conferring zeocin resistance, and transconjugants were selected on L1 agar medium supplemented with zeocin.

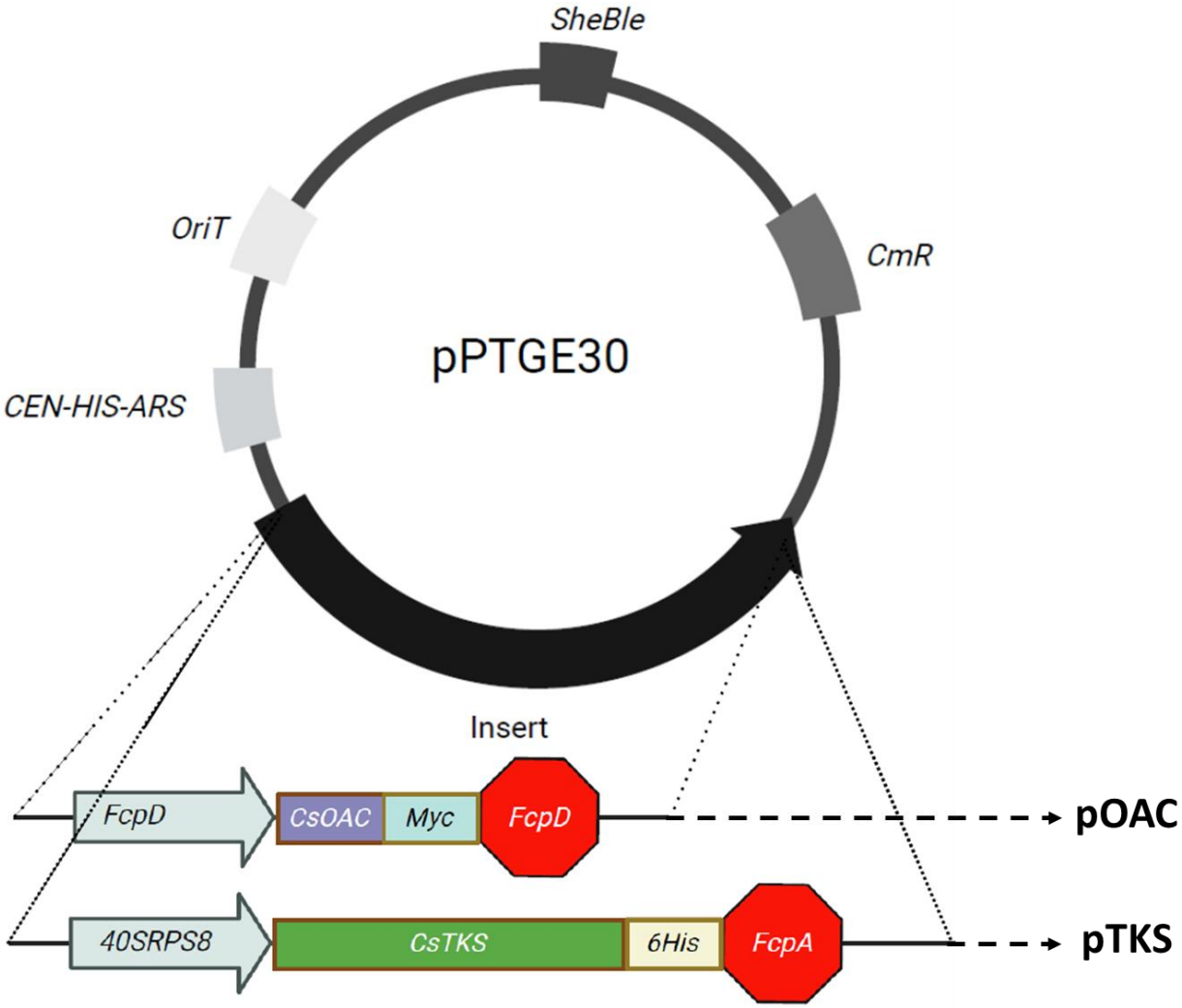


Figure 2.2. Schematic representation of pPtGE30 expression vector and insertion of CsTKS and CsOAC cassettes. The pTKS plasmid includes a construction that contains the 1971 bp cassette with *CsTKS* under control of native constitutive *40SRPS8* promoter, *6xHis* marker and a *FcpA* terminator. The pOAC plasmid harbors a construction with the 1265 bp cassette *CsOAC* under control of *FcpD* promoter, marked with a *Myc* tag, and ended with an *FcpD* terminator. The *Cen-His-Ars* region is necessary for episomal replication while the *OriT* permits the transfer of the plasmid in *P. tricornutum* during interkingdom conjugation. Finally, the *Sh ble* and *CmR* coding sequence allows for zeocin resistance when expressed in *P. tricornutum* and chloramphenicol resistance in *E. coli* respectively.

Three transconjugants of each construct were selected for plasmid rescue, and the extracted DNA was subjected to restriction digestion analysis before sequencing (Supplementary Figure 2.1). Transconjugants showing the expected digestion profiles, were further analyzed by Next Generation Sequencing. While the pTKS sequences exhibited minor substitutions, including one in the zeocin resistance gene, two in the *40SRPS8* promoter, and one in the backbone (Supplementary Table 2.1), the *Cs*OAC CDS remained intact. Despite the substitutions, all transconjugants displayed growth on zeocin-supplemented media, indicating that the resistance gene remained functional. Importantly, the heterologous expression of these enzymes did not impair the growth of the transconjugants (Supplementary Figure 2.2).

Heterologous protein production was then verified by western blot. All pTKS-containing transconjugants exhibited a band at the expected size of 42 kDa, absent in EV along with an additional smaller band likely corresponding to a truncated version of the protein (Figure 2.3). For the pOAC transconjugants, two (pOAC-5 and pOAC-7) displayed a band at the expected size of 12kDa, while one transconjugants (pOAC-6) did not. Further sequencing revealed a frameshift in the *fcpD* promoter of pOCA-6, which disrupted the myc-tag and explained the absence of a signal in the western blot (Figure 2.3).

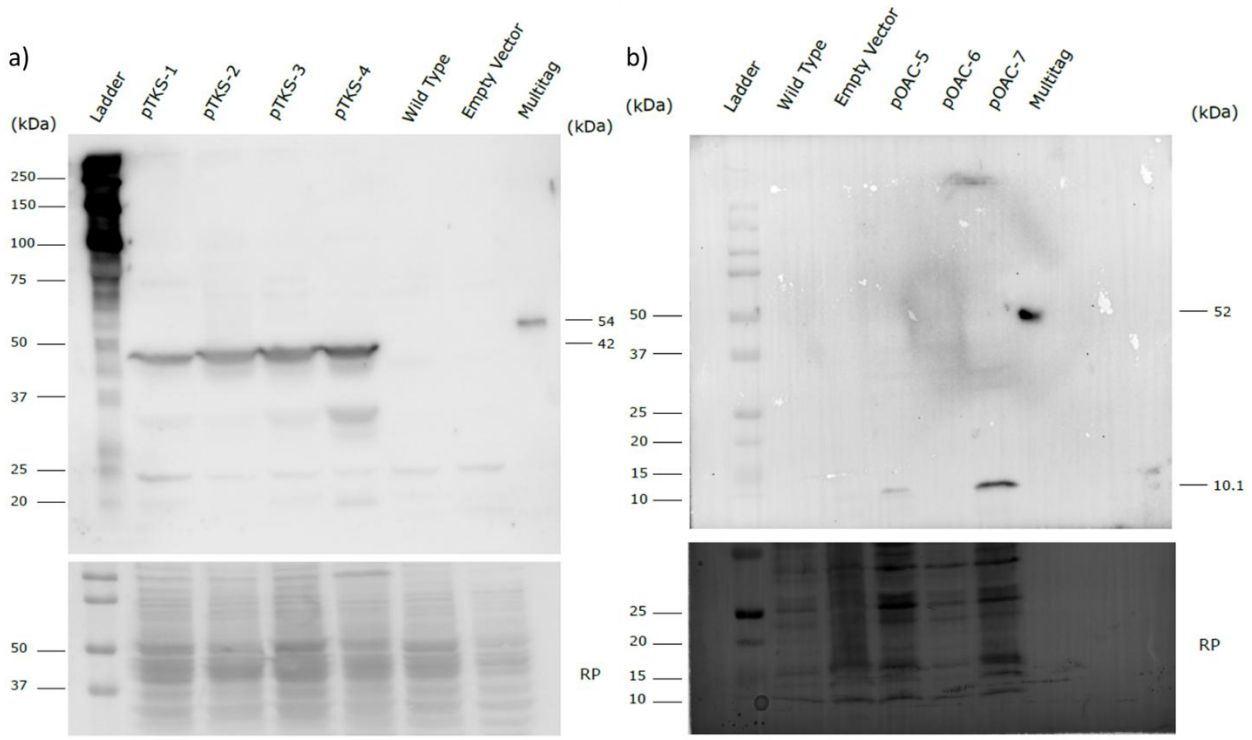


Figure 2.3. Detection of TKS and OAC enzymes in *P. tricornutum* transconjugants by western blot analysis. a) Western blot for *Cs*TKS detection in *P. tricornutum* pTKS transconjugant extracts using anti-6His antibodies. b) Western blot for *Cs*OAC detection in *P. tricornutum* pOAC transconjugant extracts using anti-Myc antibodies. In both panels, protein extracts from transconjugants are compared to those from the empty vector (EV) and wild-type *P. tricornutum*. Protein ladder sizes are indicated on the left while the expected sizes of the proteins of interest and their respective tags are shown on the right of each blot. The lower panels display red Ponceau (RP) staining to confirm protein loading.

To validate the added enzymatic activity from each of the recombinant proteins, total protein extracts from pTKS and pOAC were tested. Enzymatic assays were conducted using the extracts individually or by combining the *Cs*TKS and *Cs*OAC extracts. The resulting reactions were analyzed by HPLC-DAD, but none of the detected peaks corresponded to the OA standard (Supplementary Figure 2.3). To rule out potential

inhibitory effects from compounds present in the crude extract, protein purification was attempted. While TKS was successfully purified, purification OAC was not achieved.

Heterologous expression of CsTKS induces metabolic changes in *P. tricornutum*

The production of a tetraketide is a critical initial step in cannabinoid biosynthesis. To assess the impact of *Cs*TKS activity on the diatom metabolome, we compared the metabolic profiles of pTKS transconjugants with those of the EV control. Metabolite extracts were separated using either hydrophilic interaction liquid chromatography (HILIC) or reversed-phase chromatography (RP) followed by tandem mass spectrometry. HILIC-separated metabolites were analyzed in positive ionization mode, while RP-separated metabolites were analyzed in both positive and negative ionization modes (referred to hereafter as HILIC positive, RP positive and RP negative).

HILIC positive and RP negative analyses detected over 3 000 analytes each, with 19 and 8 metabolites annotated, respectively. RP positive analysis was more comprehensive, detecting over 7 000 analytes, of which 29 were annotated (Figure 2.4a). Among the metabolites detected, 12348 compounds have been found in all replicates of each transconjugant, 160 analytes were uniquely present in *Cs*TKS transconjugants and absent in the EV control. Each *Cs*TKS transconjugant presented a different amount of exclusive analytes ranging from 48 to 120. A notable number of 278 analytes were present in all clones but pTKS-3 while only 66 excluded from pTKS-1 and 57 from pTKS-2. A total of 57 annotated analytes have been identified with 49 shared among all clones. No exclusive analyte from a clone was annotated. Three were exclusively shared by the pTKS clones: pyocyanin, a phenazine pigment derived from the shikimate pathway in *Pseudomonas aeruginosa*; 2-benzyl-4-chlorophenol (chlorophene), a phenolic halogenated compound with an unidentified biosynthetic

route; and abietic acid, a diterpene biosynthesized from GPP. Notably, 18 analytes detected in the EV transconjugants were absent in all three pTKS transconjugants (Figure 2.4b).

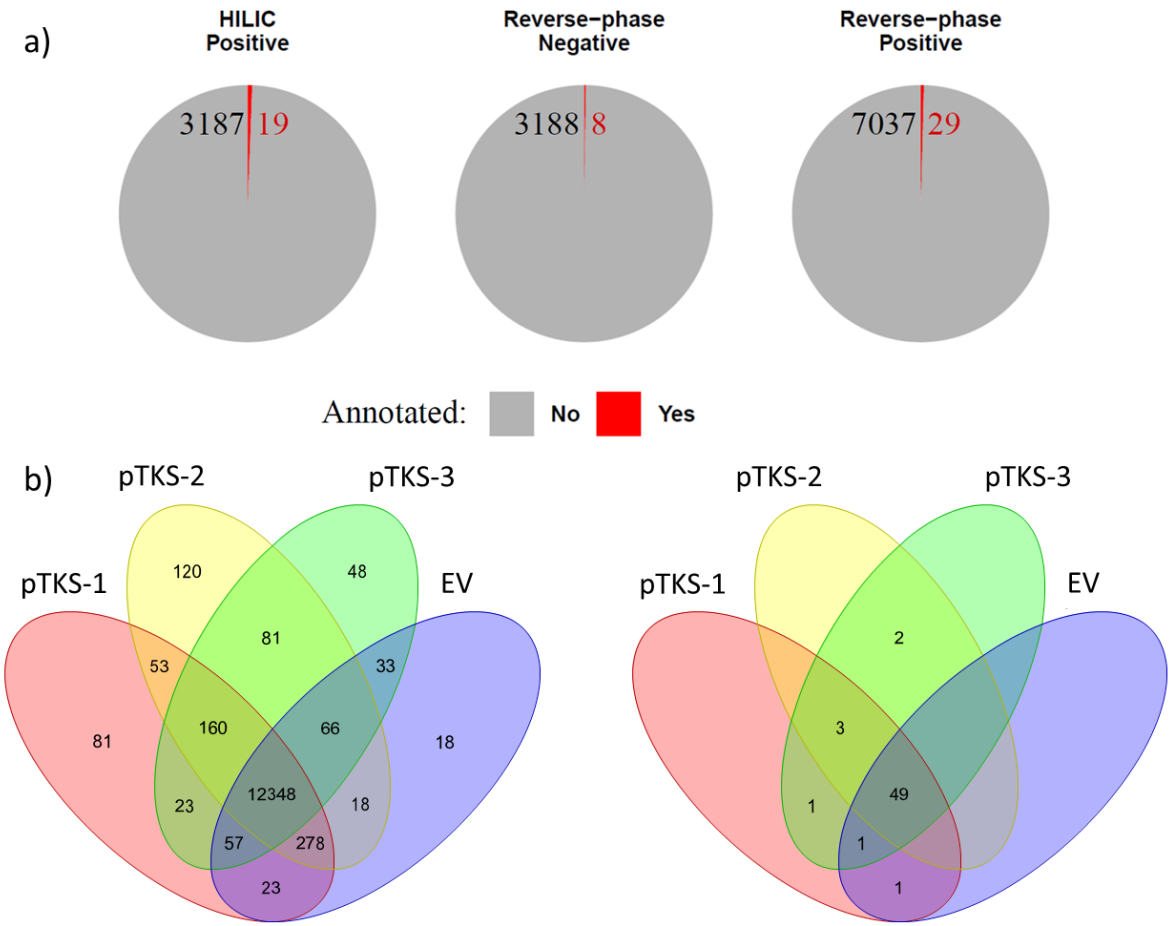


Figure 2.4. Untargeted metabolomic analysis identifies specific metabolites unique to *P. tricornutum* pTKS transconjugant lines. a) Proportion of analytes, annotated or not, detected using three different HPLC-MS/MS analytical methods: reversed-phase chromatography with positive ionization (RP ESI^+), reversed-phase chromatography with negative ionization (RP ESI), and hydrophilic interaction chromatography with positive ionization (HILIC ESI^+). b) Total analytes detected (left) and annotated (right) for each strain, combining data from all three analytical modes. Among the annotated analytes absent in EV negative control, the following were consistently detected in

pTKS transconjugant strains: pyocyanin, chlorophene and abietic acid were found in all pTKS lines; decyl dimethyl amine oxide and tri-isobutylphosphate were detected in pTKS-2 and pTKS-3; and 3-hydroxy-C6-homoserine lactone was identified in pTKS-1 and pTKS-3.

The second parameter analyzed was the quantitative impact of CsTKS expression on the diatom metabolome (Figure 2.a). Analytes were heterogeneously deregulated among the three pTKS transconjugants compared to the three replicates of EV transconjugants, with fold changes ranging from -10.3 (analyte with retention time (RT) = 3.768 min and with a *m/z* of 771.15744 in pTKS-3) to 11.2 (RT = 1.449 min; *m/z* = 431.29585 in pTKS-1). However, 48 analytes were consistently deregulated across all pTKS transconjugants triplicates (Figure 2.b).

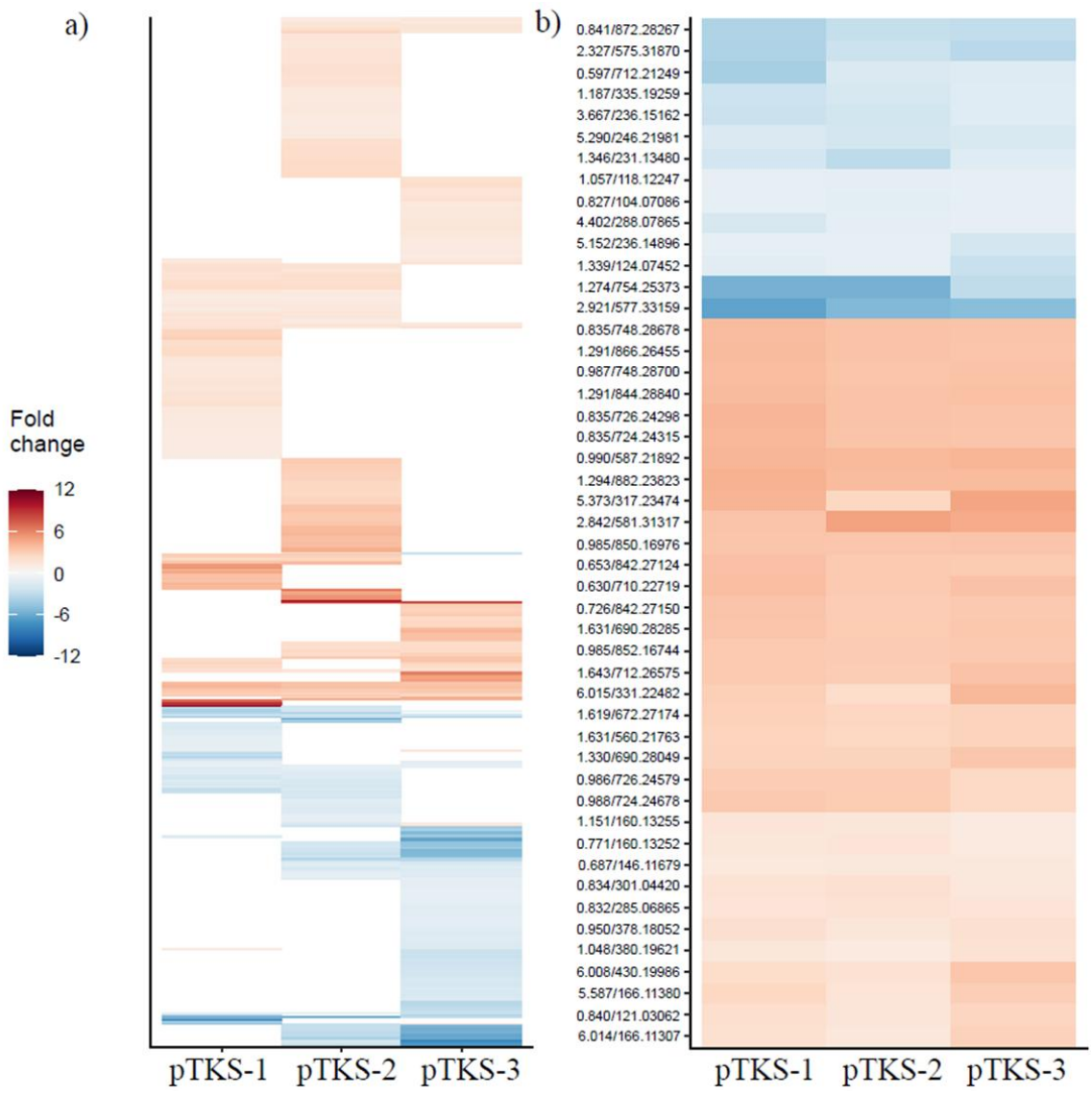


Figure 2.5. Differential metabolite abundance highlights consistent deregulation in *P. tricornutum* pTKS transconjugants compared to the empty vector. a) Overview of differentially abundant analytes detected using three analytical methods (Reversed-phased liquid chromatography coupled with ESI+/ESI- MS/MS and HILIC coupled with ESI+ MS/MS) compared to the empty vector. The y-axis lists all detected metabolites, showcasing the variability in abundance across pTKS transconjugant replicates. b) Focused analysis of 48 analytes consistently deregulated in all pTKS transconjugants compared to the empty vector. Only analytes with a (|fold change|)

greater than 1 and a significant threshold of $p < 0.01$ were included. These analytes exhibit consistent under- or overabundance across all pTKS transconjugant replicates.

Among these, the most deregulated analytes were detected at RT = 2.921 min; *m/z* = 577.33159 (fold change between -5.1 in pTKS-3 and -6.4 in pTKS-1) and RT = 2.842 min; *m/z* = 581.31317 (fold change between 3.4 in pTKS-1 and 4.9 in pTKS-2). Among the 48 deregulated analytes, 26 were annotated (Figure 2.6). Two of these annotations corresponded to 1-methyladenosine and two to thiamine. Among the 24 unique annotations, 13 were aromatic metabolites, including pinacidil, guanine, and worenine, the latter being downregulated in pTKS-2 transconjugants compared to the negative control. A single annotated metabolite, norvaline betaine, was deregulated in all pTKS transconjugants. Additionally, nine analytes were downregulated in all triplicates of at least one pTKS transconjugant. While analytes uniquely detected in the presence of *Cs*TKS (Figure 2.5b) were not consistently deregulated across all pTKS transconjugants due to variability in peak intensities among replicates, some showed significant upregulation. Among these, 3-hydroxy-C6-homoserine lactone, decyl dimethyl amine oxide, and proline betaine were the most upregulated analytes in the entire experiment.

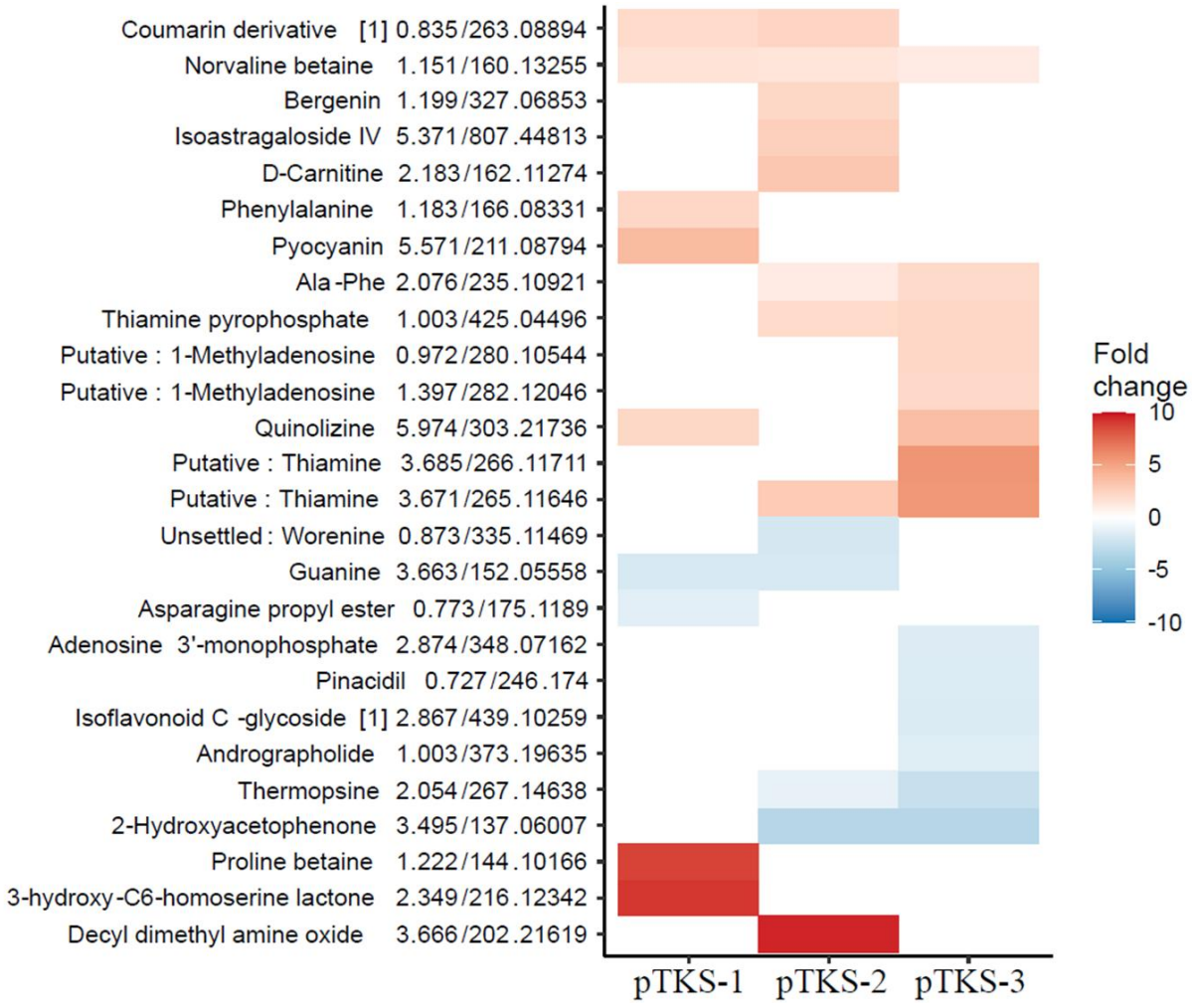


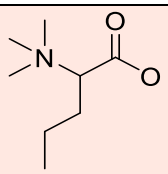
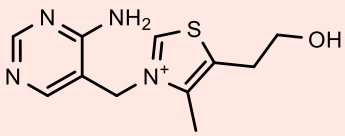
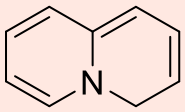
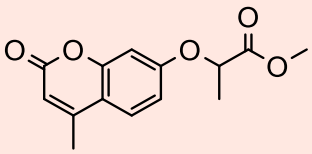
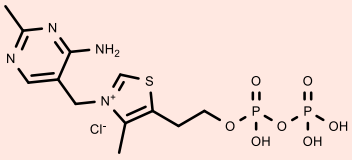
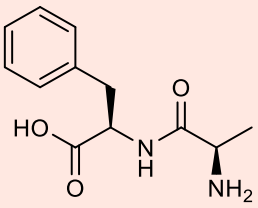
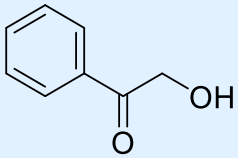
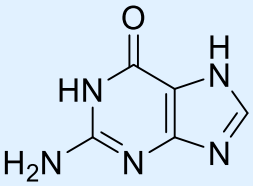
Figure 2.6. Presence of CsTKS coincides with altered abundance of cyclic metabolites in *P. tricornutum* transconjugants. Heatmap illustration the deregulation of annotated metabolites in each pTKS transconjugant replicates compared to the negative control. Only analytes with |fold change| > 1 and $p < 0.01$ were considered as deregulated. "Coumarin derivative [1]" corresponds to methyl 2-[(4-methyl-2-oxo-2H-chromen-7-yl)oxy]propanoate and "Isoflavonoid C-glycoside [1]" to Puerarin. Ala-Phe refers to alanylphenylalanine.

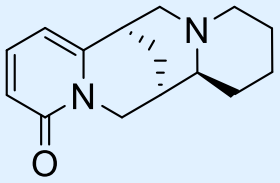
Annotated metabolites deregulated by CsTKS in *Phaeodactylum tricornutum*

To investigate the impact of *Cs*TKS on metabolism, we analyzed the structure and biosynthetic pathways of the nine annotated metabolites deregulated in at least two

transconjugant strains (Table 1). Similar metabolites or pathways were highlighted in the table to identify shared metabolic routes. Of the nine detected and annotated analytes, eight possess aromatic structures, including two dipeptides (norvaline betaine and alanylphenylalanine), two alkaloids of quinolizine and quinolizidine nature (quinoline and thermopsine), and thiamine, which was detected as both phosphorylated and not phosphorylated. The two dipeptides originate from distinct biosynthetic pathways. While the pathway for norvaline betaine biosynthesis in *P. tricornutum* remains unresolved, norvaline is a modified valine expected to derive from pyruvate metabolism via either the (2*S*)-2-isopropyl malate or (*S*)-2-acetolactate branches, both of which converge at 2-oxoisovalerate, a precursor to valine (Kanehisa et al., 2023). The betaine moiety present in norvaline betaine is likely biosynthesized through the oxidation of choline to trimethyl-glycine (Waditee et al., 2003, et al., 2000). Coumarin and alanylphenylalanine are both derived from the shikimate pathway, though the specific biosynthesis of methyl 2-[(4-methyl-2-oxo-2H-chromen-7-yl)oxy] propanoate (a coumarin derivative) remains uncharactarized. 2-Hydroxyacetophenone, alanylphenylalanine, and the coumarin derivative share a benzene ring, with the latter also containing a pyran structure. Meanwhile, guanine and thiamine are derived from the nucleotide metabolism, with guanine from the purine metabolism whereas the pyrimidine and thiazole moieties of thiamine are biosynthesized separately prior to be combined.

Table 2.1. Impact of heterologous CsTKS on *P. tricornutum*'s metabolism. The presence of *Cs*TKS alters the activity of key metabolic pathways, including the shikimate pathway, nucleotide metabolism, and amino acid metabolism. Metabolites highlighted in pink-orange are all upregulated, while those in blue are downgulated. Analytes are

Analyte	Chemical 2D structure	Family	Biosynthesis/ Metabolic Route
Norvaline Betaine		Amino Acid dipeptide	Pyruvate Isoleucine pathway and choline oxidation reactions (Kanehisa et al., 2023, Nyyssöl t al., 2000, Soini et al., 2008, Popko et al., 2016)
Thiamine*		Vitamin/Cofactor	Vitamine B6, glycine, pyruvate and purine metabolism (Llavero Pasquina, 2020, Kanehisa et al., 2023)
Quinolizine		Alkaloid / Quinolizine	Unsolved, possible pyridine derived precursor, Malonyl-CoA and a Polyketide synthase (Wang et al., 2020b)
Methyl 2-[(4-methyl-2-oxo-2H-chromen-7-yl)oxy]propanoate		Coumarine Derivative	Shikimate Pathway (Kanehisa et al., 2023, Del Mondo et al., 2022)
Thiamine Pyrophosphate		Vitamin/Cofactor	Pyruvate and purine metabolism (Llavero Pasquina, 2020, Kanehisa et al., 2023)
Alanylphenylalanine		Amino Acid dipeptide	Shikimate Pathway (Kanehisa et al., 2023, Silva et al., 2018)
2-Hydroxyacetophenone		Organic aromatic alkyl-phenylketones/Benzoic Acid	Putative Shikimate and polyketide Pathway (Singh et al., 2021)
Guanine		Nucleic Acid	Nucleotide metabolism (Kanehisa et al., 2023, Witte and

			Herde, 2020, Mojzeš et al., 2020)
Thermopsine		Alkaloid / Quinolizidine	Lysine decarboxylation (Zhang et al., 2022a, Wang et al., 2020b)

listed in descending order of fold change. The ‘’Biosynthesis/metabolic route’’ column emphasizes shared pathways or biosynthetic routes among the identified analytes.

Discussion

In this study, we successfully produced the *C. sativa* tetraketide synthase and olivetolic cyclase in *P. tricornutum*. Importantly, the heterologous expression of these enzymes did not impair the growth of the transconjugants (Supplementary Figure 2.2). Due to the unavailability of 3,5,7-trioxododecanoyl-CoA, both as a substrate and as a standard, coupled with its high molecular weight of 991.22 $g.mol^{-1}$ we were unable to directly test the enzymatic activities of *Cs*TKS and *Cs*OAC. Alternative strategies were therefore employed to evaluate their functionality. Analysis by HPLC-MS/MS (High pressure liquid chromatography with tandem mass spectrometry) of protein extracts containing *Cs*TKS and *Cs*OAC, either assessed alone or combined, did not yield a peak corresponding to olivetol or olivetolic acid standards. Several hypotheses could explain the absence of olivetolic acid in the extracts. Western blot analyses confirmed that both *Cs*TKS and *Cs*OAC were accumulating as soluble proteins in *P. tricornutum*, suggesting that protein misfolding is unlikely to account for the lack of product detection (Reynaud, 2010). Instead, the failure to detect olivetol or olivetolic acid may be attributed to the utilization of key substrates or intermediates by competing endogenous pathways. Another plausible explanation is that *Cs*TKS did not dimerize

correctly in *P. tricornutum* cellular environment, as proper dimerization is critical for its activity (Kearsey et al., 2020).

To investigate the broader impact of *Cs*TKS expression on the diatom metabolome, we performed an untargeted metabolomic analysis. This revealed significant differences in the metabolite profiles of pTKS transconjugants, with each replicate containing 50 to 120 exclusive metabolites that were absent from the control EV. A previous research study has suggested that clone consistency is rare in *P. tricornutum* (Diaz-Garza et al., 2024) supporting the variability in metabolite content and abundance observed in this study. Furthermore, 160 analytes were consistently detected in all three *CsTKS*-expressing replicates but were absent from the control EV. While only a fraction of these metabolites could be annotated, it is plausible that many of the unidentified analytes include polyketide-derived metabolites influenced *Cs*TKS activity. Previous untargeted metabolomic studies in *P. tricornutum* using Fourier-transform ion cyclotron-resonance mass spectrometer resulted in the detection of several hundred of annotated analytes (Duarte et al., 2022). However, the lower resolution of the UPLC-qTOF-MS used in our study limited the number of annotated metabolites. Despite these limitations, clear patterns emerged in the metabolite abundance data. Specifically, 48 metabolites consistently deregulated across all *CsTKS*-expressing transconjugant replicates, indicating a potential influence of *Cs*TKS on intracellular metabolite content. These results suggest that *Cs*TKS expression induces metabolic shifts in *P. tricornutum*, possibly altering the biosynthesis or accumulation of metabolites.

The structure of each annotated metabolites was analyzed to assess the potential involvement of TKS in their biosynthesis (Table 1). Polyketides are a diverse class of specialized metabolites characterized by the presence of C=O—CH_2 units or derived from such molecules, which complicates their structural identification. To address this

challenge, we relied on structural information corroborated by literature evidence. Among the deregulated metabolites, seven out of nine that were altered in two or more replicates and nine out of 16 altered in only one replicate were aromatic (Table 1 and Supplementary Table 3). In *C. sativa*, OAC catalyzes the formation of a phenolic structure from trioxododecanoyl-CoA. In *P. tricornutum*, it is plausible that a different enzyme takes over the biosynthesis of unidentified compounds from a polyketide-derived products generated by the *Cs*TKS enzyme.

Pyocyanin, one of the three exclusive annotated metabolites, was detected in all pTKS transconjugant replicates and was significantly upregulated in pTKS-1 transconjugants (Figure 2.4 and Figure 2.6). Pyocyanin is a well-characterized virulence factor in *Pseudomonas aeruginosa* (Marey et al., 2024, Jabłońska et al., 2023) and is classified as a phenazine pigment produced from chorismic acid (Zhou et al., 2022). Given that culture conditions were carefully controlled and remained stable over time and across transconjugants, the observed upregulation of pyocyanin is likely attributable to a direct or indirect modification in its biosynthetic pathway as PKS are capable of synthesizing pigments, rather than being driven by environmental factors, such as changes in light conditions or contamination (Heydarizadeh et al., 2017, Günther et al., 2022, Lebeau et al., 2017).

As previously mentioned, most of the detected metabolites exhibit a phenolic structure. Among them, a coumarin-derivative, methyl 2-[(4-methyl-2-oxo-2Hchromen-7-yl)oxy]propanoate, was upregulated in two pTKS transconjugant lines. While its precise biosynthetic pathway in vivo remains unclear, it likely originates from coumarin or coumaroyl-CoA, intermediates produced via the phenylpropanoid pathway. The formation of coumarin core is thought to involve condensation reactions between phenolic and carbonyl metabolites, which may also contribute to the biosynthesis of

other aromatic metabolites detected in this study (Smolecule, 2024). The potential involvement of a polyketide synthase such as *Cs*TKS in its synthesis remains uncertain. It has been reported that 4-coumaric acid can serve as substrate for type three polyketide synthase (PKSIII), which can catalyze the production of stilbene or chalcone metabolites (Watts et al., 2006). Although *Cs*TKS is a PKSIII, it is not expected to produce chalcone products, and no stilbene metabolites were detected among the 16 upregulated annotated metabolite. This is unexpected, as *Cs*TKS is hypothesized to generate stilbene through linear polyketides synthesis (Kearsey et al., 2020). Additionally, 3-hydroxy-C6-homoserine lactone, which is known to play a role in quorum sensing, was identified. While its structure resembles that of polyketides, its classification as a polyketide has not been confirmed (Weiland-Br uer et al., 2016, Liu et al., 2007). Another, metabolites, alanylphenylalanine, is a modified amino acid that could serve as a substrate for polyketide extension (Ray and Moore, 2016). It is plausible that the activity of *Cs*TKS stimulates the synthesis of alanylphenylalanine. Overall, we observed a general increase in metabolites derived from the shikimate pathway, including bergenin, pyocyanin, phenylalanine and coumarin. This suggest that *Cs*TKS may either influence the shikimate pathway or its products may affect light perception, thereby increasing the synthesis of metabolites capable of light absorption (Heydarizadeh et al., 2017). It has not been reported in literature that heterologous production of enzymes in *P. tricornutum* has by itself an impact on the activity of the shikimate pathway, although few of those study perform metabolomic analysis.

The only analyte consistently upregulated across all transconjugants in this study is norvaline betaine. The biological role and biosynthetic pathway of this modified amino acid remains largely unknown. Additionally, a significant upregulation of proline betaine was observed in TKS-1 transconjugant strains (Figure 2.6), suggesting that the

presence of *Cs*TKS might enhance betaine biosynthesis. Betaine, in the form of trimethylglycine, along with proline betaine and carnitine (upregulated in pTKS-1 and pTKS-2 lines, respectively as shown in Supplementary Table 3), are well-known quaternary ammonium compounds. These metabolites function as osmolytes and are typically upregulated in saline stress (Nikitashina et al., 2022). Beyond their osmoprotective properties, betaines are also involved in epigenetic regulation and can act as methyl donors, although the in vivo mechanism of amino acid betainization is still unclear (Tuomainen et al., 2019, et al., 2000). The observed upregulation of glycine could be associated with increased thiamine production, though the underlying causal relationship remains speculative.

Quinolizine, thermopsine, and worenine, which belong to the quinolizine family, were also identified. The biosynthesis of quinolizine in vivo is not well-characterized, though a synthetic pathway has been successfully achieved using a PKSIII enzyme from *Huperzia serrata* to produce 2-hydroxy-4H-quinolizin-4-one (Wang et al., 2020b). This raises the possibility that *Cs*TKS may contribute to quinolizine biosynthesis in *P. tricornutum*. However, thermopsine was downregulated in two pTKS transconjugant strains and worenine, an isoquinoline, was downregulated in one strain (Supplementary Table 3). This suggests that while *Cs*TKS might be involved in the production of quinolizine-related metabolites, its precise role and the regulatory mechanism influencing these metabolites remains unclear.

Hydroxyacetophones, such as karanjin, are suspected to be produced via the polyketide pathway (Gao et al., 2021). If the biosynthesis of 2-hydroxtyacetophenone follows a similar process in *P. tricornutum*, its reduced concentration in this study could be due to substrate channeling by the heterologous *Cs*TKS, which may utilize precursors typically employed by native polyketide synthases.

Conclusions

Significant differences were identified among transconjugant clones, with 278 analytes shared by EV (p*Pt*GE30), pTKS-1 and pTKS-2, but absent from pTKS-3, which exhibited a distinct metabolic profile (Supplementary Figure 2.4). Moreover, 160 metabolites were exclusive to *Cs*TKS transconjugants, supporting the hypothesis that *Cs*TKS is metabolically active in the pTKS transconjugant clones. Notably, most annotated analytes were present in all transconjugants, and no annotated metabolites was exclusive to a single clone. Among the metabolites detected exclusively in pTKS transconjugant lines, pyocyanin, abietic acid, and chlorophene were identified. However, these metabolites are not expected to be derived from the polyketide pathway. These findings underscore the complexity of *P. tricornutum* metabolism and suggest that achieving controllable and sustainable olivetolic acid production in this organism requires a more comprehensive understanding of its metabolic network.

List of abbreviations

THC: Δ^9-tetrahydrocannabinol

CBD: Cannabidiol

THCAS: Tetrahydrocannabinolic acid synthase

THCA: Δ^9-tetrahydrocannabinolic acid

CBDAS: Cannabidiolic acid synthase

CBDA: Cannabidiolic acid

TKS: Tetraketide synthase

OAC: Olivetolic acid cyclase

OA: Olivetolic acid

GPP: Geranyl pyrophosphate

APT: Aromatic prenyl transferase

HILIC: Hydrophilic interaction liquid chromatography

RP: Reversed-phase

MS: Mass spectrometry

PCR: Polymerase chain reaction

EV: Empty vector

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no conflict of interest.

Funding statement

Canada Research Chair on plant specialized metabolism Award No CRC-2018-00137 to I.D-P. Thanks are extended to the Canadian taxpayers and to the Canadian government for supporting the Canada Research Chairs Program. Additional support in the form of scholarship to K.C.G.d.S by Mitacs-Elevate program grant #IT28769 to I.D-P.

Data availability statement

Data generated or analyzed during this study are provided in full within the published article and its supplementary materials and from the corresponding author upon reasonable request.

Acknowledgement

The authors would like to thank Melodie B. Plourde for her professional assistance and support and all lab members for their teamwork. *Phaeodactylum tricornutum* strain used in this study was kindly provided by Professor Bogumil Karas (Western University, London, ON, Canada). Also, we would like to acknowledge the use of ChatGPT version 3.5, a free AI language model developed by OpenAI, for assistance in revising and refining the text of our manuscript. During the preparation of this work, the authors used ChatGPT 4.0 in order to correct grammatical errors and enhance readability. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

1. Śledziński P, Nowak-Terpiłowska A, Zeyland J: **Cannabinoids in medicine: cancer, immunity, and microbial diseases**. *International journal of molecular sciences* 2020, **22**(1):263.
2. Zou S, Kumar U: **Cannabinoid receptors and the endocannabinoid system: signaling and function in the central nervous system**. *International journal of molecular sciences* 2018, **19**(3):833.
3. Crocq M-A: **History of cannabis and the endocannabinoid system**. *Dialogues in clinical neuroscience* 2020, **22**(3):223-228.
4. Punja ZK, Sutton DB, Kim T: **Glandular trichome development, morphology, and maturation are influenced by plant age and genotype in high THC-containing cannabis (*Cannabis* sativa L.) inflorescences**. *Journal of Cannabis Research* 2023, **5**(1):12.
5. Conneely LJ, Mauleon R, Mieog J, Barkla BJ, Kretzschmar T: **Characterization of the *Cannabis* sativa glandular trichome proteome**. *PLoS One* 2021, **16**(4):e0242633.
6. Micale V, Mazzola C, Drago F: **Endocannabinoids and neurodegenerative diseases**. *Pharmacological Research* 2007, **56**(5):382-392.
7. Farag S, Kayser O: **The cannabis plant: Botanical aspects**. In: *Handbook of cannabis and related pathologies.* Elsevier; 2017: 3-12.

8. Gagne SJ, Stout JM, Liu E, Boubakir Z, Clark SM, Page JE: **Identification of olivetolic acid cyclase from *Cannabis* sativa reveals a unique catalytic route to plant polyketides**. *Proceedings of the National Academy of Sciences* 2012, **109**(31):12811-12816.
9. Luo X, Reiter MA, d'Espaux L, Wong J, Denby CM, Lechner A, Zhang Y, Grzybowski AT, Harth S, Lin W: **Complete biosynthesis of cannabinoids and their unnatural analogues in yeast**. *Nature* 2019, **567**(7746):123-126.
10. Tahir MN, Shahbazi F, Rondeau-Gagné S, Trant JF: **The biosynthesis of the cannabinoids**. *Journal of cannabis research* 2021, **3**:1-12.
11. Xie Z, Mi Y, Kong L, Gao M, Chen S, Chen W, Meng X, Sun W, Chen S, Xu Z: ***Cannabis* sativa: origin and history, glandular trichome development, and cannabinoid biosynthesis**. *Horticulture Research* 2023, **10**(9):uhad150.
12. Hammond D, Goodman S, Wadsworth E, Freeman TP, Kilmer B, Schauer G, Pacula RL, Hall W: **Trends in the use of cannabis products in Canada and the USA, 2018–2020: Findings from the International *Cannabis* Policy Study**. *International Journal of Drug Policy* 2022, **105**:103716.
13. ***Cannabis* Market Size, Share & COVID-19 Impact Analysis, By Type (Flowers/Buds and Concentrates), By Application (Medical, Recreational (Edibles and Topicals), and Industrial Hemp) By Component (THC-Dominant, Balanced THC & CBD, and CBD Dominant), and Regional Forecast, 2023-2030** [https://www.fortunebusinessinsights.com/industry-reports/cannabis-marijuana-market-100219]
14. Lewis MM, Yang Y, Wasilewski E, Clarke HA, Kotra LP: **Chemical profiling of medical cannabis extracts**. *ACS omega* 2017, **2**(9):6091-6103.
15. Backer R, Schwinghamer T, Rosenbaum P, McCarty V, Eichhorn Bilodeau S, Lyu D, Ahmed MB, Robinson G, Lefsrud M, Wilkins O: **Closing the yield gap for cannabis: a meta-analysis of factors determining cannabis yield**. *Frontiers in plant science* 2019, **10**:495.
16. Wiles D, Shanbhag BK, O'Brien M, Doblin MS, Bacic A, Beddoe T: **Heterologous production of *Cannabis* sativa-derived specialised metabolites of medicinal significance–Insights into engineering strategies**. *Phytochemistry* 2022, **203**:113380.
17. Zirpel B, Degenhardt F, Martin C, Kayser O, Stehle F: **Engineering yeasts as platform organisms for cannabinoid biosynthesis**. *Journal of biotechnology* 2017, **259**:204-212.
18. Kumar G, Shekh A, Jakhu S, Sharma Y, Kapoor R, Sharma TR: **Bioengineering of microalgae: recent advances, perspectives, and regulatory challenges for industrial application**. *Frontiers in Bioengineering and Biotechnology* 2020, **8**:914.
19. Gao J, Jiang L, Lian J: **Development of synthetic biology tools to engineer Pichia pastoris as a chassis for the production of natural products**. *Synthetic and systems biotechnology* 2021, **6**(2):110-119.
20. Nouemssi SB, Ghribi M, Beauchemin R, Meddeb-Mouelhi F, Germain H, Desgagné-Penix I: **Rapid and efficient colony-PCR for high throughput screening of genetically transformed Chlamydomonas reinhardtii**. *Life* 2020, **10**(9):186.
21. Laban A: **Cannabinoid production in algae**. In.: Google Patents; 2021.
22. Kassaw TK, Paton AJ, Peers G: **Episome-based gene expression modulation platform in the model diatom *Phaeodactylum tricornutum***. *ACS Synthetic Biology* 2022, **11**(1):191-204.
23. George J, Kahlke T, Abbriano RM, Kuzhiumparambil U, Ralph PJ, Fabris M: **Metabolic engineering strategies in diatoms reveal unique phenotypes and genetic configurations with implications for algal genetics and synthetic biology**. *Frontiers in Bioengineering and Biotechnology* 2020, **8**:513.
24. Pudney A, Gandini C, Economou CK, Smith R, Goddard P, Napier JA, Spicer A, Sayanova O: **Multifunctionalizing the marine diatom *Phaeodactylum tricornutum***

for sustainable co-production of omega-3 long chain polyunsaturated fatty acids and recombinant phytase. *Scientific reports* 2019, **9**(1):11444.
25. Xue J, Niu Y-F, Huang T, Yang W-D, Liu J-S, Li H-Y: **Genetic improvement of the microalga *Phaeodactylum tricornutum* for boosting neutral lipid accumulation**. *Metabolic engineering* 2015, **27**:1-9.
26. Lazarjani MP, Young O, Kebede L, Seyfoddin A: **Processing and extraction methods of medicinal cannabis: A narrative review**. *Journal of cannabis research* 2021, **3**:1-15.
27. Fajardo AR, Cerdán LE, Medina AR, Fernández FGA, Moreno PAG, Grima EM: **Lipid extraction from the microalga *Phaeodactylum tricornutum***. *European Journal of Lipid Science and Technology* 2007, **109**(2):120-126.
28. Cui Y, Thomas-Hall SR, Schenk PM: ***Phaeodactylum tricornutum* microalgae as a rich source of omega-3 oil: Progress in lipid induction techniques towards industry adoption**. *Food chemistry* 2019, **297**:124937.
29. Fabris M, George J, Kuzhiumparambil U, Lawson CA, Jaramillo-Madrid AC, Abbriano RM, Vickers CE, Ralph P: **Extrachromosomal genetic engineering of the marine diatom *Phaeodactylum tricornutum* enables the heterologous production of monoterpenoids**. *ACS synthetic biology* 2020, **9**(3):598-612.
30. **hexanoyl-CoA** [https://biocyc.org/compound?orgid=PHAEO&id=HEXANOYL-COA#RXNS]
31. Fantino E, Awwad F, Merindol N, Garza AMD, Gélinas S-E, Robles GCG, Custeau A, Meddeb-Mouelhi F, Desgagné-Penix I: **Bioengineering *Phaeodactylum tricornutum*, a marine diatom, for cannabinoid biosynthesis**. *Algal Research* 2024, **77**:103379.
32. Fantino E, Messaabi A, Mérindol N, Awwad F, Sene N, Gélinas S-E, Custeau A, Rhéaume K-L, Meddeb-Mouelhi F, Desgagné-Penix I: **Extrachromosomal expression of functional *Cannabis* sativa cannabidiolic acid synthase in Phaedodactylum *tricornutum***. *Algal Research* 2024:103889.
33. Awwad F, Fantino EI, Héneault M, Diaz-Garza AM, Merindol N, Custeau A, Gélinas S-E, Meddeb-Mouelhi F, Li J, Lemay J-F: **Bioengineering of the Marine Diatom *Phaeodactylum tricornutum* with *Cannabis* Genes Enables the Production of the Cannabinoid Precursor, Olivetolic Acid**. *International Journal of Molecular Sciences* 2023, **24**(23):16624.
34. Karas BJ, Molparia B, Jablanovic J, Hermann WJ, Lin Y-C, Dupont CL, Tagwerker C, Yonemoto IT, Noskov VN, Chuang R-Y *et al*: **Assembly of eukaryotic algal chromosomes in yeast**. *Journal of Biological Engineering* 2013, **7**(1):30.
35. Diamond A, Diaz-Garza AM, Li J, Slattery SS, Merindol N, Fantino E, Meddeb-Mouelhi F, Karas BJ, Barnabé S, Desgagné-Penix I: **Instability of extrachromosomal DNA transformed into the diatom *Phaeodactylum tricornutum***. *Algal Research* 2023, **70**:102998.
36. Karas BJ, Diner RE, Lefebvre SC, McQuaid J, Phillips APR, Noddings CM, Brunson JK, Valas RE, Deerinck TJ, Jablanovic J *et al*: **Designer diatom episomes delivered by bacterial conjugation**. *Nature Communications* 2015, **6**(1):6925.
37. Lai Z, Tsugawa H, Wohlgemuth G, Mehta S, Mueller M, Zheng Y, Ogiwara A, Meissen J, Showalter M, Takeuchi K *et al*: **Identifying metabolites by integrating metabolome databases with mass spectrometry cheminformatics**. *Nature Methods* 2018, **15**(1):53-56.
38. Tsugawa H, Cajka T, Kind T, Ma Y, Higgins B, Ikeda K, Kanazawa M, VanderGheynst J, Fiehn O, Arita M: **MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis**. *Nature Methods* 2015, **12**(6):523-526.
39. Tsugawa H, Nakabayashi R, Mori T, Yamada Y, Takahashi M, Rai A, Sugiyama R, Yamamoto H, Nakaya T, Yamazaki M *et al*: **A cheminformatics approach to characterize metabolomes in stable-isotope-labeled organisms**. *Nature Methods* 2019, **16**(4):295-298.

40. **Databank** [https://mona.fiehnlab.ucdavis.edu/]
41. Pang Z, Xu L, Viau C, Lu Y, Salavati R, Basu N, Xia J: **MetaboAnalystR 4.0: a unified LC-MS workflow for global metabolomics**. *Nature Communications* 2024, **15**(1).
42. Schwender H: **siggenes: Multiple Testing using SAM and Efron's Empirical Bayes Approaches**. In., 1.76.0 edn; 2023.
43. Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M: **KEGG for taxonomy-based analysis of pathways and genomes**. *Nucleic acids research* 2023, **51**(D1):D587-D592.
44. Waditee R, Tanaka Y, Aoki K, Hibino T, Jikuya H, Takano J, Takabe T, Takabe T: **Isolation and Functional Characterization ofN-Methyltransferases That Catalyze Betaine Synthesis from Glycine in a Halotolerant Photosynthetic Organism Aphanothece halophytica**. *Journal of Biological Chemistry* 2003, **278**(7):4932-4942.
45. Nyyssöl A, Kerovuo J, Kaukinen P, von Weymarn N, Reinikainen T: **Extreme halophiles synthesize betaine from glycine by methylation**. *Journal of Biological Chemistry* 2000, **275**(29):22196-22201.
46. Soini J, Falschlehner C, Liedert C, Bernhardt J, Vuoristo J, Neubauer P: **Norvaline is accumulated after a down-shift of oxygen in Escherichia coli W3110**. *Microbial cell factories* 2008, **7**:1-14.
47. Popko J, Herrfurth C, Feussner K, Ischebeck T, Iven T, Haslam R, Hamilton M, Sayanova O, Napier J, Khozin-Goldberg I: **Metabolome analysis reveals betaine lipids as major source for triglyceride formation, and the accumulation of sedoheptulose during nitrogen-starvation of *Phaeodactylum tricornutum***. *PLoS One* 2016, **11**(10):e0164673.
48. Llavero Pasquina, M. (2020). *Molecular biology of B vitamin metabolism genes and their regulation in Phaeodactylum tricornutum* [Apollo - University of Cambridge Repository]. https://doi.org/10.17863/CAM.61337.
49. Wang J, Ding N, Wu Y, Shi X, Qi B, Liu X, Wang X, Li J, Tu P, Shi S: **Enzymatic synthesis of 2-hydroxy-4H-quinolizin-4-one scaffolds by integrating coenzyme a ligases and a type III PKS from Huperzia serrata**. *RSC advances* 2020, **10**(40):23566-23572.
50. Del Mondo A, Sansone C, Brunet C: **Insights into the biosynthesis pathway of phenolic compounds in microalgae**. *Computational and Structural Biotechnology Journal* 2022, **20**:1901-1913.
51. Silva C, da Silva Filho J, Pinheiro G, Teixeira A, Freire P: **Vibrational and structural properties of L-Alanyl-L-Phenylalanine dipeptide by Raman spectroscopy, infrared and DFT calculations**. *Vibrational spectroscopy* 2018, **98**:128-133.
52. Singh A, Bhatt G, Gujre N, Mitra S, Swaminathan R, Limaye A, Rangan L: **Karanjin**. *Phytochemistry* 2021, **183**:112641.
53. Witte C-P, Herde M: **Nucleotide metabolism in plants**. *Plant physiology* 2020, **182**(1):63-78.
54. Mojzeš P, Gao L, Ismagulova T, Pilátová J, Moudříková Š, Gorelova O, Solovchenko A, Nedbal L, Salih A: **Guanine, a high-capacity and rapid-turnover nitrogen reserve in microalgal cells**. *Proceedings of the National Academy of Sciences* 2020, **117**(51):32722-32730.
55. Zhang P, An Q, Yi P, Cui Y, Zou J-B, Yuan C-M, Zhang Y, Gu W, Huang L-J, Zhao L-H: **Thermlanseedlines A–G, seven thermopsine-based alkaloids with antiviral and insecticidal activities from the seeds of Thermopsis lanceolata R. Br**. *Fitoterapia* 2022, **158**:105140.
56. Reynaud E: **Protein misfolding and degenerative diseases**. *Nature Education* 2010, **3**(9):28.
57. Kearsey LJ, Prandi N, Karuppiah V, Yan C, Leys D, Toogood H, Takano E, Scrutton NS: **Structure of the *Cannabis* sativa olivetol-producing enzyme reveals**

cyclization plasticity in type III polyketide synthases. *The FEBS Journal* 2020, **287**(8):1511-1524.
58. Diaz-Garza AM, Merindol N, Dos Santos KCG, Lavoie-Marchand F, Ingalls B, Desgagne-Penix I: **No two clones are alike: characterization of heterologous subpopulations in a transgenic cell line of the model diatom *Phaeodactylum tricornutum***. *Microb Cell Fact* 2024, **23**(1):286.
59. Duarte B, Feijão E, Cruz de Carvalho R, Duarte IA, Marques AP, Maia M, Hertzog J, Matos AR, Cabrita MT, Caçador I: **Untargeted metabolomics reveals antidepressant effects in a marine photosynthetic organism: The diatom *Phaeodactylum tricornutum* as a case study**. *Biology* 2022, **11**(12):1770.
60. Marey MA, Abozahra R, El-Nikhely NA, Kamal MF, Abdelhamid SM, El-Kholy MA: **Transforming microbial pigment into therapeutic revelation: extraction and characterization of pyocyanin from Pseudomonas aeruginosa and its therapeutic potential as an antibacterial and anticancer agent**. *Microbial Cell Factories* 2024, **23**(1):174.
61. Jabłońska J, Augustyniak A, Dubrowska K, Rakoczy R: **The two faces of pyocyanin-why and how to steer its production?** *World Journal of Microbiology and Biotechnology* 2023, **39**(4):103.
62. Zhou H, Yang Y, Shang W, Rao Y, Chen J, Peng H, Huang J, Hu Z, Zhang R, Rao X: **Pyocyanin biosynthesis protects Pseudomonas aeruginosa from nonthermal plasma inactivation**. *Microbial biotechnology* 2022, **15**(6):1910-1921.
63. Heydarizadeh P, Boureba W, Zahedi M, Huang B, Moreau B, Lukomska E, Couzinet-Mossion A, Wielgosz-Collin G, Martin-Jézéquel V, Bougaran G: **Response of CO2-starved diatom *Phaeodactylum tricornutum* to light intensity transition**. *Philosophical Transactions of the Royal Society B: Biological Sciences* 2017, **372**(1728):20160396.
64. Günther M, Reimer C, Herbst R, Kufs JE, Rautschek J, Ueberschaar N, Zhang S, Peschel G, Reimer L, Regestein L: **Yellow polyketide pigment suppresses premature hatching in social amoeba**. *Proceedings of the National Academy of Sciences* 2022, **119**(43):e2116122119.
65. Lebeau J, Venkatachalam M, Fouillaud M, Petit T, Vinale F, Dufossé L, Caro Y: **Production and new extraction method of polyketide red pigments produced by ascomycetous fungi from terrestrial and marine habitats**. *Journal of Fungi* 2017, **3**(3):34.
66. **methyl 2-[(4-methyl-2-oxo-2H-chromen-7-yl)oxy]propanoate** [https://www.smolecule.com/products/s12619316#pos-top]
67. Watts KT, Lee PC, Schmidt-Dannert C: **Biosynthesis of plant-specific stilbene polyketides in metabolically engineered Escherichia coli**. *BMC biotechnology* 2006, **6**:1-12.
68. Weiland-Br r N, Kisch MJ, Pinnow N, Liese A, Schmitz RA: **Highly effective inhibition of biofilm formation by the first metagenome-derived AI-2 quenching enzyme**. *Frontiers in microbiology* 2016, **7**:1098.
69. Liu X, Bimerew M, Ma Y, Müller H, Ovadis M, Eberl L, Berg G, Chernin L: **Quorum-sensing signaling is required for production of the antibiotic pyrrolnitrin in a rhizospheric biocontrol strain of Serratia plymuthica**. *FEMS microbiology letters* 2007, **270**(2):299-305.
70. Ray L, Moore BS: **Recent advances in the biosynthesis of unusual polyketide synthase substrates**. *Natural product reports* 2016, **33**(2):150-161.
71. Nikitashina V, Stettin D, Pohnert G: **Metabolic adaptation of diatoms to hypersalinity**. *Phytochemistry* 2022, **201**:113267.
72. Tuomainen M, K kk nen O n J, Auriola S, Lehtonen M, Savolainen MJ, Hermansen K, Risérus U, Åkesson B, Thorsdottir I: **Quantitative assessment of betainized compounds and associations with dietary and metabolic biomarkers in the randomized study of the healthy Nordic diet (SYSDIET)**. *The American journal of clinical nutrition* 2019, **110**(5):1108-1118.

CHAPTER III: Heterologous expression of the high molecular weight *Dictyostelium discoideum* hybrid enzyme SteelyA in the diatom *Phaeodactylum tricornutum*

Nicolas Sene [1], Basanta Lamichhane [1], Sarah-Eve Gélinas [1], Alexandre Custeau [1], Natacha Merindol [1], Fatma Meddeb-Mouelhi [1], and Isabel Desgagné-Penix [1,*]

[1]Department of biochemistry, chemistry, physics and forensic science, Université du Québec à Trois-Rivières, Trois-Rivières, QC, Canada ; nicolas.sene@uqtr.ca ; basanta.lamichhane@uqtr.ca ; alexandre.custeau@uqtr.ca ; natacha.merindol@uqtr.ca ; fatma.meddeb@uqtr.ca

Correspondence: Isabel.Desgagne-Penix@uqtr.ca

In revision for publication.

Author Contributions: NS: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. BL: Formal analysis, Data curation, Investigation, Writing – review & editing. NM: Conceptualization, Formal analysis, Investigation, Writing – review & editing. SEG: Formal analysis, Writing – review & editing. AC: Formal analysis, Validation. FM: Formal analysis, Investigation, Methodology, Writing – review & editing. HG: Resources, Supervision, Writing – review & editing. IDP: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing. All authors have read and agreed to the published version of the manuscript.

Résumé :

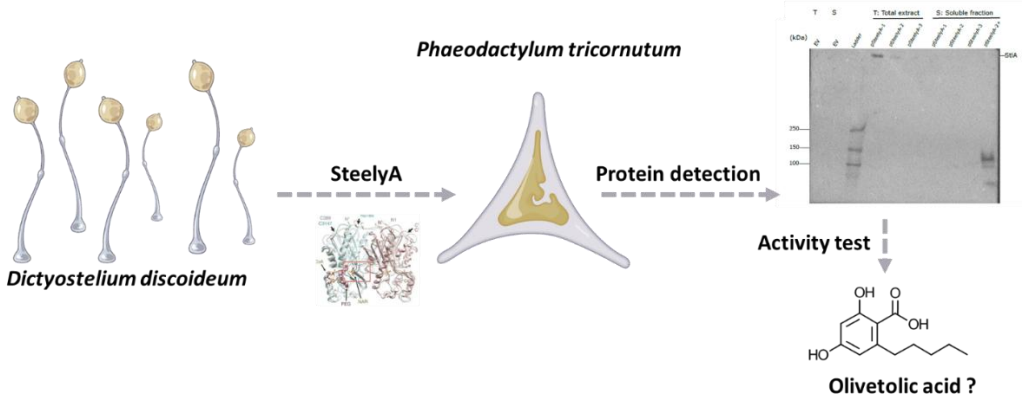
La bioproduction de molecules à haute valeur offre une alternative durable à la l'extraction conventionnelle ou à la synthèse chimique, tout particulièrement pour des métabolites complexes tels que les cannabinoïdes (CBs), possédant des propriétés thérapeutiques pour traiter les maladies neurodégénératives. La diatomée marine *Phaeodactylum tricornutum* présente un châssis prometteur pour la synthèse de CBs dû à son haut contenu en lipides, élément essentiel pour synthétiser des CBs. Dans cette étude nous explorons la possibilité de produire de l'acide olivetolique (OA), précurseur clé, en utilisant une acide gras synthase hybride polycétide synthase ; SteelyA de *Dictyostelium discoideum*. Contrairement aux enzymes natives de *Cannabis sativa*, la tetraketide synthase et l'acide olivetolique cyclase, qui présentent une faible productivité et stabilité dans la diatomée, SteelyA promettait d'offrir une voie alternative. Son expression hétérologue dans *P. tricornutum* a résulté dans l'expression stable de la protéine de très haut poids moléculaire (352 kDa) pendant 6 mois sans affecter la croissance cellulaire. Toutefois, les analyses HPLC-MS n'ont pas détecté d'OA ou de dérivés in vitro, ou in vivo, suggérant une inactivité de l'enzyme ou une limite métabolique. Ces découvertes soulignent le besoin d'investiguer les prérequis métaboliques et protéiques pour la biosynthèse d'OA, et guident les stratégies d'optimisation future pour la production durable de cannabinoïdes.

Abstract: The bioproduction of high-value molecules offers a sustainable and cost-effective alternative to traditional extraction and chemical synthesis, particularly for complex metabolites like cannabinoids (CBs), which have therapeutic potential for neurodegenerative diseases. The marine diatom *Phaeodactylum tricornutum* presents a promising chassis for CB biosynthesis due to its high lipid content, essential

building blocks to biosynthesize CBs. In this study, we explored the feasibility of producing olivetolic acid (OA), the key CB precursor, using a hybrid type polyketide synthase, SteelyA, from *Dictyostelium discoideum*. Unlike the native *Cannabis sativa* enzymes—tetraketide synthase and olivetolic acid cyclase—which exhibit low productivity and stability in diatoms, SteelyA was expected to offer an alternative biosynthetic route. Heterologous expression in *P. tricornutum* resulted in stable production of the very high-molecular-weight (352 kDa) SteelyA protein over six months without affecting cellular growth. However, HPLC-MS analysis did not detect OA or its derivatives in vivo and in vitro, suggesting enzymatic inactivity or metabolic limitations. These findings highlight the need for further investigation into the metabolic and proteomic requirements for CB precursor biosynthesis in diatoms, guiding future optimization strategies for sustainable cannabinoids production.

Keywords: metabolic engineering; cannabinoids; microalgae; diatom bioengineering; heterologous production; high-molecular-weight protein; olivetolic acid.

Graphical Abstract



1. Introduction

Cannabinoids (CBs), including Δ9-tetrahydrocannabinol (THC) and cannabidiol (CBD), are coveted for the effects stemming from their interaction with the human endocannabinoid system (Śledziński et al., 2020, Crocq, 2020a). They are naturally produced in the glandular trichomes of *Cannabis* plants. Tetraketide synthase (TKS) and olivetolic acid cyclase (OAC) from hexanoyl-CoA and malonyl-CoA lead to the biosynthesis of olivetolic acid (OA), a common intermediate of most cannabinoids (Luo et al., 2019a). The current cannabinoids supply relies on the extraction from cultivated *Cannabis sativa,* however, the cost of cultivation and difficult separation processes led to the search for alternatives (Lewis et al., 2017, Backer et al., 2019).

In recent years, bioengineering has emerged as a powerful alternative to conventional extraction and chemical synthesis for obtaining a variety of valuable compounds. Based on using and enhancing natural reactions to produce complex molecules, bioproduction converts the need for costly, energy-intensive, and low-yield extraction processes into more accessible organism cultivation and metabolite purification (Hara et al., 2014, Clomburg et al., 2017, Wilding et al., 2018). Biofuels, solvents, pharmaceuticals, and nutraceuticals are now abundantly produced using

bioengineering (Lozano Terol et al., 2019, Parapouli et al., 2020, Overkamp et al., 2002). Achieving the bioproduction of molecules of diverse nature and size demands master knowledge and techniques for each suitable model organism.

With an endogenously high lipid content, the microalgae *Phaeodactylum tricornutum* is a prominent choice as a green platform to produce biofuels, nutraceuticals, and even pharmaceuticals (Hempel et al., 2011, Branco-Vieira et al., 2020, Dhaouadi et al., 2020, Cui et al., 2021, Awwad et al., 2023a, Fantino et al., 2024a). The toolbox and data available to work on the diatom is rapidly growing, and is therefore gaining an increasing interest in the bioproduction industry (Butler et al., 2022). As it naturally produces malonyl-CoA and is predicted to produce hexanoyl-CoA, two substrates involved in the biosynthesis of fatty acids, it has been tested as a heterologous host to support the synthesis of cannabinoids (Fantino et al., 2024a, Awwad et al., 2023a, Cui et al., 2019, BioCyC, 2024). The in vitro production of CBGA by enzymatic assay using protein extracts has been successfully achieved in *P. tricornutum* using geranyl pyrophosphate and OA (Fantino et al., 2024a). However, the heterologous expression of *C. sativa* tetraketide synthase and olivetolic acid cyclase to yield OA showed limited productivity and stability in the diatom expression systems (Awwad et al., 2023). Thus, alternative enzymes, notably from other organisms, have been investigated to circumvent this necessary step of the cannabinoid biosynthesis pathway.

One potential solution is the utilization of *Dictyostelium discoideum* enzymes. *D. discoideum* is a unicellular eukaryote that undergoes multicellular aggregation under nutrient deprivation, forming a fruiting body-like structure reminiscent of fungal development (Fortunato et al., 2003, Aubry and Firtel, 1999). During this differentiation process, *D. discoideum* produces SteelyA, a hybrid enzyme combining

features of type I fatty acid synthases and type III polyketide synthases. In its native host, SteelyA catalyzes the biosynthesis of 4-methyl-5-pentylbenzene-1,3-diol (MPBD) and methyl-olivetol from acyl-CoA precursors (Figure 3.1) (Narita et al., 2011, Reimer et al., 2022). Interestingly, deletion of the enzyme's methyltransferase domain results in the production of olivetol, the direct precursor to OA and other cannabinoids (Anjard et al., 2011, Ghosh et al., 2008, Reimer et al., 2022). When expressed heterologously in *Saccharomyces cerevisiae*, SteelyA primarily produces methyl-olivetol (Mookerjee et al., 2018, Mookerjee et al., 2022) (Figure 1), with additional side products such as acylpyrones when two malonyl-CoA moieties are incorporated into the reaction.

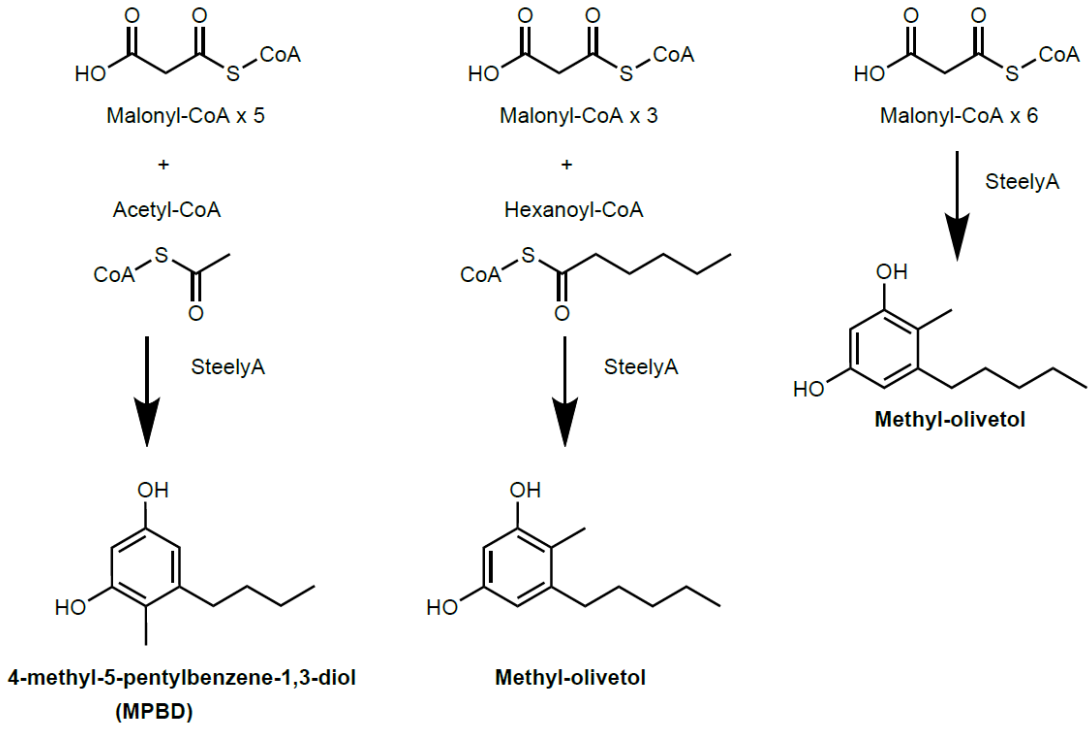


Figure 3.1. Proposed reactions catalyzed by the SteelyA enzyme. Bolded molecules indicate products observed from SteelyA activity. *Left reaction*: The native reaction of SteelyA in *Dictyostelium discoideum* as reported by (Ghosh et al.,

2008), where SteelyA utilizes acetyl-CoA and five malonyl-CoA molecules to produce MPBD. *Middle reaction*: The reaction described by (Reimer et al., 2022) in *D. discoideum* showing that SteelyA can use a hexanoyl-CoA starter unit along with three malonyl-CoA molecules to generate methyl-olivetol. *Right reaction:* represents the reaction observed by heterologous expression of SteelyA in *Saccharomyces cerevisiae* as described by (Mookerjee et al., 2018).

To address the limitations of TKS/OAC expression and assess an alternative strategy for OA biosynthesis, we investigated the heterologous production of SteelyA in *P. tricornutum*. This study represents a significant milestone in diatom-based metabolic engineering, as it involves the expression of the largest heterologous protein ever produced in *P. tricornutum* to date, a 352 kDa multifunctional enzyme. We inserted a genetic cassette containing the *steelyA* gene into *P. tricornutum* and evaluated both the production and enzymatic activity of the expressed protein. By assessing the feasibility of expressing such a large polyketide synthase/fatty acid synthase hybrid, this work provides new insights into the potential of *P. tricornutum* as a chassis for the biosynthesis of complex metabolites, expanding its applications in synthetic biology and metabolic engineering.

2. Materials and Methods

2.1. Microbial strains and growth conditions

E. coli strains 10β (New England Biolab®, ON, Canada) and Epi300 (TransforMax Epi300 Biosearch™ Technologies, Novato, CA, USA) were grown in Luria Broth (LB) supplemented with chloramphenicol (35 µg. mL^{-1}) (Thermo Fisher Scientific, Ottawa, ON, Canada) when harboring the p*Pt*GE30 plasmid, and gentamicin (50 µg. mL^{-1}) (Thermo Fisher Scientific, Ottawa, ON, Canada) when containing pTA-MOB

plasmid. All these strains were grown at 37°C and under agitation at 220 rpm when in liquid culture. *P. tricornutum* strain (CCAP 1055/1, Culture Collection of Algae and Protozoa) was grown in L1 medium (Diamond et al., 2023b) at 18°C under white lights (75 µE m^{-2} s^{-1}) at 130 rpm with a photoperiod of 16h of light. Each strain containing p*Pt*GE30 backbone was grown with 50 µg. mL^{-1} of zeocin (Invitrogen, Burlington, ON, Canada).

2.2. Plasmid constructions

For the expression plasmid constructs, two parts of the *steelyA* cassettes were synthesized by Genewiz® From Azenta Life Sciences (Waltham, MA, USA) in fragments 1MN and 1Q using various primers (Appendix A – Table A1 and Table A2) and assembled with the p*Pt*GE30 vector (named EV hereafter) by Gibson assembly using the NEBuilder® HiFi DNA Assembly Bundle for Large Fragments (New England Biolabs, Whitby, ON, Canada) and then transformed into *E. coli* 10β strain. The final assembled construct contains a *Sh Ble* bleomycin resistance gene used for the selection, a *Highly abundant secreted protein 1* (*Hasp1*) promoter, and an *FcpA* terminator were used to drive the expression of the *steelyA* gene from the patent (Mookerjee et al., 2018). A myc tag was introduced into the C-terminal of the *steelyA* gene. Plasmid DNA sequence integrity was verified by Next Generation Sequencing (NGS) (Appendix A - Table A2).

2.3. Phaeodactylum tricornutum conjugation and screening

To introduce the *steelyA* constructs into *P. tricornutum*, each plasmid was transformed into *E. coli* Epi 300 strain with pTA-MOB plasmid. Conjugation was performed as described in (Karas et al., 2015a) following exactly the steps described in (Fantino et al., 2024a) up to the final liquid culture of the transconjugants in L1 with zeocin.

Transconjugants were restreaked on a square plate and left to grow seven more days. Colonies were then taken using a Fisherbrand® Loop from selective plates then used to inoculate 150 µL of autoclaved deionized water. The cells were mixed by pipetting and vortexing for 30 sec. Samples were heated at 95°C for 12 min and vortexed a second time before being centrifuged at 13 700 rpm at 4°C for 1 min. The supernatant was then transferred to a new Eppendorf tube. A volume of 5 µL supernatant was used for PCR reactions and followed the Taq polymerase protocol for amplification and migrated by agarose electrophoresis.

2.4. Growth kinetic strains and growth conditions

Each *P. tricornutum* PCR-positive transconjugant was cultured in triplicate for 21 days in L1 medium with zeocin. Every two days 250 µL of cell culture was used to measure the OD at 730 nm using a black Greiner microplate at the Synergy H1 BioTek microplate reader (Agilent, Santa Clara, CA, USA). Curves were obtained using GraphPad Prism software. A t-test was performed using a Compare Groups of Growth Curves (CGGC) permutation test from Walter and Eliza Hall Institute of Medical Research with 10 000 permutations (Elso et al., 2004).

2.5. Plasmid rescue

To extract plasmid DNA from *P. tricornutum* transconjugants, 5 mL of *P. tricornutum* culture at day 7 cells were pelleted by centrifugation at 4000 g for 10 min. The cell pellets were resuspended in 235 µL of PDL1 Large Plasmid Extraction Kit with 5 µL lysozymes (100 mg. mL^{-1}), and 5 µL of hemicellulase (25 mg. mL^{-1}). The mixed solutions were incubated at 37°C for 30 min then 250 µL of PDL2 solution was added and mixed by inverting 5 to 10 times at room temperature. The lysis cells were neutralized by the addition of 375 µL of PDL3 buffer mixed by inversion 5 to 10

times and incubated for a 2 min at room temperature. The solution mixture was then centrifuged for 3 min on a microcentrifuge at maximal speed at room temperature. The manufacturing protocol of the Large Plasmid Extraction Kit (Geneaid™, New Taipei City, Taiwan) was followed starting from step 4 until elution in 50 µL of elution buffer preheated at 80°C. DNA concentrations were measured using N60/N50 Implen NanoPhotometer®. Purified DNA plasmid was further used for transformation in *E. coli* 10β chemicompetent cells following NEB protocol. Colonies were selected on LB agar with chloramphenicol and then put in liquid LB with chloramphenicol overnight. DNA was extracted using the Large Plasmid Extraction Kit (Geneaid™, New Taipei City, Taiwan) then measuring concentration at the N60/N50 Implen NanoPhotometer®. DNA sample was sent for NGS.

2.6. Western blot analysis

For each clone, a 50 mL culture was left to grow 21 days and was then pelleted at 4 000 g for 1 h at 4°C. The supernatant was removed, and the pellets were weighed. Cell pellets were then resuspended in cell lysis buffer (SDS 0.75 mM, glycerol 10%, Tris-HCL PH 7.2 51.4 mM, EDTA 0.02 mM, Urea 8M) with a ratio of 2 µL per mg of pellet. A volume of 50 µL of PMSF and 5 µL of protein inhibitor cocktail per sample were added before protein extraction. Proteins were then extracted by sonication (Amplitude 35%, ON time 15 sec, OFF 30 sec, run time 2 min). An aliquot of 500 µL of total protein extract was taken at this step. The rest of the protein solution was centrifuged at 15 000 g, at 4°C for 30 mins before transferring the supernatant into a new tube. This step was repeated twice. The soluble protein fraction was quantified using Bio-Rad DC protein assay Reagents Package. A 50 µg of total soluble protein, 5 µL of total protein fraction, and 5 µL of Invitrogen™ HiMark™ Pre-stained Protein Standard (Fisher Scientific, Ottawa, ON, Canada) was loaded on a SurePage Bris-Tris

gel 4-12% (GenScript, Piscataway, NJ, USA). Protein migration was performed at 80 V for 25 min, then at 120 V for 2 h 45. The 0.2 µm polyvinylidene fluoride (PVDF) membrane was activated in methanol and then equilibrated in transfer buffer (0.6% Tris, 2.88% Glycine, 0.01% SDS, 20% methanol). Wet transfer was performed on ice for 1 h 50 at 120 V. After transfer, the membrane was blocked in 5% milk in TBST (Tris Buffer Saline with 0.1% Tween 20) at room temperature for 2h. After three washes with TBST for 10 minutes each, the blocked membrane was incubated with anti-myc tag monoclonal antibody from EMD Millipore (dilution 1:500) in 5% milk overnight at 4°C. The membrane was again washed in TBST and then incubated for 1 h with a 1:20000 dilution, in 5 % milk of Immun- Star Goat Anti-Rabbit (GAR)-HRP from Bio-Rad (Bio-Rad, Hercules, CA, USA). Finally, the membrane was washed three more times with TBST and then dipped in Clarity Max Western ECL Substrate-Luminol solution from Bio-Rad before revelation. Total protein transferred onto the membrane were visualized using the Ponceau Red staining by incubating the membrane for one day at room temperature and then rinsed with distilled water. The images taken for protein detection, Luminol and Ponceau Red, were acquired using ChemiDoc Imaging System with Image Lab™ Software (Bio-Rad, Hercules, CA, USA).

2.7. P. tricornutum metabolite extraction

For each sample, after 21 days of growth, 10 mL of culture was centrifugated at 4 000 g for 50 mins. Approximately 90 mg of cell pellets were obtained and resuspended in 200 µL of mobile phase (*i.e.* MilliQ water and methanol, both containing 0.1% formic acid (30:70)) and stored at -80°C. Upon analysis, the sample was centrifugated at 4 000 g for 3 minutes. An aliquot was diluted 50-fold in the same mobile phase mentioned above.

2.8. SteelyA in vitro enzymatic assay

The enzymatic assay was performed in triplicate using 14 µg of total soluble protein extract in a 100 µL volume reaction in the presence of PBS 1X, DTT 5 mM, malonyl-CoA 0.75 mM, hexanoyl-CoA 0.24 mM, and repeated at least three times. The enzymatic reaction was incubated for 16h at 22°C and stopped using 10 µL of TCA 20%. A first extraction was performed adding 300 µL ethyl acetate to the reaction mix. The organic phase was taken and filtered using Agilent Captiva Econofilter polytetrafluoroethylene (PTFE) 13mm 0.2 µm. The aqueous phase of the enzymatic reaction was diluted in 300 µL of methanol and filtered too. All samples were dried in Thermo Scientific SPD1010 Speedvac concentrator and then resuspended in 100 µL methanol prior HPLC-MS analysis.

2.9. HPLC-MS analysis

Malonyl-CoA and hexanoyl-CoA authentic standards were purchased from Millipore Sigma Canada Ltd (Oakville, ON, Canada) while the olivetolic acid and olivetol standards were purchased from Santa Cruz biotechnologies (Dallas, TX, USA).

Metabolites detection was performed using high performance liquid chromatography (Agilent, 1260 Infinity II) coupled to mass spectrometry (Agilent, MSD iQ). Metabolites separation was done on an InfinityLab Poroshell 120 EC-C18 (2.1 x 50 mm, 2.7 µm) analytical column equipped with an Agilent EC-C18 (4.6 x 5 mm, 2.7 µm) guard column maintained at 30°C. An injection volume of 5 µL was set. The flow was set at 0.5 mL. min^{-1} with an elution program as follows: 0 min, 30 % A; 7.0 min, 0 % A; 12.0 to 20.1 min, 30 % A. The MS source parameters were set as follows: Gas temperature: 220 ºC, Gas flow: 8.0 mL. min^{-1}, Nebulizer (psi): 55. Samples analyses were acquired in Scan mode in ESI+ and ESI- between m/z 100 to 500 with a

fragmentation voltage of 135 V. Data were acquired using Agilent OpenLab CDC Acquisition (version 2) and Agilent OpenLab CDC Data Analysis (version 2.7) software were used for data acquisition and processing, respectively. A set of m/z for target compounds was analyed (Table 3.1).

Table 3.1. Targeted compounds and their corresponding m/z values in positive ion mode or LC-MS analysis.

Compound Name	**Verified m/z in M+H+**
Triketide acyl pyrone (=4-hydroxy-6-propyl-pyran-2-one)	156
Olivetol	181
Triketide pyrone	182
Methyl olivetol and MPBD (4-methyl-5-pentylbenzene-1,3-diol)	195
Tetraketide acyl pyrone (=4-hydroxy-6-(2-oxo-pentyl)-pyran-2-one)	196
Olivetolic acid	225
Tetraketide pyrone	225
Naringenin chalcone	273

3. Results

3.1 Engineering P. tricornutum strains to express SteelyA and screening of strains

To heterologously express the *D. discoideum* polyketide synthase SteelyA in *P. tricornutum*, we constructed the expression episomal vector pSteely (Figure 3.2) using p*Pt*GE30 (EV). The vector harbored the s*teelyA* coding sequence of 9441 bp under the control of the native constitutive *Hasp1* promoter, with the *FcpA* as a terminator,

and a C-terminal *myc* tag for protein detection (Figure 3.2). Plasmid construction was achieved via Gibson assembly using six primer sets (Appendix A - Table A1) and subsequently transformed into *E. coli*. Five SteelyA transformants and one EV were selected for *E. coli*-mediated conjugation with *P. tricornutum*.

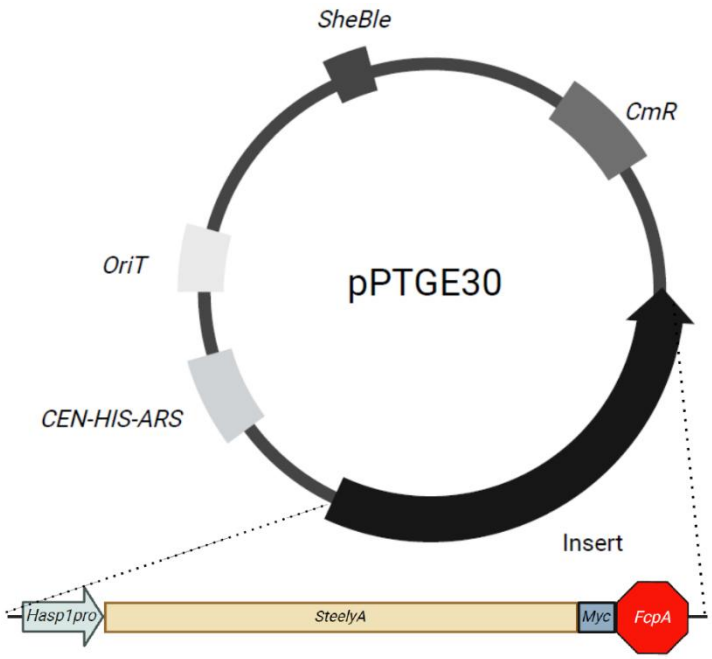


Figure 3.2. Schematic representation of the construction of the episomal plasmid pSteelyA. The p*Pt*GE30 plasmid, referred to as the Empty Vector (EV) is 16 015 bp length. The inserted cassette consists of the *SteelyA* coding sequence (brown) under the control of the native *highly abundant secreted protein 1* (*Hasp1*) promoter (light blue). The sequence is tagged with a *Myc* marker (dark blue) and includes a fucoxanthin-chlorophyll binding protein A (*FcpA*) terminator (red). The insertion cassette is 10 174 bp long, resulting in the pSteely plasmid with a total length of 26 189 bp. Abbreviations: *CEN-HIS-ARS*, centromere-histidine-autonomously replicating sequence region; *OriT*, origin of transfer; *ShBle* zeocin resistance gene; *CmR*, chloramphenicol resistance gene.

Following conjugation, transconjugants were selected on selected on L1 agar plates containing zeocin. After two weeks, 36 colonies emerged from which 15 were randomly selected for colony PCR verification of plasmid integration (Appendix A - Figure A1). Positive clones underwent plasmid rescue and were verified for DNA integrity by next-generation sequencing (Appendix A - Table A2). Sequencing analysis revealed varying degrees of mutations among the transconjugants. We retained three transconjugants for further study: pSteelyA-1, which contained a 129 bp insertion in EV backbone but no mutation in *steelyA*; pSteelyA-2, which exhibited two 129 bp and a 563 bp insertions, a 129 deletion in a repetitive backbone region, and a silent mutation in *steelyA*; and pSteelyA-3, which carried an alanine-to-aspartic acid substitution in the methyltransferase domain, a methionine-to-isoleucine substitution and a 984 bp deletion in the vector backbone (Appendix A - Table A2).

The three selected transconjugants, pSteelyA-1, pSteelyA-2, and pSteelyA-3, along with the EV control, were cultured for 21 days and their growth was monitored by measuring optical density at 730 nm. No significant differences in growth were observed among the transconjugants or between the transconjugants and the EV control (Figure 3A), indicating that the presence of the large plasmid (26 190 pb) did not adversely affect cell proliferation. However, microscopic analysis revealed a distinct morphological difference, with pSteelyA-expressing transconjugants exhibiting a shorter cell phenotype compared to the EV control with a significant reduction in cell line length however no changes in width (Figure 3.2B, 3.2C; and Appendix A - Figure A2).

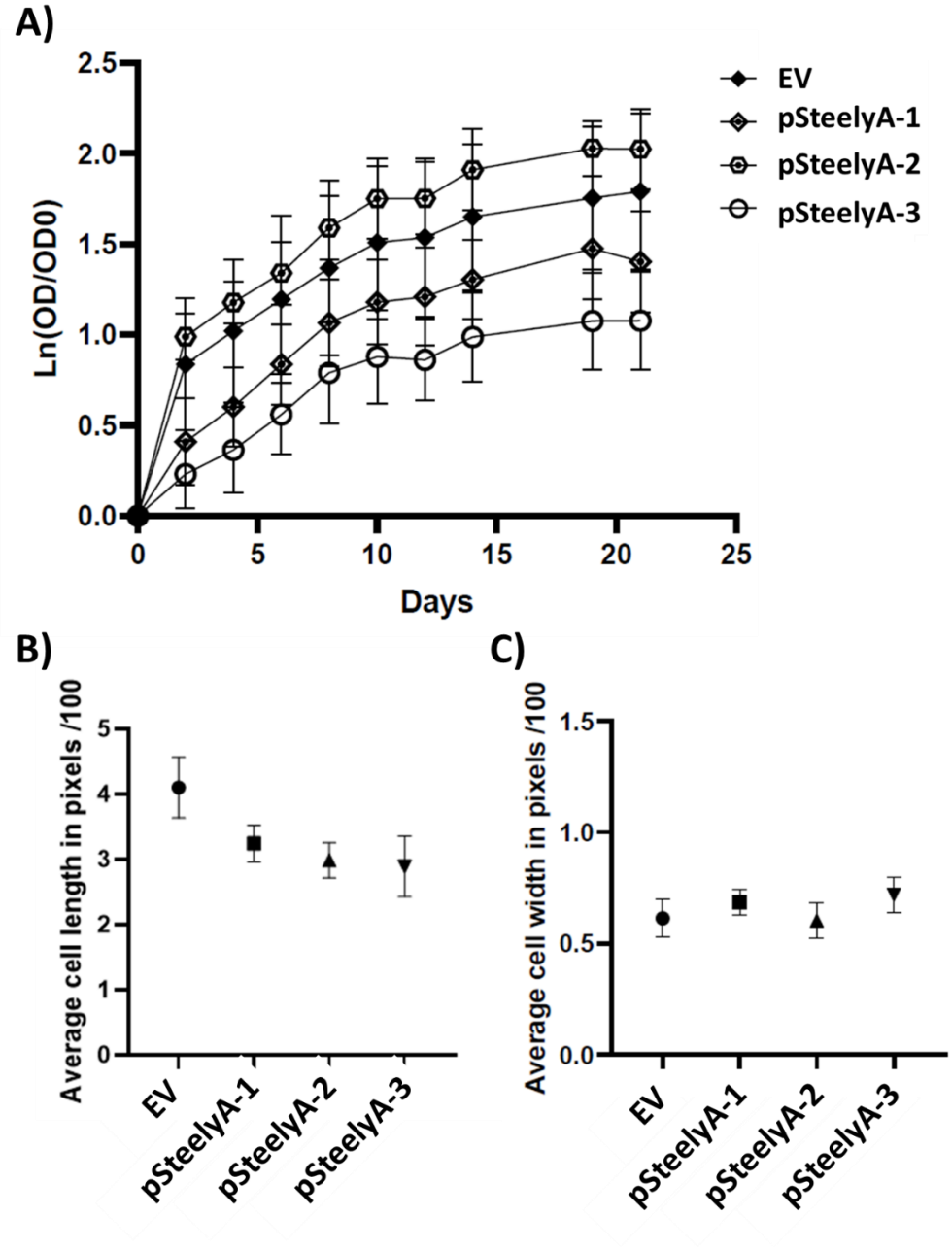


Figure 3.3. Growth and morphological analysis of *P. tricornutum* SteelyA and empty vector (EV) control transconjugant cell lines. A) Growth curves of SteelyA transconjugants and EV control cultured in L1 at 18°C. Optical density (OD) at 730 nm was monitored for 21 days. Data represent the mean of three biological replicates, with error bars indicating standard deviation. *t*-test, $p < 0.05$. B) represents the length of SteelyA transconjugant cell lines C) represents the width of SteelyA transconjugant

cell lines. Aligned dot plot of cells length and width in pixels. (n ≤ 5 cells per condition).

3.2 SteelyA production and activity

To assess SteelyA production, we performed a western blot analysis on total protein extracts and soluble fractions six months after conjugation. In the total fraction extract lanes of the pSteelyA-1,2 and 3 clones, there is a faint, but visible, diffuse band extending roughly from 350 to 400 kDa (Figure 3.4) although its broadness renders difficult to give a precise estimation of its size.A fainter band is observable in the same size vicinity in pSteelyA1 soluble fraction. A very faint signal can be detected between 268 kDa and 460 kDa in the two empty vector fractions. The strongest signals are obtained in the pSteelyA-2, pSteelyA-3 total fraction and pSteelyA-1, pSteelyA-2, and pSteelyA-3 soluble fraction at about 50 kDa. There is also a band at 50 kDa present in pSteelyA-1 total fraction but much fainter. A second band is detected in the soluble fraction of all pSteelyA samples at around 33 kDa and much fainter in the total fraction of all pSteely samples. A band around 48 kDa is observable in both empty vector fractions although more visible in the soluble fraction.

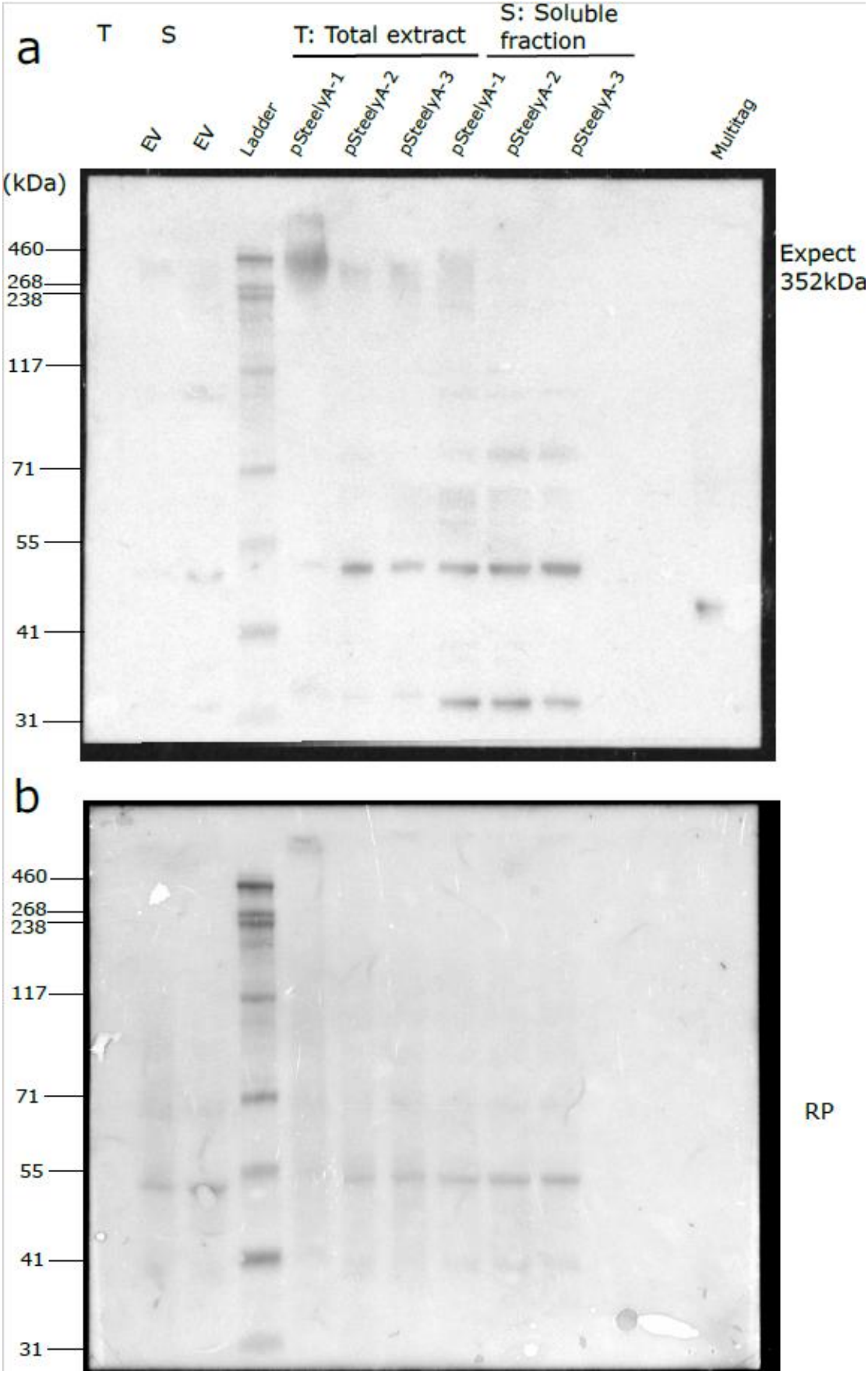

a
T
S
T: Total extract
S: Soluble fraction
EV
EV
Ladder
pSteelyA-1
pSteelyA-2
pSteelyA-3
pSteelyA-1
pSteelyA-2
pSteelyA-3
Multitag
(kDa)
460
268
238
117
71
55
41
31
Expect 352kDa
b
460
268
238
117
71
55
41
31
RP

Figure 3.4. Detection of SteelyA protein in total protein extracts. A) Western blot analysis using anti-myc on total protein extracts (T) and soluble protein fractions (S) from pSteelyA-1, pSteelyA-2, and pSteelyA-3 transconjugant cell lines. Proteins were separated on a GenScript SurePage Bris-Tris gel 4-12%. The empty vector (EV) transconjugants serve as a negative control. SteelyA (StlA) protein is expected to be approximately 352 kDa. B) Total protein staining of the blot using Ponceau Red (RP) to confirm protein transfer and loading consistency. To evaluate SteelyA enzymatic activity in vitro, total protein extracts were incubated with hexanoyl-CoA and malonyl-CoA under varying reaction conditions, including different pH levels (7, 7.5, and 8), temperatures (22°C and 25°C), and buffer systems (HEPES and Tris). LC-MS analysis of reaction products revealed no detectable levels of the expected compounds (m/z 100-500 and Table 1) in any transconjugants compared to the EV control. Similarly, no target metabolites from Table 1 were identified in total cell extracts under the same analytical conditions. These results suggest that despite successful expression and accumulation, SteelyA either lacked enzymatic activity or required additional factors for functional catalysis in *P. tricornutum*.

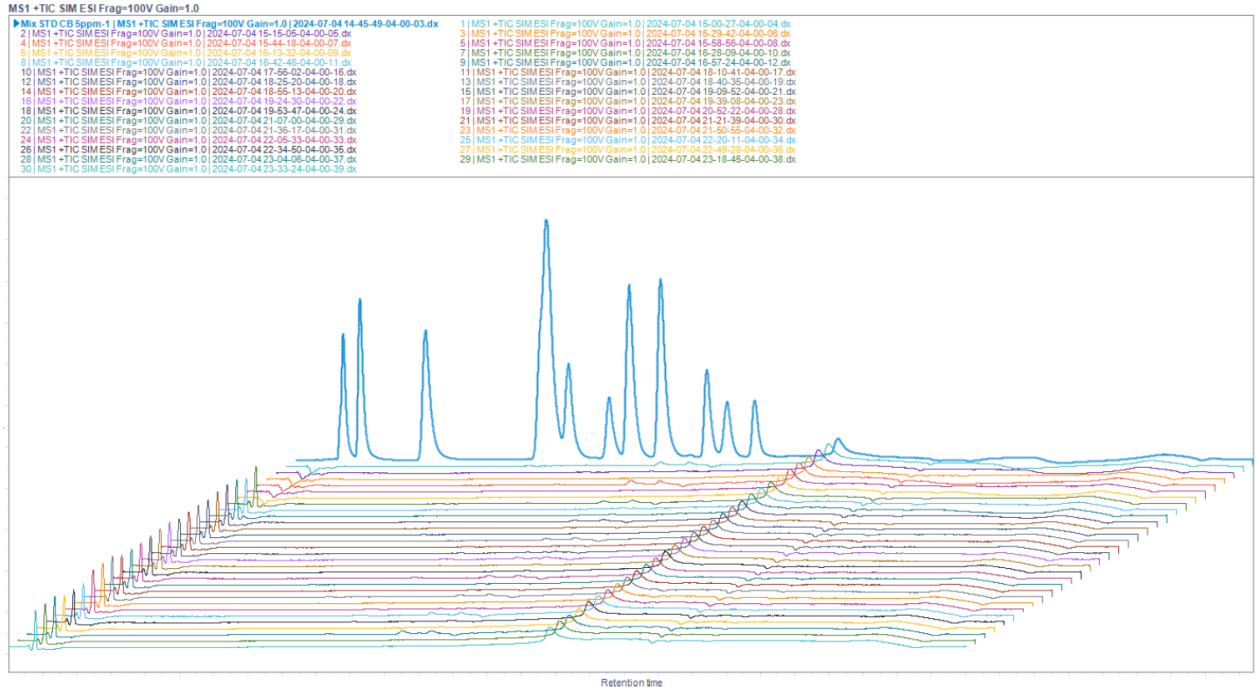


Figure 3.5. Total ion chromatogram (TIC) from LC-MS analysis in single ion monitoring (SIM) mode. The chromatogram includes a mix of standards showing expected peaks (blue) and triplicate samples from the empty vector (EV) control, and pSteelyA-1, pSteelyA-2, and pSteelyA-3 transconjugant lines (various colors). No transconjugant exhibited peaks corresponding to the expected SteelyA products. The monitored m/z values are listed in Table 1.

4. Discussion

Previous studies have demonstrated that large DNA insertions can be stably maintained in *P. tricornutum* via conjugation, such as the successful integration of a 49 kb construct (Karas et al., 2015a). Additionally, the production of heterologous proteins, including a 160 kDa antibody heavy chain, has been achieved in this diatom without adversely affecting growth (Hempel et al., 2011). In this study, we successfully introduced a 26 kb plasmid DNA into *P. tricornutum* and the obtained

transconjugant cell lines were stable over time under our experimental conditions. The presence of a band at 50 kDa in all pSteelyA clones tends to hint that the protein SteelyA is produced and degraded. A band is however slightly visible in the empty vector around 48 kDa, which could imply that the protein detected in Steely samples is actually an overproduced endogenous protein that is modified by 2 kDa peptide. A counter argument is the presence of a band of high molecular weight between 268 kDa and 460 kDa. Although they are too faint to give a precise molecular weight, they migrate in the expected range of the full SteelyA protein. The reduction of intensity of the 50 kDa and the 28 kDa band in the total extract of sample pSteelyA-1 hints toward the degradation of the protein by an endopeptidase.

Interestingly, the presence of SteelyA resulted in a distinct morphological change, with transconjugant cells displaying an average of 25% reduction in length compared to the EV control cells, while maintaining a consistent width (Figure 3 B, C and Appendix A – Figure A2). This suggests that the metabolic burden of producing such a large protein might have influenced cell morphology. However, despite the considerable size of SteelyA, more than twice the weight of previously expressed proteins, the overall growth of transconjugants remained unaffected (Figure 3A). This finding reinforces *P. tricornutum*‘s potential as a suitable host for the heterologous production of large proteins.

Despite successful expression (Figure 4), SteelyA could not be detected out of all doubt at the right size, and it is more than likely that if it is indeed produced, the protein is degraded. Besides, no expected metabolites (olivetolic acid, olivetol, MPBD, or methyl olivetol) were detected in vitro or in vivo, even with substrate supplementation of reactions (Figure 5). This suggests that Steely isor in too low abundance in its intact form after three weeks of growth to be considered active in *P.*

tricornutum. While heterologous SteelyA has been reported functional in *Saccharomyces cerevisiae* (Mookerjee et al., 2018), it is possible that *P. tricornutum* has a metabolic defense that cleaves proteins of an unexpected size. . Additionally, although the PKS domain of SteelyA has shown activity in *Escherichia coli* (Austin et al., 2006), the full- length gene could not be expressed in this host due to its size and the presence of introns.

Additionally, directing the protein to specific subcellular compartments could enhance its stability and functionality. Fluorescence microscopy could help determine SteelyA's localization and reveal whether it is trafficked to the endoplasmic reticulum (ER) and to the Golgi apparatus, where potential glycosylation could interfere with its targeting by proteases (Solá and Griebenow, 2009, Stanley, 2011, Dumontier et al., 2021, Mathieu-Rivet et al., 2014). However, SteelyA's glycosylation pattern remains uncharacterized, making this an avenue for future research.

In conclusion, this study demonstrates the stable expression of the 352-kDa recombinant SteelyA protein in *P. tricornutum*, highlighting the diatom's potential as a host for large heterologous proteins (Sharma et al., 2021). However, the observed protein insolubility and lack of activity indicate that further optimization is necessary to achieve functional SteelyA activity in this expression system. Future efforts should focus on optimizing expression conditions, improving protein folding, and targeting specific cellular compartments to enhance SteelyA's solubility and activity to enable functional cannabinoid biosynthesis in this host**.**

Funding: This research was funded by the Natural Sciences and Engineering Research Council of Canada through the Alliance Program (award nos. ALLRP

554429-20) to I.D.-P and by the Université du Québec à Trois-Rivières – Annexe C to IDP.

Institutional Review Board Statement: Not applicable**.**

Informed Consent Statement: Not applicable**.**

Data Availability Statement: Data generated or analyzed during this study are provided in full within the published article and from the corresponding author upon reasonable request.

Acknowledgments: The authors want to acknowledge the help and support of Melodie B. Plourde during this study as well as feedback from all lab members. We also extend our appreciation to the Université du Québec à Trois-Rivières for its invaluable support in facilitating this research. We sincerely appreciate the provision of laboratory space, essential infrastructure, and utilities, which were critical for the successful completion of this work. Additionally, during the preparation of this manuscript, the authors used ChatGPT -4.0, a free AI language model, to refine grammatical accuracy and enhance readability. All content was subsequently reviewed and edited by the authors, who take full responsibility for the final version of this publication.

Conflicts of Interest: The authors declare no conflicts of interest**.**

References:

ANJARD, C., SU, Y. & LOOMIS, W. F. 2011. The polyketide MPBD initiates the SDF-1 signaling cascade that coordinates terminal differentiation in Dictyostelium. *Eukaryotic cell,* 10**,** 956-963.

ATHANASAKOGLOU, A., GRYPIOTI, E., MICHAILIDOU, S., IGNEA, C., MAKRIS, A. M., KALANTIDIS, K., MASSÉ, G., ARGIRIOU, A., VERRET, F. & KAMPRANIS, S. C. 2019. Isoprenoid biosynthesis in the diatom Haslea ostrearia. *New Phytologist,* 222**,** 230-243.

AUBRY, L. & FIRTEL, R. 1999. Integration of signaling networks that regulate Dictyostelium differentiation. *Annual review of cell and developmental biology,* 15**,** 469-517.

AUSTIN, M. B., SAITO, T., BOWMAN, M. E., HAYDOCK, S., KATO, A., MOORE, B. S., KAY, R. R. & NOEL, J. P. 2006. Biosynthesis of Dictyostelium discoideum differentiation-inducing factor by a hybrid type I fatty acid–type III polyketide synthase. *Nature chemical biology,* 2**,** 494-502.

AWWAD, F., FANTINO, E. I., HÉNEAULT, M., DIAZ-GARZA, A. M., MERINDOL, N., CUSTEAU, A., GÉLINAS, S.-E., MEDDEB-MOUELHI, F., LI, J. & LEMAY, J.-F. 2023. Bioengineering of the Marine Diatom *Phaeodactylum tricornutum* with *Cannabis* Genes Enables the Production of the Cannabinoid Precursor, Olivetolic Acid. *International Journal of Molecular Sciences,* 24**,** 16624.

BACKER, R., SCHWINGHAMER, T., ROSENBAUM, P., MCCARTY, V., EICHHORN BILODEAU, S., LYU, D., AHMED, M. B., ROBINSON, G., LEFSRUD, M. & WILKINS, O. 2019. Closing the yield gap for cannabis: a meta-analysis of factors determining cannabis yield. *Frontiers in plant science,* 10**,** 495.

BIOCYC. 2024. *hexanoyl-CoA* [Online]. Available: https://biocyc.org/compound?orgid=PHAEO&id=HEXANOYL-COA#RXNS [Accessed 23-10 2024].

BRANCO-VIEIRA, M., SAN MARTIN, S., AGURTO, C., FREITAS, M. A., MARTINS, A. A., MATA, T. M. & CAETANO, N. S. 2020. Biotechnological potential of *Phaeodactylum tricornutum* for biorefinery processes. *Fuel,* 268**,** 117357.

BUTLER, T. O., PADMAPERUMA, G., LIZZUL, A. M., MCDONALD, J. & VAIDYANATHAN, S. 2022. Towards a *Phaeodactylum tricornutum* biorefinery in an outdoor UK environment. *Bioresource Technology,* 344**,** 126320.

CLOMBURG, J. M., CRUMBLEY, A. M. & GONZALEZ, R. 2017. Industrial biomanufacturing: the future of chemical production. *Science,* 355**,** aag0804.

CROCQ, M.-A. 2020. History of cannabis and the endocannabinoid system. *Dialogues in clinical neuroscience,* 22**,** 223-228.

CUI, Y., THOMAS-HALL, S. R. & SCHENK, P. M. 2019. *Phaeodactylum tricornutum* microalgae as a rich source of omega-3 oil: Progress in lipid induction techniques towards industry adoption. *Food chemistry,* 297**,** 124937.

CUI, Y., THOMAS-HALL, S. R., CHUA, E. T. & SCHENK, P. M. 2021. Development of high-level omega-3 eicosapentaenoic acid (EPA) production from *Phaeodactylum tricornutum*. *Journal of Phycology,* 57**,** 258-268.

DIAMOND, A., DIAZ-GARZA, A. M., LI, J., SLATTERY, S. S., MERINDOL, N., FANTINO, E., MEDDEB-MOUELHI, F., KARAS, B. J., BARNABÉ, S. & DESGAGNÉ-PENIX, I. 2023. Instability of extrachromosomal DNA transformed into the diatom *Phaeodactylum tricornutum*. *Algal Research,* 70**,** 102998.

FANTINO, E., AWWAD, F., MERINDOL, N., GARZA, A. M. D., GÉLINAS, S.-E., ROBLES, G. C. G., CUSTEAU, A., MEDDEB-MOUELHI, F. & DESGAGNÉ-

PENIX, I. 2024. Bioengineering *Phaeodactylum tricornutum*, a marine diatom, for cannabinoid biosynthesis. *Algal Research,* 77**,** 103379.
FORTUNATO, A., STRASSMANN, J., SANTORELLI, L. & QUELLER, D. 2003. Co-occurrence in nature of different clones of the social amoeba, Dictyostelium discoideum. *Molecular ecology,* 12**,** 1031-1038.
GHOSH, R., CHHABRA, A., PHATALE, P. A., SAMRAT, S. K., SHARMA, J., GOSAIN, A., MOHANTY, D., SARAN, S. & GOKHALE, R. S. 2008. Dissecting the functional role of polyketide synthases in Dictyostelium discoideum: biosynthesis of the differentiation regulating factor 4-methyl-5-pentylbenzene-1, 3-diol. *Journal of Biological Chemistry,* 283**,** 11348-11354.
HARA, K. Y., ARAKI, M., OKAI, N., WAKAI, S., HASUNUMA, T. & KONDO, A. 2014. Development of bio-based fine chemical production through synthetic bioengineering. *Microbial cell factories,* 13**,** 1-19.
HEMPEL, F., LAU, J., KLINGL, A. & MAIER, U. G. 2011. Algae as protein factories: expression of a human antibody and the respective antigen in the diatom *Phaeodactylum tricornutum*. *PloS one,* 6**,** e28424.
KARAS, B., DINER, R., LEFEBVRE, S., MCQUAID, J., PHILLIPS, A., NODDINGS, C., BRUNSON, J., VALAS, R., DEERINCK, T. & JABLANOVIC, J. 2015. Designer diatom episomes delivered by bacterial conjugation. Nat Commun 6: 6925.
LEWIS, M. M., YANG, Y., WASILEWSKI, E., CLARKE, H. A. & KOTRA, L. P. 2017. Chemical profiling of medical cannabis extracts. *ACS omega,* 2**,** 6091-6103.
LOZANO TEROL, G., GALLEGO-JARA, J., SOLA MARTÍNEZ, R. A., CÁNOVAS DÍAZ, M. & DE DIEGO PUENTE, T. 2019. Engineering protein production by rationally choosing a carbon and nitrogen source using *E. coli* BL21 acetate metabolism knockout strains. *Microbial cell factories,* 18**,** 1-19.
LUO, X., REITER, M. A., D'ESPAUX, L., WONG, J., DENBY, C. M., LECHNER, A., ZHANG, Y., GRZYBOWSKI, A. T., HARTH, S. & LIN, W. 2019. Complete biosynthesis of cannabinoids and their unnatural analogues in yeast. *Nature,* 567**,** 123-126.
MAXWELL, K. L., MITTERMAIER, A. K., FORMAN-KAY, J. D. & DAVIDSON, A. R. 1999. A simple in vivo assay for increased protein solubility. *Protein Science,* 8**,** 1908-1911.
MITAL, S., CHRISTIE, G. & DIKICIOGLU, D. 2021. Recombinant expression of insoluble enzymes in Escherichia coli: a systematic review of experimental design and its manufacturing implications. *Microbial cell factories,* 20**,** 1-20.
MOOKERJEE, S., CAMPBELL, A. J., WILTSHIRE, Z. D. & CHEN, K. J. 2018. Method and cell line for production of phytocannabinoids and phytocannabinoid analogues in yeast. Google Patents.
MOOKERJEE, S., CAMPBELL, A. J., WILTSHIRE, Z. D. & CHEN, K. J. 2022. Method and cell line for production of phytocannabinoids and phytocannabinoid analogues in yeast. Google Patents.
NARITA, T. B., KOIDE, K., MORITA, N. & SAITO, T. 2011. Dictyostelium hybrid polyketide synthase, SteelyA, produces 4-methyl-5-pentylbenzene-1, 3-diol and induces spore maturation. *FEMS microbiology letters,* 319**,** 82-87.
OVERKAMP, K. M., BAKKER, B TTER, P., LUTTIK, M. A., VAN DIJKEN, J. P. & PRONK, J. T. 2002. Metabolic engineering of glycerol production in Saccharomyces cerevisiae. *Applied and Environmental Microbiology,* 68**,** 2814-2821.
PARAPOULI, M., VASILEIADIS, A., AFENDRA, A.-S. & HATZILOUKAS, E. 2020. Saccharomyces cerevisiae and its industrial applications. *AIMS microbiology,* 6**,** 1.
REIMER, C., KUFS, J. E., RAUTSCHEK, J., REGESTEIN, L., VALIANTE, V. & HILLMANN, F. 2022. Engineering the amoeba Dictyostelium discoideum for biosynthesis of a cannabinoid precursor and other polyketides. *Nature biotechnology,* 40**,** 751-758.

ŚLEDZIŃSKI, P., NOWAK-TERPIŁOWSKA, A. & ZEYLAND, J. 2020. Cannabinoids in medicine: cancer, immunity, and microbial diseases. *International journal of molecular sciences,* 22**,** 263.

SOLÁ, R. J. & GRIEBENOW, K. 2009. Effects of glycosylation on the stability of protein pharmaceuticals. *Journal of pharmaceutical sciences,* 98**,** 1223-1245.

STANLEY, P. 2011. Golgi glycosylation. *Cold Spring Harbor perspectives in biology,* 3**,** a005199.

WILDING, K. M., SCHINN, S.-M., LONG, E. A. & BUNDY, B. C. 2018. The emerging impact of cell-free chemical biosynthesis. *Current opinion in biotechnology,* 53**,** 115-121.

CHAPTER IV: DISCUSSION AND CONCLUSIONS

Since the discovery of the potential of cannabinoids in treating neurodegenerative diseases, an intense effort has been made in research to produce THC, CBD, and other less abundant cannabinoids. The ability to produce pure and high amounts of molecules capable of interacting with the endocannabinoid system is expected to be a great leap towards studying and developing remedies to ease, and perhaps heal conditions linked to it. The main issue in extraction is the amount of minor cannabinoids present in low abundance as well as the presence of other molecules such as terpenes and flavonoids (Walsh et al., 2021b). It is observed that THC and CBD have stronger effects when mixed in an extract but in order to produce medicine, the effect of each molecule must be identified, tested and controlled (Williamson and Evans, 2000). Cannabinoids have also been observed to have adverse effects on human biology (Wang et al., 2008), therefore a careful control of remedy content is adamant. Purification of the major cannabinoids has therefore been a tedious process that now gives access to remedies such as Sativex, a product that has already been used to mitigate symptoms of ALS and other ailments (Barnes, 2006, Russo et al., 2007). However, numerous other molecules have been identified as cannabinoids based on their alkyl resorcinol derived structure, or similarity with endocannabinoids structure, but only a few have been isolated and tested to bind cannabinoid receptors CB1 and CB2 (Hazekamp et al., 2004, Pertwee et al., 2010). If systems have been successful at producing low quantities of olivetolic acid, THC and CBD, a sustainable and high yield platform is still being searched. Although in *Pichia pastoris* and in cell-free systems, the production of THCA and CBDA respectively reached above 1g. L^{-1}, both systems present limitation regarding their functioning or ecological costs (Blatt-Janmaat and Qu, 2021a). The cost of enzymes renders cell-free system hardly economically sustainable as for *P. pastoris* the

functioning requires organic matter and culture maintenance at 28°C thus demanding higher energy cost than *P. tricornutum,* which grows at 18°C, can grow in sea water and wastewaters, and can be grown in open ponds.

The insertion of the four enzymes required for cannabinoids production (TKS, OAC, CBGAS, CBDAs or THCAs) in *P. tricornutum* onto one episome containing four coding sequences (CDS) was attempted in our laboratory but did not yield any detectable level of THCA or CBDA (unpublished data). The investigation of each enzyme was therefore necessary to assess the localization of the bottlenecks. Further studies confirmed the production of CBGA from enzymatic assay with crude extracts but not in vivo yet (Fantino et al., 2024a).

The work made in this study to turn *P. tricornutum* into a biosynthetic platform to produce olivetolic acid was done using two strategies. The first one was based on the insertion of *Cs*TKS and *Cs*OAC in two separate clones and aimed at studying each enzyme separately and in a combined action. Based on the work of (Awwad et al., 2023a), their achievement in producing olivetolic in vivo led to think that *P. tricornutum* is indeed a suitable and promising host for the heterologous production of cannabinoids. However, the loss of detectable levels of OA through time can becausedby either progressive silencing, but this hypothesis is troubled by the lack of repeatability of the detection of OA after retransforming wild diatoms (unpublished data), by the co-expression of the two precursor producing enzymes hindering the production of the proteins, or despite the detectable levels obtained by tests carried out by different institutions, by a possible sample contamination. It has been observed that the triradiate morphotype was the only one producing cannabinoids, implying that its metabolism might be more fitting for their production, however, to date the reversion from fusiform to triradiate has never been achieved in our laboratory. The co-expression of the two

enzymes in one open reading frame (ORF) could lead to dysfunctional proteins if the cleaving sequence is not properly separating the two enzymes. However, one of (Awwad et al., 2023a)'s construction seemed to show a proper cutting of the OAC out of the TKS-T2A peptide, thus so if TKS and OAC were active originally, unless the truncation of the peptide decreased through time, the fusion should not be responsible for the disappearance of OA. In order to reproduce the conditions tested in (Awwad et al., 2023a), we first tested in chapter II the individual enzymes production by western blot. Each clone produced the expected enzyme stably for about two years; thus the two protein extracts were pooled and tested by enzymatic assay following a protocol provided and tested for cannabinoids production by Hyasynth Bio Inc. The protocol had been previously tested in our laboratory for (Awwad et al., 2023a)'s study and is the method that led to the detection of OA, but this protocol has also been validated in *Chlamydomonas reinhardtii* strains. Extracts of a *C. reinhardtii* strains containing a genome integrated construct with *Cs*TKS and *Cs*OAC in a cistron also showed production of olivetolic acid in crude extract from enzymatic assay (Majhi et al., 2025 submitted). In the chapter II, the protein extraction, quantification and enzymatic assay were done parallelly with the *P. tricornutum Cs*TKS and *Cs*OAC strains and the *C. reinhardtii* strain and in identical culture conditions (with respective medium and temperature). Eventually olivetol was only detectable in *C. reinhardtii.* Several factors could explain this difference: like in (Awwad et al., 2023a) the *Cs*TKS and *Cs*OAC in the green algae were coexpressed while in *P. tricornutum* there are in separate episomal vectors, the assessed protein content of *C. reinhardtii* was about ten times lower than in *P. tricornutum* for a biomass that was paradoxically five times higher, their metabolism is different as one is a green algae and the other a diatom, thus the consumption of hexanoyl-CoA and malonyl-CoA by metabolic pathways could be

highly different. It is possible that the integrated status of the construct in *C. reinhardtii* also reduces silencing when located in a highly expressed genomic region. The lack of information on pathways of *P. tricornutum* make it difficult to compare potential substrate and product usage with *C. reinhardtii* (Kanehisa et al., 2023). The intracellular pH was similar between the two cells (around 7.5 for both) but no data are currently available on in vivo intracellular pH of *C. sativa* disc cells, however the detection of OA in *C. reinhardtii* in those conditions indicate that this pH is favorable to OA's stability until detection. The physical proximity between the two enzymes is not required as experimentations in dialysis chambers with the two enzymes separated and metabolites diffusing led to the production of OA, therefore the cistron increasing the protein production's proximity does not seem to be a strategic advantage for the production of OA (Gagne et al., 2012).

In order to test if one of the enzymes could be active in *P. tricornutum*, the *Cs*TKS and *Cs*OAC strains had to be examined one at a time. The major drawback of the first study is the inaccessibility of the *Cs*TKS product and *Cs*OAC substrate; 3,5,7-trioxododecanoyl-CoA. Consequently, to test the activity of *Cs*TKS, three transconjugant lines and a negative control were sent in triplicates for untargeted metabolomic. As this experiment is very costly, wecould not send the *Cs*OAC samples and test the supplementation, thus only *Cs*TKS samples were sent. The samples did not yield detectable 3,5,7-trioxododecanoyl-CoA nor identifiable polyketides. The study however presents the limit that with over 13 000 detected metabolites, only 57 of the deregulated were annotated (figure 2.4). As mentioned in chapter II, the method used was probably not optimal as another study using Fourier-transform ion cyclotron-resonance mass spectrometer (FTICR-MS) yielded several hundred annotated analytes detected (Duarte et al., 2022). Allumiqs (Sherbrooke, Québec, Canada), the company

that conducted untargeted metabolomics was specialized untargeted metabolomic using an UPLC-qTOF-MS which has a greater resolution of separation of compounds of identical mass and different structures but a significatively smaller database than the one accessible when using FTICR-MS.

The metabolomic analysis showed that there were several notable differences between the *Cs*TKS cell lines and the empty vector one, both in abundance and nature of metabolites. This observation hints strongly in the activity of the *Cs*TKS enzyme into *P. tricornutum* however the absence of detection of 3,5,7-trioxododecanoyl-CoA is an obstacle. Its availability on the market would permit its use as a standard. The study still points out a rise in cyclic compounds abundance which, for now, does not correlate with any study available on *P. tricornutum*'s metabolism. Several compounds such as quinolizine or the coumarin derivative methyl 2-[(4-methyl-2-oxo-2H-chromen-7-yl)oxy]propanoate have no biosynthetic route identified in vivo yet. Besides, little knowledge has been gathered so far on *P. tricornutum*'s metabolism rendering difficult to infer on possible routes for their synthesis (Kanehisa et al., 2023), and its distant phylogenic localization from other model organisms make an inference based on comparing with existing production routes difficult.

As explained in the discussion of chapter II, the growth conditions were monitored in a high-performance incubator assessing and insuring light and temperature homogeneity. The subdivision of 200 ml of culture for each triplicate into four 50 mL flasks then pulled back together before metabolite extraction also minimized any heterogeneity that could be created by micro-conditions. Every experiment was carried out in axenic conditions and the zeocin selection reduced risks of contamination by other organisms. The presence of pyocyanin, a phenazine molecule known to be a *Pseudomonas aerugonisa*'s virulence factor (Hall et al., 2016) in every *Cs*TKS

transconjugants is thus most-likely due to its in vivo production by the algae rather than contamination. Pyocyanin has already been detected in *P. tricornutum* before, but no conclusion on its potential role has been drawn (Goedheer and Swart, 1975). Its multi cyclic structure characteristic of pigments gives it UV protective properties which can be inferred to be its main use in the diatom.

Aromatic compounds present a phenol ring capable of absorbing light (Dearden and Forbes, 1959, Bartolomei et al., 2021). Among the annotated deregulated metabolites in *Cs*TKS clones, quinolizine, alanylphenylalanine, the coumarin derivative, pyocyanin, bergenin and phenylalanine are phenolic compounds upregulated in at least one *P. tricornutum* clone. As mentioned above, light conditions were identical between clones, it is therefore unlikely that the production of phenolic compounds was as response of the transformants to a light stress otherwise the negative the control would have followed the same profile change. One hypothesis is that the TKS contributed to the biosynthesis of precursors involved in *P. tricornutum*'s pathways related to phenolic compounds, or that the polyketides produced by it generate structures that could be precursors of detected cyclic compounds. Indeed, olivetolic acid and olivetol are phenolic compound and in the absence of cyclase, 3,5,7-trioxododecanoyl-CoA spontaneously cycle into olivetol (Gagne et al., 2012). This observation is consistent with the discovery of orsellinic acid biosynthesis in lichen (another phenolic compound) coming from polyketide pathway, although phenolic compounds such as phenylalanine and coumarin are highly expected to be derived from the shikimate pathway (Birch, 1967, Talapatra et al., 2015). All those differences in metabolite deregulations and productions still show a consequent distinction between transformants and negative control, thus strongly hinting in the activity of the *Cs*TKS in *P. tricornutum.*

To date, as no other enzyme has been discovered producing 3,5,7-trioxododecanoyl-CoA it is for now impossible to replace TKS by another enzyme to feed OAC thus there is no alternative route to produce olivetolic acid. Or is there?

In chapters I and III, we discuss the slime mold *Dictyostelium discoideum* and its ability to differentiate into specialized cells. During this process, the production of the enzyme SteelyA generates different compounds among which is found olivetol and more abundantly methyl-olivetol (Ghosh et al., 2008, Reimer et al., 2022). The heterologous production of methyl-olivetol and olivetol has also been achieved in *Saccharomyces cerevisiae* (Mookerjee et al., 2018) using SteelyA proving that this enzyme can be active outside of its natural host and maintain its activity of production of olivetol-derived compounds. This achievement by Hyasynth Bio Inc motivated the company Algae-C to collaborate with the UQTR to express SteelyA in *P. tricornutum*. The metabolic proximity of the diatom with *C. sativa* was expected to cut costs in culture and genetic modifications as the work of (Mookerjee et al., 2018) required extensive modifications of the budding yeast's genome. A construct containing *steelyA* CDS under the control of the constitutive endogen promoter HASP1 (Erdene-Ochir et al., 2019b) was inserted in *P. tricornutum*. This promoter has the particularity of being highly expressed even in stationary phase, and as SteelyA is a massive 352 kDa protein it was anticipated that accumulation might be slower than regular endogenous proteins.The protein production was then tested by a first western blot and displayed a band above 250 kDa (Annex D, figure D.1). There had been technical inconveniences when the protein production had to be verified by western blot, hindering us in using an adequate ladder at that time. The western blot clearly displayed a band in clones far above the 250 kDa ladder band in transconjugants and absent in empty vector (figure 3.4) strongly hinting that the protein detected is SteelyA, however, the lack of upper

ladder band entails that the protein could be too long to actually be SteelyA and represents a consequent limit of the study carried out in chapter III. The band was only present in total protein fraction and its absence in the soluble fraction implies that the protein is insoluble.

To validate these findings, the western blot was repeated (figure D.2 and figure 3.4). The second western blot (D.2) displayed a very different profile, showing bands exclusively in the soluble fractions. One main band is visible in all soluble samples including the empty vector, indicating that it is not SteelyA or it is a cross contamination from nearby well. Two bands are observable exclusively in pSteelyA samples, one around 55 kDa in all, and one at 80 kDa in pSteelyA-2 and -3. This could imply that SteelyA is produced and truncated at specific sites leading to two major cuts (Messaabi et al., 2024b). A thrombin could be responsible for the cuts however as its theoretical recognition sequence has been increased to P2-Pro, P1-Arg, P1′-Ser/Ala/Gly/Thr, P2′-not acidic and P3′-Arg we are currently incapable of ruling out this possibility (Gallwitz et al., 2012)

Figure D.2 displays again two sets of bands around 33 kDa and 50 kDa in all pSteelyA samples. The band at 33 kDa is found between the two gels but its size seems to be slightly smaller in the soluble fraction of the empty vector. Regarding the discrepancy of size of the 55 kDa and the 50 kDa between the Figure D.2 and Figure 3.4 are hard to explain biologically as it is unlikely that truncation is occurring in different locations in the proteins of identical biological samples. One possibility is the change in the migration due to technical parameters as the gel Figure 2.D has been migrated 2h at 120 V while Figure 3.4 was migrated 1h40 at 110 V. A band is very faintly appearing also around 80 kDa in the soluble fraction of pSteelyA samples but its intensity is hard low and similar to what seems to be background noise (smear between 70 kDa and 52 kDa).

One major difference between the Figure D.2 and Figure 3.4 is the appearance of a signal at high molecular weight in the total fraction of pSteelyA samples and in the soluble fraction of pSteelyA-1. Its wide distribution and its weak signal make the estimation of its exact size difficult, but it can be estimated between 300 and 460 kDa. Interestingly, the sample showing the most intense band at high molecular weight shows the weakest one at 33 kDa and 50 kDa hinting towards the actual truncation of the protein. The biggest obstacle in the theory of the protein production is the presence of a very faint signal close to every actual signal detected in pSteelyA samples in the empty vectors. Despite the size difference of each band, the proximity is unfortunate.The Ponceau reveals that there are bands present around 48 kDa in empty vector samples as well as at 50 kDa in pSteelyA samples more abundantly expressed than the rest of the proteins. Once again, the 2 kDa difference is hard to explain and could be protein modifications of the same protein rather than actual detection of SteelyA. However, myc tag is not expected to be present in *P. tricornutum* and the change in protein mass could be explained by the activity or production of SteelyA changing the protein modification profile of the cell, either because of the production of a massive protein, or by SteelyA itself. To test this hypothesis, producing another massive protein with a totally different amino acid sequence and tag it with the myc tag would confirm if the change in protein size is entailed by SteelyA or the production of any large protein.

Despite attempts to assess enzymatic activity in crude extracts, no new metabolites were detected by LC-MS, whether or not precursor compounds were supplemented. This lack of activity likely stems from the absence of soluble full-length SteelyA, as the high MW band appears mainly in the total fraction and only weakly in one soluble sample. One remaining line of evidence supporting SteelyA production is the consistency across Western blots: each consistently shows signal with anti-Myc antibodies. Previous work

using anti-Myc in *P. tricornutum* has shown no signal near 50 kDa in EV samples (Awwad et al., 2023a, data not shown), reinforcing the notion that these recurring bands likely result from SteelyA expression. However, purification of the tagged protein has never been successfully achieved, preventing further characterization or activity assays, and leaving open the possibility that crude extract components may inhibit SteelyA activity.

4.1 Perspectives

Several objectives remain to characterize *P. tricornutum* to produce cannabinoids. On a technical aspect, when *P. tricornutum* and *C. reinhardtii* strains were cultured, their protein extracted and tested for enzymatic assay simultaneously, a fine tuning of the conditions and parameters could have been done in order to reach the same biomass or protein concentration as for *Chlamydomonas*, however, if maintained in the same volume of culture, the time of growth required would have to push the diatom in its late stationary phase. An optimization of the conditions could be done by testing different method of protein extraction on *P. tricornutum* in order to increase the protein concentration as well as changing the pH thus altering the stability of potentially fragile compounds or inactivating enzymes. The method used in this study includes a preliminary breaking of cells by putting them in liquid nitrogen as well as using sonication that could break big molecules or denature enzymes.

To confirm the activity of *Cs*TKS, one possibility is to explore other enzymes that can take 3,5,7-trioxododecanoyl-CoA as substrate and produce an identifiable and obtainable compound. This enzyme would be co-transformed in a *Cs*TKS cell lines and assed for appearance of this identifiable compounds in the transformants and absence in empty vector strain. This method would be an indirect way to prove the activity of

*Cs*TKS. Otherwise, the isolation of 3,5,7-trioxododecanoyl-CoA is currently indispensable to confirm it in a reliable way.

During the untargeted metabolomic analysis, the change in metabolite content between *CsTKS* transconjugant and the empty vector control tend to show that the heterologous enzyme is active in *P. tricornutum*. An alternative way to detect products of TKS could be to radiolabel malonyl-CoA and compare all the radioactive compounds in clones and in the empty vector. By marking a carbon, mass spectrometry would then be sensitive enough to detect the mass difference of radioactive compounds and their inert version, although to have the highest accuracy, standards of suspected polyketides would be required.

The *Cs*TKS, CsOAC as well as SteelyA are expected to dimerize to be active (Austin et al., 2006, Gagne et al., 2012, Katsuyama and Ohnishi, 2012, Kearsey et al., 2020) therefore once SteelyA has been solubilized, a native electrophoresis acrylamide gel could be done to verify their proper "in vivo" quaternary structure.

We have explored that one possible explanation on SteelyA's lack of activity could be the lack of solubility. One way to increase the solubility of SteelyA would be to use molecular chaperones to ensure correct folding and protect the protein from degradation, or to change the amino acid content to increase the protein's cytosolic solubility. However, as fatty acids are naturally produced in plastids in *P. tricornutum*, addressing the protein to this more hydrophobic compartment could also increase its solubility (Zulu et al., 2018). Finally, an inducible promoter could help tune expression in different phases of growth of the diatom and adapt the expression to the different metabolic activity. In order to confirm the presence of SteelyA and minimize background reading, a dot blot is a simple alternative to western blot losing mainly the

information of the size of the detected signal. It would give a binary information on presence or absence and could confirm SteelyA production if signal was exclusively detected in samples and not negative control. Changing the amino acid sequence close to the truncated sites could also be a possibility to reduce degradation of the protein.

Further improvements of *P. tricornutum*'s system biology are required to decipher which pathways to modulate in order to reroute resources and allocate them in an optimal manner to increase the availability of precursors, and the stability of the products. The active research on biofuel production deepens the knowledge on the fatty acid biosynthesis and putting into light relevant gene expression circuits. Large scale metabolome analysis should be conducted in the different growth phases, growth conditions and morphotypes of *P. tricornutum* in order to fully master the extend of possibilities from this pleiomorphic organism. The endogenous polyketide pathways are of a particular importance to be unraveled to achieve in vivo cannabinoids production. In the end, the accumulation of data on *P. tricornutum*'s metabolism will probably permit the establishment of relevant models allowing for the construction of strains capable of withstanding industrial scale bioproduction of cannabinoids.

4.2 Conclusion

It is hard for every PhD to be groundbreaking. This work, hybrid between academia and applied research has had several hurdles hindering a flawless experimental design and progression. A significant time has been lost in the synthesis attempts of the massive SteelyA cassette (11 kb), and like companies tend to always do, corner have often been asked to be cut. Despite this, hopes for establishing *Phaeodactylum tricornutum* as a host for cannabinoids production remain. This study proved that the stable expression of *Cs*TKS and *Cs*OAC can be achieved in *P. tricornutum*. The activity of TKS has not

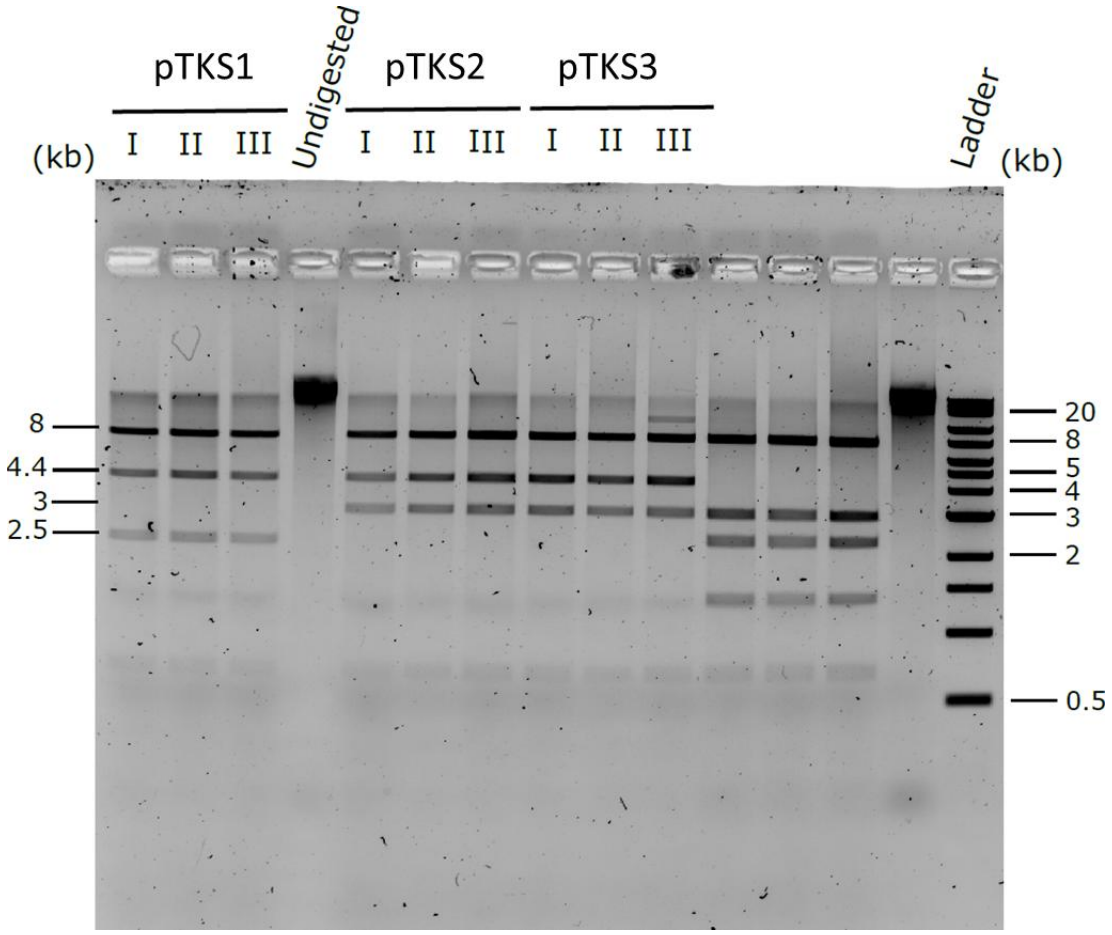
be proven out of all doubts but the results obtained still raise high hopes in building a pathway towards cannabinoids biosynthesis using *Cs*TKS and *Cs*OAC. The characterization of OAC remains but the availability of a viable substrate is a critical factor to achieve it. Significant amount of information on *P. tricornutum*'s metabolism are still missing rendering its tuning to promote heterologous expression complicated, but its naturally rich fatty acid rich content keeps it a promising host. We could also confirm previous findings about heterogeneity of clones with expected identical genomic backgrounds in the metabolomic analysis. This study has also shown the capability of the diatom to withstand the production of a 352 kDa protein over six months without impairing its growth, as well as a preliminary map on the metabolite content modifications after expression and production of a heterologous polyketide synthase.

Overall, if none of the strategy has permitted to successfully produce olivetolic acid, both have increased the knowledge on *Phaeodactylum tricornutum*'s protein production ability and metabolism, paving the way for future bioengineering strategies.

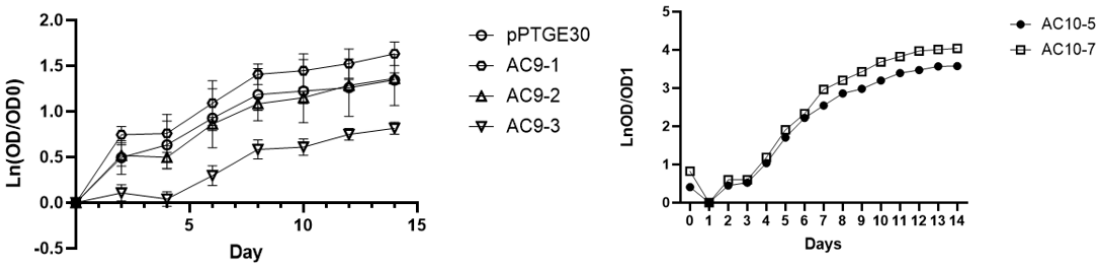
ANNEX A: Supplementary information of chapter II

Supplementary Information

Impact of heterologous expression of *Cannabis sativa* tetraketide synthase on *Phaeodactylum tricornutum* metabolic profile

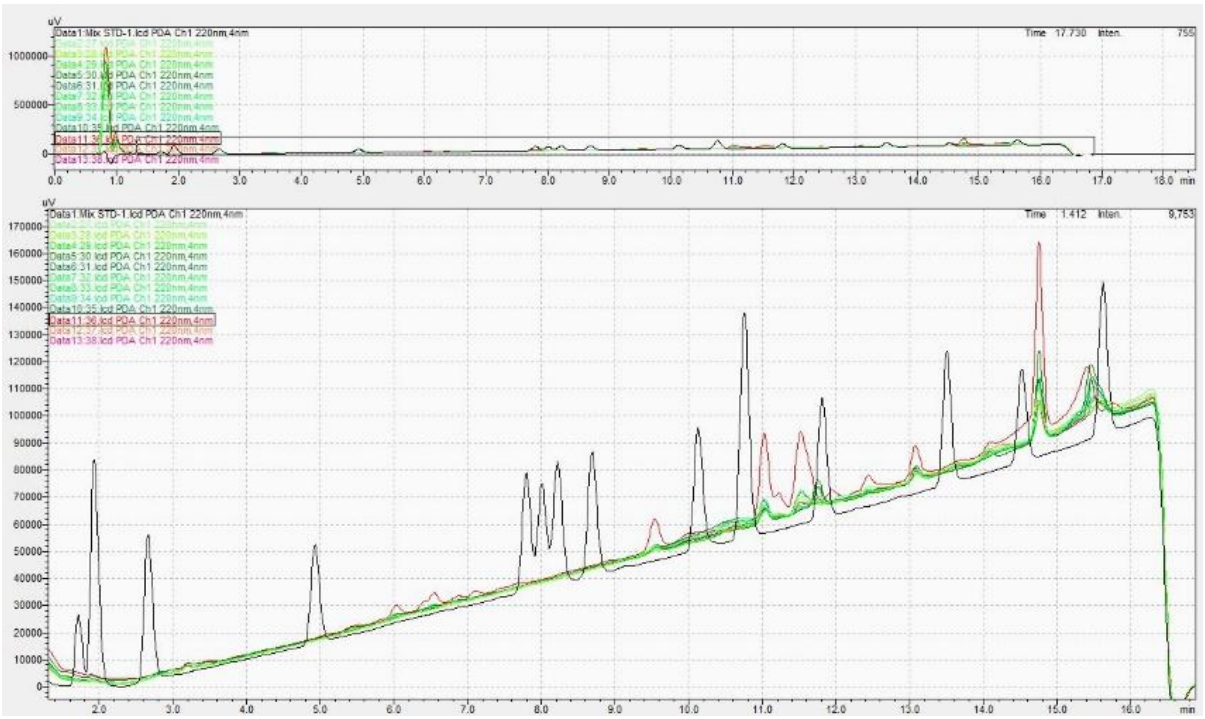


Supplementary Figure 1: *P. tricornutum* DNA plasmid digestion profiles. Agarose gel electrophoresis profile BamHI-HF digestion recombinant plasmid DNA extracted from *E. coli* after plasmid rescue from *P. tricornutum* transconjugants.

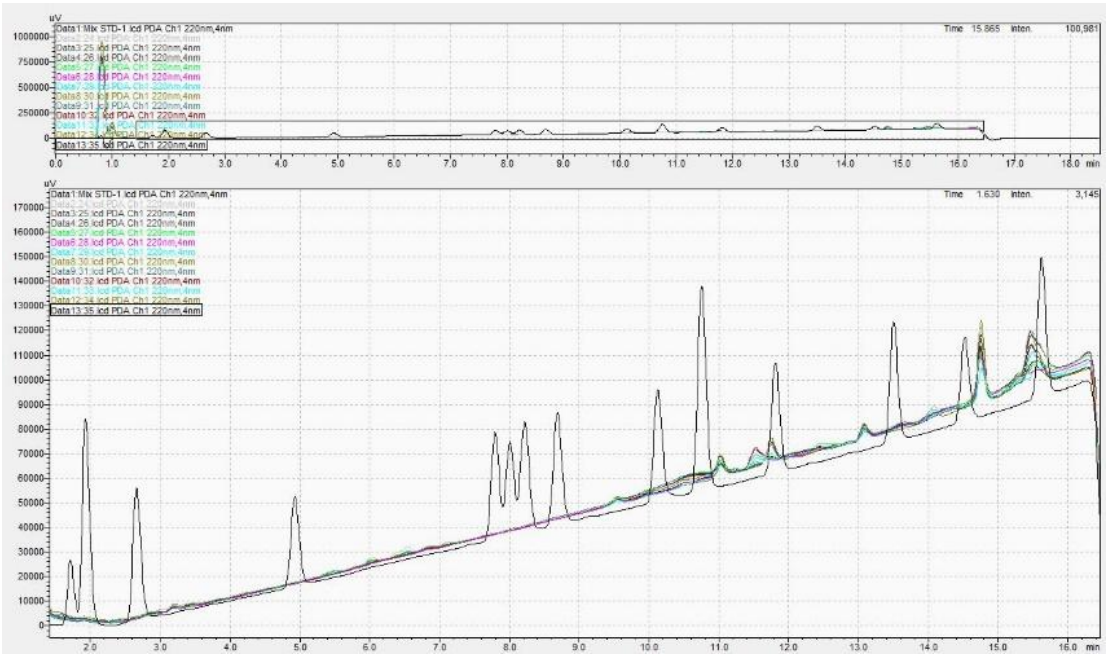


Supplementary Figure 2: *Cs*TKS and empty vector clones have a similar growth kinetic. Optic density of each culture was taken at 730 nm and a t-test was performed for statistical analysis.

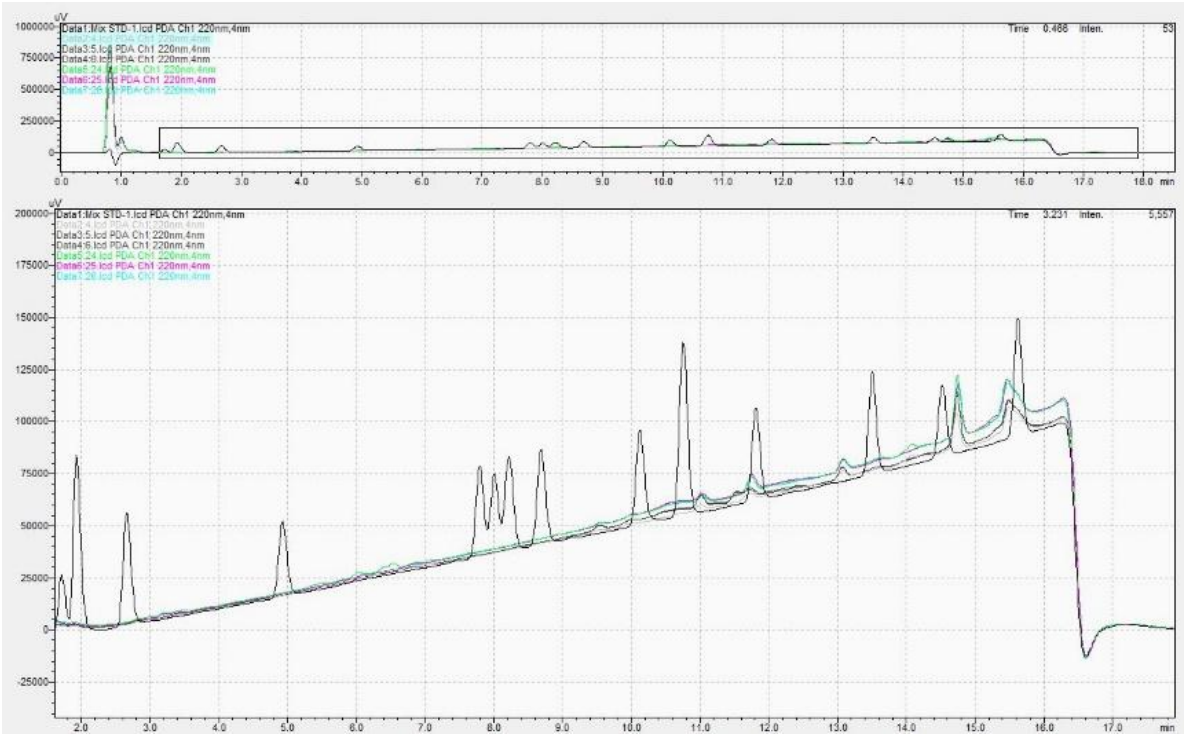
a)


uV
Data1:Mix STD-1.lcd PDA Ch1 220nm,4nm
Time 17.730 Inten. 755
Data10:35.lcd PDA Ch1 220nm,4nm
Data11:36.lcd PDA Ch1 220nm,4nm
Data13:38.lcd PDA Ch1 220nm,4nm
min
Time 1.412 Inten. 9,753

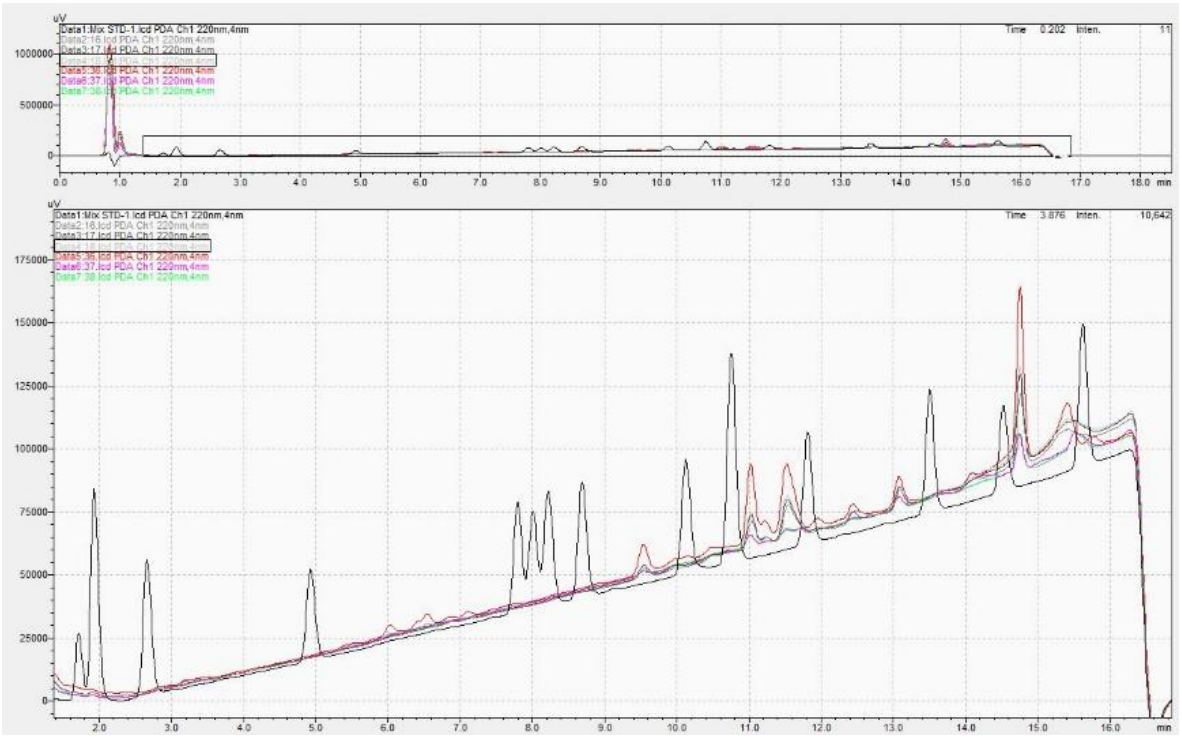
b)


uV
Data1:Mix STD-1.lcd PDA Ch1 220nm,4nm
Time 15.865 Inten. 100,981
Data10:32.lcd PDA Ch1 220nm,4nm
Data13:35.lcd PDA Ch1 220nm,4nm
min
Time 1.630 Inten. 3,145

c)

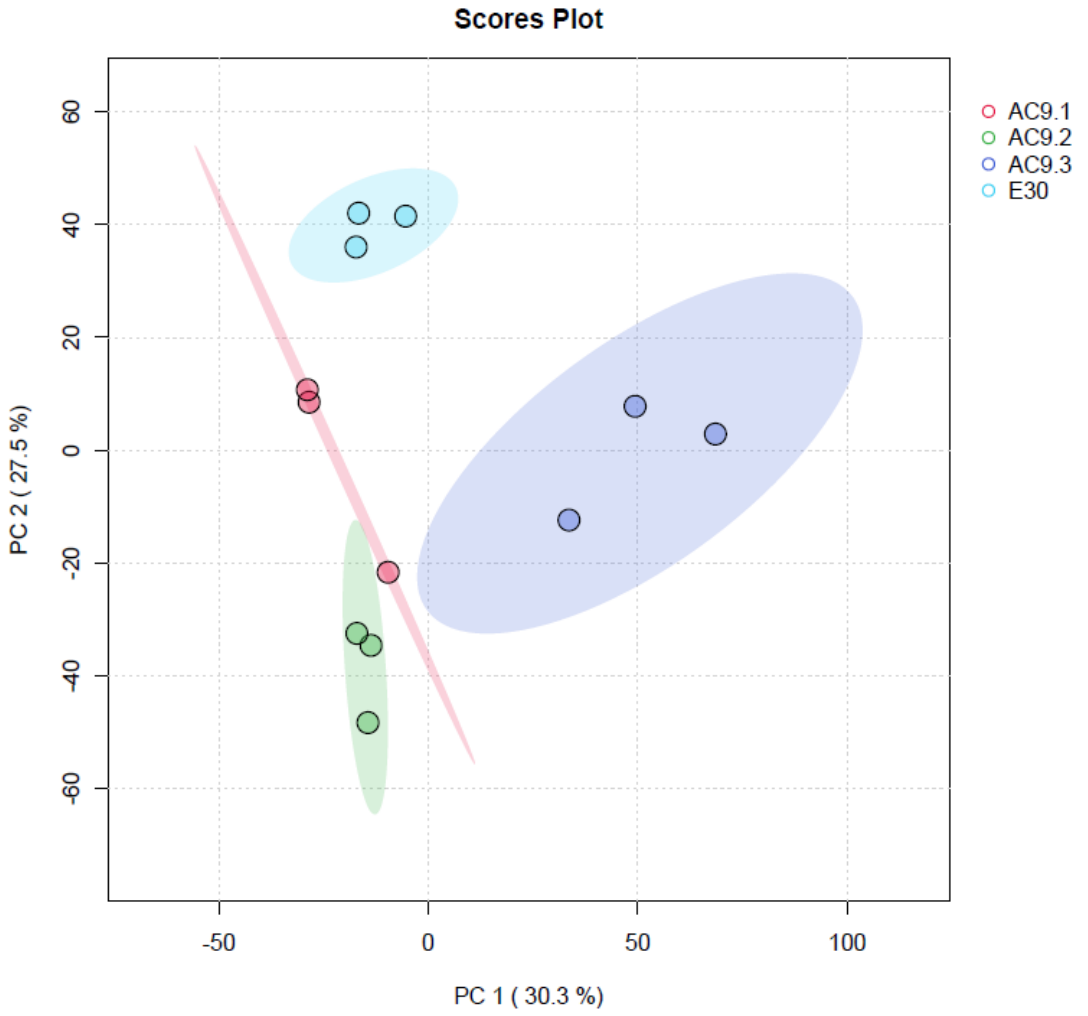


d)

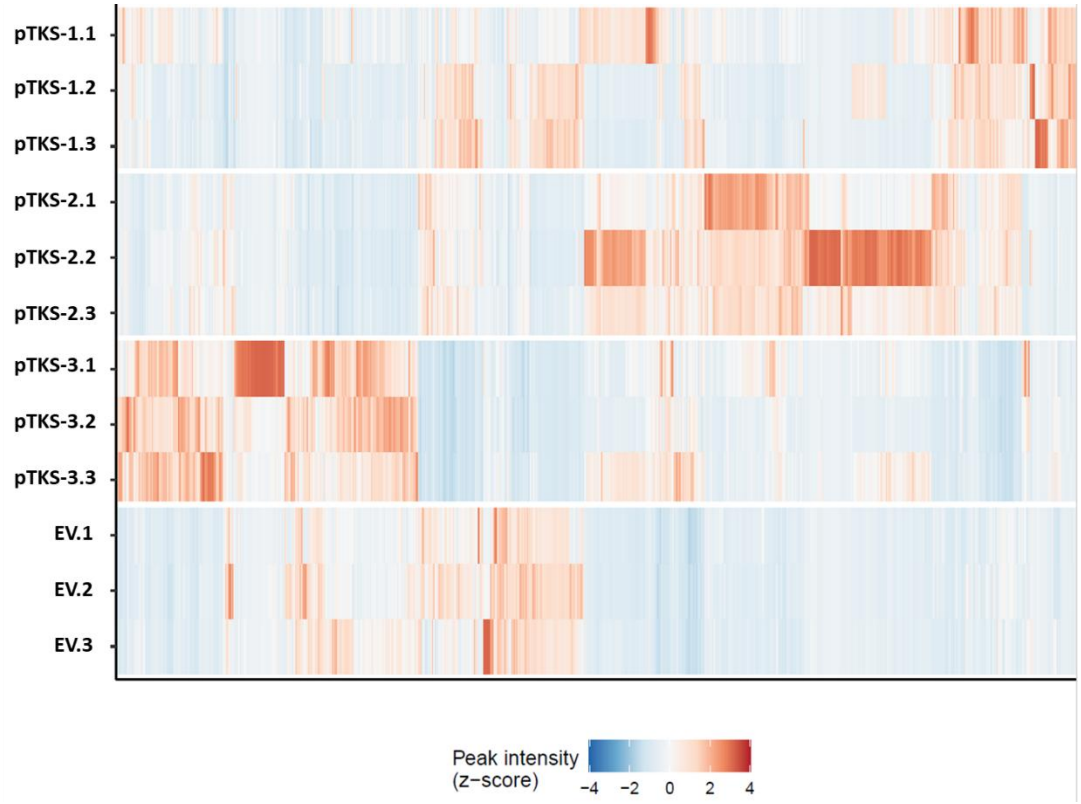


Supplementary Figure 3: HPLC chromatogram of CsTKS and CsOAC enzymatic assays. The black line corresponds to the run with cannabinoid standards. a: Chromatogram of enzymatic assay from crude extract of AC9.1, AC9.4, AC10.5

(*Cs*OAC) and the Combined (*Cs*TKS + *Cs*OAC) culture made of all three clones mixed, all supplemented with hexanoyl-CoA and malonyl-CoA. Each clone has previously been tested for protein production by western blot. The two peaks visible between 11 mins and 12 mins do not correspond to any cannabinoid standard's UV profile. b: Chromatogram of pTKS.1, pTKS.4, pOAC.5 and EV enzymatic assays All clones follow the same profile. c: Chromatogram of pPtGE30 with and without substrate. d: Chromatogram of Combined clones with and without substrate.



Supplementary Figure 4: AC9.3 has a more distant metabolite profile than the 3 other clones: Principal component analysis of every clone's metabolite profile by combining HILIC, Reverse phase positive and negative mode.



Supplementary Figure 5: All triplicates of each clone follow a similar profile. Heatmap of unannotated metabolite relative abundance in each clone compared to the mean in all samples.

Supplementary Table 1: All DNA sequences used in this study.

Identity	Sequence
pPtGE30	ACTAAGGGTACTGTTGACATTGCGAAGAGCGACAAAGATTTTGTTATCGGCTTTATTGCTCAAAGAGACATGGGTGGA AGAGATGAAGGTTACGATTGGTTGATTATGACACCCGGTGTGGGTTTAGATGACAAGGGAGACGCATTGGGTCAACA GTATAGAACCGTGGATGATGTGGTCTCTACAGGATCTGACATTATTATTGTTGGAAGAGGACTATTTGCAAAGGGAAGG GATGCTAAGGTAGAGGGTGAACGTTACAGAAAAGCAGGCTGGGAAGCATATTTGAGAAGATGCGGCCAGCAAAACT AAAAAACTGTATTATAAGTAAATGCATGTATACTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCG GCCGGGAATCTCGGTCGTAATGATTTTTATAATGACGAAAAAAAAAAAAATTGGAAAGAAAACCCCCCCCCCCCCCG CAGCGTTGGGTCCTGGCCACGGGTGCGCATGATCGTGCTCCTGTCGTTGAGGACCCGGCTAGGCTGGCGGGGTTGCC TTACTGGTTAGCAGAATGAATCACCGATACGCGAGCGAACGTGAAGCGACTGCTGCTGCAAAACGTCTGCGACCTGA GCAACAACATGAATGGTCTTCGGTTTCCGTGTTTCGTAAAGTCTGGAAACGCGGAAGTCAGCGCCCTGCACCATTATG TTCCGGATCTGCATCGCAGGATGCTGCTGGCTACCCTGTGGAACACCTACATCTGTATTAACGAAGCGCTGGCATTGAC CCTGAGTGATTTTTCTCTGGTCCCGCCGCATCCATACCGCCAGTTGTTTACCCTCACAACGTTCCAGTAACCGGGCATG TTCATCATCAGTAACCCGTATCGTGAGCATCCTCTCTCGTTTCATCGAGTTACGCTAGGGATAACAGGGTAATATAGATC TTCCGCTGCATAACCCTGCTTCGGGGTCATTATAGCGATTTTTTCGGTATATCCATCCTTTTTCGCACGATATACAGGATT TTGCCAAAGGGTTCGTGTAGACTTTCCTTGGTGTATCCAACGGCGTCAGCCGGGCAGGATAGGTGAAGTAGGCCCAC CCGCGAGCGGGTGTTCCTTCTTCACTGTCCCTTATTCGCACCTGGCGGTGCTCAACGGGAATCCTGCTCTGCGAGGCT GGCCGGCTACCGCCGGCGTAACAGATGAGGGCAAGCGGATGGCTGATGAAACCAAGCCAACCAGGAAGGGCAGCCC ACCTATCAAGGTGTACTGCCTTCCAGACGAACGAAGAGCGATTGAGGAAAAGGCGGCGGCGGCCGGCATGAGCCTG TCGGCCTACCTGCTGGCCGTCGGCCAGGGCTACAAAATCACGGGCGTCGTGGACTATGAGCACGTCCGCGAGCTGGC CCGCATCAATGGCGACCTGGGCCGCCTGGGCGGCCTGCTGAAACTCTGGCTCACCGACGACCCGCGCACGGCGCGGT TCGGTGATGCCACGATCCTCGCCCTGCTGGCGAAGATCGAAGAGAAGCAGGACGAGCTTGGCAAGGTCATGATGGGC GTGGTCCGCCCGAGGGCAGAGCCATGACTTTTTTAGCCGCTAAAACGGCCGGGGGGTGCGCGTGATTGCCAAGCACG TCCCCATGCGCTCCATCAAGAAGAGGCACTTCGAGCTGTAAGTACATCACCGACGAGCAAGGCAAGACGATCCGCGC CTGTTTTATTGAGAACGTTGTTCGTGTTGGCCTCAATGGTAGCGATGCGTCATTCAGCGAAGTTCTGGTGCTGATGATG TGGTTCGCTTTGCCACTGGTCAATGTGGTAAGCCCGTGTAATGTCAGTAACCTTTTTACTGATCTCAGCTTGAGCACGG TCGCTGATGAGCTTATCCATGGCCCCACGGTAACGGATATGATCCTCTAGGGCGTTGACAAATTCTTTGTCGGTTTTCAT GGGGTAGACGTCAGTGACCAGGGAACGTCTCCCTACAAAAAGTTGCGCATACTTTGCTCCGTTGTCGACGGCAGGGG TGTCGGACCAGATTGTATCTGTGGCAACGGCCTCATTGCGTCAATGGACCTTTAAAGCCGGGAAACGAGACTTGAAAT GTTTACGTAGAGGTGCATTATAGACCTCACGCGCATATTGAGTGGTTGCAAAAATGGTTTTGCGAACGGTATCACTGGA GACCCATGCCAGGCAAGGACGGAGGGCATCATAATCATGTTCATTGCGTTGGACAGCATGCTTGTTACACGTAAGTAT ACGGTCGAGAGCATCGTGACGTTGAGAATTACAGATGGCAACGTGTCGGTGCATGTATTGGCCGAGATTGTCAAATCG TGGCTCAATATATGGCATGCCAGGTAAGTCATGGATGTCGGTATGCCATTCAGCCTTCATGTCGATTTCGTTGTCGAGAA CTGATGGGTCCCAGTCAACGTCAGAAGTAAGAACAACATGGGGAAGGGAATGAAGTTCATGACTGGTGGGTGCGCG CATATCCATGTACGGAAGACCCTGTCGTATGTTTAGTGGGATGATGTAATCATCCAGGGTAACAATACGTTGGTTGCCG CCCACGGTACGAGACCGATCTTGAACACAATTGTGGCAATGTTCAAGTTGGGCGCTGGAATGGATGGTTTTACCTTTA CCAAGATGCGCGTATTGATGCATGATAACAACAATGGGCCCTCGTTGCGATTCGACAAGACCGGCAGCCGTGACAATG TCTAGATTGGACAAGGTATGTTCATTTATACCGGTTATGTTGGCTGAACGTCCAGTTTTGTGCAGTACGGTAACATCACT CCCCGCAAGTCCACCGTTGGCACCGCGATCGACAAGTGCAGATGTGGTATTTTGTACTGTGTGTCGGGAGACTTGGTA CTGTAACACATTTGACTGAAGTGACTTGGTGGTACCATGAGGACGGAGGGAACCCGTTGATACGGTGTTGCGTGACG CTGCAAGTACCTTACGTATGTCGCCTGGATCCATACGTCCAACCCGATCAGAAAGATGTGCTAGGAGTTCCGTTTCTGG TGCACAATCATGAAAGGTATCCACGGGAGTTCCATTGTTGTCGTCCGGGGCTGCGTCGCGACTGCCAGTTGCATGTAC ATGCGCCATCAAGGGACGTGCCAATCCATTACTAACCGGGAGAACCTTAGCTGGGGCAGCCAATCCCTGAAGTATAGC CTTAGCATCATCTGAGAGCTGGTTCCACATATCTCTAGGGATATAAGGTCGTTCACGGTTAGTGGGCGTACTATGATGTC GGTTATTAGTACAATTCTCCCTGCGGGCATGCGCATTGGCTTCATAGAGTACGGAAGGTGACAAGTCAATATTGTAGTT AACATCCGGATCAGTGTCAAAGTCTGTAGGATGGTAAGAAAGATCAGTAGAATGAATACTACGCTTTTCCTTGGGACT ACGGGAATTAGAGAAGTTGTTTCCTTTATTGTAGAGTGACGCTGAAGCAAGTAGAAGACTAAGGTAACTCTCGTAGCT AATAGGATTACCTCCTTTGGCTAGGTCAAGAGTGGCTGTGATCTTCACTTGACAAAGTTCCGGTACATTGTGGACAGC ATTCTCCAAAAGACTAAGACACAGTTGCTTTGGGAGTTGCTCAGCCATTGGTACAGTATTGTGGTAGATACAAAGGTG GTTCTTCCAATGAAGGATAAATCCTTCTGCTGTGCCTGTCCATGAGGATCCATATTTCGCCGTAGTTAGGTAACCAAGC GTGGTGGCTGAACTGATCTTCGCACTTGCTGATTCCGTATAGTGTTTGACAACTTTACAAAACACTTCTTGCGCAGTTC GCTCTAGCCTGGACAGAGAAGGGACAATCTTCGTCGTTAAGACTCGTGCGAATAGCAAAAGATCACAAAATAGCACA TCGGCACCGACCAACGATTATTTCCAAGGAAAAAAAGAATGCTTCACTACAAGAAATTGTGTCATCCCTATACAGAGT CTTGTTACTGTGACAGAAAATTGATGGAAGATGTGGCGGATTGCCTTTACACTAGCCAACTTGTTCGACTAATTGCAGC TTCTTCTGAGAGGCTTCACCGAGTAACGCGAAGAACACCGGTGTCTCGTACATGCTCGTCGGTGAACGCTCGTCCAAT GACACCCCCCCACTTTGTATCAATATCCCAACTTGGTAGTGAACTGGAATGATACATGCAATTTCGCGCCGCATCAACAG CCACGGGCACCATCGACGAATAGACTCGGTCGAGCTGGTTGCCCTCGCCGCTGGGCTGGCGGCCGTCTATGGCCCTG CAAACGCGCCAGAAACGCCGTCGAAGCCGTGTGCGAGACACCGCGGCCGGCCGCCGGCGTTGTGGATACCTCGCGG AAAACTTGGCCCTCACTGACAGATGAGGGGCGGACGTTGACACTTGAGGGGCCGACTCACCCGGCGCGGCGTTGAC AGATGAGGGGCAGGCTCGATTTCGGCCGGCGACGTGGAGCTGGCCAGCCTCGCAAATCGGCGAAAACGCCTGATTTT ACGCGAGTTTCCCACAGATGATGTGGACAAGCCTGGGGATAAGTGCCCTGCGGTATTGACACTTGAGGGGCGCGACT ACTGACAGATGAGGGGCGCGATCCTTGACACTTGAGGGGCAGAGTGCTGACAGATGAGGGGCGCACCTATTGACATT TGAGGGGCTGTCCACAGGCAGAAAATCCAGCATTTGCAAGGGTTTCCGCCCGTTTTTCGGCCACCGCTAACCTGTCTT TTAACCTGCTTTTAAACCAATATTTATAAACCTTGTTTTTAACCAGGGCTGCGCCCTGTGCGCGTGACCGCGCACGCCG AAGGGGGGTGCCCCCCCTTCTCGAACCCTCCCGGTCGAGTGAGCGAGGAAGCACCAGGGAACAGCACTTATATATTC TGCTTACACACGATGCCTGAAAAAACTTCCCTTGGGGTTATCCACTTATCCACGGGGATATTTTTATAATTATTTTTTTTA TAGTTTTTAGATCTTCTTTTTTAGAGCGCCTTGTAGGCCTTTATCCATGCTGGTTCTAGAGAAGGTGTTGTGACAAATTG CCCTTTCAGTGTGACAAATCACCCTCAAATGACAGTCCTGTCTGTGACAAATTGCCCTTAACCCTGTGACAAATTGCC CTCAGAAGAAGCTGTTTTTTCACAAAGTTATCCCTGCTTATTGACTCTTTTTTATTTAGTGTGACAATCTAAAAACTTGT

	CACACTTCACATGGATCTGTCATGGCGGAAACAGCGGTTATCAATCACAAGAAACGTAAAAATAGCCCGCGAATCGTC CAGTCAAACGACCTCACTGAGGCGGCATATAGTCTCTCCCGGGATCAAAAACGTATGCTGTATCTGTTCGTTGACCAG ATCAGAAAATCTGATGGCACCCTACAGGAACATGACGGTATCTGCGAGATCCATGTTGCTAAATATGCTGAAATATTCG GATTGACCTCTGCGGAAGCCAGTAAGGATATACGGCAGGCATTGAAGAGTTTCGCGGGGAAGGAAGTGGTTTTTTATC GCCCTGAAGAGGATGCCGGCGATGAAAAAGGCTATGAATCTTTTCCTTGGTTTATCAAACGTGCGCACAGTCCATCCA GAGGGCTTTACAGTGTACATATCAACCCATATCTCATTCCCTTCTTTATCGGGTTACAGAACCGGTTTACGCAGTTTCGG CTTAGTGAAACAAAAGAAATCACCAATCCGTATGCCATGCGTTTATACGAATCCCTGTGTCAGTATCGTAAGCCGGATG GCTCAGGCATCGTCTCTCTGAAAATCGACTGGATCATAGAGCGTTACCAGCTGCCTCAAAGTTACCAGCGTATGCCTG ACTTCCGCCGCCGCTTCCTGCAGGTCTGTGTTAATGAGATCAACAGCAGAACTCCAATGCGCCTCTCATACATTGAGA AAAAGAAAGGCCGCCAGACGACTCATATCGTATTTTCCTTCCGCGATATCACTTCCATGACGACAGGATAGTCTGAGG GTTATCTGTCACAGATTTGAGGGTGGTTCGTCACATTTGTTCTGACCTACTGAGGGTAATTTGTCACAGTTTTGCTGTT TCCTTCAGCCTGCATGGATTTTCTCATACTTTTTGAACTGTAATTTTTAAGGAAGCCAAATTTGAGGGCAGTTTGTCAC AGTTGATTTCCTTCTCTTTCCCTTCGTCATGTGACCTGATATCGGGGGTTAGTTCGTCATCATTGATGAGGGTTGATTAT CACAGTTTATTACTCTGAATTGGCTATCCGCGTGTGTACCTCTACCTGGAGTTTTTCCCACGGTGGATATTTCTTCTTGC GCTGAGCGTAAGAGCTATCTGACAGAACAGTTCTTCTTTGCTTCCTCGCCAGTTCGCTCGCTATGCTCGGTTACACGGC TGCGGCGAGCATCACGTGCTATAAAAATAATTATAATTTAAATTTTTTAATATAAATATATAAATTAAAAATAGAAAGTAA AAAAAGAAATTAAAGAAAAAATAGTTTTTGTTTTCCGAAGATGTAAAAGACTCTAGGGGGATCGCCAACAAATACTAC CTTTTACCTTGCTCTTCCTGCTCTCAGGTATTAATGCCGAATTGTTTCATCTTGTCTGTGTAGAAGACCACACACGAAA ATCCTGTGATTTTACATTTTACTTATCGTTAATCGAATGTATATCTATTTAATCTGCTTTTCTTGTCTAATAAATATATATGT AAAGTACGCTTTTTGTTGAAATTTTTTAAACCTTTGTTTATTTTTTTTTCTTCATTCCGTAACTCTTCTACCTTCTTTATTT ACTTTCTAAAATCCAAATACAAAACATAAAAATAAATAAACACAGAGTAAATTCCCAAATTATTCCATCATTAAAAGAT ACGAGGCGCGTGTAAGTTACAGGCAAGCGATCCTAGTACACTCTATATTTTTTTATGCCTCGGTAATGATTTTCATTTTT TTTTTCCACCTAGCGGATGACTCTTTTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTATACATTATATAAAGTAAT GTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAGATGACAGAGCAGAAAGCCCT AGTAAAGCGTATTACAAATGAAACCAAGATTCAGATTGCGATCTCTTTAAAGGGTGGTCCCCTAGCGATAGAGCACTC GATCTTCCCAGAAAAAGAGGCAGAAGCAGTAGCAGAACAGGCCACACAATCGCAAGTGATTAACGTCCACACAGGT ATAGGGTTTCTGGACCATATGATACATGCTCTGGCCAAGCATTCCGGCTGGTCGCTAATCGTTGAGTGCATTGGTGACT TACACATAGACGACCATCACACCACTGAAGACTGCGGGATTGCTCTCGGTCAAGCTTTTAAAGAGGCCCTAGGGGCC GTGCGTGGAGTAAAAAGGTTTGGATCAGGATTTGCGCCTTTGGATGAGGCACTTTCCAGAGCGGTGGTAGATCTTTCG AACAGGCCGTACGCAGTTGTCGAACTTGGTTTGCAAAGGGAGAAAGTAGGAGATCTCTCTTGCGAGATGATCCCGCA TTTTCTTGAAAGCTTTGCAGAGGCTAGCAGAATTACCCTCCACGTTGATTGTCTGCGAGGCAAGAATGATCATCACCG TAGTGAGAGTGCGTTCAAGGCTCTTGCGGTTGCCATAAGAGAAGCCACCTCGCCCAATGGTACCAACGATGTTCCCTC CACCAAAGGTGTTCTTATGTAGTTTTACACAGGAGTCTGGACTTGACGCTAGTGATAATAAGTGACTGAGGTATGTGCT CTTCTTATCTCCTTTTGTAGTGTTGCTCTTATTTTAAACAACTTTGCGGTTTTTTGATGACTTTGCGATTTTGTTGTTGCT TTGCAGTAAATTGCAAGATTTAATAAAAAAACGCAAAGCAATGATTAAAGGATGTTCAGAATGAAACTCATGGAAACA CTTAACCAGTGCATAAACGCTGGTCATGAAATGACGAAGGCTATCGCCATTGCACAGTTTAATGATGACAGCCCGGAA GCGAGGAAAATAACCCGGCGCTGGAGAATAGGTGAAGCAGCGGATTTAGTTGGGGTTTCTTCTCAGGCTATCAGAGA TGCCGAGAAAGCAGGGCGACTACCGCACCCGGATATGGAAATTCGAGGACGGGTTGAGCAACGTGTTGGTTATACAA TTGAACAAATTAATCATATGCGTGATGTGTTTGGTACGCGATTGCGACGTGCTGAAGACGTATTTCCACCGGTGATCGG GGTTGCTGCCCATAAAGGTGGCGTTTACAAAACCTCAGTTTCTGTTCATCTTGCTCAGGATCTGGCTCTGAAGGGGCT ACGTGTTTTGCTCGTGGAAGGTAACGACCCCCAGGGAACAGCCTCAATGTATCACGGATGGGTACCAGATCTTCATAT TCATGCAGAAGACACTCTCCTGCCTTTCTATCTTGGGGAAAAGGACGATGTCACTTATGCAATAAAGCCCACTTGCTG GCCGGGGCTTGACATTATTCCTTCCTGTCTGGCTCTGCACCGTATTGAAACTGAGTTAATGGGCAAATTTGATGAAGGT AAACTGCCCACCGATCCACACCTGATGCTCCGACTGGCCATTGAAACTGTTGCTCATGACTATGATGTCATAGTTATTG ACAGCGCGCCTAACCTGGGTATCGGCACGATTAATGTCGTATGTGCTGCTGATGTGCTGATTGTTCCCACGCCTGCTGA GTTGTTTGACTACACCTCCGCACTGCAGTTTTTCGATATGCTTCGTGATCTGCTCAAGAACGTTGATCTTAAAGGGTTC GAGCCTGATGTACGTATTTTGCTTACCAAATACAGCAATAGCAATGGCTCTCAGTCCCCGTGGATGGAGGAGCAAATTC GGGATGCCTGGGGAAGCATGGTTCTAAAAAATGTTGTACGTGAAACGGATGAAGTTGGTAAAGGTCAGATCCGGATG AGAACTGTTTTTGAACAGGCCATTGATCAACGCTCTTCAACTGGTGCCTGGAGAAATGCTCTTTCTATTTGGGAACCT GTCTGCAATGAAATTTTCGATCGTCTGATTAAACCACGCTGGGAGATTAGATAATGAAGCGTGCGCCTGTTATTCCAAA ACATACGCTCAATACTCAACCGGTTGAAGATACTTCGTTATCGACACCAGCTGCCCCGATGGTGGATTCGTTAATTGCG CGCGTAGGAGTAATGGCTCGCGGTAATGCCATTACTTTGCCTGTATGTGGTCGGGATGTGAAGTTTACTCTTGAAGTGC TCCGGGGTGATAGTGTTGAGAAGACCTCTCGGGTATGGTCAGGTAATGAACGTGACCAGGAGCTGCTTACTGAGGAC GCACTGGATGATCTCATCCCTTCTTTTCTACTGACTGGTCAACAGACACCGGCGTTCGGTCGAAGAGTATCTGGTGTC ATAGAAATTGCCGATGGGAGTCGCCGTCGTAAAGCTGCTGCACTTACCGAAAGTGATTATCGTGTTCTGGTTGGCGAG CTGGATGATGAGCAGATGGCTGCATTATCCAGATTGGGTAACGATTATCGCCCAACAAGTGCTTATGAACGTGGTCAGC GTTATGCAAGCCGATTGCAGAATGAATTTGCTGGAAATATTTCTGCGCTGGCTGATGCGGAAAATATTTCACGTAAGAT TATTACCCGCTGTATCAACACCGCCAAATTGCCTAAATCAGTTGTTGCTCTTTTTTCTCACCCCGGTGAACTATCTGCCC GGTCAGGTGATGCACTTCAAAAAGCCTTTACAGATAAAGAGGAATTACTTAAGCAGCAGGCATCTAACCTTCATGAGC AGAAAAAAGCTGGGGTGATATTTGAAGCTGAAGAAGTTATCACTCTTTTAACTTCTGTGCTTAAAACGTCATCTGCATC AAGAACTAGTTTAAGCTCACGACATCAGTTTGCTCCTGGAGCGACAGTATTGTATAAGGGCGATAAAATGGTGCTTAA CCTGGACAGGTCTCGTGTTCCAACTGAGTGTATAGAGAAAATTGAGGCCATTCTTAAGGAACTTGAAAAGCCAGCAC CCTGATGCGACCTCGTTTTAGTCTACGTTTATCTGTCTTTACTTAATGTCCTTTGTTACAGGCCAGAAAGCATAACTGGC CTGAATATTCTCTCTGGGCCCACTGTTCCACTTGTATCGTCGGTCTGATAATCAGACTGGGACCACGGTCCCACTCGTA TCGTCGGTCTGATTATTAGTCTGGGACCATGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCACGGTCCC ACTCGTATCGTCGGTCTGATTATTAGTCTGGAACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCAC GGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCACGATCCCACTCGTGTTGTCGGTCTGATTATCGGTCTG GGACCACGGTCCCACTTGTATTGTCGATCAGACTATCAGCGTGAGACTACGATTCCATCAATGCCTGTCAAGGGCAAG TATTGACATGTCGTCGTAACCTGTAGAACGGAGTAACCTCGGTGTGCGGTTGTATGCCTGCTGTGGATTGCTGCTGTGT CCTGCTTATCCACAACATTTTGCGCACGGTTATGTGGACAAAATACCTGGTTACCCAGGCCGTGCCGGCACGTTAACC GGGCTGCATCCGATGCAAGTGTGTCGCTGTCGACGAGCTCGCGAGCTCGGACATGAGGTTGCCCCGTATTCAGTGTCG CTGATTTGTATTGTCTGAAGTTGTTTTTACGTTAAGTTGATGCAGATCAATTAATACGATACCTGCGTCATAATTGATTAT TTGACGTGGTTTGATGGCCTCCACGCACGTTGTGATATGTAGATGATAATCATTATCACTTTACGGGTCCTTTCCGGTGA TCCGACAGGTTACGGGGCGGCGACCTCGCGGGTTTTCGCTATTTATGAAAATTTTCCGGTTTAAGGCGTTTCCGTTCTT

	CTTCGTCATAACTTAATGTTTTTATTTAAAATACCCTCTGAAAAGAAAGGAAACGACAGGTGCTGAAAGCGAGCTTTT TGGCCTCTGTCGTTTCCTTTCTCTGTTTTTGTCCGTGGAATGAACAATGGAAGTCCGAGCTCATCGCTAATAACTTCGT ATAGCATACATTATACGAAGTTATATTCGATGCGGCCGCAAGGGGTTCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGC TGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCACACAGATGCGTA AGGAGAAAATACCGCATCAGGCGCCATTCGCCATTCAGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCT TCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTC ACGACGTTGTAAAACGACGGCCAGTGAATTGTAATACGACTCACTATAGGGCGAATTCGAGCTCGGTACCCGGGGATC CTCTAGAGTCGACCTGCAGGCATGCAAGCTTGAGTATTCTATAGTCTCACCTAAATAGCTTGGCGTAATCATGGTCATA GCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTG GGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTG CCAGCTGCATTAATGAATCGGCCAACGCGAACCCCTTGCGGCCGCCCGGGCCGTCGACCAATTCTCATGTTTGACAGC TTATCATCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTTTAAGGGCACCAATAA CTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACTGTTGTAATTCATTAAGCATTCTGCCGACATGGA AGCCATCACAAACGGCATGATGAACCTGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCAT GGTGAAAACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAGGGATTGG CTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGTTTTCACCGTAACACGCCACATCTTGCG AATATATGTGTAGAAACTGCCGGAAATCGTCGTGGTattcactccagagcgatgaaaacgtttcagtttgctcatggaaaacggtgtaACAAGGGTGA ACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGAAATTCCGGATGAGCATTCATCAGGCGGGCAAGAA TGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTTTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGT CTGGTTATAGGTACATTGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGGTG GTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGATAACTCAAAAAATACGCCCGGTAG TGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTTACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAG GGCTTCCCGGTATCAACAGGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGCGATA AGCTCATGGAGCGGCGTAACCGTCGCACAGGAAGGACAGAGAAAGCGCGGATCTGGGAAGTGACGGACAGAACGG TCAGGACCTGGATTGGGGAGGCGGTTGCCGCCGCTGCTGCTGACGGTGTGACGTTCTCTGTTCCGGTCACACCACAT ACGTTCCGCCATTCCTATGCGATGCACATGCTGTATGCCGGTATACCGCTGAAAGTTCTGCAAAGCCTGATGGGACATA AGTCCATCAGTTCAACGGAAGTCTACACGAAGGTTTTTGCGCTGGATGTGGCTGCCCGGCACCGGGTGCAGTTTGCG ATGCCGGAGTCTGATGCGGTTGCGATGCTGAAACAATTATCCTGAGAATAAATGCCTTGGCCTTTATATGGAAATGTGG AACTGAGTGGATATGCTGTTTTTGTCTGTTAAACAGAGAAGCTGGCTGTTATCCACTGAGAAGCGAACGAAACAGTC GGGAAAATCTCCCATTATCGTAGAGATCCGCATTATTAATCTCAGGAGCCTGTGTAGCGTTTATAGGAAGTAGTGTTCTG TCATGATGCCTGCAAGCGGTAACGAAAACGATTTGAATATGCCTTCAGGAACAATAGAAATCTTCGTGCGGTGTTACG TTGAAGTGGAGCGGATTATGTCAGCAATGGACAGAACAACCTAATGAACACAGAACCATGATGTGGTCTGTCCTTTTA CAGCCAGTAGTGCTCGCCGCAGTCGAGCGACAGGGCGAAGCCCATCGATACTAGCTTGATTGGGATATCTCGCTCGTG CTTGTCGCGTGCTATGTCTTTTAGGGTACTTTGAACCTACGTTCGTACTTGTGTAATATGATCATCGTCGTCATCGTATT ATCGTTTTTCATCCGTCCAGCGCAAAATGCATTAGCAGCTAGTCCTAGCGTGCGGAGCTACCTGTACAGGTGCATGAC GGATGCGTGTCCTTAAGTGAGTTTCTAATTAACAGTAACTTCTTTACTTATGTTTCAGTTTGTAAGAAGCGGGATGCGC TCGTCGCTTGACATCTGATTGGACTGCGTCGGCACGTGAAAACTACATTGTGAAATCTGCTAAAACTCCGGGTATCTCT GACACAAAACGATTCGGCTTCGCAATTTCAACATTACGGTCAAGGCTAACGTATCTTTCTCGGTCAACTTCAGATTATG CCGATTAAATTGTCGTAGCTTTCAAGGCGTTTTGAGTACTGCGGCAGTTGTTGAACCTGCAAGGAGAAGATCTCGACA ACAGAATAAAGCGAAAAATGGGTCTCATGCACTAACACTCAGGCCTCCCTCATAATCTCTGTTTGAGTTTACCAACAA CACATATATACATTTCGACAAAATGGCCAAGTTGACCAGTGCCGTTCCGGTGCTCACCGCGCGCGACGTCGCCGGAGC GGTCGAGTTCTGGACCGACCGGCTCGGGTTCTCCCGGGACTTCGTGGAGGACGACTTCGCCGGTGTGGTCCGGGACG ACGTGACCCTGTTCATCAGCGCGGTCCAGGACCAGGTGGTGCCAGACAACACCCTGGCCTGGGTGTGGGTGCGCGG CCTGGACGAGCTGTACGCCGAGTGGTCGGAGGTCGTGTCCACGAACTTCCGGGACGCCTCCGGGCCGGCCATGACCG AGATCGGCGAGCAGCCGTGGGGGCGGGAGTTCGCCCTGCGCGACCCGGCCGGCAACTGCGTGCACTTCGTGGCCGA GGAGCAGGACTGACCGACGCCGACCAACACCGCCGGTCCGACGCGGCCCGACGGGTCCGAGGCCTCGGAGATCTGG GCCCATGCGGCCGCAACAACTACCTCGACTTTGGCTGGGACACTTTCAGTGAGGACAAGAAGCTTCAGAAGCGTGCT ATCGAACTCAACCAGGGACGTGCGGCACAAATGGGCATCCTTGCTCTCATGGTGCACGAACAGTTGGGAGTCTCTATC CTTCCTTAAAAATTTAATTTTCATTAGTTGCAGTCACTCCGCTTTGGTTTCGTAACTATAACGGTCCTAAGGTAGCGAAC GTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGGATCGCTCGCGGCTCTTACCA GCCTAACTTCGATCATTGGACCGCTGATCGTCACGGCGATTTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCAT GGATTGTAGGCGCCGCCCTATACCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTCGACCT GAATGGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAAGAATTGGAGCCAATCAATTCTTGCGGAGAACTG TGAATGCGCAAACCAACCCTTGGCAGAACATATCCATCGCGTCCGCCATCTCCAGCAGCCGCACGCGGCGCATCGGG GGGGGGGGGGTTTCAATTCATCATTTTTTTTTTATTCTTTTTTTTGATTTCGGTTTCCTTGAAATTTTTTTGATTCGGTAA TCTCCGAACAGAAGGAAGAACGAAGGAAGGAGCACAGACTTAGATTGGTATATATACGCATATGTAGTGTTGAAGAA ACATGAAATTGCCCAGTATTCTTAACCCAACTGCACAGAACAAAAACCTGCAGGAAACGAAGATAAATCATGTCGAA AGCTACATATAAGGAACGTGCTGCTACTCATCCTAGTCCTGTTGCTGCCAAGCTATTTAATATCATGCACGAAAAGCAA ACAAACTTGTGTGCTTCATTGGATGTTCGTACCACCAAGGAATTACTGGAGTTAGTTGAAGCATTAGGTCCCAAAATTT GTTTACTAAAAACACATGTGGATATCTTGACTGATTTTTCCATGGAGGGCACAGTTAAGCCGCTAAAGGCATTATCCGC CAAGTACAATTTTTTACTCTTCGAAGACAGAAAATTTGCTGACATTGGTAATACAGTCAAATTGCAGTACTCTGCGGGT GTATACAGAATAGCAGAATGGGCAGACATTACGAATGCACACGGTGTGGTGGGCCCAGGTATTGTTAGCGGTTTGAAG CAGGCGGCGGAAGAAGTAACAAAGGAACCTAGAGGCCTTTTGATGTTAGCAGAATTGTCATGCAAGGGCTCCCTAGC TACTGGAGAATAT
TKS cassette	CCCTGCGATAGACCTTTTCCAAACTCACGCAGTCCAAGAAAACAAAGGGGTGAGAAGTATACGCACCTTTCGGTTTC GGCATAATTCTTAAACTCTTGTGGTCACTTTCTTGTGAAGAAGCTAGGGGCACTCGTTTTCCCTCAGAGCCTGCAAAC ACAAAATTCCTGCAGTCAATTGTCCCAACACTCGGCAAACCGTATGCGCAAGCAACGATGCGCAGAAGGCCGTGGAT GGATGGCGACTCGCGATATGGCTTCTTGGGTCGCCAGTGTGGTACGTCCGGCGTATGTCAATACGCGAATTCGGACGA CTGGCATCTCTAGGAGGAGGATTCCTTCTTTTATGACATGTTTATTTTATATACATTGATGCTTTCCGACAGTCGGAAGT AATAAATGAATTTATTTCAAGACTACCTATACTCCTTTGACTTGTTCGACTAATCTTACCGCTTACTAAAATCTCGAAATC ACGCTTGACCTCTCGCACGCAAATTTTTGCTGCTGGACGCTACGCACTCGGCCCAATTCTTCTCGGTCCTCGTCGTCG CAATTGTCGTTGCGTTGATCTTGCACCGAAGGAATCAGAGAATAGAATACCATGAACCACCTGCGTGCCGAAGGACCC GCCTCCGTCCTCGCCATTGGAACCGCCAACCCCGAAAACATCCTCCTCCAGGACGAATTCCCCGACTACTACTTTCGT GTCACCAAGTCCGAACACATGACCCAGTTGAAGGAAAAGTTCCGTAAGATTTGTGACAAGTCCATGATTCGTAAGCG

	TAACTGCTTCCTCAACGAAGAACACCTCAAGCAGAACCCGCGCCTCGTCGAACACGAAATGCAAACCCTCGACGCCC GTCAGGACATGTTGGTTGTTGAAGTCCCCAAGCTCGGAAAGGACGCCTGCGCCAAGGCCATTAAGGAATGGGGTCAA CCTAAGTCCAAGATCACCCACCTGATCTTTACCTCCGCCTCCACCACCGATATGCCCGGAGCAGACTACCACTGCGCT AAGCTCCTCGGACTCTCCCCCTCCGTCAAGCGCGTCATGATGTACCAGCTCGGATGCTACGGAGGTGGAACCGTCCTC CGCATTGCTAAGGACATCGCCGAAAACAACAAGGGTGCCCGTGTCTTGGCCGTCTGCTGCGACATTATGGCTTGCCTT TTTCGTGGCCCGTCCGAATCCGACCTCGAACTCCTCGTTGGCCAGGCCATCTTCGGCGACGGTGCCGCCGCCGTCATT GTCGGCGCCGAACCCGACGAATCCGTCGGCGAACGTCCCATCTTCGAACTCGTCTCCACCGGACAGACCATCCTCCC CAACTCCGAAGGAACCATTGGAGGACACATCCGCGAAGCCGGTCTCATTTTCGACTTGCACAAGGACGTCCCCATGC TCATCTCGAACAACATTGAAAAGTGTTTGATTGAAGCCTTTACCCCGATTGGAATTTCCGACTGGAACTCCATCTTCTG GATTACCCATCCGGGAGGCAAGGCCATTCTTGATAAGGTCGAAGAAAAGCTCCACCTCAAGTCCGACAAGTTCGTCG ATTCCCGTCACGTCCTCTCCGAACACGGAAACATGTCCTCGTCCACGGTCCTCTTCGTCATGGACGAACTCCGTAAGC GTTCCTTGGAAGAAGGCAAATCGACCACCGGTGACGGATTTGAATGGGGTGTCTTGTTCGGTTTTGGACCGGGTTTG ACCGTAGAACGCGTTGTTGTTCGCTCCGTCCCGATTAAGTACCACCATCACCACCACCACTAACCGCAACAACTACCT CGACTTTGGCTGGGACACTTTCAGTGAGGACAAGAAGCTTCAGAAGCGTGCTATCGAACTCAACCAGGGACGTGCG GCACAAATGGGCATCCTTGCTCTCATGGTGCACGAACAGTTGGGAGTCTCTATCCTTCCTTAAAAATTTAATTTTCATTA GTTGCAGTCACTCCGCTTTGGTTT
OAC cassette	ACTAGCTTGATTGGGATATCTCGCTCATGTTTGTCGCGTGCTATGTCTTTTTAGGTACTTTGAACCTACGTTCGTACTTG TATAATATGATCATCGTATTATCGTTTTTCATCCGTCCAGCGCAAAATGCATTAGCAGCTAGTCCTAGCGTGCGGAGCTA CCTGgACAGGTGCATGACGGATGCGTGTCCTTCAGTGAcTTTCTAATTAACAGTAACTTCTTTACTTATGTTTCAGTTTG TAAGAAGCGGGATTCGCTCGTCGGTTGACATCTGATTGGACTGCGTCGGCACaTGAAAACTACATTGTGAAATCTGCT AAAACTCCGGGTATCTCTGACACAAAACGATTCGgCtTCgCAATTTCAACATTACGGTCAAGGCTAACGTATCTTTCTCG GTCAACTTCAGATTAtGCCGAtTAAATTGTCGTAGCTTTCAAGGCGTTTTGAGTACTGCGGCAGTTGTTGAACCTGCAA GGAGAAGATCTCGACAACAGAATAaAGCGAAAAATGGGTCTCATGCACTAACACTCAGgCCTCCCTCATAATCTCTGTT tGAGTTTACCAACAACACATATATACATTTCGACAAAATGGCCGTCAAGCACTTGATCGTTCTCAAGTTCAAGGACGAA ATCACCGAAGCCCAGAAGGAAGAATTCTTCAAGACCTACGTCAACCTCGTCAACATCATTCCCGCCATGAAGGATGTT TACTGGGGAAAGGACGTCACCCAGAAGAACAAGGAAGAAGGTTACACCCACATCGTCGAAGTCACCTTTGAATCCG TTGAAACCATCCAGGACTACATTATCCACCCGGCCCACGTCGGCTTCGGAGACGTCTACCGTTCCTTTTGGGAAAAGT TGTTGATTTTCGACTACACCCCCCGTAAGGAACAGAAGTTGATTTCCGAAGAAGACCTCTAATTTTGTTACATTTACTG ACTTCAAGGAGTCGAGGAATCGATACTGCCGTCGTTTCCAGGATCCGAGGTTTCATAAACTCTGTTAACGTTATAGAA ACAGACTTACCTCTCCTACGCCATTCACGTAATATTCGCAATATGCTATTCTTCCTCTGAAGACCAGGTTTATGTGCTGC CTGAAACTATTTCAATAAGTCAGCTGCACTTGCACAGGGTTTCACAAGGAAAGCGTGTCTTTTTTTCCAACGTAGGCG TCGCTTTCGTCTGACTCTTACTCTTACATTCACAGCCAATACTTACAATTAGTAAAAAACCTGTGCTCGAGAGTGAAAA CGTC
pTKS-1 sequencing	ATGGCCAAGTTGACCAGTGCCGTTCCGGTGCTCACCGCGCGCGACGTCGCCGGAGCGGTCGAGTTCTGGACCGACCG GCTCGGGTTCTCCCGGGACTTCGTGGAGGACGACTTCGCCGGTGTGGTCCGGGACGACGTGACCCTGTTCATCAGCG CGGTCCAGGACCAGGTGGTGCCGGACAACACCCTGGCCTGGGTGTGGGTGCGCGGCCTGGACGAGCTGTACGCCGA GTGGTCGGAGGTCGTGTCCACGAACTTCCGGGACGCCTCCGGGCCGGCCATGACCGAGATCGGCGAGCAGCCGTGG GGGCGGGAGTTCGCCCTGCGCGACCCGGCCGGCAACTGCGTGCACTTCGTGGCCGAGGAGCAGGACTGATTTTGTTA CATTGACTTCAAGGAGTCGAGGAATCGATACTGCCGTCGTTTCCAGGATCCGAGGTTTCTATAGACTCTCTATAGACTC TGTTAACCTAATAGAATCAGACATACCTCTCCTGCTATTTTGTTTTTATGAATTTGGCTTTTGCCTCTCTAGTCAGATTTG AATGTTATTTTCCGCCAGGTGTGTTAGTCGGGCTCTCGTTTGAGTTACAAGAGGGATTGAGTGGCGAGGATTCACTCTA ATGTAAATATGACTGTGAACAAAACTTTAAAATTACTACGCATCTTCTTTGACTGTCAGATATTCGTCGGTGACAGCAG TCAATGCCTGCAAATTGTCCTCCTGGGTCGCAATTTGGTTTTGGATTGACCTGGTATGCATTATGAAGAAAAAAATTCG TTATTAGCCAACTGCCTAGCGTGCACATTGCATGGTTAGACCTCCTTGACGACTGTGAGCCTACATCCTTCTGCAACAA GCTGCAATCCCTGCGATAGACCTTTTCCAAACTCACGCAGTCCAAGAAAACAAAGGGGTGAGAAGTATACGCACCTT TCGGTTTCGGCATAATTCTTAAACTCTTGTGGTCACTTTCTTGTGAAGAAGCTAGGGGCACTCGTTTTCCCTCAGAGCC TGCAAACACAAAATTCCTGCAGTCAATTGTCCCAACACTCGGCAAGCCGTATGCGCAAGCAACGATGCGCAGAAGGC CGTGGATGGATGGCGACTCGCGATATGGCTTCTTGGGTCGCCAGTGTGGTATGTCCGGCGTATGTCAATACGCGAATTC GGACGACTGGCATCTCTAGGAGGAGGATTCCTTCTTTTATGACATGTTTATTTTATATACATTGATGCTTTCCGACAGTC GGAAGTAATAAATGAATTTATTTCAAGACTACCTATACTCCTTTGACTTGTTCGACTAATCTTACCGCTTACTAAAATCT CGAAATCACGCTTGACCTCTCGCACGCAAATTTTTGCTGCTGGACGCTACGCACTCGGCCCAATTCTTCTCGGTCCTC GTCGTCGCAATTGTCGTTGCGTTGATCTTGCACCGAAGGAATCAGAGAAATAGAATACCATGAACCACCTGCGTGCCGA AGGACCCGCCTCCGTCCTCGCCATTGGAACCGCCAACCCCGAAAACATTCTCCTCCAAGACGAATTCCCTGACTACTA CTTCCGCGTGACTAAGAGTGAGCACATGACGCAATTGAAAGAAAAGTTCCGTAAGATCTGTGATAAGTCCATGATCCG TAAGCGTAACTGTTTCCTGAATGAGGAACACCTTAAGCAGAACCCCCGCTTGGTCGAACATGAAATGCAGACCTTGG ATGCCCGCCAGGACATGCTCGTCGTTGAAGTCCCTAAGTTGGGTAAAGACGCTTGCGCCAAGGCCATTAAGGAATGG GGACAGCCCAAGAGTAAGATCACCCATTTGATTTTCACATCCGCCTCTACGACCGATATGCCTGGAGCCGATTACCACT GCGCCAAGCTTCTGGGACTGTCCCCCTCCGTTAAGCGCGTCATGATGTACCAGCTTGGATGTTACGGAGGTGGCACCG TCCTCCGTATTGCCAAGGACATTGCCGAAAACAACAAGGGTGCCCGTGTCCTCGCCGTCTGCTGCGACATCATGGCCT GCCTCTTCCGCGGACCCTCCGAATCCGACTTGGAACTCCTCGTGGGACAGGCCATTTTTGGTGACGGCGCCGCCGCC GTCATTGTCGGAGCTGAACCCGACGAATCCGTCGGAGAAAGGCCCATTTTTGAATTGGTCTCCACCGGACAGACCATT CTCCCGAACTCCGAAGGAACCATCGGCGGACACATTCGTGAAGCCGGTCTCATCTTTGACTTGCACAAGGACGTTCC CATGCTCATTTCCAACAACATTGAAAAGTGCCTCATTGAAGCCTTCACCCCCATTGGAATCTCCGACTGGAATAGTATC TTTTGGATTACCCACCCCGGAGGAAAGGCCATCTTGGATAAGGTGGAAGAAAAGCTCCACCTCAAGTCCGACAAGTT CGTCGACTCCCGTCACGTCTTGTCCGAACACGGAAACATGTCGTCCTCCACCGTCCTCTTTGTCATGGATGAATTGCG TAAGCGCTCGCTCGAAGAAGGAAAGTCCACCACCGGCGATGGTTTCGAATGGGGCGTTCTCTTCGGATTTGGACCCG GTCTCACCGTCGAACGTGTCGTCGTCCGTTCTGTCCCGATTAAGTACCACCATCACCACCACCACTAAAAATTTAATTT TCATTAGTTGCAGTCACTCCGCTTTGGTTTCAGGGTAATATAGATCTTCCGCTGCATAACCCTGCTTCGGGGTCATTATA GCGATTTTTTCGGTATATCCATCCTTTTTCGCACGATATACAGGATTTTGCCAAAGGGTTCGTGTAGACTTTCCTTGGTG TATCCAACGGCGTCAGCCGGGCAGGATAGGTGAAGTAGGCCCACCCGCGAGCGGGTGTTCCTTCTTCACTGTCCCTTA TTCGCACCTGGCGGTGCTCAACGGGAATCCTGCTCTGCGAGGCTGGCCGGCTACCGCCGGCGTAACAGATGAGGGCA AGCGGATGGCTGATGAAACCAAGCCAACCAGGAAGGGCAGCCCACCTATCAAGGTGTACTGCCTTCCAGACGAACG AAGAGCGATTGAGGAAAAGGCGGCGGCGGCCGGCATGAGCCTGTCGGCCTACCTGCTGGCCGTCGGCCAGGGCTAC AAAATCACGGGCGTCGTGGACTATGAGCACGTCCGCGAGCTGGCCCGCATCAATGGCGACCTGGGCCGCCTGGGCGG

	CCTGCTGAAACTCTGGCTCACCGACGACCCGCGCACGGCGCGGTTCGGTGATGCCACGATCCTCGCCCTGCTGGCGA AGATCGAAGAGAAGCAGGACGAGCTTGGCAAGGTCATGATGGGCGTGGTCCGCCCGAGGGCAGAGCCATGACTTTT TTAGCCGCTAAAACGGCCGGGGGGTGCGCGTGATTGCCAAGCACGTCCCCATGCGCTCCATCAAGAAGAGGCACTTC GAGCTGTAAGTACATCACCGACGAGCAAGGCAAGACGATCCGCGCCTGTTTTATTGAGAACGTTGTTCGTGTTGGCCT CAATGGTAGCGATGCGTCATTCAGCGAAGTTCTGGTGCTGATGATGTGGTTCGCTTTGCCACTGGTCAATGTGGTAAG CCCGTGTAATGTCAGTAACCTTTTTACTGATCTCAGCTTGAGCACGGTCGCTGATGAGCTTATCCATGGCCCCACGGTA ACGGATATGATCCTCTAGGGCGTTGACAAATTCTTTGTCGGTTTTCATGGGGTAGACGTCAGTGACCAGGGAACGTCT CCCTACAAAAAGTTGCGCATACTTTGCTCCGTTGTCGACGGCAGGGGTGTCGGACCAGATTGTATCTGTGGCAACGGC CTCATTGCGTCAATGGACCTTTAAAGCCGGGAAACGAGACTTGAAATGTTTACGTAGAGGTGCATTATAGACCTCACG CGCATATTGAGTGGTTGCAAAAATGGTTTTGCGAACGGTATCACTGGAGACCCATGCCAGGCAAGGACGGAGGGCAT CATAATCATGTTCATTGCGTTGGACAGCATGCTTGTTACACGTAAGTATACGGTCGAGAGCATCGTGACGTTGAGAATT ACAGATGGCAACGTGTCGGTGCATGTATTGGCCGAGATTGTCAAATCGTGGCTCAATATATGGCATGCCAGGTAAGTCA TGGATGTCGGTATGCCATTCAGCCTTCATGTCGATTTCGTTGTCGAGAACTGATGGGTCCCAGTCAACGTCAGAAGTA AGAACAACATGGGGAAGGGAATGAAGTTCATGACTGGTGGGTGCGCGCATATCCATGTACGGAAGACCCTGTCGTAT GTTTAGTGGGATGATGTAATCATCCAGGGTAACAATACGTTGGTTGCCGCCCACGGTACGAGACCGATCTTGAACACA ATTGTGGCAATGTTCAAGTTGGGCGCTGGAATGGATGGTTTTACCTTTACCAAGATGCGCGTATTGATGCATGATAACA ACAATGGGCCCTCGTTGCGATTCGACAAGACCGGCAGCCGTGACAATGTCTAGATTGGACAAGGTATGTTCATTTATA CCGGTTATGTTGGCTGAACGTCCAGTTTTGTGCAGTACGGTAACATCACTCCCCGCAAGTCCACCGTTGGCACCGCGA TCGACAAGTGCAGATGTGGTATTTTGTACTGTGTGTCGGGAGACTTGGTACTGTAACACATTTGACTGAAGTGACTTG GTGGTACCATGAGGACGGAGGGAACCCGTTGATACGGTGTTGCGTGACGCTGCAAGTACCTTACGTATGTCGCCTGGA TCCATACGTCCAACCCGATCAGAAAGATGTGCTAGGAGTTCCGTTTCTGGTGCACAATCATGAAAGGTATCCACGGGA GTTCCATTGTTGTCGTCCGGGGCTGCGTCGCGACTGCCAGTTGCATGTACATGCGCCATCAAGGGACGTGCCAATCCA TTACTAACCGGGAGAACCTTAGCTGGGGCAGCCAATCCCTGAAGTATAGCCTTAGCATCATCTGAGAGCTGGTTCCAC ATATCTCTAGGGATATAAGGTCGTTCACGGTTAGTGGGCGTACTATGATGTCGGTTATTAGTACAATTCTCCCTGCGGGC ATGCGCATTGGCTTCATAGAGTACGGAAGGTGACAAGTCAATATTGTAGTTAACATCCGGATCAGTGTCAAAGTCTGTA GGATGGTAAGAAAGATCAGTAGAATGAATACTACGCTTTTCCTTGGGACTACGGGAATTAGAGAAGTTGTTTCCTTTAT TGTAGAGTGACGCTGAAGCAAGTAGAAGACTAAGGTAACTCTCGTAGCTAATAGGATTACCTCCTTTGGCTAGGTCAA GAGTGGCTGTGATCTTCACTTGACAAAGTTCCGGTACATTGTGGACAGCATTCTCCAAAAGACTAAGACACAGTTGCT TTGGGAGTTGCTCAGCCATTGGTACAGTATTGTGGTAGATACAAAGGTGGTTCTTCCAATGAAGGATAAATCCTTCTGC TGTGCCTGTCCATGAGGATCCATATTTCGCCGTAGTTAGGTAACCAAGCGTGGTGGCTGAACTGATCTTCGCACTTGCT GATTCCGTATAGTGTTTGACAACTTTACAAAACACTTCTTGCGCAGTTCGCTCTAGCCTGGACAGAGAAGGGACAATC TTCGTCGTTAAGACTCGTGCGAATAGCAAAAGATCACAAAATAGCACATCGGCACCGACCAACGATTATTTCCAAGGA AAAAAAGAATGCTTCACTACAAGAAATTGTGTCATCCCTATACAGAGTCTTGTTACTGTGACAGAAAATTGATGGAAG ATGTGGCGGATTGCCTTTACACTAGCCAACTTGTTCGACTAATTGCAGCTTCTTCTGAGAGGCTTCACCGAGTAACGC GAAGAACACCGGTGTCTCGTACATGCTCGTCGGTGAACGCTCGTCCAATGACACCCCCCACTTTGTATCAATATCCCA ACTTGGTAGTGAACTGGAATGATACATGCAATTTCGCGCCGCATCAACAGCCACGGGCACCATCGACGAATAGACTCG GTCGAGCTGGTTGCCCTCGCCGCTGGGCTGGCGGCCGTCTATGGCCCTGCAAACGCGCCAGAAACGCCGTCGAAGCC GTGTGCGAGACACCGCGGCCGGCCGCCGGCGTTGTGGATACCTCGCGGAAAACTTGGCCCTCACTGACAGATGAGG GGCGGACGTTGACACTTGAGGGGCCGACTCACCCGGCGCGGCGTTGACAGATGAGGGGCAGGCTCGATTTCGGCCG GCGACGTGGAGCTGGCCAGCCTCGCAAATCGGCGAAAACGCCTGATTTTACGCGAGTTTCCCACAGATGATGTGGAC AAGCCTGGGGATAAGTGCCCTGCGGTATTGACACTTGAGGGGCGCGACTACTGACAGATGAGGGGCGCGATCCTTGA CACTTGAGGGGCAGAGTGCTGACAGATGAGGGGCGCACCTATTGACATTTGAGGGGCTGTCCACAGGCAGAAAATCC AGCATTTGCAAGGGTTTCCGCCCGTTTTTCGGCCACCGCTAACCTGTCTTTTAACCTGCTTTTAAACCAATATTTATAAA CCTTGTTTTTAACCAGGGCTGCGCCCTGTGCGCGTGACCGCGCACGCCGAAGGGGGGTGCCCCCCCTTCTCGAACCC TCCCGGTCGAGTGAGCGAGGAAGCACCAGGGAACAGCACTTATATATTCTGCTTACACACGATGCCTGAAAAAACTTC CCTTGGGGTTATCCACTTATCCACGGGGATATTTTTATAATTATTTTTTTTATAGTTTTTAGATCTTCTTTTTTAGAGCGCC TTGTAGGCCTTTATCCATGCTGGTTCTAGAGAAGGTGTTGTGACAAATTGCCCTTTCAGTGTGACAAATCACCCTCAAA TGACAGTCCTGTCTGTGACAAATTGCCCTTAACCCTGTGACAAATTGCCCTCAGAAGAAGCTGTTTTTTCACAAAGTT ATCCCTGCTTATTGACTCTTTTTTATTTAGTGTGACAATCTAAAAACTTGTCACACTTCACATGGATCTGTCATGGCGGA AACAGCGGTTATCAATCACAAGAAACGTAAAAATAGCCCGCGAATCGTCCAGTCAAACGACCTCACTGAGGCGGCAT ATAGTCTCTCCCGGGATCAAAAACGTATGCTGTATCTGTTCGTTGACCAGATCAGAAAATCTGATGGCACCCTACAGGA ACATGACGGTATCTGCGAGATCCATGTTGCTAAATATGCTGAAATATTCGGATTGACCTCTGCGGAAGCCAGTAAGGAT ATACGGCAGGCATTGAAGAGTTTCGCGGGGAAGGAAGTGGTTTTTTATCGCCCTGAAGAGGATGCCGGCGATGAAAA AGGCTATGAATCTTTTCCTTGGTTTATCAAACGTGCGCACAGTCCATCCAGAGGGCTTTACAGTGTACATATCAACCCA TATCTCATTCCCTTCTTTATCGGGTTACAGAACCGGTTTACGCAGTTTCGGCTTAGTGAAACAAAAGAAATCACCAATC CGTATGCCATGCGTTTATACGAATCCCTGTGTCAGTATCGTAAGCCGGATGGCTCAGGCATCGTCTCTCTGAAAATCGA CTGGATCATAGAGCGTTACCAGCTGCCTCAAAGTTACCAGCGTATGCCTGACTTCCGCCGCCGCTTCCTGCAGGTCTG TGTTAATGAGATCAACAGCAGAACTCCAATGCGCCTCTCATACATTGAGAAAAAGAAAGGCCGCCAGACGACTCATAT CGTATTTTCCTTCCGCGATATCACTTCCATGACGACAGGATAGTCTGAGGGTTATCTGTCACAGATTTGAGGGTGGTTC GTCACATTTGTTCTGACCTACTGAGGGTAATTTGTCACAGTTTTGCTGTTTCCTTCAGCCTGCATGGATTTTCTCATACT TTTTGAACTGTAATTTTTAAGGAAGCCAAATTTGAGGGCAGTTTGTCACAGTTGATTTCCTTCTCTTTCCCTTCGTCATG TGACCTGATATCGGGGGTTAGTTCGTCATCATTGATGAGGGTTGATTATCACAGTTTATTACTCTGAATTGGCTATCCGC GTGTGTACCTCTACCTGGAGTTTTTCCCACGGTGGATATTTCTTCTTGCGCTGAGCGTAAGAGCTATCTGACAGAACAG TTCTTCTTTGCTTCCTCGCCAGTTCGCTCGCTATGCTCGGTTACACGGCTGCGGCGAGCATCACGTGCTATAAAAATAA TTATAATTTAAATTTTTTAATATAAATATATAAATTAAAAATAGAAAGTAAAAAAAGAAATTAAAGAAAAAATAGTTTTT GTTTTCCGAAGATGTAAAAGACTCTAGGGGGATCGCCAACAAATACTACCTTTTACCTTGCTCTTCCTGCTCTCAGGTA TTAATGCCGAATTGTTTCATCTTGTCTGTGTAGAAGACCACACACGAAAATCCTGTGATTTTACATTTTACTTATCGTTA ATCGAATGTATATCTATTTAATCTGCTTTTCTTGTCTAATAAATATATATGTAAAGTACGCTTTTTGTTGAAATTTTTTAAA CCTTTGTTTATTTTTTTTTCTTCATTCCGTAACTCTTCTACCTTCTTTATTTACTTTCTAAAATCCAAATACAAAACATAAA AATAAATAAACACAGAGTAAATTCCCAAATTATTCCATCATTAAAAGATACGAGGCGCGTGTAAGTTACAGGCAAGCG ATCCTAGTACACTCTATATTTTTTTATGCCTCGGTAATGATTTTCATTTTTTTTTTCCACCTAGCGGATGACTCTTTTTTTT TCTTAGCGATTGGCATTATCACATAATGAATTATACATTATATAAAGTAATGTGATTTCTTCGAAGAATATACTAAAAAAT GAGCAGGCAAGATAAACGAAGGCAAAGATGACAGAGCAGAAAGCCCTAGTAAAGCGTATTACAAATGAAACCAAGA TTCAGATTGCGATCTCTTTAAAGGGTGGTCCCCTAGCGATAGAGCACTCGATCTTCCCAGAAAAAGAGGCAGAAGCA

	GTAGCAGAACAGGCCACACAATCGCAAGTGATTAACGTCCACACAGGTATAGGGTTTCTGGACCATATGATACATGCT CTGGCCAAGCATTCCGGCTGGTCGCTAATCGTTGAGTGCATTGGTGACTTACACATAGACGACCATCACACCACTGAA GACTGCGGGATTGCTCTCGGTCAAGCTTTTAAAGAGGCCCTAGGGGCCGTGCGTGGAGTAAAAAGGTTTGGATCAGG ATTTGCGCCTTTGGATGAGGCACTTTCCAGAGCGGTGGTAGATCTTTCGAACAGGCCGTACGCAGTTGTCGAACTTGG TTTGCAAAGGGAGAAAGTAGGAGATCTCTCTTGCGAGATGATCCCGCATTTTCTTGAAAGCTTTGCAGAGGCTAGCAG AATTACCCTCCACGTTGATTGTCTGCGAGGCAAGAATGATCATCACCGTAGTGAGAGTGCGTTCAAGGCTCTTGCGGT TGCCATAAGAGAAGCCACCTCGCCCAATGGTACCAACGATGTTCCCTCCACCAAAGGTGTTCTTATGTAGTTTTACAC AGGAGTCTGGACTTGACGCTAGTGATAATAAGTGACTGAGGTATGTGCTCTTCTTATCTCCTTTTGTAGTGTTGCTCTTA TTTTAAACAACTTTGCGGTTTTTTGATGACTTTGCGATTTTGTTGTTGCTTTGCAGTAAATTGCAAGATTTAATAAAAAA ACGCAAAGCAATGATTAAAGGATGTTCAGAATGAAACTCATGGAAACACTTAACCAGTGCATAAACGCTGGTCATGA AATGACGAAGGCTATCGCCATTGCACAGTTTAATGATGACAGCCCGGAAGCGAGGAAAATAACCCGGCGCTGGAGAA TAGGTGAAGCAGCGGATTTAGTTGGGGTTTCTTCTCAGGCTATCAGAGATGCCGAGAAAGCAGGGCGACTACCGCAC CCGGATATGGAAATTCGAGGACGGGTTGAGCAACGTGTTGGTTATACAATTGAACAAATTAATCATATGCGTGATGTGT TTGGTACGCGATTGCGACGTGCTGAAGACGTATTTCCACCGGTGATCGGGGTTGCTGCCCATAAAGGTGGCGTTTACA AAACCTCAGTTTCTGTTCATCTTGCTCAGGATCTGGCTCTGAAGGGGCTACGTGTTTTGCTCGTGGAAGGTAACGACC CCCAGGGAACAGCCTCAATGTATCACGGATGGGTACCAGATCTTCATATTCATGCAGAAGACACTCTCCTGCCTTTCTA TCTTGGGGAAAAGGACGATGTCACTTATGCAATAAAGCCCACTTGCTGGCCGGGGCTTGACATTATTCCTTCCTGTCTG GCTCTGCACCGTATTGAAACTGAGTTAATGGGCAAATTTGATGAAGGTAAACTGCCCACCGATCCACACCTGATGCTC CGACTGGCCATTGAAACTGTTGCTCATGACTATGATGTCATAGTTATTGACAGCGCGCCTAACCTGGGTATCGGCACGA TTAATGTCGTATGTGCTGCTGATGTGCTGATTGTTCCCACGCCTGCTGAGTTGTTTGACTACACCTCCGCACTGCAGTT TTTCGATATGCTTCGTGATCTGCTCAAGAACGTTGATCTTAAAGGGTTCGAGCCTGATGTACGTATTTTGCTTACCAAAT ACAGCAATAGCAATGGCTCTCAGTCCCCGTGGATGGAGGAGCAAATTCGGGATGCCTGGGGAAGCATGGTTCTAAAA AATGTTGTACGTGAAACGGATGAAGTTGGTAAAGGTCAGATCCGGATGAGAACTGTTTTTGAACAGGCCATTGATCAA CGCTCTTCAACTGGTGCCTGGAGAAATGCTCTTTCTATTTGGGAACCTGTCTGCAATGAAATTTTCGATCGTCTGATTA AACCACGCTGGGAGATTAGATAATGAAGCGTGCGCCTGTTATTCCAAAACATACGCTCAATACTCAACCGGTTGAAGA TACTTCGTTATCGACACCAGCTGCCCCGATGGTGGATTCGTTAATTGCGCGCGTAGGAGTAATGGCTCGCGGTAATGCC ATTACTTTGCCTGTATGTGGTCGGGATGTGAAGTTTACTCTTGAAGTGCTCCGGGGTGATAGTGTTGAGAAGACCTCTC GGGTATGGTCAGGTAATGAACGTGACCAGGAGCTGCTTACTGAGGACGCACTGGATGATCTCATCCCTTCTTTTCTACT GACTGGTCAACAGACACCGGCGTTCGGTCGAAGAGTATCTGGTGTCATAGAAATTGCCGATGGGAGTCGCCGTCGTA AAGCTGCTGCACTTACCGAAAGTGATTATCGTGTTCTGGTTGGCGAGCTGGATGATGAGCAGATGGCTGCATTATCCA GATTGGGTAACGATTATCGCCCAACAAGTGCTTATGAACGTGGTCAGCGTTATGCAAGCCGATTGCAGAATGAATTTGC TGGAAATATTTCTGCGCTGGCTGATGCGGAAAATATTTCACGTAAGATTATTACCCGCTGTATCAACACCGCCAAATTG CCTAAATCAGTTGTTGCTCTTTTTTCTCACCCCGGTGAACTATCTGCCCGGTCAGGTGATGCACTTCAAAAAGCCTTTA CAGATAAAGAGGAATTACTTAAGCAGCAGGCATCTAACCTTCATGAGCAGAAAAAAGCTGGGGTGATATTTGAAGCT GAAGAAGTTATCACTCTTTTAACTTCTGTGCTTAAAACGTCATCTGCATCAAGAACTAGTTTAAGCTCACGACATCAGT TTGCTCCTGGAGCGACAGTATTGTATAAGGGCGATAAAATGGTGCTTAACCTGGACAGGTCTCGTGTTCCAACTGAGT GTATAGAGAAAATTGAGGCCATTCTTAAGGAACTTGAAAAGCCAGCACCCTGATGCGACCTCGTTTTAGTCTACGTTT ATCTGTCTTTACTTAATGTCCTTTGTTACAGGCCAGAAAGCATAACTGGCCTGAATATTCTCTCTGGGCCCACTGTTCCA CTTGTATCGTCGGTCTGATAATCAGACTGGGACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCAC GGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCACGGTCCCACTCGTATCGTCGGTCTGATAATCAGACTG GGACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCATGGTCCCACTCGTATCGTCGGTCTGATTATT AGTCTGGGACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGAACCACGGTCCCACTCGTATCGTCGGTCT GATTATTAGTCTGGGACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCACGATCCCACTCGTGTTG TCGGTCTGATTATCGGTCTGGGACCACGGTCCCACTTGTATTGTCGATCAGACTATCAGCGTGAGACTACGATTCCATC AATGCCTGTCAAGGGCAAGTATTGACATGTCGTCGTAACCTGTAGAACGGAGTAACCTCGGTGTGCGGTTGTATGCCT GCTGTGGATTGCTGCTGTGTCCTGCTTATCCACAACATTTTGCGCACGGTTATGTGGACAAAATACCTGGTTACCCAGG CCGTGCCGGCACGTTAACCGGGCTGCATCCGATGCAAGTGTGTCGCTGTCGACGAGCTCGCGAGCTCGGACATGAGG TTGCCCCGTATTCAGTGTCGCTGATTTGTATTGTCTGAAGTTGTTTTTACGTTAAGTTGATGTAGATCAATTAATACGATA CCTGCGTCATAATTGATTATTTGACGTGGTTTGATGGCCTCCACGCACGTTGTGATATGTAGATGATAATCATTATCACTT TACGGGTCCTTTCCGGTGATCCGACAGGTTACGGGGCGGCGACCTCGCGGGTTTTCGCTATTTATGAAAATTTTCCGGT TTAAGGCGTTTCCGTTCTTCTTCGTCATAACTTAATGTTTTTATTTAAAATACCCTCTGAAAAGAAAGGAAACGACAGG TGCTGAAAGCGAGCTTTTTGGCCTCTGTCGTTTCCTTTCTCTGTTTTTGTCCGTGGAATGAACAATGGAAGTCCGAGCT CATCGCTAATAACTTCGTATAGCATACATTATACGAAGTTATATTCGATGCGGCCGCAAGGGGTTCGCGTCAGCGGGTGT TGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAAT ACCACACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCCATTCGCCATTCAGCTGCGCAACTGTTGGGAAGGGCG ATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCC AGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTGTAATACGACTCACTATAGGGCGAATTCGAGCT CGGTACCCGGGGATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTTGAGTATTCTATAGTCTCACCTAAATAGCTTGGC GTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATA AAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCG GGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGAACCCCTTGCGGCCGCCCGGGCCGTCGACCAATT CTCATGTTTGACAGCTTATCATCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACTGTTGTAATTCATTAAGCA TTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCTGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCG TATAATATTTGCCCATGGTGAAAACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAAC TCACCCAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGTTTTCACCGTAAC ACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGTGGTattcactccagagcgatgaaaacgtttcagtttgctcatggaaaa cggtgtaACAAGGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGAAATTCCGGATGAGCATTCAT CAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTTTCTTTACGGTCTTTAAAAAGGCCGTAAT ATCCAGCTGAACGGTCTGGTTATAGGTACATTGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGG GATATATCAACGGTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGATAACTCAAAA AATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTTACGTGCCGATCAACGTCTCATTTTCGCCA AAAGTTGGCCCAGGGCTTCCCGGTATCAACAGGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGG TATTTATTCGCGATAAGCTCATGGAGCGGCGTAACCGTCGCACAGGAAGGACAGAGAAAGCGCGGATCTGGGAAGTG

	ACGGACAGAACGGTCAGGACCTGGATTGGGGAGGCGGTTGCCGCCGCTGCTGCTGACGGTGTGACGTTCTCTGTTCC GGTCACACCACATACGTTCCGCCATTCCTATGCGATGCACATGCTGTATGCCGGTATACCGCTGAAAGTTCTGCAAAGC CTGATGGGACATAAGTCCATCAGTTCAACGGAAGTCTACACGAAGGTTTTTGCGCTGGATGTGGCTGCGTTAAGCCTA CTTATTTAGAACGAGAAGTCAGACCGAGGGGTACTAAAATTCTAAGGGTTGAGAGGTATCTTGATTCCGGGTCTATGG AAGCCCATCCTTGTTGAAGCTTGAACACGATCCTTGTGAAAGGCCGACGTTGCGCGAAAAAACAGCCTGCCGATTTC TTTCCTTCTTTCTCGTCTCAACCTATATACTTTCATAATCTCTGTTAGAGTTTACCAACAACACATATATACATTTCGACA AA
pTKS-2 sequencing	GAGCACAAGAGGTGACAAAAGCCACCGGCTGGATCGCACTTCTCGGAATTTCCCCCCTACTATCAAACAAATTCGAAT TGCCAAAGGTGAAGGGACTAACTGTAAATCCTGATCAATCAAGGTCTCAATCAAGTACAATGGGCTACAATGATATTTA GATGGGAACACAATGAAACGAATTGAAACTTCTACTGACAGGAGCGCAATTGACTTGTGTAGCTTTTCATGAGCACTT GATTGCTACCGATTGTGAACGGGATGGGGAAAGACTCGAAAAGGTGCATGCTTCCGATAATCTACTATATTTTCTAGAA TCAAATAATATTTAAATGAATGAGGTCCTCAGCGTACGTTAAGCCTACTTATTTAGAACGAGAAGTCAGACCGAGGGGT ACTAAAATTCTAAGGGTTGAGAGGTATCTTGATTCCGGGTCTATGGAAGCCCATCCTTGTTGAAGCTTGAACACGATCC TTGTGAAAGGCCGACGTTGCGCGAAAAAACAGCCTGCCGATTTCTTTCCTTCTTTCTCGTCTCAACCTATATACTTTCA TAATCTCTGTTAGAGTTTACCAACAACACATATATACATTTCGACAAAATGGCCAAGTTGACCAGTGCCGTTCCGGTGC TCACCGCGCGCGACGTCGCCGGAGCGGTCGAGTTCTGGACCGACCGGCTCGGGTTCTCCCGGGACTTCGTGGAGGA CGACTTCGCCGGTGTGGTCCGGGACGACGTGACCCTGTTCATCAGCGCGGTCCAGGACCAGGTGGTGCCGGACAAC ACCCTGGCCTGGGTGTGGGTGCGCGGCCTGGACGAGCTGTACGCCGAGTGGTCGGAGGTCGTGTCCACGAACTTCCG GGACGCCTCCGGGCCGGCCATGACCGAGATCGGCGAGCAGCCGTGGGGGCGGGAGTTCGCCCTGCGCGACCCGGCC GGCAACTGCGTGCACTTCGTGGCCGAGGAGCAGGACTGATTTTGTTACATTGACTTCAAGGAGTCGAGGAATCGATA CTGCCGTCGTTTCCAGGATCCGAGGTTTCTATAGACTCTCTATAGACTCTGTTAACCTAATAGAATCAGACATACCTCTC CTGCTATTTTGTTTTTATGAATTTGGCTTTTGCCTCTCTAGTCAGATTTGAATGTTATTTTCCGCCAGGTGTGTTAGTCGG GCTCTCGTTTGAGTTACAAGAGGGATTGAGTGGCGAGGATTCACTCTAATGTAAATATGACTGTGAACAAAACTTTAA AATTACTACGCATCTTCTTTGACTGTCAGATATTCGTCGGTGACAGCAGTCAATGCCTGCAAATTGTCCTCCTGGGTCG CAATTTGGTTTTGGATTGACCTGGTATGCATTATGAAGAAAAAAATTCGTTATTAGCCAACTGCCTAGCGTGCACATTG CATGGTTAGACCTCCTTGACGACTGTGAGCCTACATCCTTCTGCAACAAGCTGCAATCCCTGCGATAGACCTTTTCCAA ACTCACGCAGTCCAAGAAAACAAAGGGGTGAGAAGTATACGCACCTTTCGGTTTCGGCATAATTCTTAAACTCTTGTG GTCACTTTCTTGTGAAGAAGCTAGGGGCACTCGTTTTCCCTCAGAGCCTGCAAACACAAAATTCCTGCAGTCAATTGT CCCAACACTCGGCAAGCCGTATGCGCAAGCAACGATGCGCAGAAGGCCGTGGATGGATGGCGACTCGCGATATGGCT TCTTGGGTCGCCAGTGTGGTATGTCCGGCGTATGTCAATACGCGAATTCGGACGACTGGCATCTCTAGGAGGAGGATT CCTTCTTTTATGACATGTTTATTTTATATACATTGATGCTTTCCGACAGTCGGAAGTAATAAATGAATTTATTTCAAGACT ACCTATACTCCTTTGACTTGTTCGACTAATCTTACCGCTTACTAAAATCTCGAAATCACGCTTGACCTCTCGCACGCAA ATTTTTGCTGCTGGACGCTACGCACTCGGCCCAATTCTTCTCGGTCCTCGTCGTCGCAATTGTCGTTGCGTTGATCTTG CACCGAAGGAATCAGAGAATAGAATACCATGAACCACCTGCGTGCCGAAGGACCCGCCTCCGTCCTCGCCATTGGAA CCGCCAACCCCGAAAACATTCTCCTCCAAGACGAATTCCCTGACTACTACTTCCGCGTGACTAAGAGTGAGCACATGA CGCAATTGAAAGAAAAGTTCCGTAAGATCTGTGATAAGTCCATGATCCGTAAGCGTAACTGTTTCCTGAATGAGGAAC ACCTTAAGCAGAACCCCCGCTTGGTCGAACATGAAATGCAGACCTTGGATGCCCGCCAGGACATGCTCGTCGTTGAA GTCCCTAAGTTGGGTAAAGACGCTTGCGCCAAGGCCATTAAGGAATGGGGACAGCCCAAGAGTAAGATCACCCATTT GATTTTCACATCCGCCTCTACGACCGATATGCCTGGAGCCGATTACCACTGCGCCAAGCTTCTGGGACTGTCCCCCTCC GTTAAGCGCGTCATGATGTACCAGCTTGGATGTTACGGAGGTGGCACCGTCCTCCGTATTGCCAAGGACATTGCCGAA AACAACAAGGGTGCCCGTGTCCTCGCCGTCTGCTGCGACATCATGGCCTGCCTCTTCCGCGGACCCTCCGAATCCGAC TTGGAACTCCTCGTGGGACAGGCCATTTTTGGTGACGGCGCCGCCGCCGTCATTGTCGGAGCTGAACCCGACGAATC CGTCGGAGAAAGGCCCATTTTTGAATTGGTCTCCACCGGACAGACCATTCTCCCGAACTCCGAAGGAACCATCGGCG GACACATTCGTGAAGCCGGTCTCATCTTTGACTTGCACAAGGACGTTCCCATGCTCATTTCCAACAACATTGAAAAGT GCCTCATTGAAGCCTTCACCCCCATTGGAATCTCCGACTGGAATAGTATCTTTTGGATTACCCACCCCGGAGGAAAGG CCATCTTGGATAAGGTGGAAGAAAAGCTCCACCTCAAGTCCGACAAGTTCGTCGACTCCCGTCACGTCTTGTCCGAA CACGGAAACATGTCGTCCTCCACCGTCCTCTTTGTCATGGATGAATTGCGTAAGCGCTCGCTCGAAGAAGGAAAGTCC ACCACCGGCGATGGTTTCGAATGGGGCGTTCTCTTCGGATTTGGACCCGGTCTCACCGTCGAACGTGTCGTCGTCCGT TCTGTCCCGATTAAGTACCACCATCACCACCACCACTAAAAATTTAATTTTCATTAGTTGCAGTCACTCCGCTTTGGTTT CAGGGTAATATAGATCTTCCGCTGCATAACCCTGCTTCGGGGTCATTATAGCGATTTTTTCGGTATATCCATCCTTTTTCG CACGATATACAGGATTTTGCCAAAGGGTTCGTGTAGACTTTCCTTGGTGTATCCAACGGCGTCAGCCGGGCAGGATAG GTGAAGTAGGCCCACCCGCGAGCGGGTGTTCCTTCTTCACTGTCCCTTATTCGCACCTGGCGGTGCTCAACGGGAATC CTGCTCTGCGAGGCTGGCCGGCTACCGCCGGCGTAACAGATGAGGGCAAGCGGATGGCTGATGAAACCAAGCCAAC CAGGAAGGGCAGCCCACCTATCAAGGTGTACTGCCTTCCAGACGAACGAAGAGCGATTGAGGAAAAGGCGGCGGCG GCCGGCATGAGCCTGTCGGCCTACCTGCTGGCCGTCGGCCAGGGCTACAAAATCACGGGCGTCGTGGACTATGAGCA CGTCCGCGAGCTGGCCCGCATCAATGGCGACCTGGGCCGCCTGGGCGGCCTGCTGAAACTCTGGCTCACCGACGACC CGCGCACGGCGCGGTTCGGTGATGCCACGATCCTCGCCCTGCTGGCGAAGATCGAAGAGAAGCAGGACGAGCTTGG CAAGGTCATGATGGGCGTGGTCCGCCCGAGGGCAGAGCCATGACTTTTTTAGCCGCTAAAACGGCCGGGGGGTGCGC GTGATTGCCAAGCACGTCCCCATGCGCTCCATCAAGAAGAGGCACTTCGAGCTGTAAGTACATCACCGACGAGCAAG GCAAGACGATCCGCGCCTGTTTTATTGAGAACGTTGTTCGTGTTGGCCTCAATGGTAGCGATGCGTCATTCAGCGAAG TTCTGGTGCTGATGATGTGGTTCGCTTTGCCACTGGTCAATGTGGTAAGCCCGTGTAATGTCAGTAACCTTTTTACTGA TCTCAGCTTGAGCACGGTCGCTGATGAGCTTATCCATGGCCCCACGGTAACGGATATGATCCTCTAGGGCGTTGACAA ATTCTTTGTCGGTTTTCATGGGGTAGACGTCAGTGACCAGGGAACGTCTCCCTACAAAAAGTTGCGCATACTTTGCTC CGTTGTCGACGGCAGGGGTGTCGGACCAGATTGTATCTGTGGCAACGGCCTCATTGCGTCAATGGACCTTTAAAGCCG GGAAACGAGACTTGAAATGTTTACGTAGAGGTGCATTATAGACCTCACGCGCATATTGAGTGGTTGCAAAAATGGTTT TGCGAACGGTATCACTGGAGACCCATGCCAGGCAAGGACGGAGGGCATCATAATCATGTTCATTGCGTTGGACAGCAT GCTTGTTACACGTAAGTATACGGTCGAGAGCATCGTGACGTTGAGAATTACAGATGGCAACGTGTCGGTGCATGTATT GGCCGAGATTGTCAAATCGTGGCTCAATATATGGCATGCCAGGTAAGTCATGGATGTCGGTATGCCATTCAGCCTTCAT GTCGATTTCGTTGTCGAGAACTGATGGGTCCCAGTCAACGTCAGAAGTAAGAACAACATGGGGAAGGGAATGAAGTT CATGACTGGTGGGTGCGCGCATATCCATGTACGGAAGACCCTGTCGTATGTTTAGTGGGATGATGTAATCATCCAGGGT AACAATACGTTGGTTGCCGCCCACGGTACGAGACCGATCTTGAACACAATTGTGGCAATGTTCAAGTTGGGCGCTGG AATGGATGGTTTTACCTTTACCAAGATGCGCGTATTGATGCATGATAACAACAATGGGCCCTCGTTGCGATTCGACAAG ACCGGCAGCCGTGACAATGTCTAGATTGGACAAGGTATGTTCATTTATACCGGTTATGTTGGCTGAACGTCCAGTTTTG TGCAGTACGGTAACATCACTCCCCGCAAGTCCACCGTTGGCACCGCGATCGACAAGTGCAGATGTGGTATTTTGTACT

	GTGTGTCGGGAGACTTGGTACTGTAACACATTTGACTGAAGTGACTTGGTGGTACCATGAGGACGGAGGGAACCCGT TGATACGGTGTTGCGTGACGCTGCAAGTACCTTACGTATGTCGCCTGGATCCATACGTCCAACCCGATCAGAAAGATGT GCTAGGAGTTCCGTTTCTGGTGCACAATCATGAAAGGTATCCACGGGAGTTCCATTGTTGTCGTCCGGGGCTGCGTCG CGACTGCCAGTTGCATGTACATGCGCCATCAAGGGACGTGCCAATCCATTACTAACCGGGAGAACCTTAGCTGGGGCA GCCAATCCCTGAAGTATAGCCTTAGCATCATCTGAGAGCTGGTTCCACATATCTCTAGGGATATAAGGTCGTTCACGGT TAGTGGGCGTACTATGATGTCGGTTATTAGTACAATTCTCCCTGCGGGCATGCGCATTGGCTTCATAGAGTACGGAAGG TGACAAGTCAATATTGTAGTTAACATCCGGATCAGTGTCAAAGTCTGTAGGATGGTAAGAAAGATCAGTAGAATGAAT ACTACGCTTTTCCTTGGGACTACGGGAATTAGAGAAGTTGTTTCCTTTATTGTAGAGTGACGCTGAAGCAAGTAGAAG ACTAAGGTAACTCTCGTAGCTAATAGGATTACCTCCTTTGGCTAGGTCAAGAGTGGCTGTGATCTTCACTTGACAAAGT TCCGGTACATTGTGGACAGCATTCTCCAAAAGACTAAGACACAGTTGCTTTGGGAGTTGCTCAGCCATTGGTACAGTA TTGTGGTAGATACAAAGGTGGTTCTTCCAATGAAGGATAAATCCTTCTGCTGTGCCTGTCCATGAGGATCCATATTTCG CCGTAGTTAGGTAACCAAGCGTGGTGGCTGAACTGATCTTCGCACTTGCTGATTCCGTATAGTGTTTGACAACTTTACA AAACACTTCTTGCGCAGTTCGCTCTAGCCTGGACAGAGAAGGGACAATCTTCGTCGTTAAGACTCGTGCGAATAGCA AAAGATCACAAAATAGCACATCGGCACCGACCAACGATTATTTCCAAGGAAAAAAAGAATGCTTCACTACAAGAAAT TGTGTCATCCCTATACAGAGTCTTGTTACTGTGACAGAAAATTGATGGAAGATGTGGCGGATTGCCTTTACACTAGCCA ACTTGTTCGACTAATTGCAGCTTCTTCTGAGAGGCTTCACCGAGTAACGCGAAGAACACCGGTGTCTCGTACATGCTC GTCGGTGAACGCTCGTCCAATGACACCCCCCACTTTGTATCAATATCCCAACTTGGTAGTGAACTGGAATGATACATGC AATTTCGCGCCGCATCAACAGCCACGGGCACCATCGACGAATAGACTCGGTCGAGCTGGTTGCCCTCGCCGCTGGGC TGGCGGCCGTCTATGGCCCTGCAAACGCGCCAGAAACGCCGTCGAAGCCGTGTGCGAGACACCGCGGCCGGCCGCC GGCGTTGTGGATACCTCGCGGAAAACTTGGCCCTCACTGACAGATGAGGGGCGGACGTTGACACTTGAGGGGCCGA CTCACCCGGCGCGGCGTTGACAGATGAGGGGCAGGCTCGATTTCGGCCGGCGACGTGGAGCTGGCCAGCCTCGCAA ATCGGCGAAAACGCCTGATTTTACGCGAGTTTCCCACAGATGATGTGGACAAGCCTGGGGATAAGTGCCCTGCGGTAT TGACACTTGAGGGGCGCGACTACTGACAGATGAGGGGCGCGATCCTTGACACTTGAGGGGCAGAGTGCTGACAGAT GAGGGGCGCACCTATTGACATTTGAGGGGCTGTCCACAGGCAGAAAATCCAGCATTTGCAAGGGTTTCCGCCCGTTTT TCGGCCACCGCTAACCTGTCTTTTAACCTGCTTTTAAACCAATATTTATAAACCTTGTTTTTAACCAGGGCTGCGCCCTG TGCGCGTGACCGCGCACGCCGAAGGGGGGGTGCCCCCCCCTTCTCGAACCCTCCCGGTCGAGTGAGCGAGGAAGCACC AGGGAACAGCACTTATATATTCTGCTTACACACGATGCCTGAAAAAACTTCCCTTGGGGTTATCCACTTATCCACGGGG ATATTTTTATAATTATTTTTTTTATAGTTTTTAGATCTTCTTTTTTAGAGCGCCTTGTAGGCCTTTATCCATGCTGGTTCTAG AGAAGGTGTTGTGACAAATTGCCCTTTCAGTGTGACAAATCACCCTCAAATGACAGTCCTGTCTGTGACAAATTGCCC TTAACCCTGTGACAAATTGCCCTCAGAAGAAGCTGTTTTTTCACAAAGTTATCCCTGCTTATTGACTCTTTTTTATTTAG TGTGACAATCTAAAAACTTGTCACACTTCACATGGATCTGTCATGGCGGAAACAGCGGTTATCAATCACAAGAAACGT AAAAATAGCCCGCGAATCGTCCAGTCAAACGACCTCACTGAGGCGGCATATAGTCTCTCCCGGGATCAAAAACGTATG CTGTATCTGTTCGTTGACCAGATCAGAAAATCTGATGGCACCCTACAGGAACATGACGGTATCTGCGAGATCCATGTTG CTAAATATGCTGAAATATTCGGATTGACCTCTGCGGAAGCCAGTAAGGATATACGGCAGGCATTGAAGAGTTTCGCGG GGAAGGAAGTGGTTTTTTATCGCCCTGAAGAGGATGCCGGCGATGAAAAAGGCTATGAATCTTTTCCTTGGTTTATCA AACGTGCGCACAGTCCATCCAGAGGGCTTTACAGTGTACATATCAACCCATATCTCATTCCCTTCTTTATCGGGTTACAG AACCGGTTTACGCAGTTTCGGCTTAGTGAAACAAAAGAAATCACCAATCCGTATGCCATGCGTTTATACGAATCCCTGT GTCAGTATCGTAAGCCGGATGGCTCAGGCATCGTCTCTCTGAAAATCGACTGGATCATAGAGCGTTACCAGCTGCCTC AAAGTTACCAGCGTATGCCTGACTTCCGCCGCCGCTTCCTGCAGGTCTGTGTTAATGAGATCAACAGCAGAACTCCAA TGCGCCTCTCATACATTGAGAAAAAGAAAGGCCGCCAGACGACTCATATCGTATTTTCCTTCCGCGATATCACTTCCAT GACGACAGGATAGTCTGAGGGTTATCTGTCACAGATTTGAGGGTGGTTCGTCACATTTGTTCTGACCTACTGAGGGTA ATTTGTCACAGTTTTGCTGTTTCCTTCAGCCTGCATGGATTTTCTCATACTTTTTGAACTGTAATTTTTAAGGAAGCCAA ATTTGAGGGCAGTTTGTCACAGTTGATTTCCTTCTCTTTCCCTTCGTCATGTGACCTGATATCGGGGGTTAGTTCGTCAT CATTGATGAGGGTTGATTATCACAGTTTATTACTCTGAATTGGCTATCCGCGTGTGTACCTCTACCTGGAGTTTTTCCCA CGGTGGATATTTCTTCTTGCGCTGAGCGTAAGAGCTATCTGACAGAACAGTTCTTCTTTGCTTCCTCGCCAGTTCGCTC GCTATGCTCGGTTACACGGCTGCGGCGAGCATCACGTGCTATAAAAATAATTATAATTTAAATTTTTTAATATAAATATAT AAATTAAAAATAGAAAGTAAAAAAAGAAATTAAAGAAAAAATAGTTTTTGTTTTCCGAAGATGTAAAAGACTCTAGG GGGATCGCCAACAAATACTACCTTTTACCTTGCTCTTCCTGCTCTCAGGTATTAATGCCGAATTGTTTCATCTTGTCTGT GTAGAAGACCACACACGAAAATCCTGTGATTTTACATTTTACTTATCGTTAATCGAATGTATATCTATTTAATCTGCTTTT CTTGTCTAATAAATATATATGTAAAGTACGCTTTTTGTTGAAATTTTTTAAACCTTTGTTTATTTTTTTTTCTTCATTCCGT AACTCTTCTACCTTCTTTATTTACTTTCTAAAATCCAAATACAAAACATAAAAATAAATAAACACAGAGTAAATTCCCAA ATTATTCCATCATTAAAAGATACGAGGCGCGTGTAAGTTACAGGCAAGCGATCCTAGTACACTCTATATTTTTTTATGCC TCGGTAATGATTTTCATTTTTTTTTTCCACCTAGCGGATGACTCTTTTTTTTTCTTAGCGATTGGCATTATCACATAATGA ATTATACATTATATAAAGTAATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG ATGACAGAGCAGAAAGCCCTAGTAAAGCGTATTACAAATGAAACCAAGATTCAGATTGCGATCTCTTTAAAGGGTGGT CCCCTAGCGATAGAGCACTCGATCTTCCCAGAAAAAGAGGCAGAAGCAGTAGCAGAACAGGCCACACAATCGCAAG TGATTAACGTCCACACAGGTATAGGGTTTCTGGACCATATGATACATGCTCTGGCCAAGCATTCCGGCTGGTCGCTAAT CGTTGAGTGCATTGGTGACTTACACATAGACGACCATCACACCACTGAAGACTGCGGGATTGCTCTCGGTCAAGCTTT TAAAGAGGCCCTAGGGGCCGTGCGTGGAGTAAAAAGGTTTGGATCAGGATTTGCGCCTTTGGATGAGGCACTTTCCA GAGCGGTGGTAGATCTTTCGAACAGGCCGTACGCAGTTGTCGAACTTGGTTTGCAAAGGGAGAAAGTAGGAGATCTC TCTTGCGAGATGATCCCGCATTTTCTTGAAAGCTTTGCAGAGGCTAGCAGAATTACCCTCCACGTTGATTGTCTGCGAG GCAAGAATGATCATCACCGTAGTGAGAGTGCGTTCAAGGCTCTTGCGGTTGCCATAAGAGAAGCCACCTCGCCCAAT GGTACCAACGATGTTCCCTCCACCAAAGGTGTTCTTATGTAGTTTTACACAGGAGTCTGGACTTGACGCTAGTGATAAT AAGTGACTGAGGTATGTGCTCTTCTTATCTCCTTTTGTAGTGTTGCTCTTATTTTAAACAACTTTGCGGTTTTTTGATGA CTTTGCGATTTTGTTGTTGCTTTGCAGTAAATTGCAAGATTTAATAAAAAAACGCAAAGCAATGATTAAAGGATGTTCA GAATGAAACTCATGGAAACACTTAACCAGTGCATAAACGCTGGTCATGAAATGACGAAGGCTATCGCCATTGCACAGT TTAATGATGACAGCCCGGAAGCGAGGAAAATAACCCGGCGCTGGAGAATAGGTGAAGCAGCGGATTTAGTTGGGGTT TCTTCTCAGGCTATCAGAGATGCCGAGAAAGCAGGGCGACTACCGCACCCGGATATGGAAATTCGAGGACGGGTTGA GCAACGTGTTGGTTATACAATTGAACAAATTAATCATATGCGTGATGTGTTTGGTACGCGATTGCGACGTGCTGAAGAC GTATTTCCACCGGTGATCGGGGTTGCTGCCCATAAAGGTGGCGTTTACAAAACCTCAGTTTCTGTTCATCTTGCTCAGG ATCTGGCTCTGAAGGGGCTACGTGTTTTGCTCGTGGAAGGTAACGACCCCCAGGGAACAGCCTCAATGTATCACGGAT GGGTACCAGATCTTCATATTCATGCAGAAGACACTCTCCTGCCTTTCTATCTTGGGGAAAAGGACGATGTCACTTATGC AATAAAGCCCACTTGCTGGCCGGGGCTTGACATTATTCCTTCCTGTCTGGCTCTGCACCGTATTGAAACTGAGTTAATG GGCAAATTTGATGAAGGTAAACTGCCCACCGATCCACACCTGATGCTCCGACTGGCCATTGAAACTGTTGCTCATGAC

	TATGATGTCATAGTTATTGACAGCGCGCCTAACCTGGGTATCGGCACGATTAATGTCGTATGTGCTGCTGATGTGCTGAT TGTTCCCACGCCTGCTGAGTTGTTTGACTACACCTCCGCACTGCAGTTTTTCGATATGCTTCGTGATCTGCTCAAGAAC GTTGATCTTAAAGGGTTCGAGCCTGATGTACGTATTTTGCTTACCAAATACAGCAATAGCAATGGCTCTCAGTCCCCGT GGATGGAGGAGCAAATTCGGGATGCCTGGGGAAGCATGGTTCTAAAAAATGTTGTACGTGAAACGGATGAAGTTGGT AAAGGTCAGATCCGGATGAGAACTGTTTTTGAACAGGCCATTGATCAACGCTCTTCAACTGGTGCCTGGAGAAATGCT CTTTCTATTTGGGAACCTGTCTGCAATGAAATTTTCGATCGTCTGATTAAACCACGCTGGGAGATTAGATAATGAAGCG TGCGCCTGTTATTCCAAAACATACGCTCAATACTCAACCGGTTGAAGATACTTCGTTATCGACACCAGCTGCCCCGATG GTGGATTCGTTAATTGCGCGCGTAGGAGTAATGGCTCGCGGTAATGCCATTACTTTGCCTGTATGTGGTCGGGATGTGA AGTTTACTCTTGAAGTGCTCCGGGGTGATAGTGTTGAGAAGACCTCTCGGGTATGGTCAGGTAATGAACGTGACCAGG AGCTGCTTACTGAGGACGCACTGGATGATCTCATCCCTTCTTTTCTACTGACTGGTCAACAGACACCGGCGTTCGGTC GAAGAGTATCTGGTGTCATAGAAATTGCCGATGGGAGTCGCCGTCGTAAAGCTGCTGCACTTACCGAAAGTGATTATC GTGTTCTGGTTGGCGAGCTGGATGATGAGCAGATGGCTGCATTATCCAGATTGGGTAACGATTATCGCCCAACAAGTG CTTATGAACGTGGTCAGCGTTATGCAAGCCGATTGCAGAATGAATTTGCTGGAAATATTTCTGCGCTGGCTGATGCGGA AAATATTTCACGTAAGATTATTACCCGCTGTATCAACACCGCCAAATTGCCTAAATCAGTTGTTGCTCTTTTTTCTCACC CCGGTGAACTATCTGCCCGGTCAGGTGATGCACTTCAAAAAGCCTTTACAGATAAAGAGGAATTACTTAAGCAGCAGG CATCTAACCTTCATGAGCAGAAAAAAGCTGGGGTGATATTTGAAGCTGAAGAAGTTATCACTCTTTTAACTTCTGTGCT TAAAACGTCATCTGCATCAAGAACTAGTTTAAGCTCACGACATCAGTTTGCTCCTGGAGCGACAGTATTGTATAAGGG CGATAAAATGGTGCTTAACCTGGACAGGTCTCGTGTTCCAACTGAGTGTATAGAGAAAATTGAGGCCATTCTTAAGGA ACTTGAAAAGCCAGCACCCTGATGCGACCTCGTTTTAGTCTACGTTTATCTGTCTTTACTTAATGTCCTTTGTTACAGGC CAGAAAGCATAACTGGCCTGAATATTCTCTCTGGGCCCACTGTTCCACTTGTATCGTCGGTCTGATAATCAGACTGGGA CCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGT CTGGGACCACGGTCCCACTCGTATCGTCGGTCTGATAATCAGACTGGGACCACGGTCCCACTCGTATCGTCGGTCTGA TTATTAGTCTGGGACCATGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCACGGTCCCACTCGTATCGTCG GTCTGATTATTAGTCTGGAACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCACGGTCCCACTCGT ATCGTCGGTCTGATTATTAGTCTGGGACCACGATCCCACTCGTGTTGTCGGTCTGATTATCGGTCTGGGACCACGGTCC CACTTGTATTGTCGATCAGACTATCAGCGTGAGACTACGATTCCATCAATGCCTGTCAAGGGCAAGTATTGACATGTCG TCGTAACCTGTAGAACGGAGTAACCTCGGTGTGCGGTTGTATGCCTGCTGTGGATTGCTGCTGTGTCCTGCTTATCCAC AACATTTTGCGCACGGTTATGTGGACAAAATACCTGGTTACCCAGGCCGTGCCGGCACGTTAACCGGGCTGCATCCGA TGCAAGTGTGTCGCTGTCGACGAGCTCGCGAGCTCGGACATGAGGTTGCCCCGTATTCAGTGTCGCTGATTTGTATTG TCTGAAGTTGTTTTTACGTTAAGTTGATGTAGATCAATTAATACGATACCTGCGTCATAATTGATTATTTGACGTGGTTTG ATGGCCTCCACGCACGTTGTGATATGTAGATGATAATCATTATCACTTTACGGGTCCTTTCCGGTGATCCGACAGGTTAC GGGGCGGCGACCTCGCGGGTTTTCGCTATTTATGAAAATTTTCCGGTTTAAGGCGTTTCCGTTCTTCTTCGTCATAACTT AATGTTTTTATTTAAAATACCCTCTGAAAAGAAAGGAAACGACAGGTGCTGAAAGCGAGCTTTTTGGCCTCTGTCGTT TCCTTTCTCTGTTTTTGTCCGTGGAATGAACAATGGAAGTCCGAGCTCATCGCTAATAACTTCGTATAGCATACATTATA CGAAGTTATATTCGATGCGGCCGCAAGGGGTTCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCG GCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCACACAGATGCGTAAGGAGAAAATACCG CATCAGGCGCCATTCGCCATTCAGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAG CTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAAC GACGGCCAGTGAATTGTAATACGACTCACTATAGGGCGAATTCGAGCTCGGTACCCGGGGATCCTCTAGAGTCGACCT GCAGGCATGCAAGCTTGAGTATTCTATAGTCTCACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGA AATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTG AGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGA ATCGGCCAACGCGAACCCCTTGCGGCCGCCCGGGCCGTCGACCAATTCTCATGTTTGACAGCTTATCATCGAATTTCT GCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTTTAAGGGCACCAATAACTGCCTTAAAAAAAT TACGCCCCGCCCTGCCACTCATCGCAGTACTGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGG CATGATGAACCTGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAACGGGGG CGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAGGGATTGGCTGAGACGAAAAAC ATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGTTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAA ACTGCCGGAAATCGTCGTGGTattcactccagagcgatgaaaacgtttcagtttgctcatggaaaacggtgtaACAAGGGTGAACACTATCCCATATC ACCAGCTCACCGTCTTTCATTGCCATACGAAATTCCGGATGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCC GGATAAAACTTGTGCTTATTTTTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACA TTGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGGTGGTATATCCAGTGATTT TTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGATAACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTAT GGTGAAAGTTGGAACCTCTTACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAA CAGGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGCGATAAGCTCATGGAGCGGC GTAACCGTCGCACAGGAAGGACAGAGAAAGCGCGGATCTGGGAAGTGACGGACAGAACGGTCAGGACCTGGATTG GGGAGGCGGTTGCCGCCGCTGCTGCTGACGGTGTGACGTTCTCTGTTCCGGTCACACCACATACGTTCCGCCATTCCT ATGCGATGCACATGCTGTATGCCGGTATACCGCTGAAAGTTCTGCAAAGCCTGATGGGACATAAGTCCATCAGTTCAAC GGAAGTCTACACGAAGGTTTTTGCGCTGGATGTGGCTGCCCGGCACCGGGTGCAGTTTGCGATGCCGGAGTCTGATG CGGTTGCGATGCTGAAACAATTATCCTGAGAATAAATGCCTTGGCCTTTATATGGAAATGTGGAACTGAGTGGATATGC TGTTTTTGTCTGTTAAACAGAGAAGCTGGCTGTTATCCACTGAGAAGCGAACGAAACAGTCGGGAAAATCTCCCATTA TCGTAGAGATCCGCATTATTAATCTCAGGAGCCTGTGTAGCGTTTATAGGAAGTAGTGTTCTGTCATGATGCCTGCAAG CGGTAACGAAAACGATTTGAATATGCCTTCAGGAACAATAGAAATCTTCGTGCGGTGTTACGTTGAAGTGGAGCGGAT TATGTCAGCAATGGACAGAACAACCTAATGAACACAGAACCATGATGTGGTCTGTCCTTTTACAGCCAGTAGTGCTCG CCGCAGTCGAGCGACAGGGCGAAGCCCAT
pTKS-3 sequencing	GAGCACAAGAGGTGACAAAAGCCACCGGCTGGATCGCACTTCTCGGAATTTCCCCCCTACTATCAAACAAATTCGAAT TGCCAAAGGTGAAGGGACTAACTGTAAATCCTGATCAATCAAGGTCTCAATCAAGTACAATGGGCTACAATGATATTTA GATGGGAACACAATGAAACGAATTGAAACTTCTACTGACAGGAGCGCAATTGACTTGTGTAGCTTTTCATGAGCACTT GATTGCTACCGATTGTGAACGGGATGGGGAAAGACTCGAAAAGGTGCATGCTTCCGATAATCTACTATATTTTCTAGAA TCAAATAATATTTAAATGAATGAGGTCCTCAGCGTACGTTAAGCCTACTTATTTAGAACGAGAAGTCAGACCGAGGGGT ACTAAAATTCTAAGGGTTGAGAGGTATCTTGATTCCGGGTCTATGGAAGCCCATCCTTGTTGAAGCTTGAACACGATCC TTGTGAAAGGCCGACGTTGCGCGAAAAAACAGCCTGCCGATTTCTTTCCTTCTTTCTCGTCTCAACCTATATACTTTCA TAATCTCTGTTAGAGTTTACCAACAACACATATATACATTTCGACAAAATGGCCAAGTTGACCAGTGCCGTTCCGGTGC TCACCGCGCGCGACGTCGCCGGAGCGGTCGAGTTCTGGACCGACCGGCTCGGGTTCTCCCGGGACTTCGTGGAGGA

	CGACTTCGCCGGTGTGGTCCGGGACGACGTGACCCTGTTCATCAGCGCGGTCCAGGACCAGGTGGTGCCGGACAAC ACCCTGGCCTGGGTGTGGGTGCGCGGCCTGGACGAGCTGTACGCCGAGTGGTCGGAGGTCGTGTCCACGAACTTCCG GGACGCCTCCGGGCCGGCCATGACCGAGATCGGCGAGCAGCCGTGGGGGCGGGAGTTCGCCCTGCGCGACCCGGCC GGCAACTGCGTGCACTTCGTGGCCGAGGAGCAGGACTGATTTTGTTACATTGACTTCAAGGAGTCGAGGAATCGATA CTGCCGTCGTTTCCAGGATCCGAGGTTTCTATAGACTCTCTATAGACTCTGTTAACCTAATAGAATCAGACATACCTCTC CTGCTATTTTGTTTTTTATGAATTTGGCTTTTGCCTCTCTAGTCAGATTTGAATGTTATTTTCCGCCAGGTGTGTTAGTCG GGCTCTCGTTTGAGTTACAAGAGGGATTGAGTGGCGAGGATTCACTCTAATGTAAATATGACTGTGAACAAAACTTTA AAATTACTACGCATCTTCTTTGACTGTCAGATATTCGTCGGTGACAGCAGTCAATGCCTGCAAATTGTCCTCCTGGGTC GCAATTTGGTTTTGGATTGACCTGGTATGCATTATGAAGAAAAAAATTCGTTATTAGCCAACTGCCTAGCGTGCACATT GCATGGTTAGACCTCCTTGACGACTGTGAGCCTACATCCTTCTGCAACAAGCTGCAATCCCTGCGATAGACCTTTTCCA AACTCACGCAGTCCAAGAAAACAAAGGGGTGAGAAGTATACGCACCTTTCGGTTTCGGCATAATTCTTAAACTCTTGT GGTCACTTTCTTGTGAAGAAGCTAGGGGCACTCGTTTTCCCTCAGAGCCTGCAAACACAAAATTCCTGCAGTCAATTG TCCCAACACTCGGCAAGCCGTATGCGCAAGCAACGATGCGCAGAAGGCCGTGGATGGATGGCGACTCGCGATATGGC TTCTTGGGTCGCCAGTGTGGTATGTCCGGCGTATGTCAATACGCGAATTCGGACGACTGGCATCTCTAGGAGGAGGAT TCCTTCTTTTATGACATGTTTATTTTATATACATTGATGCTTTCCGACAGTCGGAAGTAATAAATGAATTTATTTCAAGAC TACCTATACTCCTTTGACTTGTTCGACTAATCTTACCGCTTACTAAAATCTCGAAATCACGCTTGACCTCTCGCACGCAA ATTTTTGCTGCTGGACGCTACGCACTCGGCCCAATTCTTCTCGGTCCTCGTCGTCGCAATTGTCGTTGCGTTGATCTTG CACCGAAGGAATCAGAGAATAGAATACCATGAACCACCTGCGTGCCGAAGGACCCGCCTCCGTCCTCGCCATTGGAA CCGCCAACCCCGAAAACATTCTCCTCCAAGACGAATTCCCTGACTACTACTTCCGCGTGACTAAGAGTGAGCACATGA CGCAATTGAAAGAAAAGTTCCGTAAGATCTGTGATAAGTCCATGATCCGTAAGCGTAACTGTTTCCTGAATGAGGAAC ACCTTAAGCAGAACCCCCGCTTGGTCGAACATGAAATGCAGACCTTGGATGCCCGCCAGGACATGCTCGTCGTTGAA GTCCCTAAGTTGGGTAAAGACGCTTGCGCCAAGGCCATTAAGGAATGGGGACAGCCCAAGAGTAAGATCACCCATTT GATTTTCACATCCGCCTCTACGACCGATATGCCTGGAGCCGATTACCACTGCGCCAAGCTTCTGGGACTGTCCCCCTCC GTTAAGCGCGTCATGATGTACCAGCTTGGATGTTACGGAGGTGGCACCGTCCTCCGTATTGCCAAGGACATTGCCGAA AACAACAAGGGTGCCCGTGTCCTCGCCGTCTGCTGCGACATCATGGCCTGCCTCTTCCGCGGACCCTCCGAATCCGAC TTGGAACTCCTCGTGGGACAGGCCATTTTTGGTGACGGCGCCGCCGCCGTCATTGTCGGAGCTGAACCCGACGAATC CGTCGGAGAAAGGCCCATTTTTGAATTGGTCTCCACCGGACAGACCATTCTCCCGAACTCCGAAGGAACCATCGGCG GACACATTCGTGAAGCCGGTCTCATCTTTGACTTGCACAAGGACGTTCCCATGCTCATTTCCAACAACATTGAAAAGT GCCTCATTGAAGCCTTCACCCCCATTGGAATCTCCGACTGGAATAGTATCTTTTGGATTACCCACCCCGGAGGAAAGG CCATCTTGGATAAGGTGGAAGAAAAGCTCCACCTCAAGTCCGACAAGTTCGTCGACTCCCGTCACGTCTTGTCCGAA CACGGAAACATGTCGTCCTCCACCGTCCTCTTTGTCATGGATGAATTGCGTAAGCGCTCGCTCGAAGAAGGAAAGTCC ACCACCGGCGATGGTTTCGAATGGGGCGTTCTCTTCGGATTTGGACCCGGTCTCACCGTCGAACGTGTCGTCGTCCGT TCTGTCCCGATTAAGTACCACCATCACCACCACCACTAAAAATTTAATTTTCATTAGTTGCAGTCACTCCGCTTTGGTTT CAGGGTAATATAGATCTTCCGCTGCATAACCCTGCTTCGGGGTCATTATAGCGATTTTTTCGGTATATCCATCCTTTTCG CACGATATACAGGATTTTGCCAAAGGGTTCGTGTAGACTTTCCTTGGTGTATCCAACGGCGTCAGCCGGGCAGGATAG GTGAAGTAGGCCCACCCGCGAGCGGGTGTTCCTTCTTCACTGTCCCTTATTCGCACCTGGCGGTGCTCAACGGGAATC CTGCTCTGCGAGGCTGGCCGGCTACCGCCGGCGTAACAGATGAGGGCAAGCGGATGGCTGATGAAACCAAGCCAAC CAGGAAGGGCAGCCCACCTATCAAGGTGTACTGCCTTCCAGACGAACGAAGAGCGATTGAGGAAAAGGCGGCGGCG GCCGGCATGAGCCTGTCGGCCTACCTGCTGGCCGTCGGCCAGGGCTACAAAATCACGGGCGTCGTGGACTATGAGCA CGTCCGCGAGCTGGCCCGCATCAATGGCGACCTGGGCCGCCTGGGCGGCCTGCTGAAACTCTGGCTCACCGACGACC CGCGCACGGCGCGGTTCGGTGATGCCACGATCCTCGCCCTGCTGGCGAAGATCGAAGAGAAGCAGGACGAGCTTGG CAAGGTCATGATGGGCGTGGTCCGCCCGAGGGCAGAGCCATGACTTTTTTAGCCGCTAAAACGGCCGGGGGGTGCGC GTGATTGCCAAGCACGTCCCCATGCGCTCCATCAAGAAGAGGCACTTCGAGCTGTAAGTACATCACCGACGAGCAAG GCAAGACGATCCGCGCCTGTTTTATTGAGAACGTTGTTCGTGTTGGCCTCAATGGTAGCGATGCGTCATTCAGCGAAG TTCTGGTGCTGATGATGTGGTTCGCTTTGCCACTGGTCAATGTGGTAAGCCCGTGTAATGTCAGTAACCTTTTTACTGA TCTCAGCTTGAGCACGGTCGCTGATGAGCTTATCCATGGCCCCACGGTAACGGATATGATCCTCTAGGGCGTTGACAA ATTCTTTGTCGGTTTTCATGGGGTAGACGTCAGTGACCAGGGAACGTCTCCCTACAAAAAGTTGCGCATACTTTGCTC CGTTGTCGACGGCAGGGGTGTCGGACCAGATTGTATCTGTGGCAACGGCCTCATTGCGTCAATGGACCTTTAAAGCCG GGAAACGAGACTTGAAATGTTTACGTAGAGGTGCATTATAGACCTCACGCGCATATTGAGTGGTTGCAAAAATGGTTT TGCGAACGGTATCACTGGAGACCCATGCCAGGCAAGGACGGAGGGCATCATAATCATGTTCATTGCGTTGGACAGCAT GCTTGTTACACGTAAGTATACGGTCGAGAGCATCGTGACGTTGAGAATTACAGATGGCAACGTGTCGGTGCATGTATT GGCCGAGATTGTCAAATCGTGGCTCAATATATGGCATGCCAGGTAAGTCATGGATGTCGGTATGCCATTCAGCCTTCAT GTCGATTTCGTTGTCGAGAACTGATGGGTCCCAGTCAACGTCAGAAGTAAGAACAACATGGGGAAGGGAATGAAGTT CATGACTGGTGGGTGCGCGCATATCCATGTACGGAAGACCCTGTCGTATGTTTAGTGGGATGATGTAATCATCCAGGGT AACAATACGTTGGTTGCCGCCCACGGTACGAGACCGATCTTGAACACAATTGTGGCAATGTTCAAGTTGGGCGCTGG AATGGATGGTTTTACCTTTACCAAGATGCGCGTATTGATGCATGATAACAACAATGGGCCCTCGTTGCGATTCGACAAG ACCGGCAGCCGTGACAATGTCTAGATTGGACAAGGTATGTTCATTTATACCGGTTATGTTGGCTGAACGTCCAGTTTTG TGCAGTACGGTAACATCACTCCCCGCAAGTCCACCGTTGGCACCGCGATCGACAAGTGCAGATGTGGTATTTTGTACT GTGTGTCGGGAGACTTGGTACTGTAACACATTTGACTGAAGTGACTTGGTGGTACCATGAGGACGGAGGGAACCCGT TGATACGGTGTTGCGTGACGCTGCAAGTACCTTACGTATGTCGCCTGGATCCATACGTCCAACCCGATCAGAAAGATGT GCTAGGAGTTCCGTTTCTGGTGCACAATCATGAAAGGTATCCACGGGAGTTCCATTGTTGTCGTCCGGGGCTGCGTCG CGACTGCCAGTTGCATGTACATGCGCCATCAAGGGACGTGCCAATCCATTACTAACCGGGAGAACCTTAGCTGGGGCA GCCAATCCCTGAAGTATAGCCTTAGCATCATCTGAGAGCTGGTTCCACATATCTCTAGGGATATAAGGTCGTTCACGGT TAGTGGGCGTACTATGATGTCGGTTATTAGTACAATTCTCCCTGCGGGCATGCGCATTGGCTTCATAGAGTACGGAAGG TGACAAGTCAATATTGTAGTTAACATCCGGATCAGTGTCAAAGTCTGTAGGATGGTAAGAAAGATCAGTAGAATGAAT ACTACGCTTTTCCTTGGGACTACGGGAATTAGAGAAGTTGTTTCCTTTATTGTAGAGTGACGCTGAAGCAAGTAGAAG ACTAAGGTAACTCTCGTAGCTAATAGGATTACCTCCTTTGGCTAGGTCAAGAGTGGCTGTGATCTTCACTTGACAAAGT TCCGGTACATTGTGGACAGCATTCTCCAAAAGACTAAGACACAGTTGCTTTGGGAGTTGCTCAGCCATTGGTACAGTA TTGTGGTAGATACAAAGGTGGTTCTTCCAATGAAGGATAAATCCTTCTGCTGTGCCTGTCCATGAGGATCCATATTTCG CCGTAGTTAGGTAACCAAGCGTGGTGGCTGAACTGATCTTCGCACTTGCTGATTCCGTATAGTGTTTGACAACTTTACA AAACACTTCTTGCGCAGTTCGCTCTAGCCTGGACAGAGAAGGGACAATCTTCGTCGTTAAGACTCGTGCGAATAGCA AAAGATCACAAAATAGCACATCGGCACCGACCAACGATTATTTCCAAGGAAAAAAAGAATGCTTCACTACAAGAAAT TGTGTCATCCCTATACAGAGTCTTGTTACTGTGACAGAAAATTGATGGAAGATGTGGCGGATTGCCTTTACACTAGCCA ACTTGTTCGACTAATTGCAGCTTCTTCTGAGAGGCTTCACCGAGTAACGCGAAGAACACCGGTGTCTCGTACATGCTC

	GTCGGTGAACGCTCGTCCAATGACACCCCCCACTTTGTATCAATATCCCAACTTGGTAGTGAACTGGAATGATACATGC AATTTCGCGCCGCATCAACAGCCACGGGCACCATCGACGAATAGACTCGGTCGAGCTGGTTGCCCTCGCCGCTGGGC TGGCGGCCGTCTATGGCCCTGCAAACGCGCCAGAAACGCCGTCGAAGCCGTGTGCGAGACACCGCGGCCGGCCGCC GGCGTTGTGGATACCTCGCGGAAAACTTGGCCCTCACTGACAGATGAGGGGCGGACGTTGACACTTGAGGGGCCGA CTCACCCGGCGCGGCGTTGACAGATGAGGGGCAGGCTCGATTTCGGCCGGCGACGTGGAGCTGGCCAGCCTCGCAA ATCGGCGAAAACGCCTGATTTTACGCGAGTTTCCCACAGATGATGTGGACAAGCCTGGGGATAAGTGCCCTGCGGTAT TGACACTTGAGGGGCGCGACTACTGACAGATGAGGGGCGCGATCCTTGACACTTGAGGGGCAGAGTGCTGACAGAT GAGGGGCGCACCTATTGACATTTGAGGGGCTGTCCACAGGCAGAAAATCCAGCATTTGCAAGGGTTTCCGCCCGTTTT TCGGCCACCGCTAACCTGTCTTTTAACCTGCTTTTAAACCAATATTTATAAACCTTGTTTTTAACCAGGGCTGCGCCCTG TGCGCGTGACCGCGCACGCCGAAGGGGGGTGCCCCCCCCTTCTCGAACCCTCCCGGTCGAGTGAGCGAGGAAGCACC AGGGAACAGCACTTATATATTCTGCTTACACACGATGCCTGAAAAAACTTCCCTTGGGGTTATCCACTTATCCACGGGG ATATTTTTATAATTATTTTTTTTATAGTTTTTAGATCTTCTTTTTTAGAGCGCCTTGTAGGCCTTTATCCATGCTGGTTCTAG AGAAGGTGTTGTGACAAATTGCCCTTTCAGTGTGACAAATCACCCTCAAATGACAGTCCTGTCTGTGACAAATTGCCC TTAACCCTGTGACAAATTGCCCTCAGAAGAAGCTGTTTTTTTCACAAAGTTATCCCTGCTTATTGACTCTTTTTTATTTA GTGTGACAATCTAAAAACTTGTCACACTTCACATGGATCTGTCATGGCGGAAACAGCGGTTATCAATCACAAGAAACG TAAAAATAGCCCGCGAATCGTCCAGTCAAACGACCTCACTGAGGCGGCATATAGTCTCTCCCGGGATCAAAAACGTAT GCTGTATCTGTTCGTTGACCAGATCAGAAAATCTGATGGCACCCTACAGGAACATGACGGTATCTGCGAGATCCATGTT GCTAAATATGCTGAAATATTCGGATTGACCTCTGCGGAAGCCAGTAAGGATATACGGCAGGCATTGAAGAGTTTCGCG GGGAAGGAAGTGGTTTTTTATCGCCCTGAAGAGGATGCCGGCGATGAAAAAGGCTATGAATCTTTTCCTTGGTTTATC AAACGTGCGCACAGTCCATCCAGAGGGCTTTACAGTGTACATATCAACCCATATCTCATTCCCTTCTTTATCGGGTTACA GAACCGGTTTACGCAGTTTCGGCTTAGTGAAACAAAAGAAATCACCAATCCGTATGCCATGCGTTTATACGAATCCCT GTGTCAGTATCGTAAGCCGGATGGCTCAGGCATCGTCTCTCTGAAAATCGACTGGATCATAGAGCGTTACCAGCTGCC TCAAAGTTACCAGCGTATGCCTGACTTCCGCCGCCGCTTCCTGCAGGTCTGTGTTAATGAGATCAACAGCAGAACTCC AATGCGCCTCTCATACATTGAGAAAAAGAAAGGCCGCCAGACGACTCATATCGTATTTTCCTTCCGCGATATCACTTCC ATGACGACAGGATAGTCTGAGGGTTATCTGTCACAGATTTGAGGGTGGTTCGTCACATTTGTTCTGACCTACTGAGGG TAATTTGTCACAGTTTTGCTGTTTCCTTCAGCCTGCATGGATTTTCTCATACTTTTTGAACTGTAATTTTTAAGGAAGCC AAATTTGAGGGCAGTTTGTCACAGTTGATTTCCTTCTCTTTCCCTTCGTCATGTGACCTGATATCGGGGGTTAGTTCGTC ATCATTGATGAGGGTTGATTATCACAGTTTATTACTCTGAATTGGCTATCCGCGTGTGTACCTCTACCTGGAGTTTTTCC CACGGTGGATATTTCTTCTTGCGCTGAGCGTAAGAGCTATCTGACAGAACAGTTCTTCTTTGCTTCCTCGCCAGTTCGC TCGCTATGCTCGGTTACACGGCTGCGGCGAGCATCACGTGCTATAAAAATAATTATAATTTAAATTTTTTAATATAAATAT ATAAATTAAAAATAGAAAGTAAAAAAAGAAATTAAAGAAAAAATAGTTTTTGTTTTCCGAAGATGTAAAAGACTCTAG GGGGATCGCCAACAAATACTACCTTTTACCTTGCTCTTCCTGCTCTCAGGTATTAATGCCGAATTGTTTCATCTTGTCTG TGTAGAAGACCACACACGAAAATCCTGTGATTTTACATTTTACTTATCGTTAATCGAATGTATATCTATTTAATCTGCTTT TCTTGTCTAATAAATATATATGTAAAGTACGCTTTTTGTTGAAATTTTTTAAACCTTTGTTTATTTTTTTTCTTCATTCCG TAACTCTTCTACCTTCTTTATTTACTTTCTAAAATCCAAATACAAAACATAAAAATAAATAAACACAGAGTAAATTCCCA AATTATTCCATCATTAAAAGATACGAGGCGCGTGTAAGTTACAGGCAAGCGATCCTAGTACACTCTATATTTTTTTATGC CTCGGTAATGATTTTCATTTTTTTTTTCCACCTAGCGGATGACTCTTTTTTTTTCTTAGCGATTGGCATTATCACATAATG AATTATACATTATATAAAGTAATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAA GATGACAGAGCAGAAAGCCCTAGTAAAGCGTATTACAAATGAAACCAAGATTCAGATTGCGATCTCTTTAAAGGGTGG TCCCCTAGCGATAGAGCACTCGATCTTCCCAGAAAAAGAGGCAGAAGCAGTAGCAGAACAGGCCACACAATCGCAA GTGATTAACGTCCACACAGGTATAGGGTTTCTGGACCATATGATACATGCTCTGGCCAAGCATTCCGGCTGGTCGCTAA TCGTTGAGTGCATTGGTGACTTACACATAGACGACCATCACACCACTGAAGACTGCGGGATTGCTCTCGGTCAAGCTT TTAAAGAGGCCCTAGGGGCCGTGCGTGGAGTAAAAAGGTTTGGATCAGGATTTGCGCCTTTGGATGAGGCACTTTCC AGAGCGGTGGTAGATCTTTCGAACAGGCCGTACGCAGTTGTCGAACTTGGTTTGCAAAGGGAGAAAGTAGGAGATCT CTCTTGCGAGATGATCCCGCATTTTCTTGAAAGCTTTGCAGAGGCTAGCAGAATTACCCTCCACGTTGATTGTCTGCGA GGCAAGAATGATCATCACCGTAGTGAGAGTGCGTTCAAGGCTCTTGCGGTTGCCATAAGAGAAGCCACCTCGCCCAA TGGTACCAACGATGTTCCCTCCACCAAAGGTGTTCTTATGTAGTTTTACACAGGAGTCTGGACTTGACGCTAGTGATAA TAAGTGACTGAGGTATGTGCTCTTCTTATCTCCTTTTGTAGTGTTGCTCTTATTTTAAACAACTTTGCGGTTTTTTGATGA CTTTGCGATTTTGTTGTTGCTTTGCAGTAAATTGCAAGATTTAATAAAAAAACGCAAAGCAATGATTAAAGGATGTTCA GAATGAAACTCATGGAAACACTTAACCAGTGCATAAACGCTGGTCATGAAATGACGAAGGCTATCGCCATTGCACAGT TTAATGATGACAGCCCGGAAGCGAGGAAAATAACCCGGCGCTGGAGAATAGGTGAAGCAGCGGATTTAGTTGGGGTT TCTTCTCAGGCTATCAGAGATGCCGAGAAAGCAGGGCGACTACCGCACCCGGATATGGAAATTCGAGGACGGGTTGA GCAACGTGTTGGTTATACAATTGAACAAATTAATCATATGCGTGATGTGTTTGGTACGCGATTGCGACGTGCTGAAGAC GTATTTCCACCGGTGATCGGGGTTGCTGCCCATAAAGGTGGCGTTTACAAAACCTCAGTTTCTGTTCATCTTGCTCAGG ATCTGGCTCTGAAGGGGCTACGTGTTTTGCTCGTGGAAGGTAACGACCCCCAGGGAACAGCCTCAATGTATCACGGAT GGGTACCAGATCTTCATATTCATGCAGAAGACACTCTCCTGCCTTTCTATCTTGGGGAAAAGGACGATGTCACTTATGC AATAAAGCCCACTTGCTGGCCGGGGCTTGACATTATTCCTTCCTGTCTGGCTCTGCACCGTATTGAAACTGAGTTAATG GGCAAATTTGATGAAGGTAAACTGCCCACCGATCCACACCTGATGCTCCGACTGGCCATTGAAACTGTTGCTCATGAC TATGATGTCATAGTTATTGACAGCGCGCCTAACCTGGGTATCGGCACGATTAATGTCGTATGTGCTGCTGATGTGCTGAT TGTTCCCACGCCTGCTGAGTTGTTTGACTACACCTCCGCACTGCAGTTTTTCGATATGCTTCGTGATCTGCTCAAGAAC GTTGATCTTAAAGGGTTCGAGCCTGATGTACGTATTTTGCTTACCAAATACAGCAATAGCAATGGCTCTCAGTCCCCGT GGATGGAGGAGCAAATTCGGGATGCCTGGGGAAGCATGGTTCTAAAAAATGTTGTACGTGAAACGGATGAAGTTGGT AAAGGTCAGATCCGGATGAGAACTGTTTTTGAACAGGCCATTGATCAACGCTCTTCAACTGGTGCCTGGAGAAATGCT CTTTCTATTTGGGAACCTGTCTGCAATGAAATTTTCGATCGTCTGATTAAACCACGCTGGGAGATTAGATAATGAAGCG TGCGCCTGTTATTCCAAAACATACGCTCAATACTCAACCGGTTGAAGATACTTCGTTATCGACACCAGCTGCCCCGATG GTGGATTCGTTAATTGCGCGCGTAGGAGTAATGGCTCGCGGTAATGCCATTACTTTGCCTGTATGTGGTCGGGATGTGA AGTTTACTCTTGAAGTGCTCCGGGGTGATAGTGTTGAGAAGACCTCTCGGGTATGGTCAGGTAATGAACGTGACCAGG AGCTGCTTACTGAGGACGCACTGGATGATCTCATCCCTTCTTTTCTACTGACTGGTCAACAGACACCGGCGTTCGGTC GAAGAGTATCTGGTGTCATAGAAATTGCCGATGGGAGTCGCCGTCGTAAAGCTGCTGCACTTACCGAAAGTGATTATC GTGTTCTGGTTGGCGAGCTGGATGATGAGCAGATGGCTGCATTATCCAGATTGGGTAACGATTATCGCCCAACAAGTG CTTATGAACGTGGTCAGCGTTATGCAAGCCGATTGCAGAATGAATTTGCTGGAAATATTTCTGCGCTGGCTGATGCGGA AAATATTTCACGTAAGATTATTACCCGCTGTATCAACACCGCCAAATTGCCTAAATCAGTTGTTGCTCTTTTTTCTCACC CCGGTGAACTATCTGCCCGGTCAGGTGATGCACTTCAAAAAGCCTTTACAGATAAAGAGGAATTACTTAAGCAGCAGG CATCTAACCTTCATGAGCAGAAAAAAGCTGGGGTGATATTTGAAGCTGAAGAAGTTATCACTCTTTTAACTTCTGTGCT

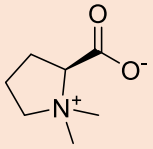
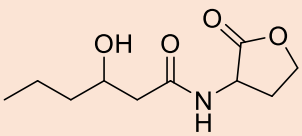
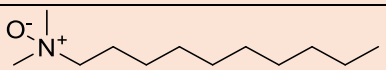
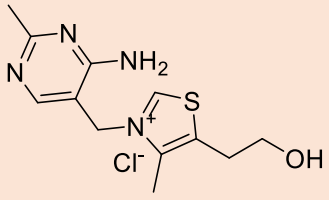
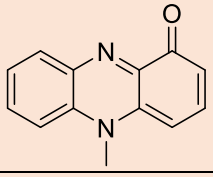
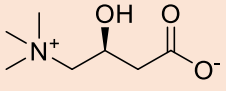
	TAAAACGTCATCTGCATCAAGAACTAGTTTAAGCTCACGACATCAGTTTGCTCCTGGAGCGACAGTATTGTATAAGGG CGATAAAATGGTGCTTAACCTGGACAGGTCTCGTGTTCCAACTGAGTGTATAGAGAAAATTGAGGCCATTCTTAAGGA ACTTGAAAAGCCAGCACCCTGATGCGACCTCGTTTTAGTCTACGTTTATCTGTCTTTACTTAATGTCCTTTGTTACAGGC CAGAAAGCATAACTGGCCTGAATATTCTCTCTGGGCCCACTGTTCCACTTGTATCGTCGGTCTGATAATCAGACTGGGA CCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGT CTGGGACCACGGTCCCACTCGTATCGTCGGTCTGATAATCAGACTGGGACCACGGTCCCACTCGTATCGTCGGTCTGA TTATTAGTCTGGGACCATGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCACGGTCCCACTCGTATCGTCG GTCTGATTATTAGTCTGGAACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCACGGTCCCACTCGT ATCGTCGGTCTGATTATTAGTCTGGGACCACGATCCCACTCGTGTTGTCGGTCTGATTATCGGTCTGGGACCACGGTCC CACTTGTATTGTCGATCAGACTATCAGCGTGAGACTACGATTCCATCAATGCCTGTCAAGGGCAAGTATTGACATGTCG TCGTAACCTGTAGAACGGAGTAACCTCGGTGTGCGGTTGTATGCCTGCTGTGGATTGCTGCTGTGTCCTGCTTATCCAC AACATTTTGCGCACGGTTATGTGGACAAAATACCTGGTTACCCAGGCCGTGCCGGCACGTTAACCGGGCTGCATCCGA TGCAAGTGTGTCGCTGTCGACGAGCTCGCGAGCTCGGACATGAGGTTGCCCCGTATTCAGTGTCGCTGATTTGTATTG TCTGAAGTTGTTTTTACGTTAAGTTGATGTAGATCAATTAATACGATACCTGCGTCATAATTGATTATTTGACGTGGTTTG ATGGCCTCCACGCACGTTGTGATATGTAGATGATAATCATTATCACTTTACGGGTCCTTTCCGGTGATCCGACAGGTTAC GGGGCGGCGACCTCGCGGGTTTTCGCTATTTATGAAAATTTTCCGGTTTAAGGCGTTTCCGTTCTTCTTCGTCATAACTT AATGTTTTTATTTAAAATACCCTCTGAAAAGAAAGGAAACGACAGGTGCTGAAAGCGAGCTTTTTGGCCTCTGTCGTT TCCTTTCTCTGTTTTTTGTCCGTGGAATGAACAATGGAAGTCCGAGCTCATCGCTAATAACTTCGTATAGCATACATTATA CGAAGTTATATTCGATGCGGCCGCAAGGGGTTCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCG GCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCACACAGATGCGTAAGGAGAAAATACCG CATCAGGCGCCATTCGCCATTCAGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAG CTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAAC GACGGCCAGTGAATTGTAATACGACTCACTATAGGGCGAATTCGAGCTCGGTACCCGGGGATCCTCTAGAGTCGACCT GCAGGCATGCAAGCTTGAGTATTCTATAGTCTCACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGA AATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTG AGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGA ATCGGCCAACGCGAACCCCTTGCGGCCGCCCGGGCCGTCGACCAATTCTCATGTTTGACAGCTTATCATCGAATTTCT GCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTTTAAGGGCACCAATAACTGCCTTAAAAAAAT TACGCCCCGCCCTGCCACTCATCGCAGTACTGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGG CATGATGAACCTGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAACGGGGG CGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAGGGATTGGCTGAGACGAAAAAC ATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGTTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAA ACTGCCGGAAATCGTCGTGGTattcactccagagcgatgaaaacgtttcagtttgctcatggaaaacggtgtaACAAGGGTGAACACTATCCCATATC ACCAGCTCACCGTCTTTCATTGCCATACGAAATTCCGGATGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCC GGATAAAACTTGTGCTTATTTTTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACA TTGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGGTGGTATATCCAGTGATTT TTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGATAACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTAT GGTGAAAGTTGGAACCTCTTACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAA CAGGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGCGATAAGCTCATGGAGCGGC GTAACCGTCGCACAGGAAGGACAGAGAAAGCGCGGATCTGGGAAGTGACGGACAGAACGGTCAGGACCTGGATTG GGGAGGCGGTTGCCGCCGCTGCTGCTGACGGTGTGACGTTCTCTGTTCCGGTCACACCACATACGTTCCGCCATTCCT ATGCGATGCACATGCTGTATGCCGGTATACCGCTGAAAGTTCTGCAAAGCCTGATGGGACATAAGTCCATCAGTTCAAC GGAAGTCTACACGAAGGTTTTTGCGCTGGATGTGGCTGCCCGGCACCGGGTGCAGTTTGCGATGCCGGAGTCTGATG CGGTTGCGATGCTGAAACAATTATCCTGAGAATAAATGCCTTGGCCTTTATATGGAAATGTGGAACTGAGTGGATATGC TGTTTTTGTCTGTTAAACAGAGAAGCTGGCTGTTATCCACTGAGAAGCGAACGAAACAGTCGGGAAAATCTCCCATTA TCGTAGAGATCCGCATTATTAATCTCAGGAGCCTGTGTAGCGTTTATAGGAAGTAGTGTTCTGTCATGATGCCTGCAAG CGGTAACGAAAACGATTTGAATATGCCTTCAGGAACAATAGAAATCTTCGTGCGGTGTTACGTTGAAGTGGAGCGGAT TATGTCAGCAATGGACAGAACAACCTAATGAACACAGAACCATGATGTGGTCTGTCCTTTTACAGCCAGTAGTGCTCG CCGCAGTCGAGCGACAGGGCGAAGCCCAT
pOAC-5 sequencing Forward	GGCCANNGCCNTGTTTGTCGCGTGCTATGTCTTTTTAGGTACTTTGAACCTACGTTCGTACTTGTATAATATGATCATCG TATTATCGTTTTTTCATCCGTCCAGCGCAAAATGCATTAGCAGCTAGTCCTAGCGTGCGGAGCTACCTGGACAGGTGCAT GACGGATGCGTGTCCTTCAGTGACTTTCTAATTAACAGTAACTTCTTTACTTATGTTTCAGTTTGTAAGAAGCGGGATT CGCTCGTCGGTTGACATCTGATTGGACTGCGTCGGCACATGAAAACTACATTGTGAAATCTGCTAAAACTCCGGGTAT CTCTGACACAAAACGATTCGGCTTCGCAATTTCAACATTACGGTCAAGGCTAACGTATCTTTCTCGGTCAACTTCAGAT TATGCCGATTAAATTGTCGTAGCTTTCAAGGCGTTTTGAGTACTGCGGCAGTTGTTGAACCTGCAAGGAGAAGATCTC GACAACAGAATAAAGCGAAAAATGGGTCTCATGCACTAACACTCAGGCCTCCCTCATAATCTCTGTTTGAGTTTACCA ACAACACATATATACATTTCGACAAAATGGCCGTCAAGCACTTGATCGTTCTCAAGTTCAAGGACGAAATCACCGAAG CCCAGAAGGAAGAATTCTTCAAGACCTACGTCAACCTCGTCAACATCATTCCCGCCATGAAGGATGTTTACTGGGGGA AAGGACGTCACCCAGAAGAACAAGGAAGAAGGTTACACCCACATCGTCGAAGTCACCTTTGAATCCGTTGAAACCAT CCAGACTACATTATCCACCCNGCCCACGTNGGCTTCGNAGACGTCTACCGTTCTTTTGGNAAAGTGTTGATTTTCGAC TACACCCCCCCGTAANGANANAGTGANTNNGAGAGNNNNNNNTAATTTGGTTANNNCNCNN
pOAC-5 sequencing Reverse	GGCCANNGCCNTGTTTGTCGCGTGCTATGTCTTTTTAGGTACTTTGAACCTACGTTCGTACTTGTATAATATGATCATCG TATTATCGTTTTTTCATCCGTCCAGCGCAAAATGCATTAGCAGCTAGTCCTAGCGTGCGGAGCTACCTGGACAGGTGCAT GACGGATGCGTGTCCTTCAGTGACTTTCTAATTAACAGTAACTTCTTTACTTATGTTTCAGTTTGTAAGAAGCGGGATT CGCTCGTCGGTTGACATCTGATTGGACTGCGTCGGCACATGAAAACTACATTGTGAAATCTGCTAAAACTCCGGGTAT CTCTGACACAAAACGATTCGGCTTCGCAATTTCAACATTACGGTCAAGGCTAACGTATCTTTCTCGGTCAACTTCAGAT TATGCCGATTAAATTGTCGTAGCTTTCAAGGCGTTTTGAGTACTGCGGCAGTTGTTGAACCTGCAAGGAGAAGATCTC GACAACAGAATAAAGCGAAAAATGGGTCTCATGCACTAACACTCAGGCCTCCCTCATAATCTCTGTTTGAGTTTACCA ACAACACATATATACATTTCGACAAAATGGCCGTCAAGCACTTGATCGTTCTCAAGTTCAAGGACGAAATCACCGAAG CCCAGAAGGAAGAATTCTTCAAGACCTACGTCAACCTCGTCAACATCATTCCCGCCATGAAGGATGTTTACTGGGGGA AAGGACGTCACCCAGAAGAACAAGGAAGAAGGTTACACCCACATCGTCGAAGTCACCTTTGAATCCGTTGAAACCAT CCAGACTACATTATCCACCCNGCCCACGTNGGCTTCGNAGACGTCTACCGTTCTTTTGGNAAAGTGTTGATTTTCGAC TACACCCCCCCGTAANGANANAGTGANTNNGAGAGNNNNNNNTAATTTGGTTANNNCNCNN

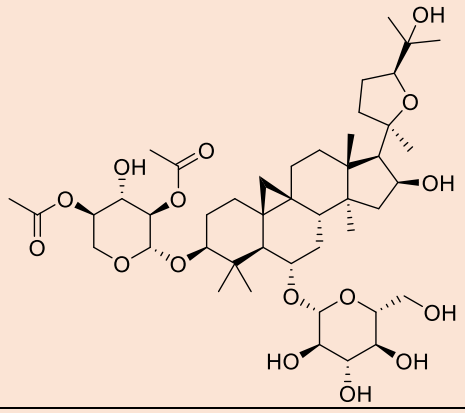
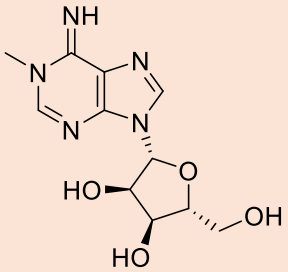
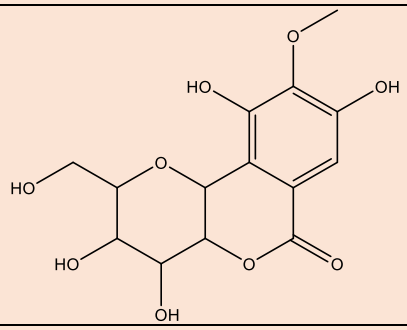
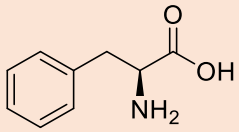
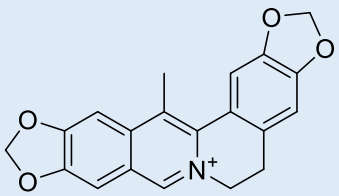
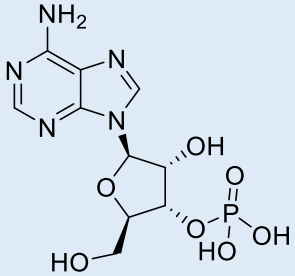
pOAC-7 sequencing Forward	CGNNNCGCCGTTTGTCGCGTGCTATGTCTTTTTAGGTACTTTGAACCTACGTTCGTACTTGTATAATATGATCAACGTAT TATCGTTTTTCATCCGTCCAGCGCAAAATGCATTAGCAGCTAGTCCTAGCGTGCGGAGCTACCTGGACAGGTGCATGA CGGATGCGTGTCCTTCAGTGACTTTCTAATTAACAGTAACTTCTTTACTTATGTTTCAGTTTGTAAGAAGCGGGATTCGC TCGTCGGTTGACATCTGATTGGACTGCGTCGGCACATGAAAACTACATTGTGAAATCTGCTAAAACTCCGGGTATCTCT GACACAAAACGATTCGGCTTCGCAATTTCAACATTACGGTCAAGGCTAACGTATCTTTCTCGGTCAACTTCAGATTATG CCGATTAAATTGTCGTAGCTTTCAAGGCGTTTTGAGTACTGCGGCAGTTGTTGAACCTGCAAGGAGAAGATCTCGACA ACAGAATAAAGCGAAAAATGGGTCTCATGCACTAACACTCAGGCCTCCCTCATAATCTCTGTTTGAGTTTACCAACAA CACATATATACATTTCGACAAAATGGCCGTCAAGCACTTGATCGTTCTCAAGTTCAAGGACGAAATCACCGAAGCCCA GAAGGAAGAATTCTTCAAGACCTACGTCAACCTCGTCAACATCATTCCCGCCATGAAGGATGTTTACTGGGGGAAAG GACGTCACCCAGAAGAACAAAGAAGAAAGTTACACCCACATCGTCGAAGTCACCTTTGAATCCGTTGAAACCATCCA GACTACATTATCCACCCNNCCCACGTNGGCTTNNGAGACGTCTACCGTTCCTTTTGGNAAAGTTGTGATTTTCGACTAC ANCCCCCCGTAACNNCAGAGNNGANNCGAGAGNNNNCNNNNATTTGGTTAAATNNCCCNNANN
pOAC-7 sequencing Reverse	CNNTNNTGAATTCTTACGGGGGGTGTAGTCGAAATCAACAACTTTTCCCAAAAGGAACGGTAGACGTCTCCGAAGCC GACGTGGGCCGGGTGGATAATGTAGTCCTGGATGGTTTCAACGGATTCAAAGGTGACTTCGACGATGTGGGTGTAACC TTCTTCCTTGTTCTTCTGGGTGACGTCCTTTCCCCAGTAAACATCCTTCATGGCGGGAATGATGTTGACGAGGTTGACG TAGGTCTTGAAGAATTCTTCCTTCTGGGCTTCGGTGATTTCGTCCTTGAACTTGAGAACGATCAAGTGCTTGACGGCC ATTTTGTCGAAATGTATATATGTGTTGTTGGTAAACTCAAACAGAGATTATGAGGGAGGCCTGAGTGTTAGTGCATGAG ACCCATTTTTCGCTTTATTCTGTTGTCGAGATCTTCTCCTTGCAGGTTCAACAACTGCCGCAGTACTCAAAACGCCTTG AAAGCTACGACAATTTAATCGGCATAATCTGAAGTTGACCGAGAAAGATACGTTAGCCTTGACCGTAATGTTGAAATT GCGAAGCCGAATCGTTTTGTGTCAGAGATACCCGGAGTTTTAGCAGATTTCACAATGTAGTTTTCATGTGCCGACGCA GTCCAATCAGATGTCAACCGACGAGCGAATCCCGCTTCTTACAAACTGAAACATAAGTAAAGAAGTTACTGTTAATTA GAAAGTCACTGAAGGACACGCATCCGTCATGCACCTGTCCAGGTAGCTCCGCACGCTAGGACTAGCTGCTAATGCATT TTGCGCTGGACGGATGAAAAACGATAATACGATGATCATATTATACAAGTACGAANGTNNNTTCAAAGTACCTAAAAA GACATAGCACGCGACAAACATGANCNNNANATCNNNTCAGCTANNNANTATTNGCNGNCATAAAANNNGNTTNNN

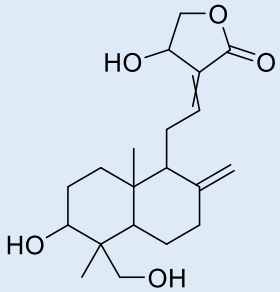
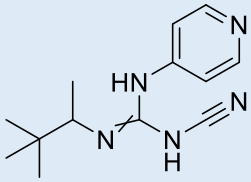
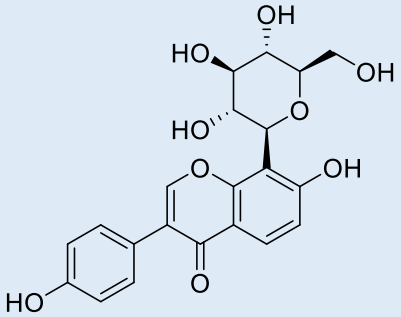
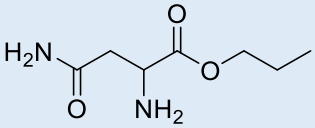
Supplementary Table 2: All primers sequences used in this study.

Identity	**Sequence**
OAC F	AAGCTGCAATACTAGCTTGA
OAC R	AACAAAATTAGAGGTCTTCT
TKS-F	ATGAACCACCTGCGTGCCGA
TKS-R	GTACTTAATCGGGACGGAGCGAACAACA

Supplementary Table 3: Annotated analytes deregulated in transconjuguants. The presence of the heterologous *Cs*TKS in *Phaeodactylum tricornutum*'s metabolism displays variations in the activity of the shikimate pathway, the nucleotide metabolism, and amino acids metabolism. Compounds in pink-orange are all upregulated while compounds in blue are deregulated. They are displayed in order of fold change from top to bottom. Biosynthesis / metabolic route section focuses on the common compounds or pathways between all analytes' biosynthesis possible routes.

Analyte	Chemical 2D structure	Compound Family	Biosynthetic Route
Proline Betaine		Dipeptide, quaternary ammonium amino acid	Pyruvate, glutamate, choline oxidation (Kanehisa et al., 2023, Khanna-Chopra et al., 2019)
3-hydroxy-C6-homoserine lactone		Acyl-homoserine-lactone (AHL)	AHL synthase on S-Adenosyl-Methionine (Hunt, 2010)
Decyl dimethyl amine oxide		Amine oxide	Unelucidated
Thiamine		Vitamin/cofactor	Pyruvate and purine metabolism (Llavero Pasquina, 2020, Kanehisa et al., 2023)
Pyocyanine		Phenazine, pigment	Chorismate, pyruvate (Marey et al., 2024)
D-Carnitine		Quarternary ammonium amino acid	Trimethyllisine, pathway. Fatty acid metabolism

			(Strijbis et al., 2010, Bourdin et al., 2007)
Isoastragaloside IV		Triterpenoid, saponin	Mevalonate (MVA) and methyerythriol phosphate (MEP) pathways (Dong et al., 2023, Xu et al., 2016)
1-methyladenosine		Purine	Purine metabolism (Kanehisa et al., 2023)
Bergenin		Lactone, isocoumarin	Shikimate pathway (Xu et al., 2016)
Phenylalanine		Amino acid	Shikimate pathway (Kanehisa et al., 2023)
Worenine**		Isoquinoline	Unelucidated
Adenosine 3'-monophosphate		Purine	Purine metabolism (Kanehisa et al., 2023)

Andrographolide		Diterpenoid	Mevalonate pathway (Srivastava and Akhila, 2010)
Pinacidil		N-cyanoguanidine	Unelucidated
Puerarin (Isoflavonoid C-glycoside)		Isoflavonoid	L-phenylalanine (Wang et al., 2024)
Asparagine propyl ester		Modified amino acid	Unelucidated

References

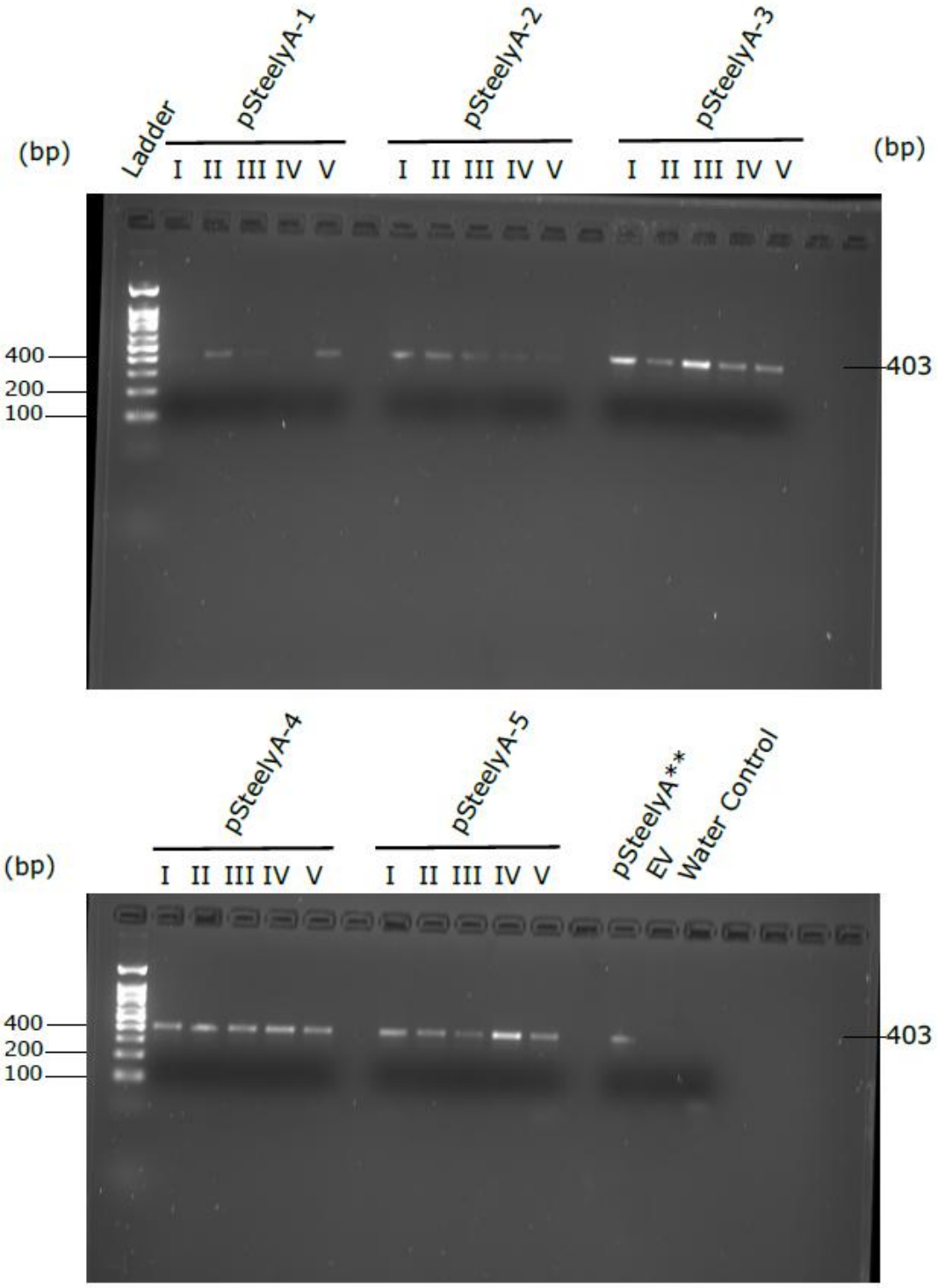
1. Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. KEGG for taxonomy-based analysis of pathways and genomes. Nucleic acids research. 2023;51(D1):D587-D92.
2. Khanna-Chopra R, Semwal VK, Lakra N, Pareek A. Proline–A key regulator conferring plant tolerance to salinity and drought. Plant tolerance to environmental stress: CRC Press; 2019. p. 59-80.
3. Hunt JM. Shiga toxin–producing Escherichia coli (STEC). Clinics in laboratory medicine. 2010;30(1):21-45.
4. Llavero Pasquina, M. (2020). *Molecular biology of B vitamin metabolism genes and their regulation in Phaeodactylum tricornutum* [Apollo - University of Cambridge Repository]. https://doi.org/10.17863/CAM.61337.
5. Marey MA, Abozahra R, El-Nikhely NA, Kamal MF, Abdelhamid SM, El-Kholy MA. Transforming microbial pigment into therapeutic revelation: extraction and characterization of pyocyanin from Pseudomonas aeruginosa and its therapeutic potential as an antibacterial and anticancer agent. Microbial Cell Factories. 2024;23(1):174.
6. Strijbis K, Vaz FM, Distel B. Enzymology of the carnitine biosynthesis pathway. IUBMB life. 2010;62(5):357-62.
7. Bourdin B, Adenier H, Perrin Y. Carnitine is associated with fatty acid metabolism in plants. Plant Physiology and Biochemistry. 2007;45(12):926-31.
8. Dong M, Li J, Yang D, Li M, Wei J. Biosynthesis and pharmacological activities of flavonoids, Triterpene Saponins and polysaccharides derived from Astragalus membranaceus. Molecules. 2023;28(13):5018.

9. Xu C, Li H, Yang X, Gu C, Mu H, Yue Y, et al. Cloning and expression analysis of MEP pathway enzyme-encoding genes in Osmanthus fragrans. Genes. 2016;7(10):78.
10. Srivastava N, Akhila A. Biosynthesis of andrographolide in Andrographis paniculata. Phytochemistry. 2010;71(11-12):1298-304.
11. Wang L, Li C, Luo K. Biosynthesis and metabolic engineering of isoflavonoids in model plants and crops: A review. Frontiers in Plant Science. 2024;15:1384091.

Supplementary Information

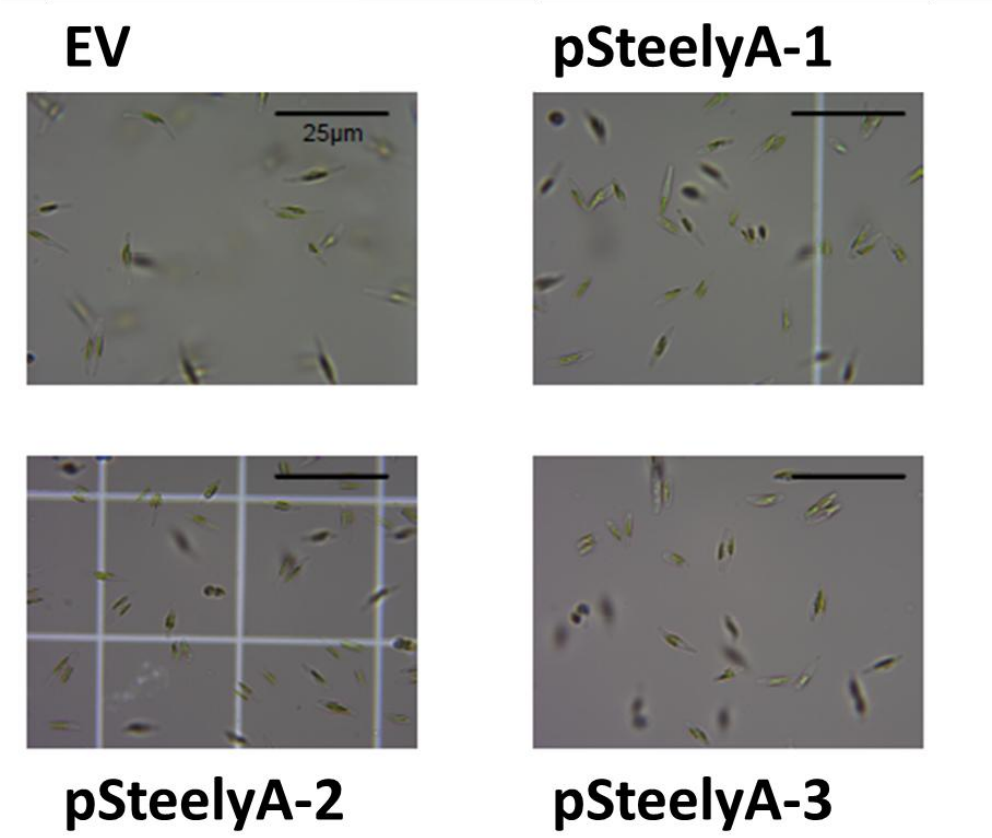
Heterologous expression of the high molecular weight *Dictyostelium discoideum* hybrid enzyme SteelyA in the diatom *Phaeodactylum*

tricornutum

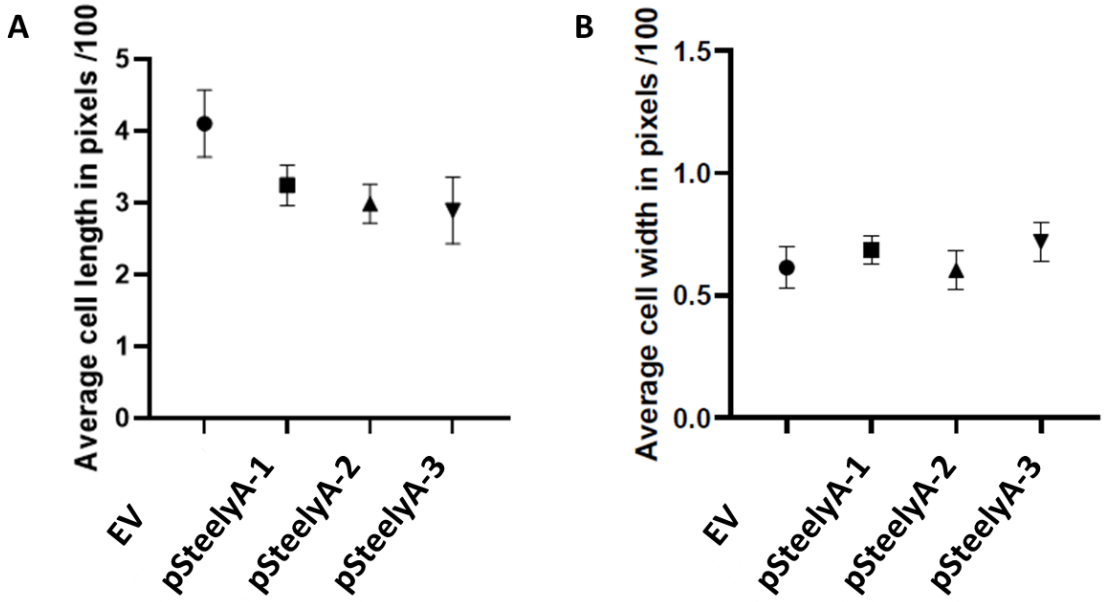


Supplementary Figure 1: *P. tricornutum* SteelyA transconjugants colony PCR. Agarose gel (1%) of five clones from five independent clones. Amplicons were

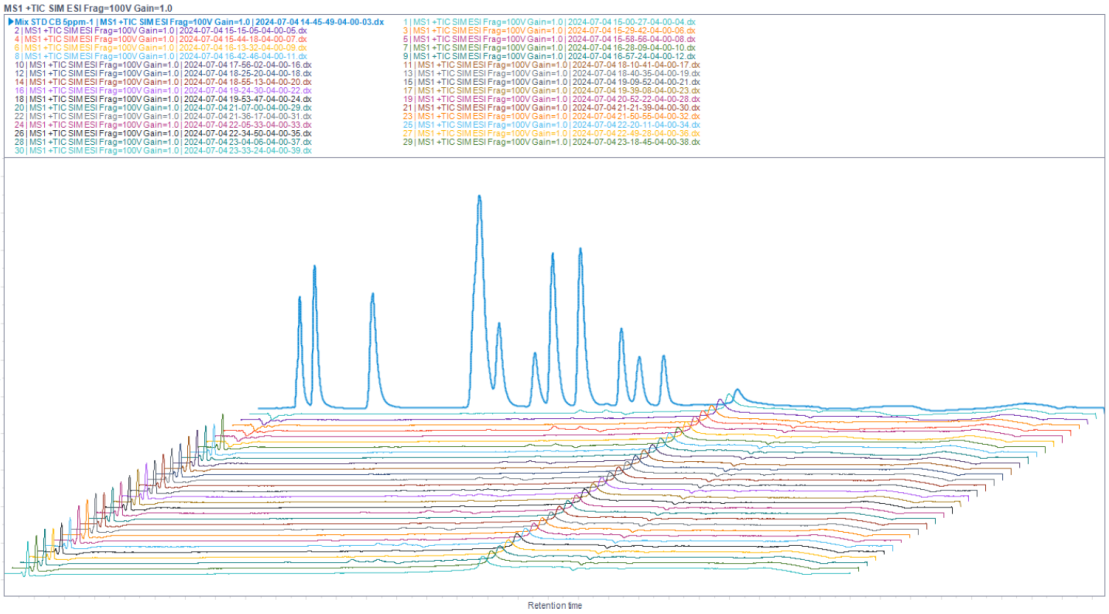
obtained using primers binding in the methyl-transferase domain expected to amplify a 403 bp fragmetn. Transconjugant pSteelyA** is DNA extracted from a previous assembly verified by NGS that displayed a premature stop codon.



Supplementary Figure 2: *P. tricornutum* steely clones show a smaller phenotype by 25% on average compared to pPTGE30. Observation through optic microscope magnification x10.



Supplementary Figure 3: The production of SteelyA reduces *Phaeodactylum tricornutum* length. Aligned dot plot of cells length and width in pixels. A minimum of n=5 cells were measured.



Supplementary Figure 4: No clone displays a peak corresponding to an expect SteelyA product. Total ion chromatogram in single ion monitoring mode of all

tripiclates of EV, pSteelyA-1, pSteelyA-2 and pSteelyA-3. The m/z monitored were 156,181, 182, 195, 196, 225, 273.

Supplementary Table 1: Table of all primers used in this study.

Primer	Sequences
E30-Stshble-F	TCATTAGTTGCAGTCACTCCGCTTTGGTTTCGTAACTATAACGGTCCTAAGGTAGCGAAC
E30-Stshble-R	GTTCGCTACCTTAGGACCGTTATAGTTACGAAACCAAAGCGGAGTGACTGCAACTAATGA
Steely-E30-F	AGGTGTTCTTATGTAGTTTTACACAGGAGTCTGGACTTGACATACAGTGAATGTAACTTT
Steely-E30-R	TAAGAAGAGCACATACCTCAGTCACTTATTATCACTAGCGAAACCAAAGCGGAGTGACTG
F-St-met-seq	GAGAAAAGAGGGTGTTCAGA
R-St-met-seq	GACTTAGGTTCAATACACAA
StlP1-R	CAATTTGTTGATTGGCGTAAACGTGAGTAGCAGACAAATC
StlP2-F	TCTAATTTCAAGTCCTTGAGATCCTTCGGTAGATTGATTG
StlP1-R 30	ATTGGCGTAAACGTGAGTAGCAGACAAATC
StlP2-F 30	AGTCCTTGAGATCCTTCGGTAGATTGATTG
St-E30shble-F	TCATTAGTTGCAGTCACTCCGCTTTGGTTTCATACAGTGAATGTAACTTTCGAATTGACA
St-E30shble-R	GTTCGCTACCTTAGGACCGTTATAGTTACGAAACCAAAGCGGAGTGACTGCAACTAATGA

Supplementary Table 2: All DNA sequences used during the study**.**

DNA Sequence

Genewiz fragment 1 (1MN)

CATACAGTGAATGTAACTTTCGAATTGACAGTATTAGTAGTCGTATTGACAGTGAGGCACGCCCCTCAATGTGCGAGGTGGAAAATATACCAGCA
TGACAATGAATCTTGGAGATTCTTTTGCTGTCATCAAGATTCACCGCCAAATCTTCAGGAACCTATCACGTCCACAGGCGATGTTAATTCTTGAGT
CGTCAAAACAAAGTCCTGTCCTACCTGTAGAAGTTGACAGCGAGCAATTGTATGCAAACTTCTGACTTTGTTATAATAACATTAAAGGTAATTAA
GTATCTTCAATTAGGCATTTTGTCACTGTCAGTCCGTTCCGACAATATAGGTAGATTTGGAATGAATCTTTTCTATGCTGCTGCGAATCTTGTACAC
CTTTGAGGCCGTAGATTCTGTCCGACGAAGCGATAATTATTGCAAAATACATGGACTCATTATTTTGATTCGATTTCTTTTTGGTATCCGACTCGAA
AAGATCCATCACGGCGAGCGCCACCATGAACAAGAACTCCAAAATCCAGTCCCCAAACTCTTCTGATGTTGCTGTTATTGGTGTTGGTTTTAGAT
TCCCAGGTAACTCTAATGACCCAGAATCTTTGTGGAACAACTTGTTGGATGGTTTCGATGCTATTACCCAAGTCCCAAAAGAAAGATGGGCTACT
TCTTTTAGAGAGATGGGTTTGATCAAGAACAAGTTCGGTGGTTTCTTGAAGGATTCTGAATGGAAGAATTTCGACCCTTTGTTCTTTGGTATCGGT
CCAAAAGAAGCTCCATTCATTGATCCACAACAAAGGTTGTTGTTGTCCATCGTTTGGGAATCTTTGGAAGATGCTTACATCAGACCAGATGAATT
GAGAGGTTCTAACACTGGTGTTTTCATCGGTGTTTCTAACAACGATTACACCAAGTTGGGTTTCCAAGACAACTACTCTATTTCTCCATACACTAT
GACCGGCTCTAACTCTTCATTGAACTCCAACAGAATTTCCTACTGCTTCGATTTTAGAGGTCCATCCATTACTGTTGATACCGCTTGTTCTTCTTCC
TTGGTTTCTGTTAATTTGGGTGTCCAATCCATCCAAATGGGTGAATGTAAGATTGCTATTTGCGGTGGTGTTAACGCTTTGTTTGATCCATCTACAT
CTGTTGCCTTTTCCAAGTTGGGTGTTTTGTCTGAAAATGGCAGATGCAACTCTTTTAGTGATCAAGCCTCTGGTTACGTTAGATCTGAAGGTGCTG
GTGTTGTTGTTTTGAAGTCTTTGGAACAAGCTAAGTTGGATGGTGATAGAATCTACGGTGTTATCAAGGGTGTTTCCTCTAATGAAGATGGTGCTT
CTAATGGTGACAAGAACTCTTTGACTACTCCATCTTGTGAAGCCCAATCCATTAACATTTCTAAGGCTATGGAAAAGGCCTCCTTGTCTCCATCTG
ATATCTATTACATTGAAGCCCATGGTACTGGTACTCCAGTTGGTGATCCAATTGAAGTTAAGGCCTTGTCCAAGATCTTCTCCAACTCTAACAACA
ACCAGTTGAACAACTTCTCTACCGATGGTAATGATAACGATGATGATGATGACGATAACACCTCTCCAGAACCATTATTGATTGGCTCATTCAAGT
CCAACATCGGTCATTTGGAATCTGCTGCTGGTATTGCTTCTTTGATTAAGTGTTGCTTGATGTTGAAGAACAGGATGTTGGTTCCATCCATTAACT
GCTCTAATTTGAACCCATCCATTCCATTCGATCAGTACAACATCTCCGTTATCAGAGAAATCAGACAATTCCCAACCGATAAGTTGGTTAACATCG
GTATCAATTCTTTCGGTTTCGGTGGTTCTAACTGCCATTTGATTATTCAAGAGTACAACAACAACTTCAAGAACAACTCTACCATCTGCAATAACA
ACAACAACAATAACAACATCGACTACTTGATCCCAATCTCCTCTAAGACTAAGAAGTCCTTGGATAAGTACTTGATTTTGATCAAGACCAAC
TCCAACTACCACAAGGATATTTCTTTCGATGACTTCGTCAAGTTCCAAATCAAGTCTAAGCAGTACAACTTGTCCAACAGAATGACTACCATTGC
TAACGATTGGAACTCCTTCATTAAGGGTTCTAACGAATTCCACAACTTGATCGAATCTAAGGATGGTGAAGGTGGTTCTTCATCTTCTAACAGAG
GTATTGATTCCGCCAATCAAATCAACACTACTACCTCTACCATCAACGATATCGAACCTTTGTTGGTTTTCGTTTTCTGTGGTCAAGGTCCACA
ATGGAATGGTATGATTAAGACCTTGTACAACTCCGAGAACGTTTTCAAGAACACCGTTGATCATGTTGACAGCATCTTGTACAAGTACTTCGGTTA
CTCCATTTTGAACGTCTTGTCTAAGATCGATGATAACGACGATTCCATCAACCATCCAATAGTTGCTCAACCATCTTTGTTCTTGTTGCAAATTGGT
TTGGTCGAGTTGTTTAAGTACTGGGGTATCTACCCATCTATCTCTGTTGGTCATTCTTTCGGTGAAGTCTCTTCTTATTACTTGTCCGGTATCATCTC
TTTGGAAACCGCTTGTAAAATCGTCTACGTCAGATCCTCTAATCAGAACAAAACTATGGGTTCCGGTAAGATGTTGGTTGTTTCTATGGGTTTTAA
GCAATGGAACGATCAATTCTCTGCTGAATGGTCCGATATTGAAATTGCTTGTTACAACGCTCCAGATTCCATAGTTGTTACTGGTAACGAAGAAAG
ATTGAAAGAATTGTCCATCAAGTTGTCCGACGAATCCAATCAAATTTTCAACACCTTCTTGAGGTCCCCATGTTCTTTTCATTCTTCCCATCAAGA
AGTCATCAAGGGTTCTATGTTCGAAGAGTTGTCTAACTTGCAATCTACTGGTGAAACCGAAATCCCTTTGTTCTCTACTGTTACTGGTAGACAAGT
TTTGTCTGGTCATGTTACTGCTCAACACATCTACGATAATGTTAGAGAACCAGTCTTGTTCCAAAAGACGATTGAATCCATTACCTCCTACATCAA
GTCTCACTACCCATCCAATCAAAAGGTTATCTACGTTGAAATTGCTCCACACCCAACCTTGTTTTCATTGATCAAAAAGTCCATCCATCCTCCAA
CAAGAATTCCTCTTCTGTTTTGTGTCCATTGAACAGAAAAGAAAACTCCAACAACTCCTACAAGAAGTTCGTTTCTCAGTTGTACTTCAACGGTG
TTAACGTTGACTTCAACTTCCAGTTGAACTCCATTTGCGATAACGTTAACAACGATCACCATTTGAACAACGTCAAGCAAAACTCCTTCAAAGAG
ACTACCAATTCCTTGCCAAGATACCAATGGGAACAAGATGAATATTGGTCCGAACCATTGATCTCCAGAAAGAATAGATTGGAAGGTCCAACTAC
TTCCTTGTTGGGTCATAGAATTATCTACAGCTTCCCAGTTTTCCAATCCGTTTTGGACTTGCAATCTGACAACTACAAATACTTGTTGGACCACTTG
GTTAACGGTAAGCCAGTTTTTCCAGGTGCTGGTTATTTGGATATCATCATCGAATTCTTCGACTACCAAAAGCAGCAGTTGAATTCCTCTGATTCC
TCTAACTCCTACATCATCAACGTTGACAAGATCCAATTCTTGAACCCAATTCACTTGACCGAAAACAAGTTGCAAACCTTGCAATCTTCTTTCGA
ACCTATCGTTACTAAGAAGTCTGCCTTCTCTGTTAACTTCTTCATCAAGGATACCGTCGAGGATCAATCTAAGGTTAAGTCTATGTCTGACGAAAC
TTGGACTAACACTTGTAAGGCTACCATTTCCTTGGAACAACAACAGCCATCTCCATCTTCTACTTTGACTTTGTCTAAGAAGCAAGACTTGCAGA
TCTTGAGAAACAGATGCGATATTAGCAAGCTAGACAAGTTTGAGTTGTACGACAAGATCTCTAAGAATTTGGGCTTGCAGTACAACTCCTTGTTT
CAAGTTGTTGATACCATCGAAACTGGTAAGGATTGCTCTTTTGCTACTTTGTCTTTGCCAGAAGATACTTTGTTCACCACCATTTTGAACCCATGC
TTGTTGGATAACTGTTTCCATGGTTTGTTGACCTTGATCAACGAAAAGGGTTCTTTCGTTGTCGAGTCCATTTCTTCTGTTTCTATCTACTTGGAGA
ACATCGGTTCCTTCAATCAAACTTCTGTTGGTAACGTCCAGTTCTACTTGTACACCACTATTTCTAAAGCCACCTCCTTTAGTTCTGAAGGTACTT
GTAAGTTGTTCACCAAGGATGGTTCCTTGATTTTGTCTATCGGTAAGTTCATCATCAAGTCCACCAATCCAAAGTCTACTAAGACCAACGAAACTA
TCGAATCTCCATTGGACGAAACCTTCTCTATTGAATGGCAATCTAAGGATTCTCCAATTCCAACCCCACAACAAATCCAACAACAATCTCCATTGA
ACTCTAACCCATCCTTCATTAGATCTACCATCTTGAAGGACATCCAGTTCGAACAATACTGCTCCTCCATTATCCACAAAGAATTGATCAACCACG
AAAAGTACAAGAACCAGCAATCCTTCGATATCAACTCCTTGGAAAACCACTTGAACGATGACCAATTGATGGAATCCTTGTCCATCTCCAAAGA
ATACTTGAGATTCTTCACCAGGATCATCTCCATCATTAAGCAATACCCAAAGATCTTGAACGAAAAAGAGCTAAAAGAATTGAAAGAAATCATCG
AATTGAAGTACCCATCCGAAGTTCAGTTGTTGGAATTCGAAGTTATCGAGAAGGTGTCCATGATTATCCCAAAGTTGTTGTTCGAAAACGACAAG
CAATCTTCCATGACCTTGTTCCAAGATAACTTGTTGACCAGGTTCTACTCCAATTCTAACTCTACCAGATTCTACTTGGAAAGGGTTTCCGAAATG
GTCTTGGAATCTATTAGACCAATCGTCAGAGAAAAGAGGGTGTTCAGAATTTTGGAAATTGGTGCTGATACAGCCTCTTTGTCTAATGTTGTTTTG
ACTAAGTTGAACACCTACTTGTCCACCTTGAATTCTAATGGTGGTTCTGGTTACAACATCATCATTGAGTACACCTTCACCGATATTTCCGCCAAC
TTCATTATTGGTGAAATCCAAGAAACCATGTGCAACTTGTACCCAAACGTTACTTTCAAGTTCTCCGTCTTGGACTTGGAGAAAGAGATTATTAA
CTCCTCCGATTTCTTGATGGGTGATTACGATATAGTTTTGATGGCCTACGTTATCCATGCCGTTTCTAACATTAAGTTCTCCATCGAACAGTTGTACA
AGTTGTTGTCTCCAAGAGGTTGGTTGTTGTGTATTGAACCTAAGTCCAACGTTGTGTTCTCCGATTTGGTTTTCGGTTGTTTTAATCAGTGGTGGA
ACTACTACGATGATATTAGAACTACCCACTGCTCCTTGTCTGAATCTCAATGGAATCAGTTGTTGTTGAACCAGTCCTTGAACAACGAATCCTCTT
CTTCTTCTAACTGTTACGGTGGTTTCTCCAACGTTTCTTTTATTGGTGGTGAAAAGGATGTCGACTCCCATTCTTTCATATTGCACTGCCAAAAAG
AATCCATCTCCCAAATGAAGTTAGCCACCACTATTAACAACGGTTTGTCATCTGGTTCCATCGTTATCGTTTTGAACTCTCAACAATTGACCAACA
TGAAGTCCTACCCAAAGGTTATTGAGTATATTCAAGAGGCTACCTCTTTGTGCAAGACCATTGAAATTATCGATTCCAAGGACGTCTTGAACTCTA
CCAATTCAGTTTTGGAAAAGATCCAAAAGTCCTTGTTGGTGTTCTGTTTGTTGGGTTATGACTTGTTGGAGAACAACTACCAAGAACAGTCTTTC
GAATACGTTAAGTTGTTGAACTTGATCTCTACTACCGCCTCTTCATCTAATGATAAGAAACCACCAAAGGTCTTGTTGATCACCAAGCAATCTGAA
AGAATCTCCAGGTCTTTCTACTCCAGATCCTTGATTGGTATTTCCAGAACCTCTATGAACGAGTACCCAAATTTGTCCATTACCTCTATCGATTTGG
ATACCAACGACTACTCATTGCAGTCTTTGTTGAAGCCAATCTTCAGCAACTCTAAGTTTTCCGACAACGAGTTCATCTTCAAAAAGGGCTTGATG
TTCGTGTCCAGGATCTTTAAGAACAAGCAGTTGCTAGAATCCTCCAACGCTTTTGAAACTGACTCTTCTAACTTGTACTGTAAGGCCTCTTCTGA
CTTGTCTTACAAGTACGCTATTAAGCAGTCTATGTTGACCGAAAATCAGATCGAAATCAAGGTTGAATGCGTCGGTATTAACTTCAAGGACAACC
TATTCTACAAGGGCTTGTTGCCACAAGAAATTTTCAGAATGGGTGACATCTACAATCCACCATATGGTTGGAATGCTCTGGTGTTATTACCAGAA
TTGGTTCTAACGTCACCGAATACTCAGTTGGTCAAAATGTTTTTGGTTTCGCCAGACATTCTTTGGGTTCTCATGTTGTTACCAACAAGGATTTGG
TTATCTTGAAGCCAGATACCATCTCATTTTCTGAAGCTGCTTCTATCCCAGTTGTTTACTGTACTGCTTGGTACTCCTTGTTCAACATTGGTCAGTT
GTCTAACGAAGAATCCATCCTAATTCATTCTGCTACTGGTGGTGTAGGTTTGGCTTCTTTGAATTTGTTGAAAATGAAGAATCAGCAACAGCAAC
CATTGACCAATGTTTATGCTACTGTTGGCTCTAACGAGAAGAAGAAGGTTCTTGATCGATAACTTCAACAACTTGTTCAAAGAGGACGGCGAAAA
CATTTTCTCTACCAGAGACAAAGAATACTCCAACCAGTTGGAATCCAAGATCGATGTTATTTTGAACACCTTGTCCGGTGAATTCGTCGAATCTAA
TTTCAAGTCCTTGAGATCCTTCGGTAGATTGATTGATTTGTCTGCTACTCACGTTTGGATCCGGAAC

Genewiz Fragment 2 (1Q)

ATCCTTCGGTAGATTGATTGATTTGTCTGCTACTCACGTTTACGCCAATCAACAAATTGGTCTAGGTAACTTCAAGTTCGACCACTTGTATTCTGCT
GTTGACTTGGAAAGATTGATCGACGAAAAACCTAAGTTGTTGCAGTCCATCTTGCAAAGAATTACCAACTCTATCGTCAACGGTTCCTTGGAAA
AAATTCCAATTACCATCTTCCCATCCACCGAAACTAAGGATGCTATCGAATTATTGTCCAAGAGATCCCATATCGGTAAAGTTGTTGTAGATTGCAC
CGATATCTCTAAGTGTAATCCTGTTGGTGATGTGATCACCAACTTCTCTATGAGATTGCCAAAGCCAAACTACCAGTTGAATTTGAACTCCACCTT
GTTGATTACTGGTCAGTCTGGTTTGTCTATCCCTTTGTTGAATTGGTTGTTGTCTAAGTCTGGTGGTAACGTTAAGAACGTTGTCATCATTTCTAAG
TCCACCATGAAGTGGAAGTTGCAGACTATGATTTCCCATTTCGTTTCCGGTTTCGGTATCCATTTTAACTACGTTCAAGTCGACATCTCCAACTAC
GATGCTTTGTCTGAAGCTATTAAGCAATTGCCATCTGATTTGCCACCAATCACCTCTGTTTTTCATTTGGCTGCTATCTACAACGATGTTCCAATGG
ATCAAGTTACCATGTCTACCGTTGAATCTGTTCATAACCCTAAAGTTTTGGGTGCCGTTAACTTGCATAGAATCTCTGTTTCTTTTGGTTGGAAGTT
GAACCACTTCGTCTTGTTCTCTTCTATTACTGCTATTACCGGTTACCCAGACCAATCTATCTACAATTCTGCCAACCTCTATTTTGGACGCTTTGTCCA
ACTTTAGAAGGTTTATGGGTTTGCCATCCTTCTCCATTAACTTGGGTCCAATGAAGGATGAAGGTAAGGTTTCTACCAACAAGAGCATCAAGAAG
CTATTCAAGTCTAGAGGTTTGCCAAGCCTATCCTTGAACAAGTTATTTGGTTTGTTGGAGGTCGTCATCAACAACCCATCTAATCATGTTATCCCAT
CCCAATTGATTTGCTCCCCAATCGATTTCAAGACCTACATCGAATCTTTCTCAACTATGAGGCCAAAGTTGTTACACTTGCAACCTACCATTTCCA
AGCAGCAATCTTCTATCATTAACGATTCTACCAAGGCTTCCTCCAACATTTCATTGCAAGATAAGATCACCTCCAAGGTGTCTGATTTGTTGTCCAT
TCCAATCTCCAAGATCAACTTCGATCATCCATTGAAACACTACGGCTTGGATTCTTTGTTGACCGTTCAATTCAAATCCTGGATCGACAAAGAATT
CGAAAAGAACTTGTTCACCCATATCCAATTGGCCACCATCTCTATTAACTCATTCTTGGAAAAGGTGAACGGCTTGTCTACAAAACGATAACAACA
ACAACAATTCCAACGTCAAGTCCTCTCCATCCATTGTCAAAGAAGAAATCGTTACCTTGGACAAGGATCAACAACCATTGCTATTGAAAGAACA
CCAGCACATTATCATCTCCCCAGATATTAGAATCAACAAGCCAAAGAGGGAATCCTTGATTAGAACCCCAATCTTGAACAAATTCAACCAGATCA
CCGAATCCATTATCACTCCATCTACACCATCTTTGTCCCAATCCGATGTTTTGAAAACTCCACCAATCAAGTCTTTGAACAACACTAAGAACTCCA
GCTTGATTAACACCCCACCAATTCAATCTGTCCAACAACATCAAAAGCAACAACAAAAGGTCCAAGTCATCCAACAACAGCAACAACCATTATC
CAGATTGTCCTACAAGAGCAACAACAACTCTTTCGTTTTGGGTATCGGTATTTCTGTTCCAGGTGAACCTATTTCCCAACAATCCTTGAAAGACTC
CATCTCCAATGACTTTTCTGATAAGGCTGAAACTAACGAGAAGGTCAAGAGAATCTTTGAGCAATCTCAAATCAAGACCAGACACTTGGTTAGA
GATTACACTAAGCCAGAGAACTCCATCAAGTTCAGACATTTGGAAACCATTACCGATGTGAACAACCAGTTCAAGAAAGTTGTTCCAGATTTGG
CTCAACAAGCCTGTTTGAGAGCTTTGAAAGATTGGGGTGGTGATAAGGGTGATATTACCCATATAGTTTCTGTTACCTCCACCGGTATTATCATCCC
AGATGTTAATTTCAAGTTGATCGACTTGTTGGGCTTGAACAAGGATGTTGAAAGAGTGTCTTTGAACCTAATGGGTTGTTTGGCTGGTTTGAGTT
CTTTGAGAACTGCTGCTTCTTTGGCTAAGGCTTCTCCAAGAAATAGAATTTTGGTTGTCTGTACCGAAGTCTGCTCCTTGCATTTTTCTAATACTG

pPTGE30 sequencing

ATGGTGGTGATCAAATGGTCGCCTCTTCTATTTTTGCTGATGGTTCTGCTGCTTACATTATTGGTTGTAACCCAAGAATTGAAGAAACCCCATTATA
CGAAGTCATGTGCTCCATTAACAGATCTTTCCCAAATACCGAAAACGCCATGGTTTGGGATTTGGAAAAAGAAGGTTGGAACTTGGGTTTGGAT
GCTTCTATTCCAATTGTCATTGGTTCTGGTATTGAAGCCTTCGTTGATACTTTGTTGGATAAGGCTAAGTTGCAAACTTCCACTGCTATTTCTGCTA
AGGATTGCGAATTCTTGATTCATACTGGTGGCAAGTCCATCTTGATGAACATCGAAAATTCCTTGGGTATCGACCCAAAGCAAACTAAGAATACTT
GGGATGTTTACCATGCCTACGGCAATATGTCATCTGCCTCTGTTATTTTCGTTATGGATCATGCCAGAAAGTCCAAGTCTTTGCCAACTTACTCAAT
TTCTTTGGCTTTTGGTCCAGGTTTGGCTTTTGAAGGTTGTTTCTTGAAGAACGTCGTCGAACAGAAGCTCATTTCCGAAGAGGACTTGTAACCGC
AACAACTACCTCGACTTTGGCTGGGACACTTTCAGTGAGGACAAGAAGCTTCAGAAGCGTGCTATCGAACTCAACCAGGGACGTGCGGCACAA
ATGGGCATCCTTGCTCTCATGGTGCACGAACAGTTGGGAGTCTCTATCCTTCCTTAAAAATTAATTTTCATTAGTTGCAGTCACTCCGCTTTGGTT
T

ACTAAGGGTACTGTTGACATTGCGAAGAGCGACAAAGATTTTGTTATCGGCTTTATTGCTCAAAGAGACATGGGTGGAAGAGATGAAGGTTACG
ATTGGTTGATTATGACACCCGGTGTGGGTTTAGATGACAAGGGAGACGCATTGGGTCAACAGTATAGAACCGTGGATGATGTGGTCTCTACAGGA
TCTGACATTATTATTGTTGGAAGAGGACTATTTGCAAAGGGAAGGGATGCTAAGGTAGAGGGTGAACGTTACAGAAAAGCAGGCTGGGAAGCAT
ATTTGAGAAGATGCGGCCAGCAAAACTAAAAAACTGTATTATAAGTAAATGCATGTATACTAAACTCACAAATTAGAGCTTCAATTTAATTATATC
AGTTATTACCCGGCCGGGAATCTCGGTCGTAATGATTTTTATAATGACGAAAAAAAAAAAAATTGGAAAGAAAACCCCCCCCCCCCCCGCAGCG
TTGGGTCCTGGCCACGGGTGCGCATGATCGTGCTCCTGTCGTTGAGGACCCGGCTAGGCTGGCGGGGTTGCCTTACTGGTTAGCAGAATGAATC
ACCGATACGCGAGCGAACGTGAAGCGACTGCTGCTGCAAAACGTCTGCGACCTGAGCAACAACATGAATGGTCTTCGGTTTCCGTGTTTCGTAA
AGTCTGGAAACGCGGAAGTCAGCGCCCTGCACCATTATGTTCCGGATCTGCATCGCAGGATGCTGCTGGCTACCCTGTGGAACACCTACATCTGT
ATTAACGAAGCGCTGGCATTGACCCTGAGTGATTTTTCTCTGGTCCCGCCGCATCCATACCGCCAGTTGTTTACCCTCACAACGTTCCAGTAACCG
GGCATGTTCATCATCAGTAACCCGTATCGTGAGCATCCTCTCTCGTTTCATCGAGTTACGCTAGGGATAACAGGGTAATATAGATCTTCCGCTGCAT
AACCCTGCTTCGGGGTCATTATAGCGATTTTTTCGGTATATCCATCCTTTTTCGCACGATATACAGGATTTTGCCAAAGGGTTCGTGTAGACTTTCC
TTGGTGTATCCAACGGCGTCAGCCGGGCAGGATAGGTGAAGTAGGCCCACCCGCGAGCGGGTGTTCCTTCTTCACTGTCCCTTATTCGCACCTGG
CGGTGCTCAACGGGAATCCTGCTCTGCGAGGCTGGCCGGCTACCGCCGGCGTAACAGATGAGGGCAAGCGGATGGCTGATGAAACCAAGCCAA
CCAGGAAGGGCAGCCCACCTATCAAGGTGTACTGCCTTCCAGACGAACGAAGAGCGATTGAGGAAAAGGCGGCGGCGGCCGGCATGAGCCTG
TCGGCCTACCTGCTGGCCGTCGGCCAGGGCTACAAAATCACGGGCGTCGTGGACTATGAGCACGTCCGCGAGCTGGCCCGCATCAATGGCGACC
TGGGCCGCCTGGGCGGCCTGCTGAAACTCTGGCTCACCGACGACCCGCGCACGGCGCGGTTCGGTGATGCCACGATCCTCGCCCTGCTGGCGA
AGATCGAAGAGAAGCAGGACGAGCTTGGCAAGGTCATGATGGGCGTGGTCCGCCCGAGGGCAGAGCCATGACTTTTTTAGCCGCTAAAACGGC
CGGGGGGTGCGCGTGATTGCCAAGCACGTCCCCATGCGCTCCATCAAGAAGAGGCACTTCGAGCTGTAAGTACATCACCGACGAGCAAGGCAA
GACGATCCGCGCCTGTTTTATTGAGAACGTTGTTCGTGTTGGCCTCAATGGTAGCGATGCGTCATTCAGCGAAGTTCTGGTGCTGATGATGTGGTT
CGCTTTGCCACTGGTCAATGTGGTAAGCCCGTGTAATGTCAGTAACCTTTTTACTGATCTCAGCTTGAGCACGGTCGCTGATGAGCTTATCCATGG
CCCCACGGTAACGGATATGATCCTCTAGGGCGTTGACAAATTCTTTGTCGGTTTTCATGGGGTAGACGTCAGTGACCGAAGGAAACGTCTCCCTACA
AAAAGTTGCGCATACTTTGCTCCGTTGTCGACGGCAGGGGTGTCGGACCAGATTGTATCTGTGGCAACGGCCTCATTGCGTCAATGGACCTTTAA
AGCCGGGAAACGAGACTTGAAATGTTTACGTAGAGGTGCATTATAGACCTCACGCGCATATTGAGTGGTTGCAAAAATGGTTTTGCGAACGGTAT
CACTGGAGACCCATGCCAGGCAAGGACGGAGGGCATCATAATCATGTTCATTGCGTTGGACAGCATGCTTGTTACACGTAAGTATACGGTCGAGA
GCATCGTGACGTTGAGAATTACAGATGGCAACGTGTCGGTGCATGTATTGGCCGAGATTGTCAAATCGTGGCTCAATATATGGCATGCCAGGTAA
GTCATGGATGTCGGTATGCCATTCAGCCTTCATGTCGATTTCGTTGTCGAGAACTGATGGGTCCCAGTCAACGTCAGAAGTAAGAACAACATGGG
GAAGGGAATGAAGTTCATGACTGGTGGGTGCGCGCATATCCATGTACGGAAGACCCTGTCGTATGTTTAGTGGGATGATGTAATCATCCAGGGTA
ACAATACGTTGGTTGCCGCCCACGGTACGAGACCGATCTTGAACACAATTGTGGCAATGTTCAAGTTGGGCGCTGGAATGGATGGTTTTACCTTT
ACCAAGATGCGCGTATTGATGCATGATAACAACAATGGGCCCTCGTTGCGATTCGACAAGACCGGCAGCCGTGACAATGTCTAGATTGGACAAG
GTATGTTCATTTATACCGGTTATGTTGGCTGAACGTCCAGTTTTGTGCAGTACGGTAACATCACTCCCCGCAAGTCCACCGTTGGCACCGCGATCG
ACAAGTGCAGATGTGGTATTTTGTACTGTGTGTCGGGAGACTTGGTACTGTAACACATTTGACTGAAGTGACTTGGTGGTACCATGAGGACGGA
GGGAACCCGTTGATACGGTGTTGCGTGACGCTGCAAGTACCTTACGTATGTCGCCTGGATCCATACGTCCAACCCGATCAGAAAGATGTGCTAGG
AGTTCCGTTTCTGGTGCACAATCATGAAAGGTATCCACGGGAGTTCCATTGTTGTCGTCCGGGGCTGCGTCGCGACTGCCAGTTGCATGTACATG
CGCCATCAAGGGACGTGCCAATCCATTACTAACCGGGAGAACCTTAGCTGGGGCAGCCAATCCCTGAAGTATAGCCTTAGCATCATCTGAGAGCT
GGTTCCACATATCTCTAGGGATATAAGGTCGTTCACGGTTAGTGGGCGTACTATGATGTCGGTTATTAGTACAATTCTCCCTGCGGGCATGCGCATT
GGCTTCATAGAGTACGGAAGGTGACAAGTCAATATTGTAGTTAACATCCGGATCAGTGTCAAAGTCTGTAGGATGGTAAGAAAGATCAGTAGAAT
GAATACTACGCTTTTCCTTGGGACTACGGGAATTAGAGAAGTTGTTTCCTTTATTGTAGAGTGACGCTGAAGCAAGTAGAAGACTAAGGTAACTC
TCGTAGCTAATAGGATTACCTCCTTTGGCTAGGTCAAGAGTGGCTGTGATCTTCACTTGACAAAGTTCCGGTACATTGTGGACAGCATTCTCCAA
AAGACTAAGACACAGTTGCTTTGGGAGTTGCTCAGCCATTGGTACAGTATTGTGGTAGATACAAAGGTGGTTCTTCCAATGAAGGATAAATCCTT
CTGCTGTGCCTGTCCATGAGGATCCATATTTCGCCGTAGTTAGGTAACCAAGCGTGGTGGCTGAACTGATCTTCGCACTTGCTGATTCCGTATAGT
GTTTGACAACTTTACAAAACACTTCTTGCGCAGTTCGCTCTAGCCTGGACAGAGAAGGGACAATCTTCGTCGTTAAGACTCGTGCGAATAGCAA
AAGATCACAAAATAGCACATCGGCACCGACCAACGATTATTTCCAAGGAAAAAAAGAATGCTTCACTACAAGAAATTGTGTCATCCCTATACAG
AGTCTTGTTACTGTGACAGAAAATTGATGGAAGATGTGGCGGATTGCCTTTACACTAGCCAACTTGTTCGACTAATTGCAGCTTCTTCTGAGAGG
CTTCACCGAGTAACGCGAAGAACACCGGTGTCTCGTACATGCTCGTCGGTGAACGCTCGTCCAATGACACCCCCCACTTTGTATCAATATCCCAA
CTTGGTAGTGAACTGGAATGATACATGCAATTTCGCGCCGCATCAACAGCCACGGGCACCATCGACGAATAGACTCGGTCGAGCTGGTTGCCCT
CGCCGCTGGGCTGGCGGCCGTCTATGGCCCTGCAAACGCGCCAGAAACGCCGTCGAAGCCGTGTGCGAGACACCGCGGCCGGCCGCCGGCGTT
GTGGATACCTCGCGGAAAACTTGGCCCTCACTGACAGATGAGGGGCGGACGTTGACACTTGAGGGGCCGACTCACCCGGCGCGGCGTTGAACAG
ATGAGGGGCAGGCTCGATTTCGGCCGGCGACGTGGAGCTGGCCAGCCTCGCAAATCGGCGAAAACGCCTGATTTTACGCGAGTTTCCCACAGAT
GATGTGGACAAGCCTGGGGATAAGTGCCCTGCGGTATTGACACTTGAGGGGCGCGACTACTGACAGATGAGGGGCGCGATCCTTGACACTTGA
GGGGCAGAGTGCTGACAGATGAGGGGCGCACCTATTGACATTTGAGGGGCTGTCCACAGGCAGAAAATCCAGCATTTGCAAGGGTTTCCGCCC
GTTTTTCGGCCACCGCTAACCTGTCTTTTAACCTGCTTTTAAACCAATATTTATAAACCTTGTTTTTAACCAGGGCTGCGCCCTGTGCGCGTGACC
GCGCACGCCGAAGGGGGGTGCCCCCCCTTCTCGAACCCTCCCGGTCGAGTGAGCGAGGAAGCACCAGGGAACAGCACTTATATATTCTGCTTAC
ACACGATGCCTGAAAAAACTTCCCTTGGGGTATCCACTTATCCACGGGGATATTTTTATAATTATTTTTTTATAGTTTTAGATCTTCTTTTTTAGA
GCGCCTTGTAGGCCTTTATCCATGCTGGTTCTAGAGAAGGTGTTGTGACAAATTGCCCTTTCAGTGTGACAAATCACCCTCAAATGACAGTCCTG
TCTGTGACAAATTGCCCTTAACCCTGTGACAAATTGCCCTCAGAAGAAGCTGTTTTTTCACAAAGTTATCCCTGCTTATTGACTCTTTTTTATTTAG
TGTGACAATCTAAAAACTTGTCACACTTCACATGGATCTGTCATGGCGGAAACAGCGGTTATCAATCACAAGAAACGTAAAAATAGCCCGCGAA
TCGTCCAGTCAAACGACCTCACTGAGGCGGCATATAGTCTCTCCCGGGATCAAAAACGTATGCTGTATCTGTTCGTTGACCAGATCAGAAAATCT
GATGGCACCCTACAGGAACATGACGGTATCTGCGAGATCCATGTTGCTAAATATGCTGAAATATTCGGATTGACCTCTGCGGAAGCCAGTAAGGA
TATACGGCAGGCATTGAAGAGTTTCGCGGGGAAGGAAGTGGTTTTTTATCGCCCTGAAGAGGATGCCGGCGATGAAAAAGGCTATGAATCTTTT
CCTTGGTTTATCAAACGTGCGCACAGTCCATCCAGAGGGCTTTACAGTGTACATATCAACCCATATCTCATTCCCTTCTTTATCGGGTTACAGAACC
GGTTTACGCAGTTTCGGCTTAGTGAAACAAAAGAAATCACCAATCCGTATGCCATGCGTTTATACGAATCCCTGTGTCAGTATCGTAAGCCGGATG
GCTCAGGCATCGTCTCTCTGAAAATCGACTGGATCATAGAGCGTTACCAGCTGCCTCAAAGTTACCAGCGTATGCCTGACTTCCGCCGCCGCTTC
CTGCAGGTCTGTGTTAATGAGATCAACAGCAGAACTCCAATGCGCCTCTCATACATTGAGAAAAAGAAAGGCCGCCAGACGACTCATATCGTATT
TTCCTTCCGCGATATCACTTCCATGACGACAGGATAGTCTGAGGGTTATCTGTCACAGATTTGAGGGTGGTTCGTCACATTTGTTCTGACCTACTG
AGGGTAATTTGTCACAGTTTTGCTGTTTCCTTCAGCCTGCATGGATTTTCTCATACTTTTTGAACTGTAATTTTTAAGGAAGCCAAATTTGAGGGC
AGTTTGTCACAGTTGATTTCCTTCTCTTTCCCTTCGTCATGTGACCTGATATCGGGGGTTAGTTCGTCATCATTGATGAGGGTTGATTATCACAGTT
TATTACTCTGAATTGGCTATCCGCGTGTGTACCTCTACCTGGAGTTTTTCCCACGGTGGATATTTCTTCTTGCGCTGAGCGTAAGAGCTATCTGACA
GAACAGTTCTTCTTTGCTTCCTCGCCAGTTCGCTCGCTATGCTCGGTTACACGGCTGCGGCGAGCATCACGTGCTATAAAAATAATTATAATTTAA
ATTTTTTAATATAAATATATAAATTAAAAATAGAAAGTAAAAAAAGAAATTAAAGAAAAAATAGTTTTTGTTTTCCGAAGATGTAAAAGACTCTAG
GGGGATCGCCAACAAATACTACCTTTTACCTTGCTCTTCCTGCTCTCAGGTATTAATGCCGAATTGTTTCATCTTGTCTGTGTAGAAGACCACACA
CGAAAATCCTGTGATTTTACATTTACTTATCGTTAATCGAATGTATATCTATTTAATCTGCTTTTCTTGTCTAATAAATATATATGTAAAGTACGCTTT
TTGTTGAAATTTTTTAAACCTTTGTTTATTTTTTTTTCTTCATTCCGTAACTCTTCTACCTTCTTTATTTACTTTCTAAAATCCAAATACAAAACATAA
AAATAAATAAACACAGAGTAAATTCCCAAATTATTCCATCATTAAAAGATACGAGGGCGCGTGTAAGTTACAGGCAAGCGATCCTAGTACACTCTAT
ATTTTTTTATGCCTCGGTAATGATTTTCATTTTTTTTTTCCACCTAGCGGATGACTCTTTTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTATA
CATTATATAAAGTAATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAGATGACAGAGCAGAAAGCCCTA
GTAAAGCGTATTACAAATGAAACCAAGATTCAGATTGCGATCTCTTTAAAGGGTGGTCCCCTAGCGATAGAGCACTCGATCTTCCCAGAAAAAGA
GGCAGAAGCAGTAGCAGAACAGGCCACACAATCGCAAGTGATTAACGTCCACACAGGTATAGGGTTTCTGGACCATATGATACATGCTCTGGCC
AAGCATTCCGGCTGGTCGCTAATCGTTGAGTGCATTGGTGACTTACACATAGACGACCATCACACCACTGAAGACTGCGGGATTGCTCTCGGTCA
AGCTTTTAAAGAGGCCCTAGGGGCCGTGCGTGGAGTAAAAAGGTTTGGATCAGGATTTGCGCCTTTGGATGAGGCACTTTCCAGAGCGGTGGTA
GATCTTTCGAACAGGCCGTACGCAGTTGTCGAACTTGGTTTGCAAAGGGAGAAAGTAGGAGATCTCTCTTGCGAGATGATCCCGCATTTTCTTGA
AAGCTTTGCAGAGGCTAGCAGAATTACCCTCCACGTTGATTGTCTGCGAGGCAAGAATGATCATCACCGTAGTGAGAGTGCGTTCAAGGCTCTT
GCGGTTGCCATAAGAAGAAGCCACCTCGCCCAATGGTACCAACGATGTTCCCTCCCACCAAAGGTGTTCTTATGTAGTTTTACACAGGAGTCTGGAC
TTGACGCTAGTGATAATAAGTGACTGAGGTATGTGCTCTTCTTATCTCCTTTTGTAGTGTTGCTCTTATTTTAAACAACTTTGCGGTTTTTTGATGAC
TTTGCGATTTTGTTGTTGCTTTGCAGTAAATTGCAAGATTTAATAAAAAAACGCAAAGCAATGATTAAAGGATGTTCAGAATGAAACTCATGGAA
ACACTTAACCAGTGCATAAACGCTGGTCATGAAATGACGAAGGCTATCGCCATTGCACAGTTTAATGATGACAGCCCGGAAGCGAGGAAATAA
CCCGGCGCTGGAGAATAGGTGAAGCAGCGGATTTAGTTGGGGTTTCTTCTCAGGCTATCAGAGATGCCGAGAAAGCAGGGCGACTACCGCACCC
GGATATGGAAATTCGAGGACGGGTTGAGCAACGTGTTGGTTATACAATTGAACAAATTAATCATATGCGTGATGTGTTTGGTACGCGATTGCGACG
TGCTGAAGACGTATTTCCACCGGTGATCGGGGTTGCTGCCCATAAAGGTGGCGTTTACAAAACCTCAGTTTCTGTTCATCTTGCTCAGGATCTGG
CTCTGAAGGGGCTACGTGTTTTGCTCGTGGAAGGTAACGACCCCCAGGGAACAGCCTCAATGTATCACGGATGGGTACCAGATCTTCATATTCAT
GCAGAAGACACTCTCCTGCCTTTCTATCTTGGGGAAAAGGACGATGTCACTTATGCAATAAAGCCCACTTGCTGGCCGGGGCTTGACATTATTCC

pSteelyA-1 sequencing

TTCCTGTCTGGCTCTGCACCGTATTGAAACTGAGTTAATGGGCAAATTTGATGAAGGTAAACTGCCCACCGATCCACACCTGATGCTCCGACTGG
CCATTGAAACTGTTGCTCATGACTATGATGTCATAGTTATTGACAGCGCGCCTAACCTGGGTATCGGCACGATTAATGTCGTATGTGCTGCTGATGT
GCTGATTGTTCCCACGCCTGCTGAGTTGTTTGACTACACCTCCGCACTGCAGTTTTTCGATATGCTTCGTGATCTGCTCAAGAACGTTGATCTTAA
AGGGTTCGAGCCTGATGTACGTATTTTGCTTACCAAATACAGCAATAGCAATGGCTCTCAGTCCCCGTGGATGGAGGAGCAAATTCGGGATGCCT
GGGGAAGCATGGTTCTAAAAAATGTTGTACGTGAAACGGATGAAGTTGGTAAAGGTCAGATCCGGATGAGAACTGTTTTTGAACAGGCCATTGA
TCAACGCTCTTCAACTGGTGCCTGGAGAAATGCTCTTTCTATTTGGGAACCTGTCTGCAATGAAATTTTCGATCGTCTGATTAAACCACGCTGGG
AGATTAGATAATGAAGCGTGCGCCTGTTATTCCAAAACATACGCTCAATACTCAACCGGTTGAAGATACTTCGTTATCGACACCAGCTGCCCCGAT
GGTGGATTCGTTAATTGCGCGCGTAGGAGTAATGGCTCGCGGTAATGCCATTACTTTGCCTGTATGTGGTCGGGATGTGAAGTTTACTCTTGAAGT
GCTCCGGGGTGATAGTGTTGAGAAGACCTCTCGGGTATGGTCAGGTAATGAACGTGACCAGGAGCTGCTTACTGAGGACGCACTGGATGATCTC
ATCCCTTCTTTTCTACTGACTGGTCAACAGACACCGGCGTTCGGTCGAAGAGTATCTGGTGTCATAGAAATTGCCGATGGGAGTCGCCGTCGTAA
AGCTGCTGCACTTACCGAAAGTGATTATCGTGTTCTGGTTGGCGAGCTGGATGATGAGCAGATGGCTGCATTATCCAGATTGGGTAACGATTATCG
CCCAACAAGTGCTTATGAACGTGGTCAGCGTTATGCAAGCCGATTGCAGAATGAATTTGCTGGAAATATTTCTGCGCTGGCTGATGCGGAAAATA
TTTCACGTAAGATTATTACCCGCTGTATCAACACCGCCAAATTGCCTAAATCAGTTGTTGCTCTTTTTTCTCACCCCGGTGAACTATCTGCCCGGTC
AGGTGATGCACTTCAAAAAGCCTTTACAGATAAAGAGGAATTACTTAAGCAGCAGGCATCTAACCTTCATGAGCAGAAAAAAGCTGGGGTGATA
TTTGAAGCTGAAGAAGTTATCACTCTTTTAACTTCTGTGCTTAAAACGTCATCTGCATCAAGAACTAGTTTAAGCTCACGACATCAGTTTGCTCCT
GGAGCGACAGTATTGTATAAGGGCGATAAAATGGTGCTTAACCTGGACAGGTCTCGTGTTCCAACTGAGTGTATAGAGAAAATTGAGGCCATTCT
TAAGGAACTTGAAAAGCCAGCACCCTGATGCGACCTCGTTTTAGTCTACGTTTATCTGTCTTTACTTAATGTCCTTTGTTACAGGCCAGAAAGCAT
AACTGGCCTGAATATTCTCTCTGGGCCCACTGTTCCACTTGTATCGTCGGTCTGATAATCAGACTGGGACCACGGTCCCACTCGTATCGTCGGTCT
GATTATTAGTCTGGGACCATGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTC
TGGAACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCACG
ATCCCACTCGTGTTGTCGGTCTGATTATCGGTCTGGGACCACGGTCCCACTTGTATTGTCGATCAGACTATCAGCGTGAGACTACGATTCCATCAA
TGCCTGTCAAGGGCAAGTATTGACATGTCGTCGTAACCTGTAGAACGGAGTAACCTCGGTGTGCGGTTGTATGCCTGCTGTGGATTGCTGCTGTG
TCCTGCTTATCCACAACATTTTGCGCACGGTTATGTGGACAAAATACCTGGTTACCCAGGCCGTGCCGGCACGTTAACCGGGCTGCATCCGATGC
AAGTGTGTCGCTGTCGACGAGCTCGCGAGCTCGGACATGAGGTTGCCCCGTATTCAGTGTCGCTGATTTGTATTGTCTGAAGTTGTTTTACGTTA
AGTTGATGCAGATCAATTAATACGATACCTGCGTCATAATTGATTATTTGACGTGGTTTGATGGCCTCCACGCACGTTGTGATATGTAGATGATAAT
CATTATCACTTTACGGGTCCTTTCCGGTGATCCGACAGGTTACGGGGCGGCGACCTCGCGGGTTTTCGCTATTTATGAAAATTTTCCGGTTTAAGG
CGTTTCCGTTCTTCTTCGTCATAACTTAATGTTTTTATTTAAAATACCCTCTGAAAAGAAAGGAAACGACAGGTGCTGAAAGCGAGCTTTTTGGC
CTCTGTCGTTTCCTTTCTCTGTTTTTGTCCGTGGAATGAACAATGGAAGTCCGAGCTCATCGCTAATAAACTTCGTATAGCATACATTATACGAAGTT
ATATTCGATGCGGCCGCAAGGGGTTCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAG
AGTGCACCATATGCGGTGTGAAATACCACACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCCATTCGCCATTCAGCTGCGCAACTGTTGGG
AAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTC
CCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTGTAATACGACTCACTATAGGGCGAATTCGAGCTCGGTACCCGGGGATCCTCTAGAGTC
GACCTGCAGGCATGCAAGCTTGAGTATTCTATAGTCTCACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCG
CTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTC
ACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGAACCCCTTGCGGCCGCCCGGGCCGTCGACCAA
TTCTCATGTTTGACAGCTTATCATCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTTTAAGGGCACCAATAA
CTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACTGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGC
ATGATGAACCTGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAACGGGGGCGAAGAAGTTGTCCATAT
TGGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAG
GTTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGTGGTATTCACTCCAGAGCGATGAAAACGTTTCAG
TTTGCTCATGGAAAACGGTGTAACAAGGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGAAATTCCGGATGAGCATTCA
TCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTTTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTC
TGGTTATAGGTACATTGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGGTGGTATATCCAGTGATTTTTT
TCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGATAACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCT
TACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACAGGGACACCAGGATTTATTTATTCTGCGAAGTGAT
CTTCCGTCACAGGTATTTATTCGCGATAAGCTCATGGAGCGGCGTAACCGTCGCACAGGAAGGACAGAGAAAGCGCGGATCTGGGAAGTGACG
GACAGAACGGTCAGGACCTGGATTGGGGAGGCGGTTGCCGCCGCTGCTGCTGACGGTGTGACGTTCTCTGTTCCGGTCACACCACATACGTTCC
GCCATTCCTATGCGATGCACATGCTGTATGCCGGTATACCGCTGAAAGTTCTGCAAAGCCTGATGGGACATAAGTCCATCAGTTCAACGGAAGTC
TACACGAAGGTTTTTGCGCTGGATGTGGCTGCCCGGCACCGGGTGCAGTTTGCGATGCCGGAGTCTGATGCGGTTGCGATGCTGAAACAATTATC
CTGAGAATAAATGCCTTGGCCTTTATATGGAAATGTGGAACTGAGTGGATATGCTGTTTTTGTCTGTTAAACAGAGAAGCTGGCTGTTATCCACTG
AGAAGCGAACGAAACAGTCGGGAAAATCTCCCATTATCGTAGAGATCCGCATTATTAATCTCAGGAGCCTGTGTAGCGTTTATAGGAAGTAGTGT
TCTGTCATGATGCCTGCAAGCGGTAACGAAAACGATTTGAATATGCCTTCAGGAACAATAGAAATCTTCGTGCGGTGTTACGTTGAAGTGGAGCG
GATTATGTCAGCAATGGACAGAACAACCTAATGAACACAGAACCATGATGTGGTCTGTCCTTTTACAGCCAGTAGTGCTCGCCGCAGTCGAGCG
ACAGGGCGAAGCCCATCGATACTAGCTTGATTGGGATATCTCGCTCGTGCTTGTCGCGTGCTATGTCTTTTAGGGTACTTTGAACCTACGTTCGT
ACTTGTGTAATATGATCATCGTCGTCATCGTATTATCGTTTTTCATCCGTCCAGCGCAAAATGCATTAGCAGCTAGTCCTAGCGTGCGGAGCTACCT
GTACAGGTGCATGACGGATGCGTGTCCTTAAGTGAGTTTCTAATTAACAGTAACTTCTTTACTTATGTTTCAGTTTGTAAGAAGCGGGATGCGCTC
CTTCCGTCACAGGTATTTATTCGCGATAAGCTCATGGAGCGGCGTAACCGTCGCACAGGAAGGACAGAGAAAGCGCGGATCTGGGAAGTGACG
GACAGAACGGTCAGGACCTGGATTGGGGAGGCGGTTGCCGCCGCTGCTGCTGACGGTGTGACGTTCTCTGTTCCGGTCACACCACATACGTTCCGC
GCCATTCCTATGCGATGCACATGCTGTATGCCGGTATACCGCTGAAAGTTCTGCAAAGCCTGATGGGACATAAGTCCATCAGTTCAACGGAAGTCTAC
GTACTGCGGCAGTTGTTGAACCTGCAAGGAGAAGATCTCGACAACAGAATAAAGCGAAAAATGGGTCTCATGCACTAACACTCAGGCCTCCCT
CATAATCTCTGTTTGAGTTTACCAACAACACATATATACATTTCGACAAAATGGCCAAGTTGACCAGTGCCGTTCCGGTGCTCACCGCGCGCGAC
GTCGCCGGAGCGGTCGAGTTCTGGACCGACCGGCTCGGGTTCTCCCGGGACTTCGTGGAGGACGACTTCGCCGGTGTGGTCCGGGACGACGTG
ACCCTGTTCATCAGCGCGGTCCAGGACCAGGTGGTGCCAGACAACACCCTGGCCTGGGTGTGGGTGCGCGGCCTGGACGAGCTGTACGCCGAG
TGGTCGGAGGTCGTGTCCACGAACTTCCGGGACGCCTCCGGGCCGGCCATGACCGAGATCGGCGAGCAGCCGTGGGGGCGGGAGTTCGCCCTG
CGCGACCCGGCCGGCAACTGCGTGCACTTCGTGGCCGAGGAGCAGGACTGACCGACGCCGACCAACACCGCCGGTCCGACGCGGCCCGACGG
GTCCGAGGCCTCGGAGATCTGGGCCCATGCGGCCGCAACAACTACCTCGACTTTGGCTGGGACACTTTCAGTGAGGACAAGAAGCTTCAGAAG
CGTGCTATCGAACTCAACCAGGGACGTGCGGCACAAATGGGCATCCTTGCTCTCATGGTGCACGAACAGTTGGGAGTCTCTATCCTTCCTTAAAA
ATTTAATTTTCATTAGTTGCAGTCACTCCGCTTTGGTTTCGTAACTATAACGGTCCTAAGGTAGCGAACGTTGCAGGCCATGCTGTCCAGGCAGGT
AGATGACGACCATCAGGGACAGCTTCAAGGATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCATTGGACCGCTGATCGTCACGGCGATTTATG
CCGCCTCGGCGAGCACATGGAACGGGTTGGCATGGATTGTAGGCGCCGCCCTATACCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAG
CCGGGCCACCTCGACCTGAATGGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAAGAATTGGAGCCAATCAATTCTTGCGGAGAACTG
TGAATGCGCAAACCAACCCTTGGCAGAACATATCCATCGCGTCCGCCATCTCCAGCAGCCGCACGCGGCGCATCGGGGGGGGGGGGGTTTCAAT
TCATCATTTTTTTTTTATTCTTTTTTTTGATTTCGGTTTCCTTGAAATTTTTTTGATTCGGTAATCTCCGAACAGAAGGAAGAACGAAGGAAGGAGC
ACAGACTTAGATTGGTATATATACGCATATGTAGTGTTGAAGAAACATGAAATTGCCCAGTATTCTTAACCCAACTGCACAGAACAAAAACCTGC
AGGAAACGAAGATAAATCATGTCGAAAGCTACATATAAGGAAACGTGCTGCTACTCATCCTAGTCCTGTTGCTGCCAAGCTATTTAATATCATGCAC
GAAAAGCAAACAAACTTGTGTGCTTCATTGGATGTTCGTACCACCAAGGAATTACTGGAGTTAGTTGAAGCATTAGGTCCCAAAATTTGTTTACT
AAAAACACATGTGGATATCTTGACTGATTTTTCCATGGAGGGCACAGTTAAGCCGCTAAAGGCATTATCCGCCAAGTACAATTTTTTACTCTTCGA
AGACAGAAAATTTGCTGACATTGGTAATACAGTCAAATTGCAGTACTCTGCGGGTGTATACAGAATAGCAGAATGGGCAGACATTACGAATGCAC
ACGGTGTGGTGGGCCCAGGTATTGTTAGCGGTTTGAAGCAGGCGGCGGAAGAAGTAACAAAGGAACCTAGAGGCCTTTTGATGTTAGCAGAAT
TGTCATGCAAGGGCTCCCTAGCTACTGGAGAATAT

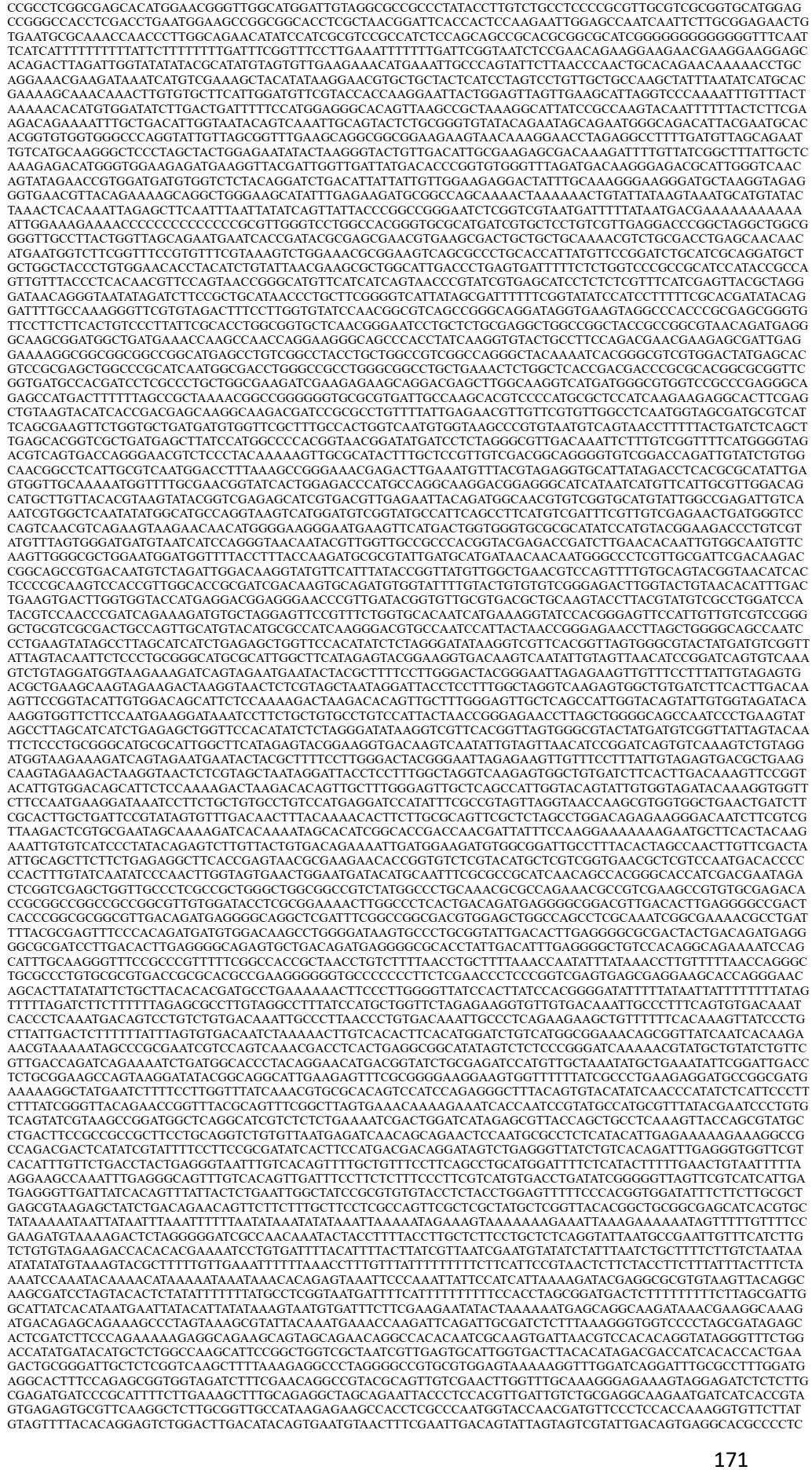
ACGAAGGTTTTTGCGCTGGATGTGGCTGCCCGGCACCGGGTGCAGTTTGCGATGCCGGAGTCTGATGCGGTTGCGATGCTGAAACAATTATCCT
GAGAATAAATGCCTTGGCCTTTATATGGAAATGTGGAACTGAGTGGATATGCTGTTTTTGTCTGTTAAACAGAGAAGCTGGCTGTTATCCACTGAG
AAGCGAACGAAACAGTCGGGAAAATCTCCCATTATCGTAGAGATCCGCATTATTAATCTCAGGAGCCTGTGTAGCGTTTATAGGAAGTAGTGTTC
TGTCATGATGCCTGCAAGCGGTAACGAAAACGATTTGAATATGCCTTCAGGAACAATAGAAATCTTCGTGCGGTGTTACGTTGAAGTGGAGCGG
ATTATGTCAGCAATGGACAGAACAACCTAATGAACACAGAACCATGATGTGGTCTGTCCTTTTACAGCCAGTAGTGCTCGCCGCAGTCGAGCGA
CAGGGCGAAGCCCATCGATACTAGCTTGATTGGGATATCTCGCTCGTGCTTGTCGCGTGCTATGTCTTTTTAGGGTACTTTGAACCTACGTTCGTA
CTTGTGTAATATGATCATCGTCGTCATCGTATTATCGTTTTTCATCCGTCCAGCGCAAAATGCATTAGCAGCTAGTCCTAGCGTGCGGAGCTACCTG
TACAGGTGCATGACGGATGCGTGTCCTTAAGTGAGTTTCTAATTAACAGTAACTTCTTTACTTATGTTTCAGTTTGTAAGAAGCGGGATGCGCTCG
TGCCTTGACATCTGATTGGACTGCGTCGGCACGTGAAAACTACATTGTGAAATCTGCTAAAACTCCGGGTATCTCTGACACAAAACGATTCGGCT
TCGCAATTTCAACATTACGGTCAAGGCTAACGTATCTTTCTCGGTCAACTTCAGATTATGCCGATTAAATTGTCGTAGCTTTCAAGGCGTTTTGAG
TACTGCGGCAGTTGTTGAACCTGCAAGGAGAAGATCTCGACAACAGAATAAAGCGAAAAATGGGTCTCATGCACTAACACTCAGGCCTCCCTC
ATAATCTCTGTTTGAGTTTACCAACAACACATATATACATTTCGACAAAATGGCCAAGTTGACCAGTGCCGTTCCGGTGCTCACCGCGCGCGACG
TCGCCGGAGCGGTCGAGTTCTGGACCGACCGGCTCGGGTTCTCCCGGGACTTCGTGGAGGACGACTTCGCCGGTGTGGTCCGGGACGACGTGA
CCCTGTTCATCAGCGCGGTCCAGGACCAGGTGGTGCCAGACAACACCCTGGCCTGGGTGTGGGTGCGCGGCCTGGACGAGCTGTACGCCGAGT
GGTCGGAGGTCGTGTCCACGAACTTCCGGGACGCCTCCGGGCCGGCCATGACCGAGATCGGCGAGCAGCCGTGGGGGCGGGAGTTCGCCCTGC
GCGACCCGGCCGGCAACTGCGTGCACTTCGTGGCCGAGGAGCAGGACTGACCGACGCCGACCAACACCGCCGGTCCGACGCGGCCCGACGGG
TCCGAGGCCTCGGAGATCTGGGCCCATGCGGCCGCAACAACTACCTCGACTTTGGCTGGGACACTTTCAGTGAGGACAAGAAGCTTCAGAAGC
GTGCTATCGAACTCAACCAGGGACGTGCGGCACAAATGGGCATCCTTGCTCTCATGGTGCACGAACAGTTGGGAGTCTCTATCCTTCCTTAAAA
ATTTAATTTTCATTAGTTGCAGTCACTCCGCTTTGGTTTCGTAACTATAACGGTCCTAAGGTAGCGAACGTTGCAGGCCATGCTGTCCAGGCAGGT
AGATGACGACCATCAGGGACAGCTTCAAGGATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCATTGGACCGCTGATCGTCACGGCGATTTATG
CCGCCTCGGCGAGCACATGGAACGGGTTGGCATGGATTGTAGGCGCCGCCCTATACCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAG
CCGGGCCACCTCGACCTGAATGGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAAGAATTGGAGCCAATCAATTCTTGCGGAGAACTG
TGAATGCGCAAACCAACCCTTGGCAGAACATATCCATCGCGTCCGCCATCTCCAGCAGCCGCACGCGGCGCATCGGGGGGGGGGGGGGGGGTTT
CAATTCATCATTTTTTTTATTCTTTTTTTGATTTCGGTTTCCTTGAAATTTTTTTGATTCGGTAATCTCCGAACAGAAGGAAGAACGAAGGAAGG
AGCACAGACTTAGATTGGTATATATACGCATATGTAGTGTTGAAGAAACATGAAATTGCCCAGTATTCTTAACCCAACTGCACAGAACAAAAACC
TGCAGGAAACGAAGATAAATCATGTCGAAAGCTACATATAAGGAACGTGCTGCTACTCATCCTAGTCCTGTTGCTGCCAAGCTATTTAATATCATG
CACGAAAAGCAAACAAACTTGTGTGCTTCATTGGATGTTCGTACCACCAAGGAATTACTGGAGTTAGTTGAAGCATTAGGTCCCAAAATTTGTTT
ACTAAAAACACATGTGGATATCTTGACTGATTTTTCCATGGAGGGCACAGTTAAGCCGCTAAAGGCATTATCCGCCAAGTACAATTTTTTACTCTT
CGAAGACAGAAAATTTGCTGACATTGGTAATACAGTCAAATTGCAGTACTCTGCGGGTGTATACAGAATAGCAGAATGGGCAGACATTACGAATG
CACACGGTGTGGTGGGCCCAGGTATTGTTAGCGGTTTGAAGCAGGCGGCGGAAGAAGTAACAAAGGAACCTAGAGGCCTTTTGATGTTAGCAG
AATTGTCATGCAAGGGCTCCCTAGCTACTGGAGAATATACTAAGGGTACTGTTGACATTGCGAAGAGCGACAAAGATTTTGTTATCGGCTTTATTG
CTCAAAGAGACATGGGTGGAAGAGATGAAGGTTACGATTGGTTGATTATGACACCCGGTGTGGGTTTAGATGACAAGGGAGACGCATTGGGTCA
ACAGTATAGAACCGTGGATGATGTGGTCTCTACAGGATCTGACATTATTATTGTTGGAAGAGGACTATTTGCAAAGGGAAGGGATGCTAAGGTAG
AGGGTGAACGTTACAGAAAAGCAGGCTGGGAAGCATATTTGAGAAGATGCGGCCAGCAAAACTAAAAAACTGTATTATAAGTAAATGCATGTAT
ACTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCGGCCGGGAATCTCGGTCGTAATGATTTTTATAATGACGAAAAAAAAA
AAAATTGGAAAGAAAACCCCCCCCCCCCCCGCAGCGTTGGGTCCTGGCCACGGGTGCGCATGATCGTGCTCCTGTCGTTGAGGACCCGGCTA
GGCTGGCGGGGTTGCCTTACTGGTTAGCAGAATGAATCACCGATACGCGAGCGAACGTGAAGCGACTGCTGCTGCAAAACGTCTGCGACCTGA
GCAACAACATGAATGGTCTTCGGTTTCCGTGTTTCGTAAAGTCTGGAAACGCGGAAGTCAGCGCCCTGCACCATTATGTTCCGGATCTGCATCGC
AGGATGCTGCTGGCTACCCTGTGGAACACCTACATCTGTATTAACGAAGCGCTGGCATTGACCCTGAGTGATTTTTCTCTGGTCCCGCCGCATCC
ATACCGCCAGTTGTTTACCCTCACAACGTTCCAGTAACCGGGCATGTTCATCATCAGTAACCCGTATCGTGAGCATCCTCTCTCGTTTCATCGAGTT
TACGCTAGGGATAACAGGGTAATATAGATCTTCCGCTGCATAACCCTGCTTCGGGGTCATTATAGCGATTTTTTCGGTATATCCATCCTTTTTCGCAC
GATATACAGGATTTTGCCAAAGGGTTCGTGTAGACTTTCCTTGGTGTATCCAACGGCGTCAGCCGGGCAGGATAGGTGAAGTAGGCCCACCCGC
GAGCGGGTGTTCCTTCTTCACTGTCCCTTATTCGCACCTGGCGGTGCTCAACGGGAATCCTGCTCTGCGAGGCTGGCCGGCTACCGCCGGCGTA
ACAGATGAGGGCAAGCGGATGGCTGATGAAACCAAGCCAACCAGGAAGGGCAGCCCACCTATCAAGGTGTACTGCCTTCCAGACGAACGAAG
AGCGATTGAGGAAAAGGCGGCGGCGGCCGGCATGAGCCTGTCGGCCTACCTGCTGGCCGTCGGCCAGGGCTACAAAATCACGGGCGTCGTGGA
CTATGAGCACGTCCGCGAGCTGGCCCGCATCAATGGCGACCTGGGCCGCCTGGGCGGCCTGCTGAAACTCTGGCTCACCGACGACCCGCGCAC
GGCGCGGTTCGGTGATGCCACGATCCTCGCCCTGCTGGCGAAGATCGAAGAGAAGCAGGACGAGCTTGGCAAGGTCATGATGGGCGTGGTCCG
CCCGAGGGCAGAGCCATGACTTTTTAGCCGCTAAAACGGCCGGGGGGTGCGCGTGATTGCCAAGCACGTCCCCATGCGCTCCATCAAGAAGA
GGCACTTCGAGCTGTAAGTACATCACCGACGAGCAAGGCAAGACGATCCGCGCCTGTTTTATTGAGAACGTTGTTCGTGTTGGCCTCAATGGTA
GCGATGCGTCATTCAGCGAAGTTCTGGTGCTGATGATGTGGTTCGCTTTGCCACTGGTCAATGTGGTAAGCCCGTGTAATGTCAGTAACCTTTTA
CTGATCTCAGCTTGAGCACGGTCGCTGATGAGCTTATCCATGGCCCCACGGTAACGGATATGATCCTCTAGGGCGTTGACAAATTCTTTGTCGGTT
TTCATGGGGTAGACGTCAGTGACCAGGGAACGTCTCCCTACAAAAAGTTGCGCATACTTTGCTCCGTTGTCGACGGCAGGGGTGTCGGACCAGA
TTGTATCTGTGGCAACGGCCTCATTGCGTCAATGGACCTTTAAAGCCGGGAAACGAGACTTGAAATGTTTACGTAGAGGTGCATTATAGACCTCA
CGCGCATATTGAGTGGTTGCAAAAATGGTTTTGCGAACGGTATCACTGGAGACCCATGCCAGGCAAGGACGGAGGGCATCATAATCATGTTCATT
GCGTTGGACAGCATGCTTGTTACACGTAAGTATACGGTCGAGAGCATCGTGACGTTGAGAATTACAGATGGCAACGTGTCGGTGCATGTATTGGC
CGAGATTGTCAAATCGTGGCTCAATATATGGCATGCCAGGTAAGTCATGGATGTCGGTATGCCATTCAGCCTTCATGTCGATTTCGTTGTCGAGAA
CTGATGGGTCCCAGTCAACGTCAGAAGTAAGAACAACATGGGGAAGGGAATGAAGTTCATGACTGGTGGGTGCGCGCATATCCATGTACGGAA
GACCCTGTCGTATGTTTAGTGGGATGATGTAATCATCCAGGGTAACAATACGTTGGTTGCCGCCCACGGTACGAGACCGATCTTGAACACAATTG
TGGCAATGTTCAAGTTGGGCGCTGGAATGGATGGTTTTACCTTTACCAAGATGCGCGTATTGATGCATGATAACAACAATGGGCCCTCGTTGCGAT
TCGACAAGACCGGCAGCCGTGACAATGTCTAGATTGGACAAGGTATGTTCATTTATACCGGTTATGTTGGCTGAACGTCCAGTTTTGTGCAGTAC
GGTAACATCACTCCCCGCAAGTCCACCGTTGGCACCGCGATCGACAAGTGCAGATGTGGTATTTTGTACTGTGTGTCGGGAGACTTGGTACTGTA
ACACATTTGACTGAAGTGACTTGGTGGTACCATGAGGACGGAGGGAACCCGTTGATACGGTGTTGCGTGACGCTGCAAGTACCTTACGTATGTC
GCCTGGATCCATACGTCCAACCCGATCAGAAAGATGTGCTAGGAGTTCCGTTTCTGGTGCACAATCATGAAAGGTATCCACGGGAGTTCCATTGT
TGTCGTCCGGGGCTGCGTCGCGACTGCCAGTTGCATGTACATGCGCCATCAAGGGACGTGCCAATCCATTACTAACCGGGAGAACCTTAGCTGG
GGCAGCCAATCCCTGAAGTATAGCCTTAGCATCATCTGAGAGCTGGTTCCACATATCTCTAGGGATATAAGGTCGTTCACGGTTAGTGGGCGTACT
ATGATGTCGGTTATTAGTACAATTCTCCCTGCGGGCATGCGCATTGGCTTCATAGAGTACGGAAGGTGACAAGTCAATATTGTAGTTAACATCCGG
ATCAGTGTCAAAGTCTGTAGGATGGTAAGAAAGATCAGTAGAATGAATACTACGCTTTTCCTTGGGACTACGGGAATTAGAGAAGTTGTTTCCTT
TATTGTAGAGTGACGCTGAAGCAAGTAGAAGACTAAGGTAACTCTCGTAGCTAATAGGATTACCTCCTTTGGCTAGGTCAAGAGTGGCTGTGATC
TTCACTTGACAAAGTTCCGGTACATTGTGGACAGCATTCTCCAAAAGACTAAGACACAGTTGCTTTGGGAGTTGCTCAGCCATTGGTACAGTATT
GTGGTAGATACAAAGGTGGTTCTTCCAATGAAGGATAAATCCTTCTGCTGTGCCTGTCCATGAGGATCCATATTTCGCCGTAGTTAGGTAACCAAG
CGTGGTGGCTGAACTGATCTTCGCACTTGCTGATTCCGTATAGTGTTTGACAACTTTACAAAACACTTCTTGCGCAGTTCGCTCTAGCCTGGACA
GAGAAGGGACAATCTTCGTCGTTAAGACTCGTGCGAATAGCAAAAGATCACAAAATAGCACATCGGCACCGACCAACGATTATTTCCAAGGAAA
AAAAGAATGCTTCACTACAAGAAATTGTGTCATCCCTATACAGAGTCTTGTTACTGTGACAGAAAATTGATGGAAGATGTGGCGGATTGCCTTTA
CACTAGCCAACTTGTTCGACTAATTGCAGCTTCTTCTGAGAGGCTTCACCGAGTAACGCGAAGAACACCGGTGTCTCGTACATGCTCGTCGGTG
AACGCTCGTCCAATGACACCCCCCCACTTTGTATCAATATCCCAACTTGGTAGTGAACTGGAATGATACATGCAATTTCGCGCCGCATCAACAGCC
ACGGGCACCATCGACGAATAGACTCGGTCGAGCTGGTTGCCCTCGCCGCTGGGCTGGCGGCCGTCTATGGCCCTGCAAACGCGCCAGAAACGC
CGTCGAAGCCGTGTGCGAGACACCGCGGCCGGCCGCCGGCGTTGTGGATACCTCGCGGAAAACTTGGCCCTCACTGACAGATGAGGGGCGGAC
GTTGACACTTGAGGGGCCGACTCACCCGGCGCGGCGTTGACAGATGAGGGGCAGGCTCGATTTCGGCCGGCGACGTGGAGCTGGCCAGCCTCG
CAAATCGGCGAAAACGCCTGATTTTACGCGAGTTTCCCACAGATGATGTGGACAAGCCTGGGGATAAGTGCCCTGCGGTATTGACACTTGAGGG
GCGCGACTACTGACAGATGAGGGGCGCGATCCTTGACACTTGAGGGGCAGAGTGCTGACAGATGAGGGGCGCACCTATTGACATTTGAGGGGC
TGTCCACAGGCAGAAAATCCAGCATTTGCAAGGGTTTCCGCCCGTTTTTCGGCCACCGCTAACCTGTCTTTTAACCTGCTTTTAAACCAATATTTA
TAAACCTTGTTTTTAACCAGGGCTGCGCCCTGTGCGCGTGACCGCGCACGCCGAAGGGGGGTGCCCCCCCTTCTCGAACCCTCCCGGTCGAGTG
AGCGAGGAAGCACCAGGGAACAGCACTATTATATTCTGCTTACACACGATGCCTGAAAAAACTTCCCTTGGGGTTATCCACTTATCCACGGGGAT
ATTTTTATAATTATTTTTTTTATAGTTTTTAGATCTTCTTTTTTAGAGCGCCTTGTAGGCCTTTATCCATGCTGGTTCTAGAGAAGGTGTTGTGACAA
ATTGCCCTTTCAGTGTGACAAATCACCCTCAAATGACAGTCCTGTCTGTGACAAATTGCCCTTAACCCTGTGACAAATTGCCCTCAGAAGAAGCT
GTTTTTTCACAAAGTTATCCCTGCTTATTGACTCTTTTTTATTTAGTGTGACAATCTAAAAACTTGTCACACTTCACATGGATCTGTCATGGCGGAA
ACAGCGGTTATCAATCACAAGAAACGTAAAAATAGCCCGCGAATCGTCCAGTCAAACGACCTCACTGAGGCGGCATATAGTCTCTCCCGGGATC
AAAAACGTATGCTGTATCTGTTCGTTGACCAGATCAGAAAATCTGATGGCACCCTACAGGAACATGACGGTATCTGCGAGATCCATGTTGCTAAA
TATGCTGAAATATTCGGATTGACCTCTGCGGAAGCCAGTAAGGATATACGGCAGGCATTGAAGAGTTTCGCGGGGAAGGAAGTGGTTTTTTATCG
CCCTGAAGAGGATGCCGGCGATGAAAAAGGCTATGAATCTTTTCCTTGGTTTATCAAACGTGCGCACAGTCCATCCAGAGGGCTTTACAGTGTAC
ATATCAACCCATATCTCATTCCCTTCTTTATCGGGTTACAGAACCGGTTTACGCAGTTTCGGCTTAGTGAAACAAAAGAAATCACCAATCCGTATG
CCATGCGTTTATACGAATCCCTGTGTCAGTATCGTAAGCCGGATGGCTCAGGCATCGTCTCTCTGAAAATCGACTGGATCATAGAGCGTTACCAGC
TGCCTCAAAGTTACCAGCGTATGCCTGACTTCCGCCGCCGCTTCCTGCAGGTCTGTGTTAATGAGATCAACAGCAGAACTCCAATGCGCCTCTCA
TACATTGAGAAAAAGAAAGGCCGCCAGACGACTCATATCGTATTTTCCTTCCGCGATATCACTTCCATGACGACAGGATAGTCTGAGGGTTATCTG
TCACAGATTTGAGGGTGGTTCGTCACATTTGTTCTGACCTACTGAGGGTAATTTGTCACAGTTTTGCTGTTTCCTTCAGCCTGCATGGATTTTCTC
ATACTTTTTGAACTGTAATTTTTAAGGAAGCCAAATTTGAGGGCAGTTTGTCACAGTTGATTTCCTTCTCTTTCCCTTCGTCATGTGACCTGATATC
GGGGGTTAGTTCGTCATCATTGATGAGGGTTGATTATCACAGTTTATTACTCTGAATTGGCTATCCGCGTGTGTACCTCTACCTGGAGTTTTTCCCA
CGGTGGATATTTCTTCTTGCGCTGAGCGTAAGAGCTATCTGACAGAACAGTTCTTCTTTGCTTCCTCGCCAGTTCGCTCGCTATGCTCGGTTACAC
GGCTGCGGCGAGCATCACGTGCTATAAAAATAATTATAATTTAAATTTTTTAATATAAATATATAAATTAAAAATAGAAAGTAAAAAAAGAAATTAA

AGAAAAAATAGTTTTTGTTTTCCGAAGATGTAAAAGACTCTAGGGGGATCGCCAACAAATACTACCTTTTACCTTGCTCTTCCTGCTCTCAGGTAT
TAATGCCGAATTGTTTCATCTTGTCTGTGTAGAAGACCACACACGAAAATCCTGTGATTTTACATTTTACTTATCGTTAATCGAATGTATATCTATTT
AATCTGCTTTTCTTGTCTAATAAATATATATGTAAAGTACGCTTTTTGTTGAAATTTTTTAAACCTTTGTTTATTTTTTTTTCTTCATTCCGTAACTCTT
CTACCTTCTTTATTTACTTTCTAAAATCCAAATACAAAACATAAAAATAAATAAACACAGAGTAAATTCCCAAATTATTCCATCATTAAAAGATACG
AGGCGCGTGTAAGTTACAGGCAAGCGATCCTAGTACACTCTATATTTTTTTATGCCTCGGTAATGATTTTCATTTTTTTTTTCCACCTAGCGGATGA
CTCTTTTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTATACATTATATAAAGTAATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAG
GCAAGATAAACGAAGGCAAAGATGACAGAGCAGAAAGCCCTAGTAAAGCGTATTACAAATGAAACCAAGATTCAGATTGCGATCTCTTTAAAGG
GTGGTCCCCTAGCGATAGAGCACTCGATCTTCCCAGAAAAAGAGGCAGAAGCAGTAGCAGAACAGGCCACACAATCGCAAGTGATTAACGTCC
ACACAGGTATAGGGTTTCTGGACCATATGATACATGCTCTGGCCAAGCATTCCGGCTGGTCGCTAATCGTTGAGTGCATTGGTGACTTACACATAG
ACGACCATCACACCACTGAAGACTGCGGGATTGCTCTCGGTCAAGCTTTTAAAGAGGCCCTAGGGGCCGTGCGTGGAGTAAAAAGGTTTGGAT
CAGGATTTGCGCCTTTGGATGAGGCACTTTCCAGAGCGGTGGTAGATCTTTCGAACAGGCCGTACGCAGTTGTCGAACTTGGTTTGCAAAGGGA
GAAAGTAGGAGATCTCTCTTGCGAGATGATCCCGCATTTTCTTGAAAGCTTTGCAGAGGCTAGCAGAATTACCCTCCACGTTGATTGTCTGCAAG
GCAAGAATGATCATCACCGTAGTGAGAGTGCGTTCAAGGCTCTTGCGGTTGCCATAAGAGAAGCCACCTCGCCCAATGGTACCAACGATGTTCC
CTCCACCAAAGGTGTTCTTATGTAGTTTTACACAGGAGTCTGGACTTGACATACAGTGAATGTAACTTTCGAATTGACAGTATTAGTAGTCGTATT
GACAGTGAGGCACGCCCCTCAATGTGCGAGGTGGAAAATATACCAGCATGACAATGAATCTTGGAGATTCTTTTGCTGTCATCAAGATTCACCGC
CAAATCTTCAGGAACCTATCACGTCCACAGGCGATGTTAATTCTTGAGTCGTCAAAACAAAGTCCTGTCCTACCTGTAGAAGTTGACAGCGAGC
AATTGTATGCAAACTTCTGACTTTGTTATAATAACATTAAAGGTAATTAAGTATCTTCAATTAGGCATTTTGTCACTGTCAGTCCGTTCCGACAATAT
AGGTAGATTTGGAATGAATCTTTTCTATGCTGCTGCGAATCTTGTACACCTTTGAGGCCGTAGATTCTGTCCGACGAAGCGATAATTATTGCAAAA
TACATGGACTCATTATTTTGATTCGATTTCTTTTTGGTATCCGACTCGAAAAGATCCATCACGGCGAGCGCCACCATGAACAAGAACTCCAAAATC
CAGTCCCCAAACTCTTCTGATGTTGCTGTTATTGGTGTTGGTTTTAGATTCCCAGGTAACTCTAATGACCCAGAATCTTTGTGGAACAACTTGTTG
GATGGTTTCGATGCTATTACCCAAGTCCCAAAAGAAAGATGGGCTACTTCTTTTAGAGAGATGGGTTTGATCAAGAACAAGTTCGGTGGTTTCTT
GAAGGATTCTGAATGGAAGAATTTCGACCCTTTGTTCTTTGGTATCGGTCCAAAAGAAGCTCCATTCATTGATCCACAACAAAGGTTGTTGTTGT
CCATCGTTTGGGAATCTTTGGAAGATGCTTACATCAGACCAGATGAATTGAGAGGTTCTAACACTGGTGTTTTCATCGGTGTTTCTAACAACGATT
ACACCAAGTTGGGTTTCCAAGACAACTACTCTATTTCTCCATACACTATGACCGGCTCTAACTCTTCATTGAACTCCAACAGAATTTCCTACTGCT
TCGATTTTAGAGGTCCATCCATTACTGTTGATACCGCTTGTTCTTCTTCCTTGGTTTCTGTTAATTTGGGTGTCCAATCCATCCAAATGGGTGAATG
TAAGATTGCTATTTGCGGTGGTGTTAACGCTTTGTTTGATCCATCTACATCTGTTGCCTTTTCCAAGTTGGGTGTTTTGTCTGAAAATGGCAGATGC
AACTCTTTTAGTGATCAAGCCTCTGGTTACGTTAGATCTGAAGGTGCTGGTGTTGTTGTTTTGAAGTCTTTGGAACAAGCTAAGTTGGATGGTGA
TAGAATCTACGGTGTTATCAAGGGTGTTTCCTCTAATGAAGATGGTGCTTCTAATGGTGACAAGAACTCTTTGACTACTCCATCTTGTGAAGCCCA
ATCCATTAACATTTCTAAGGCTATGGAAAAGGCCTCCTTGTCTCCATCTGATATCTATTACATTGAAGCCCATGGTACTGGTACTCCAGTTGGTGAT
CCAATTGAAGTTAAGGCCTTGTCCAAGATCTTCTCCAACTCTAACAACAACCAGTTGAACAACTTCTCTACCGATGGTAATGATAACGACGATTCCA
TGATGACGATAACACCTCTCCAGAACCATTATTGATTGGCTCATTCAAGTCCAACATCGGTCATTTGGAATCTGCTGCTGGTATTGCTTCTTTGATT
AAGTGTTGCTTGATGTTGAAGAACAGGATGTTGGTTCCATCCATTAACTGCTCTAATTTGAACCCATCCATTCCATTCGATCAGTACAACATCTCC
GTTATCAGAGAAATCAGACAATTCCCAACCGATAAGTTGGTTAACATCGGTATCAATTCTTTCGGTTTCGGTGGTTCTAACTGCCATTTGATTATTC
AAGAGTACAACAACAACTTCAAGAACAACTCTACCATCTGCAATAACAACAACAACAACAATAACAACATCGACTACTTGATCCCAATCTCCTC
TAAGACTAAGAAGTCCTTGGATAAGTACTTGATTTTGATCAAGACCAACTCCAACTACCACAAGGATATTTCTTTCGATGACTTCGTCAAGTTCCA
AATCAAGTCTAAGCAGTACAACTTGTCCAACAGAATGACTACCATTGCTAACGATTGGAACTCCTTCATTAAGGGTTCTAACGAATTCCACAACT
TGATCGAATCTAAGGATGGTGAAGGTGGTTCTTCATCTTCTAACAGAGGTATTGATTCCGCCAATCAAATCAACACTACTACTACCTCTACCATCA
ACGATATCGAACCTTTGTTGGTTTTCGTTTTCTGTGGTCAAGGTCCACAATGGAATGGTATGATTAAGACCTTGTACAACTCCGAGAACGTTTTCA
AGAACACCGTTGATCATGTTGACAGCATCTTGTACAAGTACTTCGGTTACTCCATTTTGAACGTCTTGTCTAAGATCGATGATAACGACGATTCCA
TCAACCATCCAATAGTTGCTCAACCATCTTTGTTCTTGTTGCAAATTGGTTTGGTCGAGTTGTTTAAGTACTGGGGTATCTACCCATCTATCTCTGT
TGGTCATTCTTTCGGTGAAGTCTCTTCTTATTACTTGTCCGGTATCATCTCTTTGGAAACCGCTTGTAAAATCGTCTACGTCAGATCCTCTAATCAG
AACAAAACTATGGGTTCCGGTAAGATGTTGGTTGTTTCTATGGGTTTTAAGCAATGGAACGATCAATTCTCTGCTGAATGGTCCGATATTGAAATT
GCTTGTTACAACGCTCCAGATTCCATAGTTGTTACTGGTAACGAAGAAAGATTGAAAGAATTGTCCATCAAGTTGTCCGACGAATCCAATCAAAT
TTTCAACACCTTCTTGAGGTCCCCATGTTCTTTTCATTCTTCCCATCAAGAAGTCATCAAGGGTTCTATGTTCGAAGAGTTGTCTAACTTGCAATC
TACTGGTGAAACCGAAATCCCTTTGTTCTCTACTGTTACTGGTAGACAAGTTTTGTCTGGTCATGTTACTGCTCAACACATCTACGATAATGTTAG
AGAACCAGTCTTGTTCCAAAAGACGATTGAATCCATTACCTCCTACATCAAGTCTCACTACCCATCCAATCAAAAGGTTATCTACGTTGAAATTGC
TCCACACCCAACCTTGTTTTCATTGATCAAAAAGTCCATCCCATCCTCCAACAAGAATTCCTCTTCTGTTTTGTGTCCATTGAACAGAAAAGAAA
ACTCCAACAACTCCTACAAGAAGTTCGTTTCTCAGTTGTACTTCAACGGTGTTAACGTTGACTTCAACTTCCAGTTGAACTCCATTTGCGATAAC
GTTAACAACGATCACCATTTGAACAACGTCAAGCAAAACTCCTTCAAAGAGACTACCAATTCCTTGCCAAGATACCAATGGGAACAAGATGAAT
ATTGGTCCGAACCATTGATCTCCAGAAAGAATAGATTGGAAGGTCCAACTACTTCCTTGTTGGGTCATAGAATTATCTACAGCTTCCCAGTTTCC
AATCCGTTTTGGACTTGCAATCTGACAACTACAAATACTTGTTGGACCACTTGGTTAACGGTAAGCCAGTTTTCCAGGTGCTGGTTATTTGGATA
TCATCATCGAATTCTTCGACTACCAAAAGCAGCAGTTGAATTCCTCTGATTCCTCTAACTCCTACATCATCAACGTTGACAAGATCCAATTCTTGA
ACCCAATTCACTTGACCGAAAACAAGTCTTGCAATCTTCTTTCGAACCTATCGTTACTAAGAAGTCTGCCTTCTCTGTTAACTTCTTCA
TCAAGGATACCGTCGAGGATCAATCTAAGGTTAAGTCTATGTCTGACGAAACTTGGACTAACACTTGTAAGGCTACCATTTCCTTGGAACAACAA
CAGCCATCTCCATCTTCTACTTTGACTTTGTCTAAGAAGCAAGACTTGCAGATCTTGAGAAACAGATGCGATATTAGCAAGCTAGACAAGTTTGA
GTTGTACGACAAGATCTCTAAGAATTTGGGCTTGCAGTACAACTCCTTGTTTCAAGTTGTTGATACCATCGAAACTGGTAAGGATTGCTCTTTTGC
TACTTTGTCTTTGCCAGAAGATACTTTGTTCACCACCATTTTGAACCCATGCTTGTTGGATAACTGTTTCCATGGTTTGTTGACCTTGATCAACGA
AAAGGGTTCTTTCGTTGTCGAGTCCATTTCTTCTGTTTCTATCTACTTGGAGAACATCGGTTCCTTCAATCAAACTTCTGTTGGTAACGTCCAGTT
CTACTTGTACACCACTATTTCTAAAGCCACCTCCTTTAGTTCTGAAGGTACTTGTAAGTTGTTCACCAAGGATGGTTCCTTGATTTTGTCTATCGGT
AAGTTCATCATCAAGTCCACCAATCCAAAGTCTACTAAGACCAACGAAACTATCGAATCTCCATTGGACGAAACCTTCTCTATTGAATGGCAATC
TAAGGATTCTCCAATTCCAACCCCACAACAAATCCAACAACAATCTCCATTGAACTCTAACCCATCCTTCATTAGATCTACCATCTTGAAGGACAT
CCAGTTCGAACAATACTGCTCCTCCATTATCCACAAAGAATTGATCAACCACGAAAAGTACAAGAACCAGCAATCCTTCGATATCAACTCCTTGG
AAAACCACTTGAACGATGACCAATTGATGGAATCCTTGTCCATCTCCAAAGAATACTTGAGATTCTTCACCAGGATCATCTCCATCATTAAGCAAT
ACCCAAAGATCTTGAACGAAAAAGAGCTAAAAGAATTGAAAGAAATCATCGAATTGAAGTACCCATCCGAAGTTCAGTTGTTGGAATTCGAAGT
TATCGAGAAGGTGTCCATGATTATCCCAAAGTTGTTGTTCGAAAACGACAAGCAATCTTCCATGACCTTGTTCCAAGATAACTTGTTGACCAGGT
TCTACTCCAATTCTAACTCTACCAGATTCTACTTGGAAAGGGTTTCCGAAATGGTCTTGGAATCTATTAGACCAATCGTCAGAGAAAAGAGGGTG
TTCAGAATTTTGGAAATTGGTGCTGATACAGCCTCTTTGTCTAATGTTGTTTTGACTAAGTTGAACACCTACTTGTCCACCTTGAATTCTAATGGTG
GTTCTGGTTACAACATCATCATTGAGTACACCTTCACCGATATTTCCGCCAACTTCATTATTGGTGAAATCCAAGAAACCATGTGCAACTTGTACC
CAAACGTTACTTTCAAGTTCTCCGTCTTGGACTTGGAGAAAGAGATTATTAACTCCTCCGATTTCTTGATGGGTGATTACGATATAGTTTTGATGGC
CTACGTTATCCATGCCGTTTCTAACATTAAGTTCTCCATCGAACAGTTGTACAAGTTGTTGTCTCCAAGAGGTTGGTTGTTGTGTATTGAACCTAA
GTCCAACGTTGTGTTCTCCGATTTGGTTTTCGGTTGTTTTAATCAGTGGTGGAACTACTACGATGATATTAGAACTACCCACTGCTCCTTGTCTGA
ATCTCAATGGAATCAGTTGTTGTTGAACCAGTCCTTGAACAACGAATCCTCTTCTTCTTCTAACTGTTACGGTGGTTTTTCCAACGTTTCTTTTATT
GGTGGTGAAAAGGATGTCGACTCCCATTCTTTCATATTGCACTGCCAAAAAGAATCCATCTCCCAAATGAAGTTAGCCACCACTATTAACAACGG
TTTGTCATCTGGTTCCATCGTTATCGTTTTGAACTCTCAACAATTGACCAACATGAAGTCCTACCCAAAGGTTATTGAGTATATTCAAGAGGCTAC
CTCTTTGTGCAAGACCATTGAAATTATCGATTCCAAGGACGTCTTGAACTCTACCAATTCAGTTTTGGAAAAGATCCAAAAGTCCTTGTTGGTGT
TCTGTTTGTTGGGTTATGACTTGTTGGAGAACAACTACCAAGAACAGTCTTTCGAATACGTTAAGTTGTTGAACTTGATCTCTACTACCGCCTCTT
CATCTAATGATAAGAAACCACCAAAGGTCTTGTTGATCACCAAGCAATCTGAAAGAATCTCCAGGTCTTTCTACTCCAGATCCTTGATTGGTATTT
CCAGAACCTCTATGAACGAGTACCCAAATTTGTCCATTACCTCTATCGATTTGGATACCAACGACTACTCATTGCAGTCTTTGTTGAAGCCAATCT
TCAGCAACTCTAAGTTTTCCGACAACGAGTTCATCTTCAAAAAGGGCTTGATGTTCGTGTCCAGGATCTTTAAGAACAAGCAGTTGCTAGAATCC
TCCAACGCTTTGAAACTGACTCTTCTAACTTGTACTGTAAGGCCTCTTCTGACTTGTCTTACAAGTACGCTATTAAGCAGTCTATGTTGACCGAA
AATCAGATCGAAATCAAGGTTGAATGCGTCGGTATTAACTTCAAGGACAACCTATTCTACAAGGGCTTGTTGCCACAAGAAATTTTCAGAATGGG
TGACATCTACAATCCACCATATGGTTTGGAATGCTCTGGTGTTATTACCAGAATTGGTTCTAACGTCACCGAATACTCAGTTGGTCAAAATGTTTTT
GGTTTCGCCAGACATTCTTTGGGTTCTCATGTTGTTACCAACAAGGATTTGGTTATCTTGAAGCCAGATACCATCTCATTTTCTGAAGCTGCTTCTA
TCCCAGTTGTTTACTGTACTGCTTGGTACTCCTTGTTCAACATTGGTCAGTTGTCTAACGAAGAATCCATCCTAATTCATTCTGCTACTGGTGGTGT
AGGTTTGGCTTCTTTGAATTTGTTGAAAATGAAGAATCAGCAACAGCAACCATTGACCAATGTTTATGCTACTGTTGGCTCTAACGAGAAGAAGA
AGTTCTTGATCGATAACTTCAACAACTTGTTCAAAGAGGACGGCGAAAACATTTTCTCTACCAGAGACAAAGAATACTCCAACCAGTTGGAATC
CAAGATCGATGTTATTTTGAACACCTTGTCCGGTGAATTCGTCGAATCTAATTTCAAGTCCTTGAGATCCTTCGGTAGATTGATTGATTTGTCTGCT
ACTCACGTTTACGCCAATCAACAAATTGGTCTAGGTAACTTCAAGTTCGACCACTTGTATTCTGCTGTTGACTTGGAAAGATTGATCGACGAAAA
ACCTAAGTTGTTGCAGTCCATCTTGCAAAGAATTACCAACTCTATCGTCAACGGTTCCTTGGAAAAAATTCCAATTACCATCTTCCCATCCACCGA
AACTAAGGATGCTATCGAATTATTGTCCAAGAGATCCCATATCGGTAAAGTTGTTGTAGATTGCACCGATATCTCTAAGTGTAATCCTGTTGGTGAT
GTGATCACCAACTTCTCTATGAGATTGCCAAAGCCAAACTACCAGTTGAATTTGAACTCCACCTTGTTGATTACTGGTCAGTCTGGTTTGTCTATC
CCTTTGTTGAATTGGTTGTTGTCTAAGTCTGGTGGTAACGTTAAGAACGTTGTCATCATTTCTAAGTCCACCATGAAGTGGAAGTTGCAGACTATG
ATTTCCCATTTCGTTTCCGGTTTCGGTATCCATTTTAACTACGTTCAAGTCGACATCTCCAACTACGATGCTTTGTCTGAAGCTATTAAGCAATTGC
CATCTGATTTGCCACCAATCACCTCTGTTTTTCATTTGGCTGCTATCTACAACGATGTTCCAATGGATCAAGTTACCGTGTCTACCGTTGAATCTGT
TCATAACCCTAAAGTTTTGGGTGCCGTTAACTTGCATAGAATCTCTGTTTCTTTTGGTTGGAAGTTGAACCACTTCGTCTTGTTCTCTTCTATTACT
GCTATTACCGGTTACCCAGACCAATCTATCTACAATTCTGCCAACTCTATTTTGGACGCTTTGTCCAACTTTAGAAGGTTTATGGGTTTGCCATCCT
TCTCCATTAACTTGGGTCCAATGAAGGATGAAGGTAAGGTTTCTACCAACAAGAGCATCAAGAAGCTATTCAAGTCTAGAGGTTTGCCAAGCCTA
TCCTTGAACAAGTTATTTGGTTTGTTGGAGGTCGTCATCAACAACCCATCTAATCATGTTATCCCATCCCAATTGATTTGCTCCCCAATCGATTTCA
AGACCTACATCGAATCTTTCTCAACTATGAGGCCAAAGTTGTTACACTTGCAACCTACCATTTCCAAGCAGCAATCTTCTATCATTAACGATTCTA

CCAAGGCTTCCTCCAACATTTCATTGCAAGATAAGATCACCTCCAAGGTGTCTGATTTGTTGTCCATTCCAATCTCCAAGATCAACTTCGATCATC
CATTGAAACACTACGGCTTGGATTCTTTGTTGACCGTTCAATTCAAATCCTGGATCGACAAAGAATTCGAAAAGAACTTGTTCACCCATATCCAAT
TGGCCACCATCTCTATTAACTCATTCTTGGAAAAGGTGAACGGCTTGTCTACAAACAATAACAACAACAACAATTCCAACGTCAAGTCCTCTCCA
TCCATTGTCAAAGAAGAAATCGTTACCTTGGACAAGGATCAACAACCATTGCTATTGAAAGAACACCAGCACATTATCATCTCCCCAGATATTAG
AATCAACAAGCCAAAGAGGGAATCCTTGATTAGAACCCCAATCTTGAACAAATTCAACCAGATCACCGAATCCATTATCACTCCATCTACACCAT
CTTTGTCCCAATCCGATGTTTTGAAAACTCCACCAATCAAGTCTTTGAACAACACTAAGAACTCCAGCTTGATTAACACCCCACCAATTCAATCT
GTCCAACAACATCAAAAGCAACAACAAAAGGTCCAAGTCATCCAACAACAGCAACAACCATTATCCAGATTGTCCTACAAGAGCAACAACAAC
TCTTTCGTTTTGGGTATCGGTATTTCTGTTCCAGGTGAACCTATTTCCCAACAATCCTTGAAAGACTCCATCTCCAATGACTTTTCTGATAAGGCTG
AAACTAACGAGAAGGTCAAGAGAATCTTTGAGCAATCTCAAATCAAGACCAGACACTTGGTTAGAGATTACACTAAGCCAGAGAAACTCCATCA
AGTTCAGACATTTGGAAACCATTACCGATGTGAACAACCAGTTCAAGAAAGTTGTTCCAGATTTGGCTCAACAAGCCTGTTTGAGAAGCTTTGAA
AGATTGGGGTGGTGATAAGGGTGATATTACCCATATAGTTTCTGTTACCTCCACCGGTATTATCATCCCAGATGTTAATTTCAAGTTGATCGACTTG
TTGGGCTTGAACAAGGATGTTGAAAGAGTGTCTTTGAACCTAATGGGTTGTTTGGCTGGTTTGAGTTCTTTGAGAACTGCTGCTTCTTTGGCTAA
GGCTTCTCCAAGAAATAGAATTTTGGTTGTCTGTACCGAAGTCTGCTCCTTGCATTTTTCTAATACTGATGGTGGTGATCAAATGGTCGCCTCTTCT
ATTTTTGCTGATGGTTCTGCTGCTTACATTATTGGTTGTAACCCAAGAATTGAAGAAACCCCATTATACGAAGTCATGTGCTCCATTAACAGATCTT
TCCCAAATACCGAAAACGCCATGGTTTGGGATTTGGAAAAAGAAGGTTGGAACTTGGGTTTGGATGCTTCTATTCCAATTGTCATTGGTTCTGGT
ATTGAAGCCTTCGTTGATACTTTGTTGGATAAGGCTAAGTTGCAAACTTCCACTGCTATTTCTGCTAAGGATTGCGAATTCTTGATTCATACTGGTG
GCAAGTCCATCTTGATGAACATCGAAAATTCCTTGGGTATCGACCCAAAGCAAACTAAGAATACTTGGGATGTTTACCATGCCTACGGCAATATGT
CATCTGCCTCTGTTATTTTCGTTATGGATCATGCCAGAAAGTCCAAGTCTTTGGCCAACTTACTCAATTTCTTTGGCTTTTGGTCCAGGTTTGGCTTT
TGAAGGTTGTTTCTTGAAGAACGTCGTCGAACAGAAGCTCATTTCCGAAGAGGACTTGTAACCGCAACAACTACCTCGACTTTGGCTGGGACAC
TTTCAGTGAGGACAAGAAGCTTCAGAAGCGTGCTATCGAACTCAACCAGGGACGTGCGGCACAAATGGGCATCCTTGCTCTCATGGTGCACGA
ACAGTTGGGAGTCTCTATCCTTCCTTAAAAATTAATTTTCATTAGTTGCAGTCACTCCGCTTTGGTTTCGCTAGTGATAATAAGTGACTGAGGTAT
GTGCTCTTCTTATCTCCTTTTGTAGTGTTGCTCTTATTTTAAACAACTTTGCGGTTTTTTGATGACTTTGCGATTTTGTTGTTGCTTTGCAGTAAATT
GCAAGATTTAATAAAAAAACGCAAAGCAATGATTAAAGGATGTTCAGAATGAAACTCATGGAAACACTTAACCAGTGCATAAACGCTGGTCATG
AAATGACGAAGGCTATCGCCATTGCACAGTTTAATGATGACAGCCCGGAAGCGAGGAAAATAACCCGGCGCTGGAGAATAGGTGAAGCAGCGG
ATTTAGTTGGGGTTTCTTCTCAGGCTATCAGAGATGCCGAGAAAGCAGGGCGACTACCGCACCCGGATATGGAAATTCGAGGACGGGTTGAGCA
ACGTGTTGGTTATACAATTGAACAAATTAATCATATGCGTGATGTGTTTGGTACGCGATTGCGACGTGCTGAAGACGTATTTCCACCGGTGATCGG
GGTTGCTGCCCATAAAGGTGGCGTTTACAAAACCTCAGTTTCTGTTCATCTTGCTCAGGATCTGGCTCTGAAGGGGCTACGTGTTTTGCTCGTGG
AAGGTAACGACCCCCAGGGAACAGCCTCAATGTATCACGGATGGGTACCAGATCTTCATATTCATGCAGAAGACACTCTCCTGCCTTTCTATCTT
GGGGAAAAGGACGATGTCACTTATGCAATAAAGCCCACTTGCTGGCCGGGGCTTGACATTATTCCTTCCTGTCTGGCTCTGCACCGTATTGAAAC
TGAGTTAATGGGCAAATTTGATGAAGGTAAACTGCCCACCGATCCACACCTGATGCTCCGACTGGCCATTGAAACTGTTGCTCATGACTATGATG
TCATAGTTATTGACAGCGCGCCTAACCTGGGTATCGGCACGATTAATGTCGTATGTGCTGCTGATGTGCTGATTGTTCCCACGCCTGCTGAGTTGT
TTGACTACACCTCCGCACTGCAGTTTTTCGATATGCTTCGTGATCTGCTCAAGAACGTTGATCTTAAAGGGTTCGAGCCTGATGTACGTATTTTGC
TTACCAAATACAGCAATAGCAATGGCTCTCAGTCCCCGTGGATGGAGGAGCAAATTCGGGATGCCTGGGGAAGCATGGTTCTAAAAAATGTTGT
ACGTGAAACGGATGAAGTTGGTAAAGGTCAGATCCGGATGAGAACTGTTTTTGAACAGGCCATTGATCAACGCTCTTCAACTGGTGCCTGGAGA
AATGCTCTTTCTATTTGGGAACCTGTCTGCAATGAAATTTTCGATCGTCTGATTAAACCACGCTGGGAGATTAGATAATGAAGCGTGCGCCTGTTA
TTCCAAAACATACGCTCAATACTCAACCGGTTGAAGATACTTCGTTATCGACACCAGCTGCCCCGATGGTGGATTCGTTAATTGCGCGCGTAGGA
GTAATGGCTCGCGGTAATGCCATTACTTTGCCTGTATGTGGTCGGGATGTGAAGTTTACTCTTGAAGTGCTCCGGGGTGATAGTGTTGAGAAGAC
CTCTCGGGTATGGTCAGGTAATGAACGTGACCAGGAGCTGCTTACTGAGGACGCACTGGATGATCTCATCCCTTCTTTTCTACTGACTGGTCAAC
AGACACCGGCGTTCGGTCGAAGAGTATCTGGTGTCATAGAAATTGCCGATGGGAGTCGCCGTCGTAAAGCTGCTGCACTTACCGAAAGTGATTA
TCGTGTTCTGGTTGGCGAGCTGGATGATGAGCAGATGGCTGCATTATCCAGATTGGGTAACGATTATCGCCCAACAAGTGCTTATGAACGTGGTC
AGCGTTATGCAAGCCGATTGCAGAATGAATTTGCTGGAAATATTTCTGCGCTGGCTGATGCGGAAAATATTTCACGTAAGATTATTACCCGCTGTA
TCAACACCGCCAAATTGCCTAAATCAGTTGTTGCTCTTTTTTCTCACCCCGGTGAACTATCTGCCCGGTCAGGTGATGCACTTCAAAAAGCCTTT
ACAGATAAAGAGGAATTACTTAAGCAGCAGGCATCTAACCTTCATGAGCAGAAAAAAGCTGGGGTGATATTTGAAGCTGAAGAAGTTATCACTC
TTTTAACTTCTGTGCTTAAAACGTCATCTGCATCAAGAACTAGTTTAAGCTCACGACATCAGTTTGCTCCTGGAGCGACAGTATTGTATAAGGGCG
ATAAAATGGTGCTTAACCTGGACAGGTCTCGTGTTCCAACTGAGTGTATAGAGAAAATTGAGGCCATTCTTAAGGAACTTGAAAAGCCAGCACC
CTGATGCGACCTCGTTTTAGTCTACGTTTATCTGTCTTTACTTAATGTCCTTTGTTACAGGCCAGAAAGCATAACTGGCCTGAATATTCTCTCTGGG
CCCACTGTTCCACTTGTATCGTCGGTCTGATAATCAGACTGGGACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCACGGTCC
CACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCACGGTCCCACTCGTATCGTCGGTCTGATAATCAGACTGGGACCACGGTCCCACTCGTATC
GTCGGTCTGATTATTAGTCTGGGACCATGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCACGGTCCCACTCGTATCGTCGGTCTGAT
TATTAGTCTGGAACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGG
GACCACGATCCCACTCGTGTTGTCGGTCTGATTATCGGTCTGGGACCACGGTCCCACTTGTATTGTCGATCAGACTATCAGCGTGAGACTACGATT
CCATCAATGCCTGTCAAGGGCAAGTATTGACATGTCGTCGTAACCTGTAGAACGGAGTAACCTCGGTGTGCGGTTGTATGCCTGCTGTGGATTGC
TGCTGTGTCCTGCTTATCCACAACATTTTGCGCACGGTTATGTGGACAAAATACCTGGTTACCCAGGCCGTGCCGGCACGTTAACCGGGCTGCAT
CCGATGCAAGTGTGTCGCTGTCGACGAGCTCGCGAGCTCGGACATGAGGTTGCCCCGTATTCAGTGTCGCTGATTTGTATTGTCTGAAGTTGTTT
TTACGTTAAGTTGATGCAGATCAATTAATACGATACCTGCGTCATAATTGATTATTTGACGTGGTTTGATGGCCTCCACGCACGTTGTGATATGTAG
ATGATAATCATTATCACTTTACGGGTCCTTTCCGGTGATCCGACAGGTTACGGGCGGCGACCTCGCGGGTTTTCGCTATTTATGAAAATTTTCCG
GTTTAAGGCGTTTCCGTTCTTCTTCGTCATAACTTAATGTTTTTATTTAAAATACCCTCTGAAAAGAAAGGAAACGACAGGTGCTGAAAGCGAGC
TTTTTGGCCTCTGTCGTTTCCTTTCTCTGTTTTTGTCCGTGGAATGAACAATGGAAGTCCGAGCTCATCGCTAATAACTTCGTATAGCATACATTAT
ACGAAGTTATA

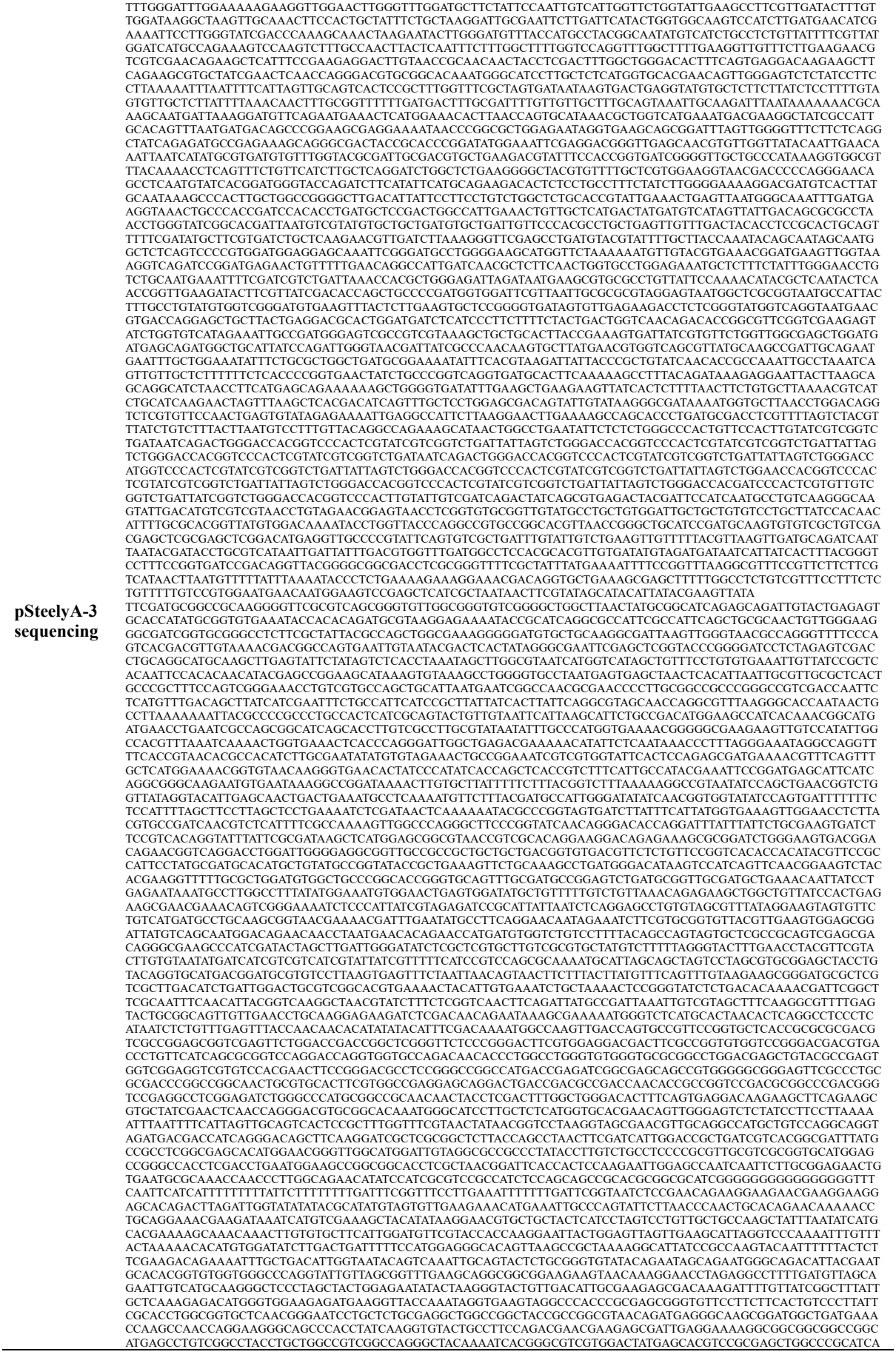
pSteelyA-2 sequencing

TTCGATGCGGCCGCAAGGGGTTCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGT
GCACCATATGCGGTGTGAAATACCACACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCCATTCGCCATTCAGCTGCGCAACTGTTGGGAAG
GGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCA
GTCACGACGTTGTAAAACGACGGCCAGTGAATTGTAATACGACTCACTATAGGGCGAATTCGAGCTCGGTACCCGGGGATCCTCTAGAGTCGAC
CTGCAGGCATGCAAGCTTGAGTATTCTATAGTCTCACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTC
ACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACT
GCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGAACCCCTTGCGGCCGCCCGGGCCGTCGACCAATTC
TCATGTTTGACAGCTTATCATCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTTTAAGGGCACCAATAACTG
CCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACTGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATG
ATGAACCTGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAACGGGGGCGAAGAAGTTGTCCATATTGGC
CCACGTTTAAATCAAAACTGGTGAAACTCACCCAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGTT
TTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGTGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTT
GCTCATGGAAAACGGTGTAACAAGGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGAAATTCCGGATGAGCATTCATC
AGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTTTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTG
GTTATAGGTACATTGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGGTGGTATATCCAGTGATTTTTTC
TCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGATAACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTTA
CGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACAGGGACACCAGGATTTATTTATTCTGCGAAGTGATCT
TCCGTCACAGGTATTTATTCGCGATAAGCTCATGGAGCGGCGTAACCGTCGCACAGGAAGGACAGAGAAAGCGCGGATCTGGGAAGTGACGGA
CAGAACGGTCAGGACCTGGATTGGGGAGGCGGTTGCCGCCGCTGCTGCTGACGGTGTGACGTTCTCTGTTCCGGTCACACCACATACGTTCCGC
CATTCCTATGCGATGCACATGCTGTATGCCGGTATACCGCTGAAAGTTCTGCAAAGCCTGATGGGACATAAGTCCATCAGTTCAACGGAAGTCTAC
ACGAAGGTTTTTGCGCTGGATGTGGCTGCCCGGCACCGGGTGCAGTTTGCGATGCCGGAGTCTGATGCGGTTGCGATGCTGAAACAATTATCCT
GAGAATAAATGCCTTGGCCTTTATATGGAAATGTGGAACTGAGTGGATATGCTGTTTTTGTCTGTTAAACAGAGAAGCTGGCTGTTATCCACTGAG
AAGCGAACGAAACAGTCGGGAAAATCTCCCATTATCGTAGAGATCCGCATTATTAATCTCAGGAGCCTGTGTAGCGTTTATAGGAAGTAGTGTTC
TGTCATGATGCCTGCAAGCGGTAACGAAACGATTTGAATATGCCTTCAGGAACAATAGAAATCTTCGTGCGGTGTTACGTTGAAGTGGAGCGG
ATTATGTCAGCAATGGACAGAACAACCTAATGAACACAGAACCATGATGTGGTCTGTCCTTTTACAGCCAGTAGTGCTCGCCGCAGTCGAGCGA
CAGGGCGAAGCCCATCGATACTAGCTTGATTGGGATATCTCGCTCGTGCTTGTCGCGTGCTATGTCTTTTAGGGTACTTTGAACCTACGTTCGTA
CTTGTGTAATATGATCATCGTCGTCATCGTATTATCGTTTTTCATCCGTCCAGCGCAAAATGCATTAGCAGCTAGTCCTAGCGTGCGGAGCTACCTG
TACAGGTGCATGACGGATGCGTGTCCTTAAGTGAGTTTCTAATTAACAGTAACTTCTTTACTTATGTTTCAGTTTGTAAGAAGCGGGATGCGCTCG
TCGCTTGACATCTGATTGGACTGCGTCGGCACGTGAAAACTACATTGTGAAATCTGCTAAAACTCCGGGTATCTCTGACACAAAACGATTCGGCT
TCGCAATTTCAACATTACGGTCAAGGCTAACGTATCTTTCTCGGTCAACTTCAGATTATGCCGATTAAATTGTCGTAGCTTTCAAGGCGTTTTGAG
TACTGCGGCAGTTGTTGAACCTGCAAGGAGAAGATCTCGACAACAGAATAAAGCGAAAAATGGGTCTCATGCACTAACACTCAGGCCTCCCTC
ATAATCTCTGTTTGAGTTTACCAACAACACATATATACATTTCGACAAAATGGCCAAGTTGACCAGTGCCGTTCCGGTGCTCACCGCGCGCGACG
TCGCCGGAGCGGTCGAGTTCTGGACCGACCGGCTCGGGTTCTCCCGGGACTTCGTGGAGGACGACTTCGCCGGTGTGGTCCGGGACGACGTGA
CCCTGTTCATCAGCGCGGTCCAGGACCAGGTGGTGCCAGACAACACCCTGGCCTGGGTGTGGGTGCGCGGCCTGGACGAGCTGTACGCCGAGT
GGTCGGAGGTCGTGTCCACGAACTTCCGGGACGCCTCCGGGCCGGCCATGACCGAGATCGGCGAGCAGCCGTGGGGCGGGAGTTCGCCCTGC
GCGACCCGGCCGGCAACTGCGTGCACTTCGTGGCCGAGGAGCAGGACTGACCGACGCCGACCAACACCGCCGGTCCGACGCGGCCCGACGGG
TCCGAGGCCTCGGAGATCTGGGCCCATGCGGCCGCAACAACTACCTCGACTTTGGCTGGGACACTTTCAGTGAGGACAAGAAGCTTCAGAAGC
GTGCTATCGAACTCAACCAGGGACGTGCGGCACAAATGGGCATCCTTGCTCTCATGGTGCACGAACAGTTGGGAGTCTCTATCCTTCCTTAAAA
ATTTAATTTTCATTAGTTGCAGTCACTCCGCTTTGGTTTCGTAACTATAACGGTCCTAAGGTAGCGAACGTTGCAGGCCATGCTGTCCAGGCAGGT
AGATGACGACCATCAGGGACAGCTTCAAGGATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCATTGGACCGCTGATCGTCACGGCGATTTATG



AATGTGCGAGGTGGAAAATATACCAGCATGACAATGAATCTTGGAGATTCTTTTGCTGTCATCAAGATTCACCGCCAAATCTTCAGGAACCTATC
ACGTCCACAGGCGATGTTAATTCTTGAGTCGTCAAAACAAAGTCCTGTCCTACCTGTAGAAGTTGACAGCGAGCAATTGTATGCAAACTTCTGAC
TTTGTTATAATAACATTAAAGGTAATTAAGTATCTTCAATTAGGCATTTTGTCACTGTCAGTCCGTTCCGACAAATATAGGTAGATTTGGAATGAATC
TTTTCTATGCTGCTGCGAATCTTGTACACCTTTGAGGCCGTAGATTCTGTCCGACGAAGCGATAATTATTGCAAAATACATGGACTCATTATTTTGA
TTCGATTTCTTTTTGGTATCCGACTCGAAAAGATCCATCACGGCGAGCGCCACCATGAACAAGAACTCCAAAATCCAGTCCCCAAACTCTTCTGA
TGTTGCTGTTATTGGTGTTGGTTTTAGATTCCCAGGTAACTCTAATGACCCAGAATCTTTGTGGAACAACTTGTTGGATGGTTTCGATGCTATTACC
CAAGTCCCAAAAGAAAGATGGGCTACTTCTTTTAGAGAGATGGGTTTGATCAAGAACAAGTTCGGTGGTTTCTTGAAGGATTCTGAATGGAAGA
ATTTCGACCCTTTGTTCTTTGGTATCGGTCCAAAAGAAGCTCCATTCATTGATCCACAACAAAGGTTGTTGTTGTCCATCGTTTGGGAATCTTTGG
AAGATGCTTACATCAGACCAGATGAATTGAGAGGTTCTAACACTGGTGTTTTCATCGGTGTTTCTAACAACGATTACACCAAGTTGGGTTTCCAA
GACAACTACTCTATTTCTCCATACACTATGACCGGCTCTAACTCTTCATTGAACTCCAACAGAATTTCCTACTGCTTCGATTTTAGAGGTCCATCCA
TTACTGTTGATACCGCTTGTTCTTCTTCCTTGGTTTCTGTTAATTTGGGTGTCCAATCCATCCAAATGGGTGAATGTAAGATTGCTATTTGCGGTGG
TGTTAACGCTTTGTTTGATCCATCTACATCTGTTGCCTTTTCCAAGTTGGGTGTTTTGTCTGAAAATGGCAGATGCAACTCTTTTAGTGATCAAGCC
TCTGGTTACGTTAGATCTGAAGGTGCTGGTGTTGTTGTTTTGAAGTCTTTGGAACAAGCTAAGTTGGATGGTGATAGAATCTACGGTGTTATCAAG
GGTGTTTCCTCTAATGAAGATGGTGCTTCTAATGGTGACAAGAACTCTTTGACTACTCCATCTTGTGAAGCCCAATCCATTAACATTTCTAAGGCT
ATGGAAAAGGCCTCCTTGTCTCCATCTGATATCTATTACATTGAAGCCCATGGTACTGGTACTCCAGTTGGTGATCCAATTGAAGTTAAGGCCTTG
TCCAAGATCTTCTCCAACTCTAACAACAACCAGTTGAACAACTTCTCTACCGATGGTAATGATAACGATGATGATGATGACGATAACACCTCTCCA
GAACCATTATTGATTGGCTCATTCAAGTCCAACATCGGTCATTTGGAATCTGCTGCTGGTATTGCTTCTTTGATTAAGTGTTGCTTGATGTTGAAGA
ACAGGATGTTGGTTCCATCCATTAACTGCTCTAATTTGAACCCATCCATTCCATTCGATCAGTACAACATCTCCGTTATCAGAGAAATCAGACAATT
CCCAACCGATAAGTTGGTTAACATCGGTATCAATTCTTTCGGTTTCGGTGGTTCTAACTGCCATTTGATTATTCAAGAGTACAACAACAACTTCAA
GAACAACTCTACCATCTGCAATAACAACAACAACAACAATAACAACATCGACTACTTGATCCCAATCTCCTCTAAGACTAAGAAGTCCTTGGATA
AGTACTTGATTTTGATCAAGACCAACTCCAACTACCACAAGGATATTTCTTTCGATGACTTCGTCAAGTTCCAAATCAAGTCTAAGCAGTACAAC
TTGTCCAACAGAATGACTACCATTGCTAACGATTGGAACTCCTTCATTAAGGGTTCTAACGAATTCCACAACTTGATCGAATCTAAGGATGGTGA
AGGTGGTTCTTCATCTTCTAACAGAGGTATTGATTCCGCCAATCAAATCAACACTACTACTACCTCTACCATCAACGATATCGAACCTTTGTTGGTT
TTCGTTTTCTGTGGTCAAGGTCCACAATGGAATGGTATGATTAAGACCTTGTACAACTCCGAGAACGTTTCAAGAACACCGTTGATCATGTTGA
CAGCATCTTGTACAAGTACTTCGGTTACTCCATTTTGAACGTCTTGTCTAAGATCGATGATAACGACGATTCCATCAACCATCCAATAGTTGCTCA
ACCATCTTTGTTCTTGTTGCAAATTGGTTTGGTCGAGTTGTTTAAGTACTGGGGTATCTACCCATCTATCTCTGTTGGTCATTCTTTCGGTGAAGTC
TCTTCTTATTACTTGTCCGGTATCATCTCTTTGGAAACCGCTTGTAAAATCGTCTACGTCAGATCCTCTAATCAGAACAAAACTATGGGTTCCGGTA
AGATGTTGGTTGTTTCTATGGGTTTTAAGCAATGGAACGATCAATTCTCTGCTGAATGGTCCGATATTGAAATTGCTTGTTACAACGCTCCAGATTC
CATAGTTGTTACTGGTAACGAAGAAAGATTGAAAGAATTGTCCATCAAGTTGTCCGACGAATCCAATCAAATTTTCAACACCTTCTTGAGGTCCC
CATGTTCTTTTCATTCTTCCCATCAAGAAGTCATCAAGGGTTCTATGTTCGAAGAGTTGTCTAACTTGCAATCTACTGGTGAAACCGAAATCCCTT
TGTTCTCTACTGTTACTGGTAGACAAGTTTTGTCTGGTCATGTTACTGCTCAACACATCTACGATAATGTTAGAGAACCAGTCTTGTTCCAAAAGA
CGATTGAATCCATTACCTCCTACATCAAGTCTCACTACCCATCCAATCAAAAGGTTATCTACGTTGAAATTGCTCCACACCCAACCTTGTTTTCATT
GATCAAAAAGTCCATCCCATCCTCCAACAAGAATTCCTCTTCTGTTTTGTGTCCATTGAACAGAAAAGAAAACTCCAACAACTCCTACAAGAAG
TTCGTTTCTCAGTTGTACTTCAACGGTGTTAACGTTGACTTCAACTTCCAGTTGAACTCCATTTGCGATAACGTTAACAACGATCACCATTTGAAC
AACGTCAAGCAAAACTCCTTCAAAGAGACTACCAATTCCTTGCCAAGATACCAATGGGAACAAGATGAATATTGGTCCGAACCATTGATCTCCA
GAAAGAATAGATTGGAAGGTCCAACTACTTCCTTGTTGGGTCATAGAATTATCTACAGCTTCCCAGTTTTCCAATCCGTTTTGGACTTGCAATCTG
ACAACTACAAATACTTGTTGGACCACTTGGTTAACGGTAAGCCAGTTTTTCCAGGTGCTGGTTATTTGGATATCATCATCGAATTCTTCGACTACC
AAAAGCAGCAGTTGAATTCCTCTGATTCCTCTAACTCCTACATCATCAACGTTGACAAGATCCAATTCTTGAACCCAATTCACTTGACCGAAAAC
AAGTTGCAAACCTTGCAATCTTCTTTCGAACCTATCGTTACTAAGAAGTCTGCCTTCTCTGTTAACTTCTTCATCAAGGATACCGTCGAGGATCAA
TCTAAGGTTAAGTCTATGTCTGACGAAACTTGGACTAACACTTGTAAGGCTACCATTTCCTTGGAACAACAACAGCCATCTCCATCTTCTACTTTG
ACTTTGTCTAAGAAGCAAGACTTGCAGATCTTGAGAAACAGATGCGATATTAGCAAGCTAGACAAGTTTGAGTTGTACGACAAGATCTCTAAGA
ATTTGGGCTTGCAGTACAACTCCTTGTTTCAAGTTGTTGATACCATCGAAACTGGTAAGGATTGCTCTTTTGTCTACTTTGTCTTTGCCAGAAGATA
CTTTGTTCACCACCATTTTGAACCCATGCTTGTTGGATAACTGTTTCCATGGTTTGTTGACCTTGATCAACGAAAAGGGTTCTTTCGTTGTCGAGT
CCATTTCTTCTGTTTCTATCTACTTGGAGAACATCGGTTCCTTCAATCAAACTTCTGTTGGTAACGTCCAGTTCTACTTGTACACCACTATTTCTAA
AGCCACCTCCTTTAGTTCTGAAGGTACTTGTAAGTTGTTCACCAAGGATGGTTCCTTGATTTTGTCTATCGGTAAGTTCATCATCAAGTCCACCAA
TCCAAAGTCTACTAAGACCAACGAAACTATCGAATCTCCATTGGACGAAACCTTCTCTATTGAATGGCAATCTAAGGATTCTCCAATTCCAACCCC
ACAACAAATCCAACAACAATCTCCATTGAACTCTAACCCATCCTTCATTAGATCTACCATCTTGAAGGACATCCAGTTCGAACAATACTGCTCCTC
CATTATCCACAAAGAATTGATCAACCACGAAAAGTACAAGAACCAGCAATCCTTCGATATCAACTCCTTGGAAAACCACTTGAACGATGACCAA
TTGATGGAATCCTTGTCCATCTCCAAAGAATACTTGAGATTCTTCACCAGGATCATCTCCATCATTAAGCAATACCCAAAGATCTTGAACGAAAAA
GAGCTAAAAGAATTGAAAGAAATCATCGAATTGAAGTACCCATCCGAAGTTCAGTTGTTGGAATTCGAAGTTATCGAGAAGGTGTCCATGATTAT
CCCAAAGTTGTTGTTCGAAAACGACAAGCAATCTTCCATGACCTTGTTCCAAGATAACTTGTTGACCAGGTTCTACTCCAATTCTAACTCTACCA
GATTCTACTTGGAAAGGGTTTCCGAAATGGTCTTGGAATCTATTAGACCAATCGTCAGAGAAAAGAGGGTGTTCAGAATTTTGGAAATTGGTGCT
GATACAGCCTCTTTGTCTAATGTTGTTTTGACTAAGTTGAACACCTACTTGTCCACCTTGAATTCTAATGGTGGTTCTGGTTACAACATCATCATTG
AGTACACCTTCACCGATATTTCCGCCAACTTCATTATTGGTGAAATCCAAGAAACCATGTGCAACTTGTACCCAAACGTTACTTTCAAGTTCTCCG
TCTTGGACTTGGAGAAAGAGATTATTAACTCCTCCGATTTCTTGATGGGTGATTACGATATAGTTTTGATGGCCTACGTTATCCATGCCGTTTCTAA
CATTAAGTTCCATCGAACAGTTGTACAAGTTGTTGTCTCCAAGAGGTTGGTTGTTGTGTATTGAACCTAAGTCCAACGTTGTGTTCTCCGATTT
GGTTTTCGGTTGTTTTAATCAGTGGTGGAACTACTACGATGATATTAGAACTACCCACTGCTCCTTGTCTGAATCTCAATGGAATCAGTTGTTGTTG
AACCAGTCCTTGAACAACGAATCCTCTTCTTCTTCTAACTGTTACGGTGGTTTCTCCAACGTTTCTTTTATTGGTGGTGAAAAGGATGTCGACTCC
CATTCTTTCATATTGCACTGCCAAAAAGAATCCATCTCCCAAATGAAGTTAGCCACCACTATTAACAACGGTTTGTCATCTGGTTCCATCGTTATCG
TTTTGAACTCTCAACAATTGACCAACATGAAGTCCTACCCAAAGGTTATTGAGTATATTCAAGAGGCTACCTCTTTGTGCAAGACCATTGAAATTA
TCGATTCCAAGGACGTCTTGAACTCTACCAATTCAGTTTTGGAAAAGATCCAAAAGTCCTTGTTGGTGTTCTGTTTGTTGGGTTATGACTTGTTGG
AGAACAACTACCAAGAACAGTCTTTCGAATACGTTAAGTTGTTGAACTTGATCTCTACTACCGCCTCTTCATCTAATGATAAGAAAACCACCAAAG
GTCTTGTTGATCACCAAGCAATCTGAAAGAATCTCCAGGTCTTTCTACTCCAGATCCTTGATTGGTATTTCCAGAACCTCTATGAACGAGTACCCA
AATTTGTCCATTACCTCTATCGATTTGGATACCAACGACTACTCATTGCAGTCTTTGTTGAAGCCAATCTTCAGCAACTCTAAGTTTTCCGACAAC
GAGTTCATCTTCAAAAAGGGCTTGATGTTCGTGTCCAGGATCTTTAAGAACAAGCAGTTGCTAGAATCCTCCAACGCTTTTGAAACTGACTCTTC
TAACTTGTACTGTAAGGCCTCTTCTGACTTGTCTTACAAGTACGCTATTAAGCAGTCTATGTTGACCGAAAATCAGATCGAAATCAAGGTTGAATG
CGTCGGTATTAACTTCAAGGACAACCTATTCTACAAGGGCTTGTTGCCACAAGAAATTTTCAGAATGGGTGACATCTACAATCCACCATATGGTTT
GGAATGCTCTGGTGTTATTACCAGAATTGGTTCTAACGTCACCGAATACTCAGTTGGTCAAAATGTTTTTGGTTTCGCCAGACATTCTTTGGGTTC
TCATGTTGTTACCAACAAGGATTTGGTTATCTTGAAGCCAGATACCATCTCATTTTCTGAAGCTGCTTCTATCCCAGTTGTTTACTGTACTGCTTGG
TACTCCTTGTTCAACATTGGTCAGTTGTCTAACGAAGAATCCATCCTAATTCATTCTGCTACTGGTGGTGTAGGTTTGGCTTCTTTGAATTTGTTGA
AAATGAAGAATCAGCAACAGCAACCATTGACCAATGTTTATGCTACTGTTGGCTCTAACGAGAAGAAGAAGTTCTTGATCGATAACTTCAACAA
CTTGTTCAAAGAGGACGGCGAAAACATTTTCTCTACCAGAGACAAAGAATACTCCAACCAGTTGGAATCCAAGATCGATGTTATTTTGAACACC
TTGTCCGGTGAATTCGTCGAATCTAATTTCAAGTCCTTGAGATCCTTCGGTAGATTGATTGATTTGTCTGCTACTCACGTTTACGCCAATCAACAA
ATTGGTCTAGGTAACTTCAAGTTCGACCACTTGTATTCTGCTGTTGACTTGGAAAGATTGATCGACGAAAAACCTAAGTTGTTGCAGTCCATCTT
GCAAAGAATTACCAACTCTATCGTCAACGGTTCCTTGGAAAAAATTCCAATTACCATCTTCCCATCCACCGAAACTAAGGATGCTATCGAATTATT
GTCCAAGAGATCCCATATCGGTAAAGTTGTTGTAGATTGCACCGATATCTCTAAGTGTAATCCTGTTGGTGATGTGATCACCAACTTCTCTATGAG
ATTGCCAAAGCCAAACTACCAGTTGAATTTGAACTCCACCTTGTTGATTACTGGTCAGTCTGGTTTGTCTATCCCTTTGTTGAATTGGTTGTTGTC
TAAGTCTGGTGGTAACGTTAAGAACGTTGTCATCATTTCTAAGTCCACCATGAAGTGGAAGTTGCAGACTATGATTTCCCATTTCGTTTCCGGTTT
CGGTATCCATTTTAACTACGTTCAAGTCGACATCTCCAACTACGATGCTTTGTCTGAAGCTATTAAGCAATTGCCATCTGATTTGCCACCAATCACC
TCTGTTTTTCATTTGGCTGCTATCTACAACGATGTTCCAATGGATCAAGTTACCATGTCTACCGTTGAATCTGTTCATAACCCTAAAGTTTTGGGTG
CCGTTAACTTGCATAGAATCTCTGTTTCTTTTGGTTGGAAGTTGAACCACTTCGTCTTGTTCTCTTCTATTACTGCTATTACCGGTTACCCAGACCA
ATCTATCTACAATTCTGCCAACTCTATTTTGGACGCTTTGTCCAACTTTAGAAGGTTTATGGGTTTGCCATCCTTCTCCATTAACTTGGGTCCAATG
AAGGATGAAGGTAAGGTTTCTACCAACAAGAGCATCAAGAAGCTATTCAAGTCTAGAGGTTTGCCAAGCCTATCCTTGAACAAGTTATTTGGTTT
GTTGGAGGTCGTCATCAACAACCCATCTAATCATGTTATCCCATCCCAATTGATTTGCTCCCCAATCGATTTCAAGACCTACATCGAATCTTTCTCA
ACTATGAGGCCAAAGTTGTTACACTTGCAACCTACCATTTCCAAGCAGCAATCTTCTATCATTAACGATTCTACCAAGGCTTCCTCCAACATTTCA
TTGCAAGATAAGATCACCTCCAAGGTGTCTGATTTGTTGTCCATTCCAATCTCCAAGATCAACTTCGATCATCCATTGAAACACTACGGCTTGGAT
TCTTTGTTGACCGTTCAATTCAAATCCTGGATCGACAAAGAATTCGAAAAGAACTTGTTCACCCATATCCAATTGGCCACCATCTCTATTAACTCA
TTCTTGGAAAAGGTGAACGGCTTGTCTACAAACAATAACAACAACAATTCCAACGTCAAGTCCTCCATCCATTGTCAAAGAAGAAATCG
TTACCTTGGACAAGGATCAACAACCATTGCTATTGAAAGAACACCAGCACATTATCATCTCCCCAGATATTAGAATCAACAAGCCAAAGAGGGAA
TCCTTGATTAGAACCCCAATCTTGAACAAATTCAACCAGATCACCGAATCCATTATCACTCCATCTACACCATCTTTGTCCCAATCCGATGTTTTGA
AAACTCCACCAATCAAGTCTTTGAACAACACTAAGAACTCCAGCTTGATTAACACCCCACCAATTCAATCTGTCCAACAACATCAAAAGCAACA
ACAAAAGGTCCAAGTCATCCAACAACAGCAACAACCATTATCCAGATTGTCCTACAAGAGCAACAACAACTCTTTCGTTTTGGGTATCGGTATTT
CTGTTCCAGGTGAACCTATTTCCCAACAATCCTTGAAAGACTCCATCTCCAATGACTTTTCTGATAAGGCTGAAACTAACGAGAAGGTCAAGAG
AATCTTTGAGCAATCTAAATCAAGACCAGACACTTGGTTAGAGATTACACTAAGCCAGAGAACTCCATCAAGTTCAGACATTTGGAAACCATTA
CCGATGTGAACAACCAGTTCAAGAAAGTTGTTCCAGATTTGGCTCAACAAGCCTGTTTGAGAGCTTTGAAAGATTGGGGTGGTGATAAGGGTGA
TATTACCCATATAGTTTCTGTTACCTCCACCGGTATTATCATCCCAGATGTTAATTTCAAGTTGATCGACTTGTTGGGCTTGAACAAGGATGTTGAA
AGAGTGTCTTTGAACCTAATGGGTTGTTTGGCTGGTTTGAGTTCTTTGAGAACTGCTGCTTCTTTGGCTAAGGCTTCTCCAAGAAATAGAATTTTG
GTTGTCTGTACCGAAGTCTGCTCCTTGCATTTTTCTAATACTGATGGTGGTGATCAAATGGTCGCCTCTTCTATTTTTGCTGATGGTTCTGCTGCTT
ACATTATTGGTTGTAACCCAAGAATTGAAGAAACCCCATTATACGAAGTCATGTGCTCCATTAACAGATCTTTCCCAAATACCGAAAACGCCATGG

pSteelyA-3 sequencing



ATGGCGACCTGGGCCGCCTGGGCGGCCTGCTGAAACTCTGGCTCACCGACGACCCGCGCACGGCGCGGTTCGGTGATGCCACGATCCTCGCCC
TGCTGGCGAAGATCGAAGAGAAGCAGGACGAGCTTGGCAAGGTCATGATGGGCGTGGTCCGCCCGAGGGCAGAGCCATGACTTTTTTAGCCGC
TAAAACGGCCGGGGGGTGCGCGTGATTGCCAAGCACGTCCCCATGCGCTCCATCAAGAAGAGGGCACTTCGAGCTGTAAGTACATCACCGACGA
GCAAGGCAAGACGATCCGCGCCTGTTTTATTGAGAACGTTGTTCGTGTTGGCCTCAATGGTAGCGATGCGTCATTCAGCGAAGTTCTGGTGCTGA
TGATGTGGTTCGCTTTGCCACTGGTCAATGTGGTAAGCCCGTGTAATGTCAGTAACCTTTTTACTGATCTCAGCTTGAGCACGGTCGCTGATGAGC
TTATCCATGGCCCCACGGTAACGGATATGATCCTCTAGGGCGTTGACAAATTCTTTGTCGGTTTTCATGGGGTAGACGTCAGTGACCAGGGAACG
TCTCCCTACAAAAAGTTGCGCATACTTTGCTCCGTTGTCGACGGCAGGGGTGTCGGACCAGATTGTATCTGTGGCAACGGCCTCATTGCGTCAAT
GGACCTTTAAAGCCGGGAAACGAGACTTGAAATGTTTACGTAGAGGTGCATTATAGACCTCACGCGCATATTGAGTGGTTGCAAAAATGGTTTTG
CGAACGGTATCACTGGAGACCCATGCCAGGCAAGGACGGAGGGCATCATAATCATGTTCATTGCGTTGGACAGCATGCTTGTTACACGTAAGTAT
ACGGTCGAGAGCATCGTGACGTTGAGAATTACAGATGGCAACGTGTCGGTGCATGTATTGGCCGAGATTGTCAAATCGTGGCTCAATATATGGCA
TGCCAGGTAAGTCATGGATGTCGGTATGCCATTCAGCCTTCATGTCGATTTCGTTGTCGAGAACTGATGGGTCCCAGTCAACGTCAGAAGTAAGA
ACAACATGGGGAAGGGAATGAAGTTCATGACTGGTGGGTGCGCGCATATCCATGTACGGAAGACCCTGTCGTATGTTTAGTGGGATGATGTAATC
ATCCAGGGTAACAATACGTTGGTTGCCGCCCACGGTACGAGACCGATCTTGAACACAATTGTGGCAATGTTCAAGTTGGGCGCTGGAATGGATG
GTTTTACCTTTACCAAGATGCGCGTATTGATGCATGATAACAACAATGGGCCCTCGTTGCGATTCGACAAGACCGGCAGCCGTGACAATGTCTAG
ATTGGACAAGGTATGTTCATTTATACCGGTTATGTTGGCTGAACGTCCAGTTTTGTGCAGTACGGTAACATCACTCCCCGCAAGTCCACCGTTGGC
ACCGCGATCGACAAGTGCAGATGTGGTATTTTGTACTGTGTGTCGGGAGACTTGGTACTGTAACACATTTGACTGAAGTGACTTGGTGGTACCAT
GAGGACGGAGGGAACCCGTTGATACGGTGTTGCGTGACGCTGCAAGTACCTTACGTATGTCGCCTGGATCCATACGTCCAACCCGATCAGAAAG
ATGTGCTAGGAGTTCCGTTTCTGGTGCACAATCATGAAAGGTATCCACGGGAGTTCCATTGTTGTCGTCCGGGGCTGCGTCGCGACTGCCAGTTG
CATGTACATGCGCCATCAAGGGACGTGCCAATCCATTACTAACCGGGAGAACCTTAGCTGGGGCAGCCAATCCCTGAAGTATAGCCTTAGCATCA
TCTGAGAGCTGGTTCCACATATCTCTAGGGATATAAGGTCGTTCACGGTTAGTGGGCGTACTATGATGTCGGTTATTAGTACAATTCTCCCTGCGGG
CATGCGCATTGGCTTCATAGAGTACGGAAGGTGACAAGTCAATATTGTAGTTAACATCCGGATCAGTGTCAAAGTCTGTAGGATGGTAAGAAAGA
TCAGTAGAATGAATACTACGCTTTTCCTTGGGACTACGGGAATTAGAGAAGTTGTTTCCTTTATTGTAGAGTGACGCTGAAGCAAGTAGAAGACT
AAGGTAACTCTCGTAGCTAATAGGATTACCTCCTTTGGCTAGGTCAAGAGTGGCTGTGATCTTCACTTGACAAAGTTCCGGTACATTGTGGACAG
CATTCTCCAAAAGACTAAGACACAGTTGCTTTGGGAGTTGCTCAGCCATTGGTACAGTATTGTGGTAGATACAAAGGTGGTTCTTCCAATGAAGG
ATAAATCCTTCTGCTGTGCCTGTCCATGAGGATCCATATTTCGCCGTAGTTAGGTAACCAAGCGTGGTGGCTGAACTGATCTTCGCACTTGCTGAT
TCCGTATAGTGTTTGACAACTTTACAAAACACTTCTTGCGCAGTTCGCTCTAGCCTGGACAGAGAAGGGACAATCTTCGTCGTTAAGACTCGTGC
GAATAGCAAAAGATCACAAAATAGCACATCGGCACCGACCAACGATTATTTCCAAGGAAAAAAAGAATGCTTCACTACAAGAAATTGTGTCATC
CCTATACAGAGTCTTGTTACTGTGACAGAAAATTGATGGAAGATGTGGCGGATTGCCTTTACACTAGCCAACTTGTTCGACTAATTGCAGCTTCTT
CTGAGAGGCTTCACCGAGTAACGCGAAGAACACCGGTGTCTCGTACATGCTCGTCGGTGAACGCTCGTCCAATGACACCCCCCACTTTGTATCA
ATATCCCAACTTGGTAGTGAACTGGAATGATACATGCAATTTCGCGCCGCATCAACAGCCACGGGCACCATCGACGAATAGACTCGGTCGAGCTG
GTTGCCCTCGCCGCTGGGCTGGCGGCCGTCTATGGCCCTGCAAACGCGCCAGAAACGCCGTCGAAGCCGTGTGCGAGACACCGCGGCCGGCCG
CCGGCGTTGTGGATACCTCGCGGAAAACTTGGCCCTCACTGACAGATGAGGGGCGGACGTTGACACTTGAGGGGCCGACTCACCCGGCGCGGC
GTTGACAGATGAGGGGCAGGCTCGATTTCGGCCGGCGACGTGGAGCTGGCCAGCCTCGCAAATCGGCGAAAACGCCTGATTTTACGCGAGTTT
CCCACAGATGATGTGGACAAGCCTGGGGATAAGTGCCCTGCGGTATTGACACTTGAGGGGCGCGACTACTGACAGATGAGGGGCGCGATCCTTG
ACACTTGAGGGGCAGAGTGCTGACAGATGAGGGGCGCACCTATTGACATTTGAGGGGCTGTCCACAGGCAGAAAATCCAGCATTTGCAAGGGT
TTCCGCCCGTTTTTCGGCCACCGCTAACCTGTCTTTTAACCTGCTTTTAAACCAATATTTATAAACCTTGTTTTTAACCAGGGCTGCGCCCTGTGCG
CGTGACCGCGCACGCCGAAGGGGGGTGCCCCCCCTTCTCGAACCCTCCCGGTCGAGTGAGCGAGGAAGCACCAGGGAACAGCACTTATATATT
CTGCTTACACACGATGCCTGAAAAAACTTCCCTTGGGGTTATCCACTTATCCACGGGGATATTTTTATAATTATTTTTTTTATAGTTTTTAGATCTTCT
TTTTTAGAGCGCCTTGTAGGCCTTTATCCATGCTGGTTCTAGAGAAGGTGTTGTGACAAATTGCCCTTTCAGTGTGACAAATCACCCTCAAATGAC
AGTCCTGTCTGTGACAAATTGCCCTTAACCCTGTGACAAATTGCCCTCAGAAGAAGCTGTTTTTTCACAAAGTTATCCCTGCTTATTGACTCTTTT
TTATTTAGTGTGACAATCTAAAAACTTGTCACACTTCACATGGATCTGTCATGGCGGAAACAGCGGTTATCAATCACAAGAAACGTAAAAATAGC
CCGCGAATCGTCCAGTCAAACGACCTCACTGAGGCGGCATATAGTCTCTCCCGGGATCAAAAACGTATGCTGTATCTGTTCGTTGACCAGATCAG
AAAATCTGATGGCACCCTACAGGAACATGACGGTATCTGCGAGATCCATGTTGCTAAATATGCTGAAATATTCGGATTGACCTCTGCGGAAGCCA
GTAAGGATATACGGCAGGCATTGAAGAGTTTCGCGGGGAAGGAAGTGGTTTTTTATCGCCCTGAAGAGGATGCCGGCGATGAAAAAGGCTATGA
ATCTTTTCCTTGGTTTATCAAACGTGCGCACAGTCCATCCAGAGGGCTTTACAGTGTACATATCAACCCATATCTCATTCCCTTCTTTATCGGGTTA
CAGAACCGGTTTACGCAGTTTCGGCTTAGTGAAACAAAAGAAATCACCAATCCGTATGCCATGCGTTTATACGAATCCCTGTGTCAGTATCGTAA
GCCGGATGGCTCAGGCATCGTCTCTCTGAAAATCGACTGGATCATAGAGCGTTACCAGCTGCCTCAAAGTTACCAGCGTATGCCTGACTTCCGCC
GCCGCTTCCTGCAGGTCTGTGTTAATGAGATCAACAGCAGAACTCCAATGCGCCTCTCATACATTGAGAAAAAGAAAGGCCGCCAGACGACTCA
TATCGTATTTTCCTTCCGCGATATCACTTCCATGACGACAGGATAGTCTGAGGGTTATCTGTCACAGATTTGAGGGTGGTTCGTCACATTTGTTCTG
ACCTACTGAGGGTAATTTGTCACAGTTTTGCTGTTTCCTTCAGCCTGCATGGATTTTCTCATACTTTTTGAACTGTAATTTTTAAGGAAGCCAAATT
TGAGGGCAGTTTGTCACAGTTGATTTCCTTCTCTTTCCCTTCGTCATGTGACCTGATATCGGGGGGTTAGTTCGTCATCATTGATGAGGGTTGATTAT
CACAGTTTATTACTCTGAATTGGCTATCCGCGTGTGTACCTCTACCTGGAGTTTTTCCCACGGTGGATATTTCTTCTTGCGCTGAGCGTAAGAGCTA
TCTGACAGAACAGTTCTTCTTTGCTTCCTCGCCAGTTCGCTCGCTATGCTCGGTTACACGGCTGCGGCGAGCATCACGTGCTATAAAAATAATTAT
AATTTAAATTTTTTAATATAAATATATAAATTAAAAATAGAAAGTAAAAAAAGAAATTAAAGAAAAAATAGTTTTTGTTTTCCGAAGATGTAAAAG
ACTCTAGGGGGATCGCCAACAAATACTACCTTTTACCTTGCTCTTCCTGCTCTCAGGTATTAATGCCGAATTGTTTCATCTTGTCTGTGTAGAAGAC
CACACACGAAAATCCTGTGATTTTACATTTTACTTATCGTTAATCGAATGTATATCTATTTAATCTGCTTTTCTTGTCTAATAAATATATATGTAAAGTA
CGCTTTTTGTTGAAATTTTTTAAACCTTTGTTTATTTTTTTTTCTTCATTCCGTAACTCTTCTACCTTCTTTATTTACTTTCTAAAATCCAAATACAAA
ACATAAAAATAAATAAACACAGAGTAAATTCCCAAATTATTCCATCATTAAAAGATACGAGGCGCGTGTAAGTTACAGGCAAGCGATCCTAGTAC
ACTCTATATTTTTTTATGCCTCGGTAATGATTTTCATTTTTTTTTTCCACCTAGCGGATGACTCTTTTTTTTTCTTAGCGATTGGCATTATCACATAATG
AATTATACATTATATAAAGTAATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAGATGACAGAGCAGAAA
GCCCTAGTAAAGCGTATTACAAATGAAACCAAGATTCAGATTGCGATCTCTTTAAAGGGTGGTCCCCTAGCGATAGAGCACTCGATCTTCCCAGA
AAAAGAGGCAGAAGCAGTAGCAGAACAGGCCACACAATCGCAAGTGATTAACGTCCACACAGGTATAGGGTTTCTGGACCATATGATACATGCT
CTGGCCAAGCATTCCGGCTGGTCGCTAATCGTTGAGTGCATTGGTGACTTACACATAGACGACCATCACACCACTGAAGACTGCGGGATTGCTCT
CGGTCAAGCTTTTAAAGAGGCCCTAGGGGCCGTGCGTGGAGTAAAAAGGTTTGGATCAGGATTTGCGCCTTTGGATGAGGCACTTTCCAGAGCG
GTGGTAGATCTTTCGAACAGGCCGTACGCAGTTGTCGAACTTGGTTTGCAAAGGGAGAAAGTAGGAGATCTCTCTTGCGAGATGATCCCGCATT
TTCTTGAAAGCTTTGCAGAGGCTAGCAGAATTACCCTCCACGTTGATTGTCTGCGAGGCAAGAATGATCATCACCGTAGTGAGAGTGCGTTCAA
GGCTCTTGCGGTTGCCATAAGAGAAGCCACCTCGCCCAATGGTACCAACGATGTTCCCTCCACCAAAGGTGTTCTTATGTAGTTTTACACAGGAG
TCTGGACTTGACATACAGTGAATGTAACTTTCGAATTGACAGTATTAGTAGTCGTATTGACAGTGAGGCACGCCCCTCAATGTGCGAGGTGGAAA
ATATACCAGCATGACAATGAATCTTGGAGATTCTTTTGCTGTCATCAAGATTCACCGCCAAATCTTCAGGAACCTATCACGTCCACAGGCGATGTT
AATTCTTGAGTCGTCAAAACAAAGTCCTGTCCTACCTGTAGAAGTTGACAGCGAGCAATTGTATGCAAACTTCTGACTTTGTTATAATAACATTAA
AGGTAATTAAGTATCTTCAATTAGGCATTTTGTCACTGTCAGTCCGTTCCGACAATATAGGTAGATTTGGAATGAATCTTTTCTATGCTGCTGCGAA
TCTTGTACACCTTTGAGGCCGTAGATTCTGTCCGACGAAGCGATAATTATTGCAAAATACATGGACTCATTATTTTGATTCGATTTCTTTTTGGTATC
CGACTCGAAAAGATCCATCACGGCGAGCGCCACCATGAACAAGAACTCCAAAATCCAGTCCCCAAACTCTTCTGATGTTGCTGTTATTGGTGTT
GGTTTTAGATTCCCAGGTAACTCTAATGACCCAGAATCTTTGTGGAACAACTTGTTGGATGGTTTCGATGCTATTACCCAAGTCCCAAAAGAAAG
ATGGGCTACTTCTTTTAGAGAGATGGGTTTGATCAAGAACAAGTTCGGTGGTTTCTTGAAGGATTCTGAATGGAAGAATTTCGACCCTTTGTTCT
TTGGTATCGGTCCAAAAGAAGCTCCATTCATTGATCCACAACAAAGGTTGTTGTTGTCCATCGTTTGGGAATCTTTGGAAGATGCTTACATCAGA
CCAGATGAATTGAGAGGTTCTAACAACTGGTGTTTTCATCGGTGTTTCTAACAACGATTACACCAAGTTGGGTTTCCAAGACAACTACTCTATTTCT
CCATACACTATGACCGGCTCTAACTCTTCATTGAACTCCAACAGAATTTCCTACTGCTTCGATTTTAGAGGTCCATCCATTACTGTTGATACCGCTT
GTTCTTCTTCCTTGGTTTCTGTTAATTTGGGTGTCCAATCCATCCAAATGGGTGAATGTAAGATTGCTATTTGCGGTGGTGTTAACGCTTTGTTTGA
TCCATCTACATCTGTTGCCTTTTCCAAGTTGGGTGTTTTGTCTGAAAATGGCAGATGCAACTCTTTTAGTGATCAAGCCTCTGGTTACGTTAGATCT
GAAGGTGCTGGTGTTGTTGTTTTGAAGTCTTTGGAACAAGCTAAGTTGGATGGTGATAGAATCTACGGTGTTATCAAGGGTGTTTCCTCTAATGA
AGATGGTGCTTCTAATGGTGACAAGAACTCTTTGACTACTCCATCTTGTGAAGCCCAATCCATTAACATTTCTAAGGCTATGGAAAAGGCCTCCTT
GTCTCCATCTGATATCTATTACATTGAAGCCCATGGTACTTCCAGTTGGTGATCCAATTGAAGTTAAGGCCTTGTCCAAGATCTTCTCCAAC
TCTAACAACAACCAGTTGAACAACTTCTCTACCGATGGTAATGATAACGATGATGATGATGACGATAACACCTCTCCAGAACCATTATTGATTGGC
TCATTCAAGTCCAACATCGGTCATTTGGAATCTGCTGCTGGTATTGCTTCTTTGATTAAGTGTTGCTTGATGTTGAAGAACAGGATGTTGGTTCCA
TCCATTAACTGCTCTAATTTGAACCCATCCATTCCATTCGATCAGTACAACATCTCCGTTATCAGAGAAATCAGACAATTCCCAACCGATAAGTTG
GTTAACATCGGTATCAATTCTTTCGGTTTCGGTGGTTCTAACTGCCATTTGATTATTCAAGAGTACAACAACAACTTCAAGAACAACTCTACCATC
TGCAATAACAACAACAACAATAACAACATCGACTACTTGATCCCAATCTCCTCTAAGACTAAGAAGTCCTTGGATAAGTACTTGATTTTGATC
AAGACCAACTCCAACTACCACAAGGATATTTCTTTCGATGACTTCGTCAAGTTCCAAATCAAGTCTAAGCAGTACAACTTGTCCAACAGAATGAC
TACCATTGCTAACGATTGGAACTCCTTCATTAAGGGTTCTAACGAATTCCACAACTTGATCGAATCTAAGGATGGTGAAGGTGGTTCTTCATCTTC
TAACAGAGGTATTGATTCCGCCAATCAAATCAACACTACTACCTCTACCATCAACGATATCGAACCTTTGTTGGTTTTCGTTTTCTGTGGTCAA
GGTCCACAATGGAATGGTATGATTAAGACCTTGTACAACTCCGAGAACGTTTTCAAGAACACCGTTGATCATGTTGACAGCATCTTGTACAAGTA
CTTCGGTTACTCCATTTTGAACGTCTTGTCTAAGATCGATGATAACGACGATTCCATCAACCATCCAATAGTTGCTCAACCATCTTTGTTCTTGTTG
CAAATTGGTTTGGTCGAGTTGTTTAAGTACTGGGGTATCTACCCATCTATCTCTGTTGGTCATTCTTTCGGTGAAGTCTCTTCTTATTACTTGTCCG
GTATCATCTCTTTGGAAACCGCTTGTAAAATCGTCTACGTCAGATCCTCTAATCAGAACAAAACTATGGGTTCCGGTAAGATGTTGGTTGTTTCTA
TGGGTTTTAAGCAATGGAACGATCAATTCTCTGCTGAATGGTCCGATATTGAAATTGCTTGTTACAACGCTCCAGATTCCATAGTTGTTACTGGTA
ACGAAGAAAGATTGAAAGAATTGTCCATCAAGTTGTCCGACGAATCCAATCAAATTTTCAACACCTTCTTGAGGTCCCCATGTTCTTTTCATTCTT
CCCATCAAGAAGTCATCAAGGGTTCTATGTTCGAAGAGTTGTCTAACTTGCAATCTACTGGTGAAACCGAAATCCCTTTGTTCTCTACTGTTACTG
GTAGACAAGTTTTGTCTGGTCATGTTACTGCTCAACACATCTACGATAATGTTAGAGAACCAGTCTTGTTCCAAAAGACGATTGAATCCATTACCT

CCTACATCAAGTCTCACTACCCATCCAATCAAAAGGTTATCTACGTTGAAATTGCTCCACACCCAACCTTGTTTTCATTGATCAAAAAGTCCATCC
CATCCTCCAACAAGAATTCCTCTTCTGTTTTGTGTCCATTGAACAGAAAAGAAAACTCCAACAACTCCTACAAGAAGTTCGTTTCTCAGTTGTAC
TTCAACGGTGTTAACGTTGACTTCAACTTCCAGTTGAACTCCATTTGCGATAACGTTAACAACGATCACCATTTGAACAACGTCAAGCAAAACTC
CTTCAAAGAGACTACCAATTCCTTGCCAAGATACCAATGGGAACAAGATGAATATTGGTCCGAACCATTGATCTCCAGAAAGAATAGATTGGAAG
GTCCAACTACTTCCTTGTTGGGTCATAGAATTATCTACAGCTTCCCAGTTTTCCAATCCGTTTTGGACTTGCAATCTGACAACTACAAATACTTGTT
GGACCACTTGGTTAACGGTAAGCCAGTTTTTCCAGGTGCTGGTTATTTGGATATCATCATCGAATTCTTCGACTACCAAAAGCAGCAGTTGAATTC
CTCTGATTCCTCTAACTCCTACATCATCAACGTTGACAAGATCCAATTCTTGAACCCAATTCACTTGACCGAAAACAAGTTGCAAACCTTGCAAT
CTTCTTTCGAACCTATCGTTACTAAGAAGTCTGCCTTCTCTGTTAACTTCTTCATCAAGGATACCGTCGAGGATCAATCTAAGGTTAAGTCTATGTC
TGACGAAACTTGGACTAACACTTGTAAGGCTACCATTTCCTTGGAACAACAACAGCCATCTCCATCTTCTACTTTGACTTTGTCTAAGAAGCAAG
ACTTGCAGATCTTGAGAAACAGATGCGATATTAGCAAGCTAGACAAGTTTGAGTTGTACGACAAGATCTCTAAGAATTTGGGCTTGCAGTACAAC
TCCTTGTTTCAAGTTGTTGATACCATCGAAACTGGTAAGGATTGCTCTTTTGCTACTTTGTCTTTGCCAGAAGATACTTTGTTCACCACCATTTTGA
ACCCATGCTTGTTGGATAACTGTTTCCATGGTTTGTTGACCTTGATCAACGAAAAGGGTTCTTTCGTTGTCGAGTCCATTTCTTCTGTTTCTATCTA
CTTGGAGAACATCGGTTCCTTCAATCAAACTTCTGTTGGTAACGTCCAGTTCTACTTGTACACCACTATTTCTAAAGCCACCTCCTTTAGTTCTGA
AGGTACTTGTAAGTTGTTCACCAAGGATGGTTCCTTGATTTTGTCTATCGGTAAGTTCATCATCAAGTCCACCAATCCAAAGTCTACTAAGACCAA
CGAAACTATCGAATCTCCATTGGACGAAACCTTCTCTATTGAATGGCAATCTAAGGATTCTCCAATTCCAACCCCACAACAAATCCAACAACAAT
CTCCATTGAACTCTAACCCATCCTTCATTAGATCTACCATCTTGAAGGACATCCAGTTCGAACAATACTGCTCCTCCATTATCCACAAAGAATTGAT
CAACCACGAAAAGTACAAGAACCAGCAATCCTTCGATATCAACTCCTTGGAAAACCACTTGAACGATGACCAATTGATGGAATCCTTGTCCATCT
CCAAAGAATACTTGAGATTCTTCACCAGGATCATCTCCATCATTAAGCAATACCCAAAGATCTTGAACGAAAAAGAGCTAAAAGAATTGAAAGA
AATCATCGAATTGAAGTACCCATCCGAAGTTCAGTTGTTGGAATTCGAAGTTATCGAGAAGGTGTCCATGATTATCCCAAAGTTGTTGTTCGAAA
ACGACAAGCAATCTTCCATGACCTTGTTCCAAGATAACTTGTTGACCAGGTTCTACTCCAATTCTAACTCTACCAGATTCTACTTGGAAAGGGTTT
CCGAAATGGTCTTGGAATCTATTAGACCAATCGTCAGAGAAAAGAGGGTGTTCAGAAATTTGGAAATTGGTGCTGATACAGCCTCTTTGTCTAAT
GTTGTTTTGACTAAGTTGAACACCTACTTGTCCACCTTGAATTCTAATGGTGGTTCTGGTTACAACATCATCATTGAGTACACCTTCACCGATATTT
CCGCCAACTTCATTATTGGTGAAATCCAAGAAACCATGTGCAACTTGTACCCAAACGTTACTTTCAAGTTCTCCGTCTTGGACTTGGAGAAAGAG
ATTATTAACTCCTCCGATTTCTTGATGGGTGATTACGATATAGTTTTGATGGCCTACGTTATCCATGCCGTTTCTAACATTAAGTTCTCCATCGAACA
GTTGTACAAGTTGTTGTCTCCAAGAGGTTGGTTGTTGTGTATTGAACCTAAGTCCAACGTTGTGTTCTCCGATTTGGTTTTCGGTTGTTTTAATCA
GTGGTGGAACTACTACGATGATATTAGAACTACCCACTGCTCCTTGTCTGAATCTCAATGGAATCAGTTGTTGTTGAACCAGTCCTTGAACAACG
AATCCTCTTCTTCTAACTGTTACGGTGGTTTCTCCAACGTTTCTTTTATTGGTGGTGAAAAGGATGTCGACTCCCATTCTTTCATATTGCACTG
CCAAAAAGAATCCATCTCCCAAATGAAGTTAGCCACCACTATTAACAACGGTTTGTCATCTGGTTCCATCGTTATCGTTTTGAACTCTCAACAATT
GACCAACATGAAGTCCTACCCAAAGGTTATTGAGTATATTCAAGAGGCTACCTCTTTGTGCAAGACCATTGAAATTATCGATTCCAAGGACGTCTT
GAACTCTACCAATTCAGTTTTGGAAAAGATCCAAAAGTCCTTGTTGGTGTTCTGTTTGTTGGGTTATGACTTGTTGGAGAACAACTACCAAGAAC
AGTCTTTCGAATACGTTAAGTTGTTGAACTTGATCTCTACTACCGCCTCTTCATCTAATGATAAGAAACCACCAAAGGTCTTGTTGATCACCAAGC
AATCTGAAAGAATCTCCAGGTCTTTCTACTCCAGATCCTTGATTGGTATTTCCAGAACCTCTATGAACGAGTACCCAAATTTGTCCATTACCTCTAT
CGATTTGGATACCAACGACTACTCATTGCAGTCTTTGTTGAAGCCAATCTTCAGCAACTCTAAGTTTTCCGACAACGAGTTCATCTTCAAAAAGG
GCTTGATGTTCGTGTCCAGGATCTTTAAGAACAAGCAGTTGCTAGAATCCTCCAACGCTTTTGAAACTGACTCTTCTAACTTGTACTGTAAGGCC
TCTTCTGACTTGTCTTACAAGTACGCTATTAAGCAGTCTATGTTGACCGAAAATCAGATCGAAATCAAGGTTGAATGCGTCGGTATTAACTTCAAG
GACAACCTATTCTACAAGGGCTTGTTGCCACAAGAAATTTTCAGAATGGGTGACATCTACAATCCACCATATGGTTTGGAATGCTCTGGTGTTATT
ACCAGAATTGGTTCTAACGTCACCGAATACTCAGTTGGTCAAAATGTTTTTGGTTTCGCCAGACATTCTTTGGGTTCTCATGTTGTTACCAACAAG
GATTTGGTTATCTTGAAGCCAGATACCATCTCATTTTCTGAAGCTGCTTCTATCCCAGTTGTTTACTGTACTGCTTGGTACTCCTTGTTCAACATTG
GTCAGTTGTCTAACGAAGAATCCATCCTAATTCATTCTGCTACTGGTGGTGTAGGTTTGGCTTCTTTGAATTTGTTGAAAATGAAGAATCAGCAAC
AGCAACCATTGACCAATGTTTATGCTACTGTTGGCTCTAACGAGAAGAAGAAGGTTCTTGATCGATAACTTCAACAACTTGTTCAAAGAGGACGG
CGAAAACATTTTCTCTACCAGAGACAAAGAATACTCCAACCAGTTGGAATCCAAGATCGATGTTATTTTGAACACCTTGTCCGGTGAATTCGTCG
AATCTAATTTCAAGTCCTTGAGATCCTTCGGTAGATTGATTGATTTGTCTGCTACTCACGTTTACGCCAATCAACAAATTGGTCTAGGTAACTTCAA
GTTCGACCACTTGTATTCTGCTGTTGACTTGGAAAGATTGATCGACGAAAAACCTAAGTTGTTGCAGTCCATCTTGCAAAGAATTACCAACTCTA
TCGTCAACGGTTCCTTGGAAAAAATTCCAATTACCATCTTCCCATCCACCGAAACTAAGGATGCTATCGAATTATTGTCCAAGAGATCCCATATCG
GTAAAGTTGTTGTAGATTGCACCGATATCTCTAAGTGTAATCCTGTTGGTGATGTGATCACCAACTTCTCTATGAGATTGCCAAAGCCAAACTACC
AGTTGAATTTGAACTCCACCTTGTTGATTACTGGTCAGTCTGGTTTGTCTATCCCTTTGTTGAATTGGTTGTTGTCTAAGTCTGGTGGTAACGTTAA
GAACGTTGTCATCATTTCTAAGTCCACCATGAAGTGGAAGTTGCAGACTATGATTTCCCATTTCGTTTCCGGTTTCGGTATCCATTTTAACTACGTT
CAAGTCGACATCTCCAACTACGATGCTTTGTCTGAAGCTATTAAGCAATTGCCATCTGATTTGCCACCAATCACCTCTGTTTTCATTTGGCTGCTA
TCTACAACGATGTTCCAATGGATCAAGTTACCATGTCTACCGTTGAATCTGTTCATAACCCTAAAGTTTTGGGTGCCGTTAACTTGCATAGAATCTC
TGTTTCTTTTGGTTGGAAGTTGAACCACTTCGTCTTGTTCTCTTCTATTACTGCTATTACCGGTTACCCAGACCAATCTATCTACAATTCTGCCAAC
TCTATTTTGGACGCTTTGTCCAACTTTAGAAGGTTTATGGGTTTGCCATCCTTCTCCATTAACTTGGGTCCAATTAAGGATGAAGGTAAGGTTTCTA
CCAACAAGAGCATCAAGAAGCTATTCAAGTCTAGAGGTTTGCCAAGCCTATCCTTGAACAAGTTATTTGGTTTGTTGGAGGTCGTCATCAACAAC
CCATCTAATCATGTTATCCCATCCCAATTGATTTGCTCCCCAATCGATTTCAAGACCTACATCGAATCTTTCTCAACTATGAGGCCAAAGTTGTTAC
ACTTGCAACCTACCATTTCCAAGCAGCAATCTTCTATCATTAACGATTCTACCAAGGCTTCCTCCAACATTTCATTGCAAGATAAGATCACCTCCA
AGGTGTCTGATTTGTTGTCCATTCCAATCTCCAAGATCAACTTCGATCATCCATTGAAACACTACGGCTTGGATTCTTTGTTGACCGTTCAATTCA
AATCCTGGATCGACAAAGAATTCGAAAAGAACTTGTTCACCCATATCCAATTGGCCACCATCTCTATTAACTCATTCTTGGAAAAGGTGAACGGC
TTGTCTACAAACAATAACAACAACAACAATTCCAACGTCAAGTCCTCTCCATCCATTGTCAAAGAAGAAATCGTTACCTTGGACAAGGATCAAC
AACCATTGCTATTGAAAGAACACCAGCACATTATCATCTCCCCAGATATTAGAATCAACAAGCCAAAGAGGGAATCCTTGATTAGAACCCCAATC
TTGAACAAATTCAACCAGATCACCGAATCCATTATCACTCCATCTACACCATCTTTGTCCCAATCCGATGTTTTGAAAACTCCACCAATCAAGTCT
TTGAACAACACTAAGAACTCCAGCTTGATTAACACCCCACCAATTCAATCTGTCCAACAACATCAAAAGCAACAACAAAAGGTCCAAGTCATCC
AACAACAGCAACAACCATTATCCAGATTGTCCTACAAGAGCAACAACAACTCTTTCGTTTTGGGTATCGGTATTTCTGTTCCAGGTGAACCTATTT
CCCAACAATCCTTGAAAGACTCATCTCCAATGACTTTTCTGATAAGGCTGAAACTAACGAGAAGGTCAAGAGAATCTTTGAGCAATCTCAAATC
AAGACCAGACACTTGGTTAGAGATTACACTAAGCCAGAGAACTCCATCAAGTTCAGACATTTGGAAACCATTACCGATGTGAACAACCAGTTCA
AGAAAGTTGTTCCAGATTTGGCTCAACAAGCCTGTTTGAGAGCTTTGAAAGATTGGGGTGGTGATAAGGGTGATATTACCCATATAGTTTCTGTTA
CCTCCACCGGTATTATCATCCCAGATGTTAATTTCAAGTTGATCGACTTGTTGGGCTTGAACAAGGATGTTGAAAGAGTGTCTTTGAACCTAATGG
GTTGTTTGGCTGGTTTGAGTTCTTTGAGAACTGCTGCTTCTTTGGCTAAGGCTTCTCCAAGAAATAGAATTTTGGTTGTCTGTACCGAAGTCTGCT
CCTTGCATTTTTCTAATACTGATGGTGGTGATCAAATGGTCGCCTCTTCTATTTTTGCTGATGGTTCTGCTGCTTACATTATTGGTTGTAACCCAAGA
ATTGAAGAAACCCATTATACGAAGTCATGTGCTCCATTAACAGATCTTTCCCAAATACCGAAAACGCCATGGTTTGGGATTTGGAAAAAGAAGG
TTGGAACTTGGGTTTGGATGCTTCTATTCCAATTGTCATTGGTTCTGGTATTGAAGCCTTCGTTGATACTTTGTTGGATAAGGCTAAGTTGCAAACT
TCCACTGCTATTTCTGCTAAGGATTGCGAATTCTTGATTCATACTGGTGGCAAGTCCATCTTGATGAACATCGAAAATTCCTTGGGTATCGACCCA
AAGCAAACTAAGAATACTTGGGATGTTTACCATGCCTACGGCAATATGTCATCTGCCTCTGTTATTTTCGTTATGGATCATGCCAGAAAGTCCAAG
TCTTTGCCAACTTACTCAATTTCTTTGGCTTTTGGTCCAGGTTTGGCTTTTGAAGGTTGTTTCTTGAAGAACGTCGTCGAACAGAAGCTCATTTCC
GAAGAGGACTTGTAACCGCAACAACTACCTCGACTTTGGCTGGGACACTTTCAGTGAGGACAAGAAGCTTCAGAAGCGTGCTATCGAACTCAA
CCAGGGACGTGCGGCACAAATGGGCATCCTTGCTCTCATGGTGCACGAACAGTTGGGAGTCTCTATCCTTCCTTAAAAATTTAATTTTCATTAGTT
GCAGTCACTCCGCTTTGGTTTCGCTAGTGATAATAAGTGACTGAGGTATGTGCTCTTCTTATCTCCTTTTGTAGTGTTGCTCTTATTTTAAACAACT
TTGCGGTTTTTTGATGACTTTGCGATTTTGTTGTTGCTTTGCAGTAAATTGCAAGATTTAATAAAAAAACGCAAAGCAATGATTAAAGGATGTTCA
GAATGAAACTCATGGAAACACTTAACCAGTGCATAAACGCTGGTCATGAAATGACGAAGGCTATCGCCATTGCACAGTTTAATGATGACAGCCC
GGAAGCGAGGAAAATAACCCGGCGCTGGAGAATAGGTGAAGCAGCGGATTTAGTTGGGGTTTCTTCTCAGGCTATCAGAGATGCCGAGAAAGC
AGGGCGACTACCGCACCCGGATATGGAAATTCGAGGACGGGTTGAGCAACGTGTTGGTTATACAATTGAACAAATTAATCATATGCGTGATGTGT
TTGGTACGCGATTGCGACGTGCTGAAGACGTATTTCCACCGGTGATCGGGGTTGCTGCCCATAAAGGTGGCGTTTACAAAACCTCAGTTTCTGTT
CATCTTGCTCAGGATCTGGCTCTGAAGGGGCTACGTGTTTTGCTCGTGGAAGGTAACGACCCCCAGGGAACAGCCTCAATGTATCACGGATGGG
TACCAGATCTTCATATTCATGCAGAAGACACTCTCCTGCCTTTCTATCTTGGGGAAAAGGACGATGTCACTTATGCAATAAAGCCCACTTGCTGGC
CGGGGCTTGACATTATTCCTTCCTGTCTGGCTCTGCACCGTATTGAAACTGAGTTAATGGGCAAATTTGATGAAGGTAAACTGCCCACCGATCCA
CACCTGATGCTCCGACTGGCCATTGAAACTGTTGCTCATGACTATGATGTCATAGTTATTGACAGCGCGCCTAACCTGGGTATCGGCACGATTAAT
GTCGTATGTGCTGCTGATGTTCCCACGCCTGCTGAGTTGTTTGACTACACCTCCGCACTGCAGTTTTTCGATATGCTTCGTGATCTG
CTCAAGAACGTTGATCTTAAAGGGTTCGAGCCTGATGTACGTATTTTGCTTACCAAATACAGCAATAGCAATGGCTCTCAGTCCCCGTGGATGGA
GGAGCAAATTCGGGATGCCTGGGGAAGCATGGTTCTAAAAAATGTTGTACGTGAAACGGATGAAGTTGGTAAAGGTCAGATCCGGATGAGAAC
TGTTTTTGAACAGGGCCATTGATCAACGCTCTTCAACTGGTGCCTGGAGAAATGCTCTTTCTATTTGGGAACCTGTCTGCAATGAAATTTTCGATCG
TCTGATTAAACCACGCTGGGAGATTAGATAATGAAGCGTGCGCCTGTTATTCCAAAACATACGCTCAATACTCAACCGGTTGAAGATACTTCGTTA
TCGACACCAGCTGCCCCGATGGTGGATTCGTTAATTGCGCGCGTAGGAGTAATGGCTCGCGGTAATGCCATTACTTTGCCTGTATGTGGTCGGGAT
GTGAAGTTTACTCTTGAAGTGCTCCGGGGTGATAGTGTTGAGAAGACCTCTCGGGTATGGTCAGGTAATGAACGTGACCAGGAGCTGCTTACTG
AGGACGCACTGGATGATCTCATCCCTTCTTTTCTACTGACTGGTCAACAGACACCGGCGTTCGGTCGAAGAGTATCTGGTGTCATAGAAATTGCC
GATGGGAGTCGCCGTCGTAAAGCTGCTGCACTTACCGAAAGTGATTATCGTGTTCTGGTTGGCGAGCTGGATGATGAGCAGATGGCTGCATTATC
CAGATTGGGTAACGATTATCGCCCAACAAGTGCTTATGAACGTGGTCAGCGTTATGCAAGCCGATTGCAGAATGAATTTGCTGGAAATATTTCTGC
GCTGGCTGATGCGGAAAATATTTCACGTAAGATTATTACCCGCTGTATCAACACCGCCAAATTGCCTAAATCAGTTGTTGCTCTTTTTTCTCACCCC
GGTGAACTATCTGCCCGGTCAGGTGATGCACTTCAAAAAGCCTTTACAGATAAAGAGGAATTACTTAAGCAGCAGGCATCTAACCTTCATGAGCA
GAAAAAAGCTGGGGTGATATTTGAAGCTCTTAACTTCTGTGCTTAAAACGTCATCTGCATCAAGAACTAGTTTAAGCT
CACGACATCAGTTTGCTCCTGGAGCGACAGTATTGTATAAGGGCGATAAAATGGTGCTTAACCTGGACAGGTCTCGTGTTCCAACTGAGTGTATA
GAGAAAATTGAGGCCATTCTTAAGGAACTTGAAAAGCCAGCACCCTGATGCGACCTCGTTTTAGTCTACGTTTATCTGTCTTTACTTAATGTCCTT

TGTTACAGGCCAGAAAGCATAACTGGCCTGAATATTCTCTCTGGGCCCACTGTTCCACTTGTATCGTCGGTCTGATAATCAGACTGGGACCACGG
TCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCACGGTCCCACTCGT
ATCGTCGGTCTGATAATCAGACTGGGACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCATGGTCCCACTCGTATCGTCGGTC
TGATTATTAGTCTGGGACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGAACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGT
CTGGGACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCACGATCCCACTCGTGTTGTCGGTCTGATTATCGGTCTGGGACCAC
GGTCCCACTTGTATTGTCGATCAGACTATCAGCGTGAGACTACGATTCCATCAATGCCTGTCAAGGGCAAGTATTGACATGTCGTCGTAACCTGTA
GAACGGAGTAACCTCGGTGTGCGGTTGTATGCCTGCTGTGGATTGCTGCTGTGTCCTGCTTATCCACAACATTTTGCGCACGGTTATGTGGACAA
AATACCTGGTTACCCAGGCCGTGCCGGCACGTTAACCGGGCTGCATCCGATGCAAGTGTGTCGCTGTCGACGAGCTCGCGAGCTCGGACATGAG
GTTGCCCCGTATTCAGTGTCGCTGATTTGTATTGTCTGAAGTTGTTTTTACGTTAAGTTGATGCAGATCAATTAATACGATACCTGCGTCATAATTG
ATTATTTGACGTGGTTTGATGGCCTCCACGCACGTTGTGATATGTAGATGATAATCATTATCACTTTACGGGTCCTTTCCGGTGATCCGACAGGTTA
CGGGGCGGCGACCTCGCGGGTTTTCGCTATTTATGAAAATTTTCCGGTTTAAGGCGTTTCCGTTCTTCTTCGTCATAACTTAATGTTTTTATTTAAA
ATACCCTCTGAAAAGAAAGGAAACGACAGGTGCTGAAAGCGAGCTTTTTGGCCTCTGTCGTTTCCTTTCTCTGTTTTTGTCCGTGGAATGAACA
ATGGAAGTCCGAGCTCATCGCTAATAACTTCGTATAGCATACATTATACGAAGTTATA

ANNEX C: Collaboration and contribution

Functional *Cannabis sativa* cannabidiolic acid synthase in *Phaedodactylum tricornutum*

Elisa Fantino[1,2†], Anis Messaabi[1†], Natacha Mérindol[1], Fatima Awwad[1], Nicolas Sene[1], Sarah-Eve Gélinas[1], Alexandre Custeau[1], Kimy-Li Rhéaume[1], Fatma Meddeb-Mouelhi[1,2], Isabel Desgagné-Penix[1,2*]

[1]Department of Chemistry, Biochemistry and Physics, Université du Québec à Trois-Rivières, Trois-Rivières, QC, Canada.

[2]Plant Biology Research Group, Université du Québec à Trois-Rivières, Trois-Rivières, QC, Canada.

[†]These authors contributed equally to the work.

**Correspondance:* Isabel.Desgagne-Penix@uqtr.ca

This article is under revision

Abstract

Cannabis sativa's cannabidiolic acid (CBDA) offers significant therapeutic potential without inducing psychotropic effects but is typically found as part of a complex mixture of metabolites in plant extracts. Using a heterologous expression platform could allow the production of pure CBDA. Here, we propose to express CBDA synthase (CBDAS) in *Phaeodactylum tricornutum.* Episomes carrying *CBDAS* variants, incorporating the native signal peptide (CBDAS) or the highly abundant secreted protein 1 secretory signal peptide (SP:CBDAS) were constructed. *CBDAS* variants were tagged with the yellow fluorescent protein (YFP), introduced into the marine diatom, and screened by fluorescence. Confocal microscopy revealed that CBDAS and SP:CBDAS arranged in aggregated structures indicative of secretory pathway involvement. Western blot assays confirmed whole construct accumulation intracellularly, while soluble YFP was detected extracellularly. Finally, enzymatic assays showed CBDA production by both CBDAS and SP:CBDAS strains, confirming the potential of *P. tricornutum* as a platform for cannabinoid biosynthesis.

Keywords

Microalgae; cannabinoids; biofactories; cannabidiolic acid synthase; marine platform; episome; secretory pathway.

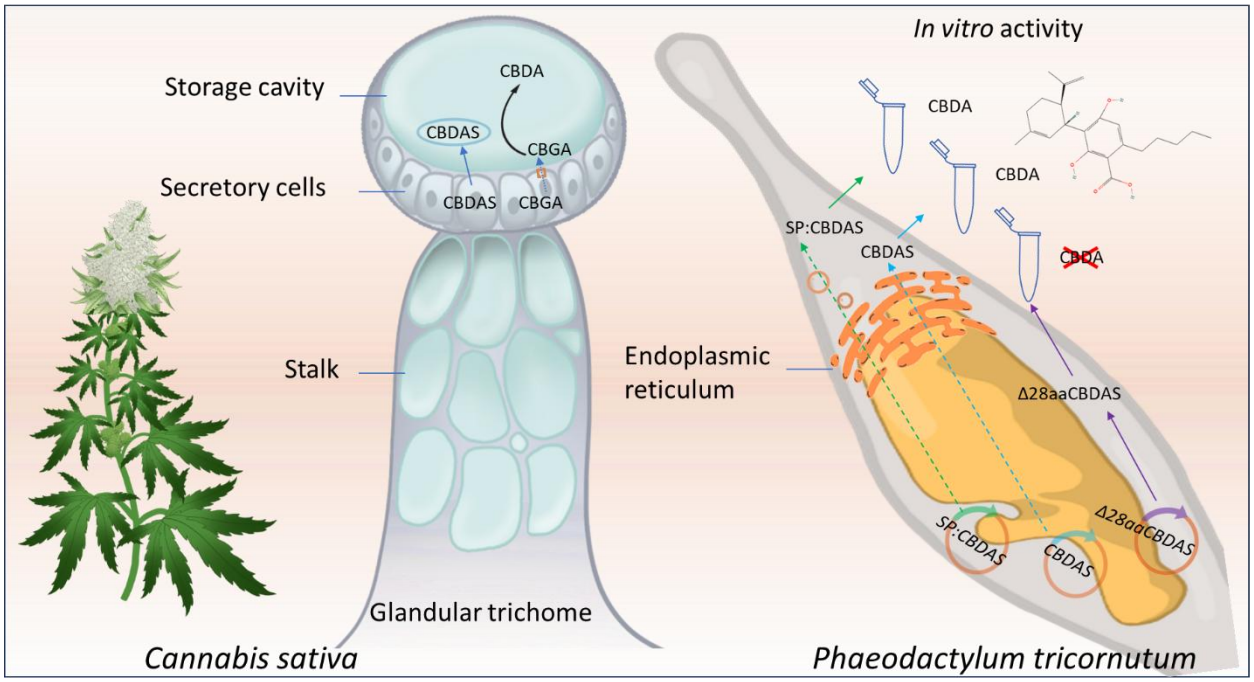
Abbreviations

Au, Absorbance units; CBD, Cannabidiol; CBDA, Cannabidiolic acid; CBDAS, Cannabidiolic acid synthase; CBGA, Cannabigerolic acid; EV, Empty Vector; ER, Endoplasmic Reticulum; FAD, Flavin Adenin Dinucleotide; HA, Hemagglutinin, HASP1, highly abundant secreted protein; OD, Optical density; SP, Signal peptide; THC, Δ9-tetrahydrocannabinol; THCA, Δ-9-tetrahydrocannabinolic acid; Thr, Thrombine; YFP, yellow fluorescent protein.

Highlights

- Microalgae production of cannabidiolic acid (CBDA), a bioactive cannabinoid (CB)
- Promising photosynthetic platform for sustainable biomanufacturing pharmaceuticals
- *Cannabis* CBDA synthase was successfully heterologously expressed in *Phaeodactylum tricornutum*
- Fluorescent tagging and microscopy used to track CBDA synthase localization
- Secretory signal peptides enhance CBDA synthase targeting and secretion pathways

Graphical abstract



Introduction

Cannabis sativa-based preparations are prescribed in the USA and Canada to treat nausea following cancer chemotherapy, appetite loss due to acquired immune deficiency syndrome, and to provide symptomatic relief from neuropathic pain in multiple sclerosis patients (Castillo-Arellano et al., 2023). Cannabinoids (CBs) are a group of metabolites long coveted for their psychoactive and medicinal properties (Stella, 2023, Wang et al., 2020a, Crocq, 2020b, Walsh et al., 2021a, Quinones et al., 2022). Δ^9-Tetrahydrocannabinol (THC) and cannabidiol (CBD), as well as their carboxylated forms Δ^9-tetrahydrocannabinolic acid (THCA) and cannabidiolic acid (CBDA), are the potent CBs associated with the therapeutic benefits of *C. sativa* (*Cs*). CBD, in particular, has been reported to reduce sustained anxiety and improve sleep while being devoid of psychotropic effect (Cristino et al., 2020a, Devinsky et al., 2016, Devinsky et al., 2017). *In planta*, cannabinoid and terpene biosynthesis pathways occur in secretory cells of the glandular trichomes of *C. sativa*. Both pathways utilize precursors such as geranyl pyrophosphate from the methylerythritol phosphate pathway and intermediate such as acetyl-CoA from fatty acid metabolisms (Desaulniers Brousseau et al., 2021). THCA and CBDA synthesis is catalyzed in the glandular trichome cavity by the THCA synthase (*Cs*THCAS) and the CBDA synthase (*Cs*CBDAS), respectively. Both secreted enzymes are members of the berberine bridge enzyme-like gene family, which contains a 28 amino acids (aa) N-terminal signal peptide and a flavin adenine dinucleotide (FAD) binding domain (Taura et al., 1996, Sirikantaramas et al., 2004a, Lim et al., 2021). These FAD-dependent oxygenases catalyze the oxidative cyclization of cannabigerolic acid (CBGA), the last step of the major CB pathway. The structural and functional properties of *Cs*CBDAS are quite similar to those of *Cs*THCAS with 84% identity in the amino acid sequences (Taura et al., 1996, Andre et al., 2016). Both possess a disulfide bond and several N-glycosylation sites (Sirikantaramas et al., 2004a). The main difference between the reactions they catalyze is the proton transfer step, *i.e. Cs*CBDAS extracts a proton from the methyl group at the end of CBGA to yield CBD, while the *Cs*THCAS enzyme extracts a hydroxyl group to produce THC (Andre et al., 2016, Dai, 2024).

The isolation of pure CBD, desirable for its specific pharmaceutical properties, from *C. sativa* is challenging, as the extraction results in formulations containing hundreds of plant chemicals and minor cannabinoids (de A. Leite et al., 2018, Ohtsuki et al., 2022). Although there has been progress, its chemical synthesis has not been proven to be a cost-effective method either, mainly due to the complexity of its structure, resulting in high production and environmental costs, and suboptimal yields (Chiurchi 2021). For these reasons, there is a need for alternative production platforms that are controlled and sustainable. To this aim, CBs pathway enzymes have been introduced in heterologous prokaryotic and eukaryotic hosts such as *E. coli* (Kearsey et al., 2019), yeast (Shoyama et al., 2012, Zirpel et al., 2015, Zirpel et al., 2017b, Zirpel et al., 2018a, Luo et al., 2019b), model plants like *Nicotiana benthamiana* (Gülck et al., 2020), filamentous fungi like *Penicillium chrysogenum* (Kosalková, 2023), and marine diatoms such as *Phaeodactylum tricornutum* (Awwad et al., 2023b, Fantino, 2024). These studies achieved variable yields of CBs and precursors, reaching up to hundreds of mg/L (Blatt-janmaat and Qu, 2021b), and revealed that CBs are more suitable to be produced in eukaryotic compared to prokaryotic organisms. For instance, expression in *E. coli* cytosol failed to yield an active *Cs*THCAS (Zirpel et al., 2015). The coding sequence of *CsTHCAS* fused to different signal peptides and tagged with the yellow fluorescent protein (YFP) was transiently heterologously expressed in *N. benthamiana* (Gülck et al., 2020, Sirikantaramas et al., 2005, Geissler et al., 2018). Interestingly, *Cs*THCAS enzyme was detected only when targeted to the endoplasmic

reticulum (ER); while cytosolic and plastid localization resulted in no detectable protein (Geissler et al., 2018). Recent studies suggested that glycosylation and other post-translational modifications of *Cs*THCAS occurring in the ER might contribute to the correct folding of the enzyme and influence its stability (Zirpel et al., 2018a, Geissler et al., 2018).

Directing *Cs*THCAS or *Cs*CBDAS to specific intracellular compartments, such as the vacuoles, was also key to their function and CBs production in yeast (Zirpel et al., 2017b, Zirpel et al., 2018a, Luo et al., 2019b, Schmidt, 2024). These studies indicated that the expression and translocation of *Cs*THCAS and *Cs*CBDAS to specific intracellular compartments might result in the stabilization of the enzyme and increase their activity, possibly involving the pH-dependency of the reaction. Moreover, the translocation of the enzymes into specific compartments might shield the cells from the accumulation of CBs in the cytosol which was shown to be toxic to *C. sativa* protoplasts and *P. tricornutum* cell cultures (Sirikantaramas et al., 2005, Fantino, 2024). Thus, an efficient *Cs*CBDAS heterologous platform would ideally allow for eukaryotic post-translational modifications through compartmentalization or secretion of the enzyme, or its product, and contain the necessary pathway precursors.

Microalgae are unicellular photosynthetic organisms emerging as heterologous hosts to produce plant specialized metabolites (Hu, 2023, Einhaus et al., 2024). Diatoms are especially resistant to diverse environments and play an important role in fixing carbon from the ocean (Sethi et al., 2020). *P. tricornutum*, in particular, is known for its rapid growth rate and richness in fatty acids (Ova Ozcan and Ovez, 2020), and has been successfully used as a platform to produce lipids, terpenoids, and geraniol (Cui, 2021, Jaramillo-Madrid et al., 2020, Fabris et al., 2020a, George et al., 2020a). Recently, we demonstrated that diatom cell cultures could support *in vivo* production of the cannabinoid precursors olivetolic acid and CBGA following heterologous expression of tetraketide synthase and olivetolic acid cyclase (Awwad et al., 2023b, Fantino, 2024). Based on these characteristics, we hypothesize that *P. tricornutum* could be modified to efficiently biosynthesize CBDA.

This study aims to address these challenges associated with functional CBDAS heterologous expression and explore the suitability of producing CBDA in a sustainable and efficient manner. We engineered diatom strains with episomes containing three versions of the *CsCBDAS* sequence driven by the *P. tricornutum* endogenous promoter from the highly abundant secreted protein 1 (HASP1) (Erdene-ochir et al., 2019a, Slattery et al., 2022). The cassettes included 1) the native *Cs*CBDAS sequence, 2) a truncated sequence lacking the 28 aa N-terminal signal peptide (SP), and 3) a hybrid sequence, replacing the native SP with HASP1 18 aa secretory SP to direct *Cs*CBDAS to the extracellular matrix, mimicking enzymes behavior *in planta*. All sequences were codon-optimized for expression in the diatom and fused to YFP at the C-terminal, to easily follow protein accumulation and subcellular localization. After transformation, *P. tricornutum* strains harboring *CsCBDAS* complete sequence (CBDAS), truncated (Δ28aaCBDAS), and secreted version (SP:CBDAS*)* were characterized by flow cytometry, immunoblotting, confocal microscopy, and enzymatic assay. SP:CBDAS transconjugants yielded more clones with higher fluorescence compared to the other two designs. Both CBDAS and SP:CBDAS formed aggregate-like structures between the two plastid lobes, suggesting secretory pathway localization. Enzymatic assays confirmed that enzymes were active in both contexts, yielding CBDA as a product, whereas constructs without signal peptide were not active. The successful heterologous expression of *CsCBDAS* in *P. tricornutum* represents an important step in the development of microalgal-based biofactories

for cannabinoid production. Moreover, our results reinforce the potential of the marine diatom, as a platform for efficient production of complex metabolites with pharmaceutical applications.

Results and Discussion

Engineering P. tricornutum strains to express CsCBDAS

To metabolically engineer the brown algae *P. tricornutum* to biosynthetize CBDA, we designed three episomes harboring different versions of the gene encoding *Cs*CBDAS (Fig. 1A). Gene expression was driven from 499-bp of the *HASP1* promoter (HASP1p); followed by 1) a complete codon optimized *CsCBDAS* gene sequence (CBDAS), 2) a truncated version (Δ28aaCBDAS) without the secretory signal peptide (SP, 84 bp), or 3) the truncated version with an additional 54 bp from the 5' coding sequence of *HASP1* (SP:CBDAS). Enzyme sequences were tagged with *YFP* and 3x hemagglutinin (HA) at their C terminal, linked by a thrombin cleavage sequence (LVPRGS) (Jenny et al., 2003). Three additional episomes were included as controls; the empty vector (EV) containing only the nourseothricin resistance cassette with the *nat* gene, a plasmid with the *YFP* reporter gene (YFP) under HASP1p, and one with *YFP* fused with the HASP1 SP (SP:YFP) at its N-terminal under HASP1p (Fig. 1A). These five constructs were successfully transformed into *P. tricornutum* by *E. coli* conjugation, and characterized for YFP signal, protein detection, and enzymatic assay (Fig. 1B).

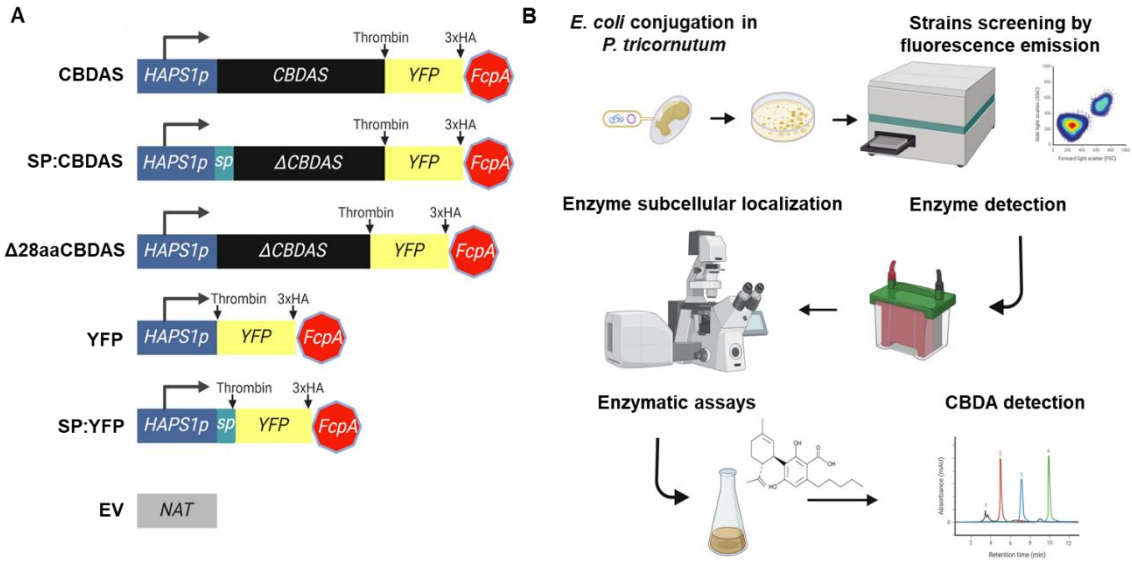


Fig.1. Expression system and platform development. A. Scheme of the recombinant cassettes expressing cannabidiolic acid synthase (*Cs*CBDAS) driven by the diatom promoter *HASP1p* and tagged with the *yellow fluorescent protein* (YFP) reporter gene. Expression cassettes included the complete codon-optimized native sequence of *Cs*CBDAS (CBDAS), a truncated version (Δ28aaCBDAS), and a truncated version plus 54 bp from the 5' coding sequence of *HASP1* (SP:CBDAS). All sequences were tagged in the C terminal with *YFP* and *3xhemagglutinin* (HA), linked to the enzyme by a thrombin cleavage sequence (LVPRGS). Moreover, three controls were used, *i.e. YFP* gene, SP:YFP both driven by HASP1p, and Empty vector (EV) encoding only the *nat* gene conferring nourseothricin resistance. B. Study workflow. *P. tricornutum* cells were transformed by *E. coli* conjugation, then clones were screened and analyzed by fluorescence emission in a microplate reader and flow cytometer, enzyme size was verified by immunoblotting assay and subcellular localization was observed in the confocal microscope. Enzyme activity was confirmed by the detection of CBDA on an HPLC-DAD. The figure was created with BioRender.com (BioRender, 2023).

Two weeks post-transformation, 36 colonies of each strain were randomly selected and grown in nourseothricin-containing plates for 10 more days, for a second antibiotic selection round. At that stage, colonies were screened by epifluorescence microscopy. The frequency of YFP^+ colonies, *i.e.* with higher fluorescence compared to the EV strain, was computed for each line (Table 1).

Table 1. Transconjugant strains screening by yellow fluorescent protein (YFP) fluorescence. Thirty-six randomly selected colonies of each strain, *Cs*CBDAS:YFP (CBDAS), *Cs*CBDAS fused to the HASP1 signal peptide (SP:CBDAS), and a truncated *Cs*CBDAS version without the first 28 aa (Δ28aaCBDAS) were screened under a fluorescence microscope. YFP^+ clones selected from the first screen were grown in L1 liquid media for 10 days in 96-well plates, and YFP fluorescence was analyzed with a microplate reader (second screen). Empty vector (EV) strain was used as a negative control to settle the basal autofluorescence.

Strain	1st screening YFP^+ clones (Fluorescence microscopy)	2nd screening YFP^+ clones $(Ex_{500}Em_{539})/OD_{680}$ nm
CBDAS	30.6% (n=11/36)	72.7% (n=8/11)
SP:CBDAS	50% (n=18/36)	100% (n=18/18)
Δ28aaCBDAS	19.4% (n=7/36)	100 % (n=7/7)

SP:CBDAS strain presented a higher percentage of YFP^+ colonies (50%) compared to CBDAS (30.55%) and Δ28aaCBDAS (19.44%) strains (Table 1). Strains that presented fluorescence were grown in liquid culture for 10 days and further analyzed in a second screening round by recording YFP emission normalized on culture OD_{680} nm in a microplate reader, using EV strain as negative control (Fig. 2A, B; Fig. S1). Fluorescence was detected in YFP, SP:YFP, CBDAS, SP:CBDAS and Δ28aaCBDAS transconjugants, and SP:CBDAS clones displayed more fluorescence compared to CBDAS (One way Anova, Tukey post-test $p<0.05$, Fig. 2A). A threshold equivalent to the mean background fluorescence detected in the EV clones, plus twice the standard deviation of that mean, was used to determine positive clones (Table 1, Fig. 2B). All three YFP and SP:YFP clones, used as positive controls, were fluorescent (Fig. 2A, B). Eight of the 11 clones of the CBDAS strain (72.7%) remained fluorescent (Table 1). All tested clones from SP:CBDAS and Δ28aaCBDAS displayed YFP fluorescence above the threshold (Fig. 2B).

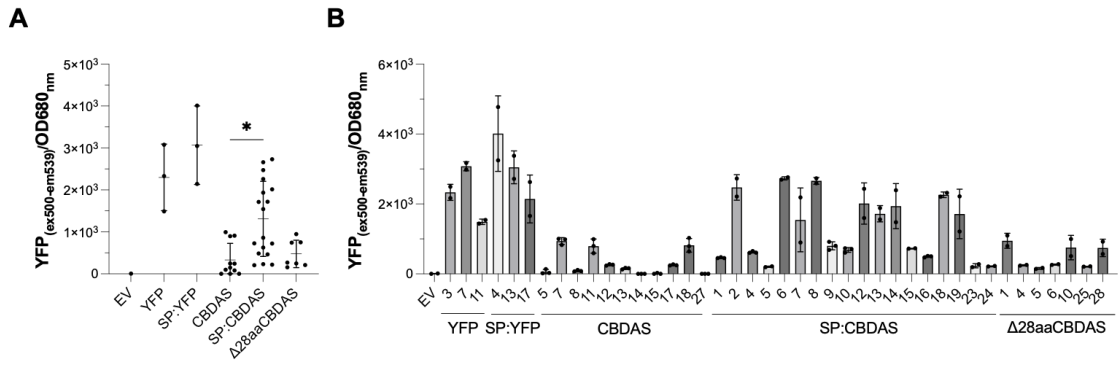


Fig 2. CBDAS strains screening for YFP fluorescence. The strains selected by the fluorescence microscope were transferred to liquid media where the YFP fluorescence intensity was measured by the microplate reader, normalized to the OD680 nm, and compared with EV strain autofluorescence (A). Mean ± SD; n=3 biologically independent samples for each strain were plotted. The clones with a YFP fluorescence higher than the EV fluorescence (mean + (SD) x 2) were determined as YFP positive (B). A One way Anova with Tukey post-test was performed (*p<0.05).

Thus, our results indicated that the constructs were successfully transformed and heterologously expressed by *P. tricornutum*, and that SP:CBDAS transconjugants presented a higher YFP fluorescence compared to CBDAS constructs. To some extent, these results are in contrast with Erdene-Ochir observations, which suggested that the addition of the SP signal to GFP decreased transcript levels compared to GFP alone constructs (Erdene-ochir et al., 2019a). The authors suggested that the presence of the SP sequence could affect the steady-state level of GFP mRNA. In our study, the addition of the SP signal had a positive impact at the fluorescence level, but we used different constructs, and episomal expression rather than randomly integrated genomic expression, limiting our ability to compare with Erdene-Ochir *et al.* (Erdene-ochir et al., 2019a).

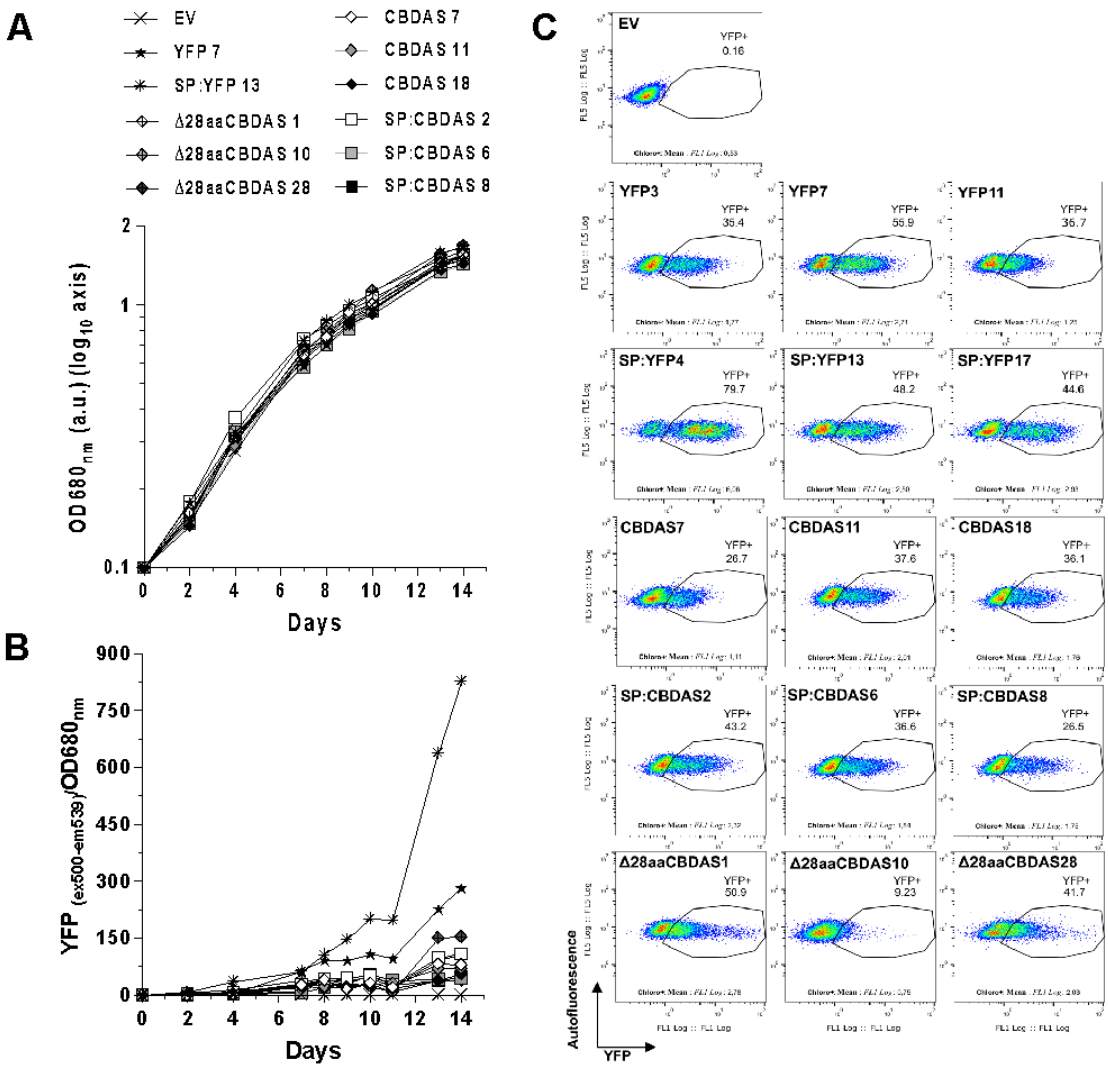


Fig. 3. Characterization of the selected transconjugants. A. Growth curves of each strain, the optical density (OD) at 680 nm was followed for 14 days. B. YFP fluorescence emission (excitation 500 nm-emission 539 nm) was measured and normalized to the OD for the 12 cultures monitored in A. C. Representative dot plots of 3 selected clones of each strain. Pseudo-color dot plots of empty vector (EV) and transformants strains; YFP fluorescence was detected at 530 nm on the x-axis, and autofluorescence at 448 nm on the y-axis. Gates and frequencies of total YFP populations were designed according to the autofluorescence of the negative control EV shown as a reference.

Characterization of the selected fluorescent clones

Three clones with the highest fluorescence for each strain were selected for further characterization. The culture growth curves and YFP fluorescence were registered for 14 days and compared to the EV strain (Fig. 3 A and B). Strains presented a similar pattern, showing that the presence of the transgene did not negatively affect the growth of *P. tricornutum* cells. The YFP fluorescence increased from day 7, as shown before (Erdene-ochir et al., 2019a, Slattery et al., 2022)[35, 36], with clone SP:YFP 13 showing the highest level at day 14.

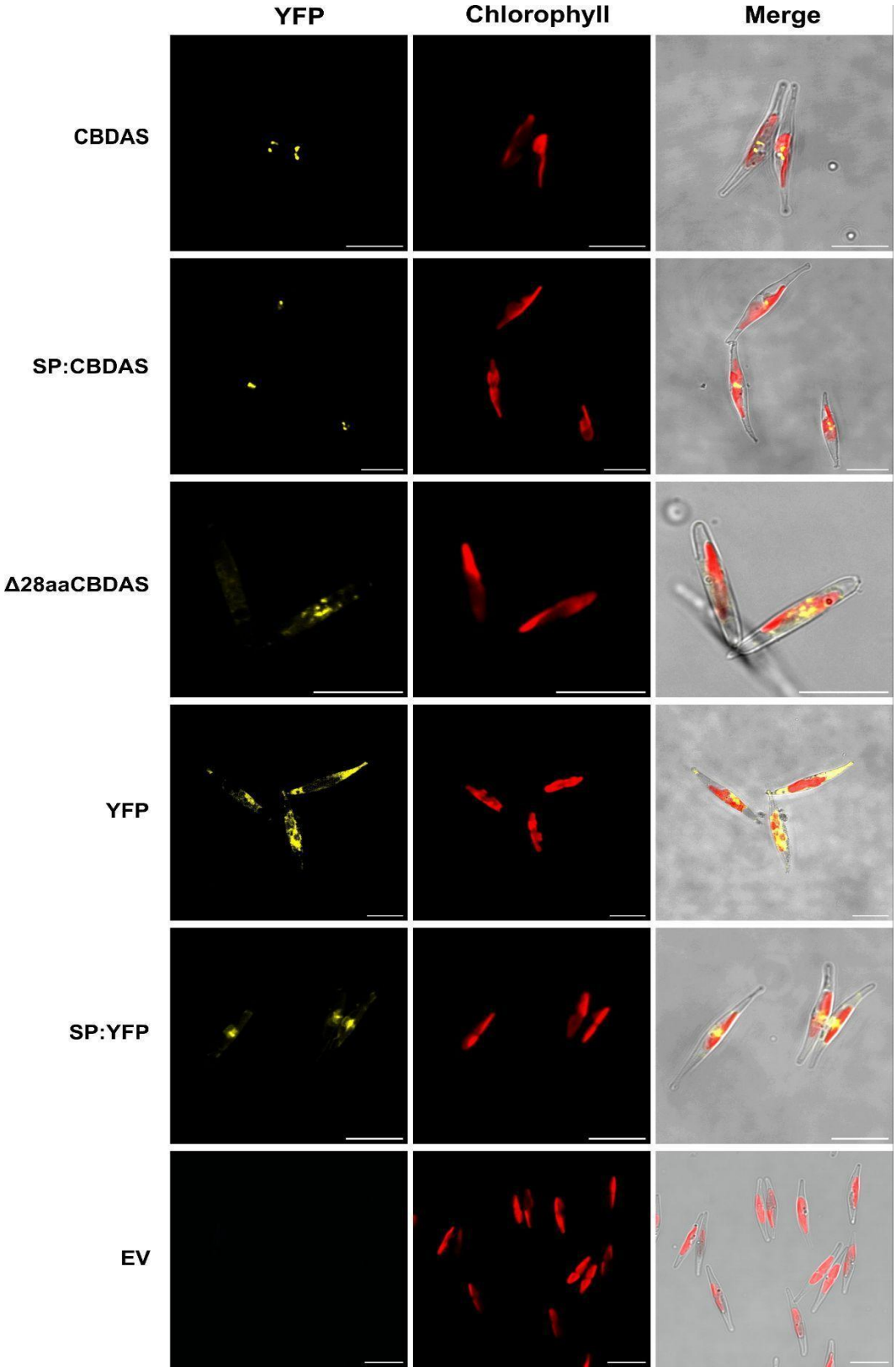
Clones were then analyzed on a flow cytometer at day 10 (Fig. 3C). The EV strain was used as a negative control. YFP clones 3, 7, and 11 presented 35.4 %, 55.9 %, and 36.7 % of YFP^+ cells, respectively, while SP:YFP clones 4, 13, and 17 showed 79.7 %, 48.2%, and 44.6 % of YFP^+ cells. The selected CBDAS clones (7, 11, and 18) and SP:CBDAS (2, 6, and 8) showed similar frequency of YFP^+ cells; 26.7 %, 37.6 % and 36.1 % *vs.* 43.2 %, 36.6 % and 26.5 %. Regarding Δ28aaCBDAS, clone 10 presented a cell population with a lower percentage (9.23 % of YFP^+ cells), compared to clones 1 and 28, with 50.9 %, and 41.7 %, respectively. These results are consistent with George *et al*, showing that episomal expression is an efficient method to generate *P. tricornutum* transconjugants that express transgenes, requiring less screening compared to random integration methods (George et al., 2020a).

Subcellular localization of CBDAS in P. tricornutum transconjugants

To assess the subcellular localization of CBDAS from the three constructs, *i.e.* without SP (Δ28aaCBDAS), with endogenous SP (CBDAS), and with HASP1 SP (SP:CBDAS), transconjugants were visualized on a confocal microscope, acquiring YFP fluorescence (yellow), and chlorophyll autofluorescence (red) using EV, YFP and SP:YFP as controls (Fig. 4). YFP fluorescence in the YFP control strain was visible throughout the cytoplasm at the early stationary phase (day 10). As previously observed, the HASP1 SP caused considerable changes in the subcellular localization of YFP, clustering in the chloroplast ER rather than the cytoplasm, as a clustered dotted structure between the two plastid lobes (Erdene-ochir et al., 2019a, Apt et al., 2002, Kilian and Kroth, 2005, Liu et al., 2016). According to Erdene-Ochir *et al*, this suggests that the HASP1 signal peptide leads to YFP entry into the secretory pathway. Interestingly, both CBDAS and SP:CBDAS strains presented a localization pattern consistent with clustered structure between the plastid lobes, similar to SP:YFP (Fig. 4, S2 and S3). CBDAS fluorescence pattern also corresponded to the signal of the medial Golgi marker protein XylT (Liu et al., 2016). A similar pattern was observed when the potential retromer subunit Vps29, the structure involved in endosomal retrograde transport to the Golgi apparatus, was heterologously expressed in *P. tricornutum*. These results suggest that CBDAS and SP:CBDAS both localized in the secretory pathway. Unexpectedly, the soluble version of Δ28aaCBDAS enzyme did not behave as a cytosolic protein, although more diffuse than CBDAS and SP:CBDAS, aggregates were observed close to the plastid (Fig. 4 and S4), in all three clones.

Overall, the localization pattern of SP:CBDAS and CBDAS suggests that they cluster in the chloroplastic ER, where post-translational modifications like N-glycosylation, required for activity, stability, and correct folding of endogenous as well as transgenic enzymes, occur (Lingg et al., 2012).

Fig. 4. SP:CBDAS and CBDAS localize at the chloroplastic endoplasmic reticulum in *P. tricornutum* transconjugants. YFP fluorescence, chlorophyll autofluorescence, and the merging of three fields are shown in transgenic lines producing CBDAS:YFP, SP:CBDAS:YFP, Δ28aaCBDAS:YFP,



SP:YFP, and YFP were visualized by confocal laser microscopy. YFP and EV cells were used as positive and negative controls. Scale bars = 10 µm.

CBDAS accumulation and secretion

To analyze the impact of HASP1 signal peptide on enzyme secretion into the extracellular matrix, we performed a western blot using protein pellets from cell extracts and supernatants from stationary phase cultures (Fig. 5A & B). As expected, the EV protein extract did not show any bands following immunoblotting with anti-YFP (Fig. 5A), while the positive control (10 ng of purified GFP) showed the predicted band at 27 kDa. YFP clones presented the expected band at 32 kDa (Thr:YFP:3HA) and a band below it (~30 kDa). The YFP clone 11, presented both bands, but the upper band showed higher intensity. In the case of the SP:YFP strains, clones showed the expected band (~34 kDa, SP:Thr:YFP:3HA), with additional smaller bands at high intensity for clone 4. For further assays, clones YFP 7 and SP:YFP13 were used. All cell extracts from CBDAS clones and SP:CBDAS produced proteins of the predicted size of 92-93 kDa, confirming intracellular accumulation. When Slattery *et al.* heterologously expressed the receptor-binding domain of the SARS-CoV-2 spike protein, under a longer version of HASP1p containing the secretory signal (642-bp), only one of the 9 selected clones produced the desired protein [36]. In our study, the fusion of CBDAS to an endogenous and recognized signal peptide by the secretory mechanism could stabilize the enzyme, as proven when targeted to the ER (Geissler et al., 2018). To verify, if CBDAS and SP:CBDAS were secreted, the extracellular media was concentrated 250 times, and analyzed by western blot (Fig. 5B). YFP clone 7 and EV strains supernatants were used as controls and no specific band was detected. However, SP:YFP, CBDAS, and SP:CBDAS displayed bands close to YFP size (~ 27 kDa), but CBDAS:YFP was not detected. Uncropped blot and red ponceau stain are shown in Fig. S5.

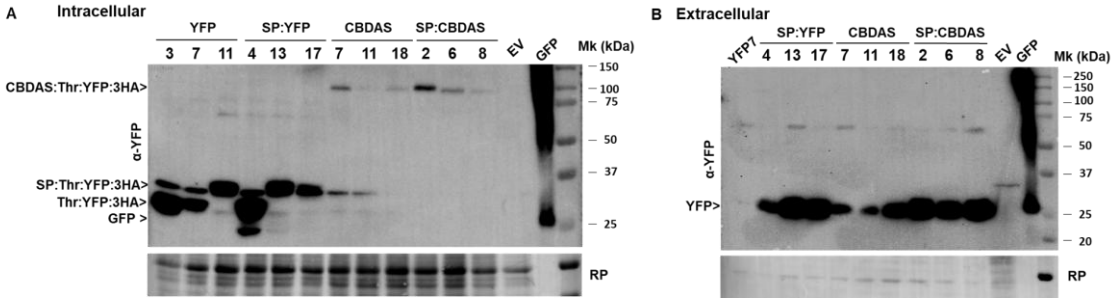


Fig. 5. CBDAS protein detection. Western blot analysis using anti-YFP antibody of total intracellular protein extract from cell lysate and extracellular culture. (A) cell lysates and (B) culture supernatants. YFP (Thr:YFP:3HA, 32 kDa), SP:YFP (SP:Thr:YFP:3HA, 34 kDa), CBDAS (CBDAS:Thr:YFP:3HA, 93 kDa), SP:CBDAS (SP:CBDAS:Thr:YFP:3HA, 92 kDa), EV and purified GFP/YFP (27 kDa). Lower panel, stained blot with red ponceau solution (RP). Full-length Western blots are presented in Supplementary Fig. S5.

Thus, we could not detect CBDAS and SP:CBDAS in the extracellular media, only the YFP cleaved portion was visible. The detection of YFP in the extracellular is consistent with Erdene-Ochir *et al.*, which detected secreted GFP from SP:GFP constructs by western blot during the stationary phase (Erdene-Ochir et al., 2016). The absence of detection of the enzymes could be due to degradation during the secretion process, and/or cleavage from YFP. Studies of protein production in yeast showed that when cells reached their maximal secretory adaptation capabilities, the subset of high protein accumulator cells trigger protein degradation in the entire

cell population, or transiently experience a secretion burnout (Sosa-Carrillo et al., 2023). Alternatively, a cleaved tag-free version of CBDAS could be present in the extracellular matrix but would remain undetectable. A previous study has suggested that the thrombin cleavage sequence could be recognized and cleaved in *P. tricornutum* (Messaabi et al., 2024a).

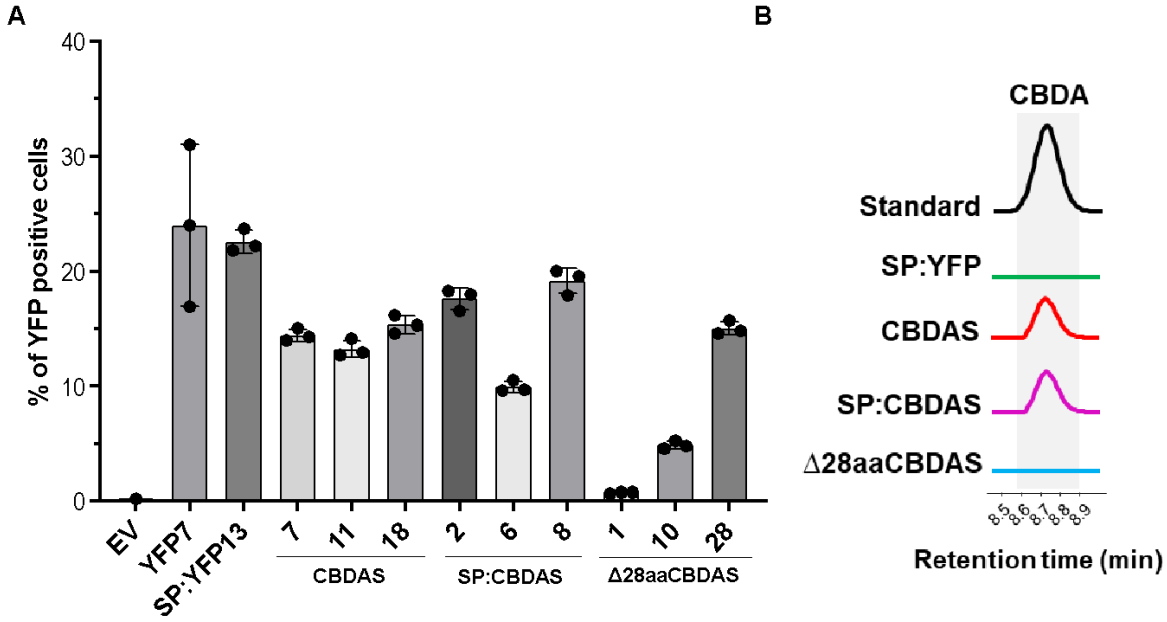


Fig. 6. *P. tricornutum* transconjugants display CBDAS activity. A) Percentage of YFP$^+$ cells of selected transconjugants after 21 days of culture. B) HPLC-DAD chromatograms of *in vitro* CBDA production. Extracts were analyzed by HPLC-DAD and signals were compared to authentic CBDA standard (Fig. S6, S7, S8, and S9).

CBDAS enzymatic activity

As the enzyme was detected in the cell extracts but not in supernatant, cell lysates were used to assess the activity of CBDAS transconjugants, similarly to Slattery *et al.* (Slattery et al., 2022).

Clones were scaled up from the L1 agar plate colonies obtained after *E. coli* conjugation, to avoid possible mutations generated by continuous transfers (Diamond et al., 2023a). The percentage of YFP$^+$ cells was measured in cell cultures grown to the stationary phase for 21 days (Fig. 6A). Total soluble protein was extracted from fluorescent CBDAS 7, 11, 18, SP:CBDAS 2, 6, 8, Δ28aaCBDAS clones 1, 10, 28, EV, YFP clone 7, and SP:YFP clone 13 transconjugants and used for enzymatic assay without protein purification, to avoid protein degradation. CBGA was used as a substrate, and the products of the reaction were analyzed and identified by HPLC-DAD (Fig. 6B, S5, S6). Neither CBDA nor any other traces of CBs were detected in cell protein extracts of SP:YFP and Δ28aaCBDAS strains. A signal eluting at 8.827 min corresponding to the CBDA standard was detected when CBDAS and SP:CBDAS transconjugants protein extracts were incubated with CBGA as substrate. The UV spectrum of the *in vitro* assay was also compared with the one of authentic reference standard CBDA (Fig. S7) and then the identification was confirmed by HPLC-MS/MS (Fig. S8). The highest yield corresponding to 0.55 mg/L of CBDA was obtained in SP:CBDAS clone 6 cell protein extract. Thus, both constructions, with endogenous signal peptide (CBDAS) and HASP1 signal peptide (SP:CBDAS) led to the *in vitro* production of CBDA, while removing the signal peptide

abolished the enzyme activity. These results add to the growing evidence that the enzyme CBDAS requires to be processed through the secretory pathway. It also adds to our previous studies suggesting that the diatom is a suitable host for CBs production (Awwad et al., 2023b, Fantino, 2024).

The first report of CBDAS heterologous production was published in 2018, using wild-type and mutant enzymes heterologously expressed in *Komagataella phaffii* (Zirpel et al., 2018b). The mutant strain CBDAS A414V+A46V+T47A was able to produce 0.42 $g.L^{-1}$ CBDA and 0.13 $g.L^{-1}$ of THCA when supplemented with CBGA (Zirpel et al., 2018a), increasing CBDAS catalytic activity by 3.3 times (Zirpel et al., 2018a). In 2019, Luo *et al.* generated strains of *S. cerevisiae* that produced different CBs from galactose. The highest reported titers of CBDA by hexanoic acid supplementation were 4.3 $\mu g.L^{-1}$ CBDA (Luo et al., 2019b). A second group successfully implemented *de novo* biosynthesis of CBD in *S. cerevisiae*, yielding 6.92 mg/L, by overexpressing the bile pigment transporter 1, which efficiently transfers CBGA from the cytoplasm to the vacuole, where catalysis takes place (Dai, 2024). Experiments in *S. cerevisiae* with co-overexpression of WT *Cs*CBDAS and transcription factor HAC1, which activates the unfolded protein response to restore ER homeostasis (Cox, 1996), resulted in increased specific activities of 11-fold, as determined by *in vitro* assays (Schmidt, 2024). Moreover, multiple copies of *Cs*CBDAS were integrated using a multicopy integration system, and the resulting strain produced 0.13 mg/L CBDA at pH 4 (Schmidt, 2024). Overall, the heterologous expression systems for CB production still require optimization and fine-tuning to be profitable to the industry. However, these studies indicate several avenues for further optimization of *Cs*CBDAS activity in *P. tricornutum.*

Here, we succeeded in engineering the diatom *P. tricornutum* with an active CBDAS, showing the importance of its localization in the ER through the encoded signal peptide. This work highlights the importance of studying linkers and signal peptide behavior in new heterologous platforms, such as the diatom *P. tricornutum*, as well as characterizing the secretory network to construct more efficient tools.

Future research should include the evaluation of medium optimization (phosphate and iron deprivation) (Slattery et al., 2022), directed mutagenesis, the use of smaller tags, different sequence of linkers (Messaabi et al., 2024a), and increasing the gene copy number on CBDA titers in *P. tricornutum*. In addition, strategies to reduce CBs cytotoxicity in *P. tricornutum* cells (Fantino, 2024) such as cell compartmentalization and the use of cannabinoid transporters must be considered.

Conclusion

We previously demonstrated the use of *P. tricornutum* as an *in vivo* heterologous production platform for olivetolic acid and CBGA, precursors of cannabinoids. In the present study, we report, for the first time, the production of one of the most critical enzymes in the cannabinoid biosynthetic pathway in microalgae. Our results showed that the production of three versions of CBDAS i.e. native, modified with an endogenous secretory signal peptide, and truncated one, can be produced in *P. tricornutum*. The YFP fluorescence-based screening method enabled

high-throughput clone selection and facilitated the detection of high fluorescence in the constructs driven by the HASP1 promoter. Enzymatic assays confirmed the activity of CBDAS, with CBDA as the product in both the native and secretory constructs. Overall, we demonstrated that CBDA production through the heterologous expression of CBDAS in *P. tricornutum,* achieves competitive levels when compared to yeast platforms. Although additional improvements are required, these findings validate the diatom as a platform for cannabinoid biosynthesis, offering a sustainable and scalable alternative to traditional production systems.

Methods

Plasmids Construction

Plasmids were constructed by Gibson assembly using the NEBuilder® HiFi DNA Assembly Bundle for Large Fragments (New England Biolabs, Canada). Amplicons used for assemblies were amplified by PCR with PrimeSTAR GXL DNA Polymerase (Takara Bio, Japon) following the manufacturer's protocol. Plasmid pPtGE31 is used as a template to construct the expression vectors (Slattery et al., 2018). The *C. sativa CBDAS* gene was fused in the C terminal to the YFP gene, linked by thrombin cleavage sequence, and a 3xHemagglutinin (HA) tag was introduced in the C terminal of the *YFP* gene. All the sequences were codon optimized for *P. tricornutum* and synthesized by Bio Basic (Markham, Ontario, Canada). These sequences were introduced in the expression cassette containing the 499 bp '*Highly abundant secreted protein 1'* (*HASP)1* promoter (Erdene-ochir et al., 2019a) and *FcpA* terminator. *N-acetyltransferase* gene (*nat*) conferring resistance to Nourseothricin was used as a selection marker, under *FcpC* promoter and *FcpC* terminator. As a result, the following plasmids were generated; the *CBDAS* complete codon sequence, *HASP1pro:CBDAS:Thr:YFP:3HA* (CBDAS), a truncated *CBDAS* version without 84 bp from the 5', *HASP1pro:Δ84CBDAS:Thr:YFP:3HA* (Δ28aaCBDAS) and the truncated *CBDAS* version with the endogenous HASP1 signal peptide SP, 54 bp in the 5', *HASP1pro:SP54:Δ84CBDAS:Thr:YFP:3HA* (SP:CBDAS). As a control, *Thr:YFP:3HA* sequence (*YFP*) and *SP:Thr:YFP:3HA* sequence with the signal peptide *SP (SP:YFP)*. To enhance protein production, a minimal Kozak sequence was placed directly before each ATG initial codon. A modified version of the pPtGE31 vector (Karas et al., 2015c), harboring *nat* gene cassette, was used as a negative control. All DNA sequences and primers used in this study are listed in Supporting Information Table S1 and S2 respectively.

Microbial Strains, Growth Conditions and Transformation

Escherichia coli (NEB® 10-beta, New England Biolabs, Canada) were grown in Luria Broth (LB) supplemented with appropriate antibiotics (chloramphenicol, 30 mg/L). Plasmids were extracted using a miniprep kit allowing the extraction of large vectors (Biobasic EZ10 miniprep kit, NY, USA), sequenced by CCIB DNA Core (Massachusetts General Hospital, United States of America) through Next-Generation sequencing platform and then amplified in Epi 300 strain containing pTA-MOB plasmid to allow conjugation with diatoms, as described in the literature (Karas et al., 2015c). Briefly, 1 mL of wild-type *P. tricornutum* was seeded on 0.5X L1, 1% agar plates and grown at 18 °C on a light/dark cycle of 16/8h for 4 days. Before transformation, 1 mL of L1 media was added to each agar plate, cells were scraped and recovered by pipetting

in a sterile tube. Cell concentration was then adjusted to 5.0×10^8 cells/mL. A volume of 25 mL *E. coli* culture containing the assembled plasmid and pTA-MOB was grown at 37°C under agitation to OD600 of 0.8, then centrifuged at 3000 x g for 10 min and resuspended in 250 µL of SOC media. Conjugation was initiated by adding 200 µL of *P. tricornutum* to 200 µL of *E. coli* cells. The cell mixture was plated on 0.5X L1, 5% LB, ~1% agar plates, incubated at 30°C for 90 min in the dark, and transferred to 18 °C in the light and grown for 2 days. After the recovery period, 1 mL of L1 media was added to the plates to collect cells by scraping. Then, cells were plated on 0.5 X L1, 1% agar plates supplemented with nourseothricin 200 mg/L for selection and incubated at 18 ° C. Transformed colonies appeared after 2 weeks.

Strains Selection

Thirty-six clones per construct were plated on new selective media supplemented with nourseothricin 200 mg/L. After 10 days colonies were observed with a magnification of 80 to 120x and screened by fluorescence under a Fluorescent Stereo Microscope Leica M165 FC with GFP filter. Then, 200 µL cell culture of the selected strains was analyzed in the Synergy H1 BioTek microplate reader to measure the YFP fluorescence at Ex/Em wavelengths of 500/539 nm (n = 3) and OD680nm. EV and YFP strains were used as negative and positive controls, respectively. Strains with fluorescence higher than the mean ratio ($[Ex_{500}/Em_{539}]/OD_{680}$) + 2 x [SD] of EV were determined as YFP positive and analyzed with a CytoFLEX S flow cytometer (Beckman) equipped with violet (405 nm), blue (488 nm), yellow-green (561 nm) and red (638 nm) lasers. In this case, 75 µL of 10-day-old cultures were filtered and transferred to clear 96-well plates with 200 µL of L1 in each well. Chlorophyll autofluorescence was detected in the PerCP channel (690/50 nm), while YFP fluorescence was detected in the FITC channel (525/40 nm). Figures and statistics were analyzed using BD FlowJo version 10 software (BD Biosciences, La Jolla, CA, USA, 2020).

Protein extraction

Fifty mL of bioengineered strains were grown for 21 days in L1 media containing nourseothricin (200 mg/L) then centrifuged at 3,500 x g for 15 min at 4ºC. The supernatant was kept and concentrated 200x using Amicon Ultra-15 30K Centrifugal Filter Units from Sigma (cat. #UFC903008). Pellets were weighed and protein extraction was performed according to Fantino *et al*. (Fantino, 2024). Concentrated supernatants containing soluble protein fraction were kept at -20°C for further use in western blot. Total protein samples were quantified using RC DC™ Protein Assay Kit I (Bio-Rad cat # 5000121).

Western blot

For each clone, 60 µg of total proteins were loaded in 10% SDS- PAGE. Transfer settings: 80 volts until proteins went through the stacking gel, then at 120 volts for 2 hours were used. Primary antibodies were incubated overnight at 4ºC. Primary anti-GFP/YFP/CFP and anti-HA from Cedarlane (Ontario L7L 5R2 Canada, cat. #CLH106AP) and ThermoFisher Scientific (Illinois 61101 USA, cat. #MA1-21316) respectively. Both were used at a 1:1000 dilution in 3% BSA. After three washes with TBST solution, the blots were incubated for 1 hour in a

1:20000 dilution, of Immun-Star Goat Anti-Mouse (GAM)-HRP Conjugate from Bio-Rad (Ontario L5T 1C9 Canada, cat. #1705047) in 5% milk. Purified GFP (10 ng) and Multiple Tag (10 ng), from GenScript (cat. #M0101) were used as a positive control for YFP and HA Tag. After three washes with TBST solution, protein detection was realized by using Clarity Max Western ECL Substrate-Luminol solution from Bio-Rad (cat #1705062S). Chemiluminescence detection and Red Ponceau stained (Glacial Acetic Acid 5% v/v, Ponceau Red dye 0.1% m/v) of the blots were visualized using ChemiDoc Imaging System with Image Lab Touch Software (Bio-Rad cat # 12003153) and Image Lab™ Software (Bio-Rad cat # 1709690).

Confocal Microscopy

Live cell Images were captured, and protein localization was visualized with a Leica TCS SP8 confocal laser scanning microscope (Leica Microsystems) with a 40×/1.30 oil immersion objective. The YFP excitation wavelength used was 488 nm and the emission of fluorescence signals was detected from 500 to 525 nm. Chlorophyll auto-fluorescence was observed with an excitation wavelength of 552 nm and the emission of fluorescence signals was detected from 630 to 670 nm. The combined images were generated using Leica Las X software.

CBDAS in vitro enzymatic assay

Enzymatic assays were carried out using 500 µg of total protein extraction, from 3 biologically independent samples, in a 200 µL volume reaction. The reaction was adapted from *Valliere et al.* (Valliere et al., 2019), composed of (pH 4.5), 25 mM $MgCl_2$, 25 mM KCl, 250 µM FAD, and 300 uM CBGA, and incubated ON at 30°C. The reactions were then extracted with 3 volumes of methanol. Metabolite extracts were filtered (0.2 µm PTFE, Agilent Technologies, cat. no. 5190-5265), then dried in a SpeedVac concentrator and resuspended in 100 µL methanol, HPLC grade, and stored at -20°C for high-performance liquid chromatography coupled with diode-array detection (HPLC-DAD) analyses. Metabolite detection was confirmed by HPLC coupled with tandem mass spectrometry (MS/MS).

Cannabinoids and precursors standards

Olivetolic acid (OA, CAS 491-72-5) and olivetol (OL, CAS 500-66-3) were purchased from Santa Cruz Biotechnologies (Dallas United States). Delta-9-tetrahydrocannabinol (THC, CAS 1972-08-3), cannabidiol (CBD, CAS 13956-29-1), cannabinol (CBN, CAS 521-35-7), *Δ*-9-tetrahydrocannabinolic acid (THCA, CAS 23978-85-0), cannabidiolic acid (CBDA, CAS 1244-58-2), cannabigerolic acid (CBGA, CAS 25555-57-1), cannabichromene (CBC, CAS 20675-51-8), cannabigerol (CBG, CAS 25654-31-3), tetrahydrocannabivarin (THCV, CAS 31262-37-0) and cannabidivarin (CBDV, CAS 24274-48-4) were purchased from Agilent Technologies (QC, Canada). Cannabinolic acid (CBNA, CAS 2808-39-1) was purchased from Sigma-Aldrich (ON, Canada).

HPLC-DAD and HPLC-MS/MS analysis

Analyses were conducted using high-performance liquid chromatography (HPLC) with diode-array detection (DAD). Chromatographic separation of analytes was performed using an

InfinityLab Poroshell 120 EC-C18 column (4.6 × 100 mm, 2.7 mm; Agilent Technologies, QC, Canada) maintained at 30°C. Ten microliters of the sample were injected.

into the analytical device. Mobile phases used during analysis were made of (A) formic acid 0.1 % v/v in milli-Q water and (B) formic acid 0.1 % v/v in methanol with a flow rate of 1 mL/min. The HPLC gradient program was set as follows: 0 min, 70 % B; 1.0 min, 70 % B; 6.0 min, 77 % B; 15.0 min, 90 % B; 15.1 min, 70 % B and 18.0 min, 70 % B. The total run time per sample was 18.5 min to allow the reconditioning of the column before the next injection. The diode array detector was set to acquire the wavelength range of 190 to 400 nm with a deuterium (D2) lamp. All the analyses were done using a UV wavelength of 220 nm. Compounds were identified by comparing retention time and maximum absorption wavelengths obtained with the ones of reference standards (Table S3). Standard calibration curves were prepared as follows to allow absolute quantification; two working solutions were prepared containing CBGA and CBDA at 10 mg/L and 100 mg/L each in HPLC grade methanol. These solutions were further diluted to prepare calibration solutions with the following concentrations in triplicate: 0.5, 1, 2, 4, 5, 10, 25, 50, and 100 mg/L. These standard solutions were injected into the HPLC-DAD system and used to generate calibration curve regressions. Fig. S1 shows the calibration curves obtained by plotting the area under the curve obtained as a function of the analyte's concentration. This allowed CBGA and CBDA quantification for the enzymatic and supplementation assays. Confirmatory analyses were performed using high-performance liquid chromatography (HPLC) coupled with tandem mass spectrometry (MS/MS) (Agilent, QC, Canada). This system has an Agilent Jet Stream ionization source, a binary pump, an autosampler, and a column compartment. Compound separation was achieved using an InfinityLab Poroshell 120 EC-C18 column (4.6 × 100 mm, 2.7 mm; Agilent Technologies, QC, Canada). *In vitro* and supplementation assays samples were centrifuged for 10 min at 12000 rpm and diluted 10-fold in the mobile phase (i.e., formic acid 0.1 % v/v in milli-Q water and formic acid 0.1 % v/v in methanol (30:70)). Five µL of each sample were injected onto the column that was set at 50°C. A gradient method made of (A) formic acid 0.1 % v/v in milli-Q water and (B) formic acid 0.1 % v/v in methanol with a flow rate of 0.5 mL/min was used to achieve chromatographic separation. The HPLC elution program was as follows: 0 min, 70 % B; 7.0 min, 100 % B; 10 min, 100 % B; 12.0 min, 70 % B. The total run time was 14 min per sample. The parameters used in the MS/MS source were set as follows: gas flow rate 8 L/min, gas temperature 220°C, nebulizer 55 psi, sheath gas flow 12 L/min, sheat gas temperature 380°C, capillary voltage 4500 V and nozzle voltage 0 V. Agilent MassHunter Data Acquisition (version 1.2) and MassHunter Qualitative Analysis (version 10.0) softwares were used for data acquisition and processing respectively. Samples analyses were carried out in triggered multiple reaction monitoring (tMRM) acquisition mode allowing compound identification using authentic standards. Table S4 shows MRM transitions and MS/MS parameters used for targeted compound identification.

Statistics and reproducibility

General data analysis (means and standard deviation) was performed primarily by GraphPad Prism 10.0.2. All experiments were performed with three biological replicates and values were expressed as means ± standard errors.

Conflicts of interest

The authors declare no conflict of interest.

Acknowledgments

Special thanks to Professor Hugo Germain and MSc. Mélodie B. Plourde from the Université du Québec à Trois -Rivières for providing us with the purified GFP to use as a control for immunoblotting of Fig. 4. We acknowledge that financial support for this research was funded by the Natural Sciences and Engineering Research Council of Canada through the Alliance program Award No ALLRP 570476-2021 to IDP. Additional support in the form of scholarships to EF, AM, FA, NS, K-LR, and AC from Mitacs-Acceleration/Globalink program grants no IT12310, IT16463, and IT19432 to IDP is also acknowledged.

CRediT authorship contribution statement

Elisa Fantino & Anis Messaabi: Conceptualization, Data curation, Formal analysis, Investigation, Laboratory work, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing; Natacha Mérindol: Data curation, Project administration, Visualization, Writing – review & editing; Fatima Awwad: Formal analysis, Writing – original draft, Writing – review & editing; Nicolas Sene: Formal analysis, Writing – original draft, Writing – review & editing; Sarah-Eve Gélinas: Investigation, Formal analysis, methodology, Writing – review & editing; Alexandre Custeau and Kimy-Li Rhéaume: Investigation, Formal analysis; Writing – review & editing; Fatma Meddeb-Mouelhi: Conceptualization, Supervision, Project administration, Writing – review & editing; Isabel Desgagné-Penix: Conceptualization, Funding acquisition, Resources, Supervision, Project administration, Writing – review & editing.

References

1. Castillo-Arellano, J., et al., *The Polypharmacological Effects of Cannabidiol.* Molecules, 2023. **28**(7).
2. Stella, N., *THC and CBD: Similarities and differences between siblings.* Neuron, 2023. **111**(3): p. 302-327.
3. Wang, B., et al., *In search of preventive strategies: novel high-CBD Cannabis sativa extracts modulate ACE2 expression in COVID-19 gateway tissues*, in *Aging*. 2020. p. 22425-22440.
4. Crocq, M.A., *History of cannabis and the endocannabinoid system*Dialogues Clin Neurosci, 2020. **22**(3): p. 223-228.
5. Walsh, K.B., A.E. McKinney, and A.E. Holmes, *Minor Cannabinoids: Biosynthesis, Molecular Pharmacology and Potential Therapeutic Uses*, in *Frontiers in Pharmacology*. 2021. p. 1-18.
6. Quinones, R., et al., *Quantification of Cannabis in Infused Consumer Products and Their Residues on Skin.* ACS Pharmacol Transl Sci, 2022. **5**(8): p. 642-651.
7. Cristino, L., T. Bisogno, and V. Di Marzo, *Cannabinoids and the expanded endocannabinoid system in neurological disorders.* Nat Rev Neurol, 2020. **16**(1): p. 9-29.

8. Devinsky, O., E. Marsh, and D. Friedman, *Cannabidiol in patients with treatment-resistant epilepsy - Authors' reply.* Lancet Neurol, 2016. **15**(6): p. 545-6.

9. Devinsky, O., et al., *Trial of Cannabidiol for Drug-Resistant Seizures in the Dravet Syndrome.* N Engl J Med, 2017. **376**(21): p. 2011-2020.

10. Desaulniers Brousseau, V., et al., *Cannabinoids and Terpenes: How Production of Photo-Protectants Can Be Manipulated to Enhance Cannabis sativa L. Phytochemistry.* Front Plant Sci, 2021. **12**: p. 620021.

11. Taura, F., et al., *Purification and characterization of cannabidiolic-acid synthase from Cannabis sativa L. Biochemical analysis of a novel enzyme that catalyzes the oxidocyclization of cannabigerolic acid to cannabidiolic acid*, in *Journal of Biological Chemistry*. 1996, Â© 1996 ASBMB. Currently published by Elsevier Inc; originally published by American Society for Biochemistry and Molecular Biology. p. 17411-17416.

12. Sirikantaramas, S., et al., *The gene controlling marijuana psychoactivity. Molecular cloning and heterologous expression of Δ1-tetrahydrocannabinolic acid synthase from Cannabis sativa L.*, in *Journal of Biological Chemistry*. 2004, Â© 2004 ASBMB. Currently published by Elsevier Inc; originally published by American Society for Biochemistry and Molecular Biology. p. 39767-39774.

13. Lim, K.J.H., et al., *Biosynthesis of nature-inspired unnatural cannabinoids*, in *Molecules*. 2021.

14. Andre, C.M., J.-f. Hausman, and G. Guerriero, *Cannabis sativa : The Plant of the Thousand and One Molecules*. 2016. p. 1-17.

15. Dai, L.N., T.; Luo, R.; Zhang, L.; Zhang, S.; Kang, Y.; Chi, J.; Feng, X.; Shi, J.; Tian, Y.; Gao, B.; Li, Z., *Improvement of cannabidiolic acid synthetase activity through molecular docking and site-directed mutagenesis.* Industrial Crops and Products, 2024.

16. de A. Leite, J., et al., *Extraction and isolation of cannabinoids from marijuana seizures and characterization by (1)H NMR allied to chemometric tools.* Sci Justice, 2018. **58**(5): p. 355-365.

17. Ohtsuki, T., et al., *Selective Preparation and High Dynamic-Range Analysis of Cannabinoids in "CBD Oil" and Other Cannabis sativa Preparations.*, in *Journal of natural products*. 2022. p. 634-646.

18. Chiurchi E.S., S.; Allegrini, P.; Ciceri, D.; Ballini, R.; Palmieri A., *A Novel and Practical Continuous Flow Chemical Synthesis of Cannabidiol (CBD) and its CBDV and CBDB Analogues.* European Journal of Organic Chemistry, 2021.

19. Kearsey, L.J., et al., *Structure of the Cannabis sativa olivetol-producing enzyme reveals cyclization plasticity in type III polyketide synthases*, in *FEBS Journal*. 2019. p. 1-14.

20. Shoyama, Y., et al., *Structure and Function of Δ 1-Tetrahydrocannabinolic Acid (THCA) Synthase , the Enzyme Controlling the Psychoactivity of Cannabis sativa*, in *Journal of Molecular Biology*. 2012, Elsevier Ltd. p. 96-105.

21. Zirpel, B., F. Stehle, and O. Kayser, *Production of Delta9-tetrahydrocannabinolic acid from cannabigerolic acid by whole cells of Pichia (Komagataella) pastoris expressing Delta9-tetrahydrocannabinolic acid synthase from Cannabis sativa L.*, in *Biotechnology letters*. 2015. p. 1869-1875.

22. Zirpel, B., et al., *Engineering yeasts as platform organisms for cannabinoid biosynthesis*, in *Journal of Biotechnology*. 2017, Elsevier. p. 204-212.

23. Zirpel, B., et al., *Optimization of Δ9-tetrahydrocannabinolic acid synthase production in Komagataella phaffii via post-translational bottleneck identification*, in *Journal of Biotechnology*. 2018, Elsevier. p. 40-47.

24. Luo, X., et al., *Complete biosynthesis of cannabinoids and their unnatural analogues in yeast*, in *Nature*. 2019, Springer US. p. 123-126.

25. Gülck, T., et al., *Synthetic Biology of Cannabinoids and Cannabinoid Glucosides in Nicotiana benthamiana and Saccharomyces cerevisiae*, in *Journal of Natural Products*. 2020. p. 2877-2893.

26. Kosalková, K.B., C.; Sánchez-Orejas, I.C.; Cueto, L.; García-Estrada, C., *Biotechnological Fungal Platforms for the Production of Biosynthetic Cannabinoids.* Journal of Fungi, 2023.

27. Awwad, F., et al., *Bioengineering of the Marine Diatom Phaeodactylum tricornutum with Cannabis Genes Enables the Production of the Cannabinoid Precursor, Olivetolic Acid.* Int J Mol Sci, 2023. **24**(23).

28. Fantino, E.A., F.; Merindol, N.; Diaz Garza, A.M.; Gelinas, S.E; Gajon Robles, G.C.; Custeau, A.; Meddeb-Mouelhi, F.; Desgagne-Penix, I., *Bioengineering Phaeodactylum tricornutum, a marine diatom, for cannabinoid biosynthesis.* Algal Research, 2024.

29. Blatt-janmaat, K. and Y. Qu, *The Biochemistry of Phytocannabinoids and Metabolic Engineering of Their Production in Heterologous Systems*. 2021.

30. Sirikantaramas, S., et al., *Tetrahydrocannabinolic Acid Synthase , the Enzyme Controlling Marijuana Psychoactivity , is Secreted into the Storage Cavity of the Glandular Trichomes*. 2005. p. 1578-1582.

31. Geissler, M., et al., *Subcellular localization defines modification and production of D 9 -tetrahydrocannabinolic acid synthase in transiently transformed Nicotiana benthamiana*, in *Biotechnology Letters*. 2018, Springer Netherlands. p. 913-919.

32. Schmidt, C.A., M.; Kayser, O., *Engineering cannabinoid production in Saccharomyces cerevisiae.* Biotechnology journal, 2024.

33. Hu, J.M., W.; Su, Y.; Qian, C.; Fu, W., *Emerging technologies for advancing microalgal photosynthesis and metabolism toward sustainable production.* Frontiers, 2023.

34. Einhaus, A., T. Baier, and O. Kruse, *Molecular design of microalgae as sustainable cell factories.* Trends Biotechnol, 2024. **42**(6): p. 728-738.

35. Sethi, D., et al., *Diatoms for Carbon Sequestration and Bio-Based Manufacturing.* Biology (Basel), 2020. **9**(8).

36. Ova Ozcan, D. and B. Ovez, *Evaluation of the interaction of temperature and light intensity on the growth of Phaeodactylum tricornutum: Kinetic modeling and optimization*, in *Biochemical Engineering Journal*. 2020, Elsevier. p. 107456.

37. Cui, Y.T.-H., S.R.; Chua, E.T.; Schenk, P.M., *Development of high-level omega-3 eicosapentaenoic acid (EPA) production from Phaeodactylum tricornutum.* Journal of Phycology, 2021.

38. Jaramillo-Madrid, A.C., et al., *Overexpression of key sterol pathway enzymes in two model marine diatoms alters sterol profiles in Phaeodactylum tricornutum*, in *Pharmaceuticals*. 2020. p. 1-19.

39. Fabris, M., et al., *Extrachromosomal Genetic Engineering of the Marine Diatom Phaeodactylum tricornutum Enables the Heterologous Production of Monoterpenoids*, in *ACS Synthetic Biology*. 2020. p. 598-612.

40. George, J., et al., *Metabolic Engineering Strategies in Diatoms Reveal Unique Phenotypes and Genetic Configurations With Implications for Algal Genetics and Synthetic Biology*, in *Frontiers in Bioengineering and Biotechnology*. 2020. p. 1-19.

41. Erdene-ochir, E., et al., *Identification and characterisation of the novel endogenous promoter HASP1 and its signal peptide from Phaeodactylum tricornutum*, in *Scientific Reports*. 2019, Springer US. p. 1-10.

42. Slattery, S.S., et al., *Phosphate - regulated expression of the SARS - CoV - 2 receptor - binding domain in the diatom Phaeodactylum tricornutum for pandemic diagnostics*, in *Scientific Reports*. 2022, Nature Publishing Group UK. p. 1-15.

43. Jenny, R.J., K.G. Mann, and R.L. Lundblad, *A critical review of the methods for cleavage of fusion proteins with thrombin and factor Xa.* Protein Expr Purif, 2003. **31**(1): p. 1-11.

44. BioRender, *BioRender*. 2023.

45. Apt, K.E., et al., *In vivo characterization of diatom multipartite plastid targeting signals.* J Cell Sci, 2002. **115**(Pt 21): p. 4061-9.

46. Kilian, O. and P.G. Kroth, *Identification and characterization of a new conserved motif within the presequence of proteins targeted into complex diatom plastids.*, in *The Plant journal : for cell and molecular biology*. 2005. p. 175-183.

47. Liu, X., et al., *Addressing various compartments of the diatom model organism Phaeodactylum tricornutum via sub-cellular marker proteins*, in *Algal Research*. 2016, Elsevier B.V. p. 249-257.

48. Lingg, N., et al., *The sweet tooth of biopharmaceuticals: importance of recombinant protein glycosylation analysis.* Biotechnol J, 2012. **7**(12): p. 1462-72.

49. Erdene-Ochir, E., et al., *Cloning of a novel endogenous promoter for foreign gene expression in Phaeodactylum tricornutum*, in *Applied Biological Chemistry*. 2016, Springer Netherlands. p. 861-867.

50. Sosa-Carrillo, S., et al., *Maximizing protein production by keeping cells at optimal secretory stress levels using real-time control approaches.* Nat Commun, 2023. **14**(1): p. 3028.

51. Messaabi, A., et al., *In vivo thrombin activity in the diatom Phaeodactylum tricornutum: biotechnological insights.* Appl Microbiol Biotechnol, 2024. **108**(1): p. 481.

52. Diamond, A., et al., *Instability of extrachromosomal DNA transformed into the diatom Phaeodactylum tricornutum*. 2023.

53. Zirpel, B., O. Kayser, and F. Stehle, *Elucidation of structure-function relationship of THCA and CBDA synthase from Cannabis sativa L .* in *Journal of Biotechnology*. 2018, Elsevier. p. 17-26.

54. Cox, J.S.W., P., *A Novel Mechanism for Regulating Activity of a Transcription Factor That Controls the Unfolded Protein Response.* 1996.

55. Slattery, S.S., et al., *An Expanded Plasmid-Based Genetic Toolbox Enables Cas9 Genome Editing and Stable Maintenance of Synthetic Pathways in Phaeodactylum tricornutum.*, in *ACS synthetic biology*. 2018. p. 328-338.

56. Karas, B.J., et al., *Designer diatom episomes delivered by bacterial conjugation*, in *Nature Communications*. 2015, Nature Publishing Group. p. 1-10.

57. Valliere, M.A., et al., *A cell-free platform for the prenylation of natural products and application to cannabinoid production*, in *Nature Communications*. 2019, Springer US. p. 1-9.

58. Sharma, N., et al., *Impact of different light characteristics on the growth and lipid content of diatom Phaeodactylum tricornutum transconjugant strains*, in *American Journal of Plant Sciences*. 2023. p. 41-63.

59. Lyska, D., et al., *pAUL: a gateway-based vector system for adaptive expression and flexible tagging of proteins in Arabidopsis.* PLoS One, 2013. **8**(1): p. e53787.

ANNEX D: SteelyA Western blots

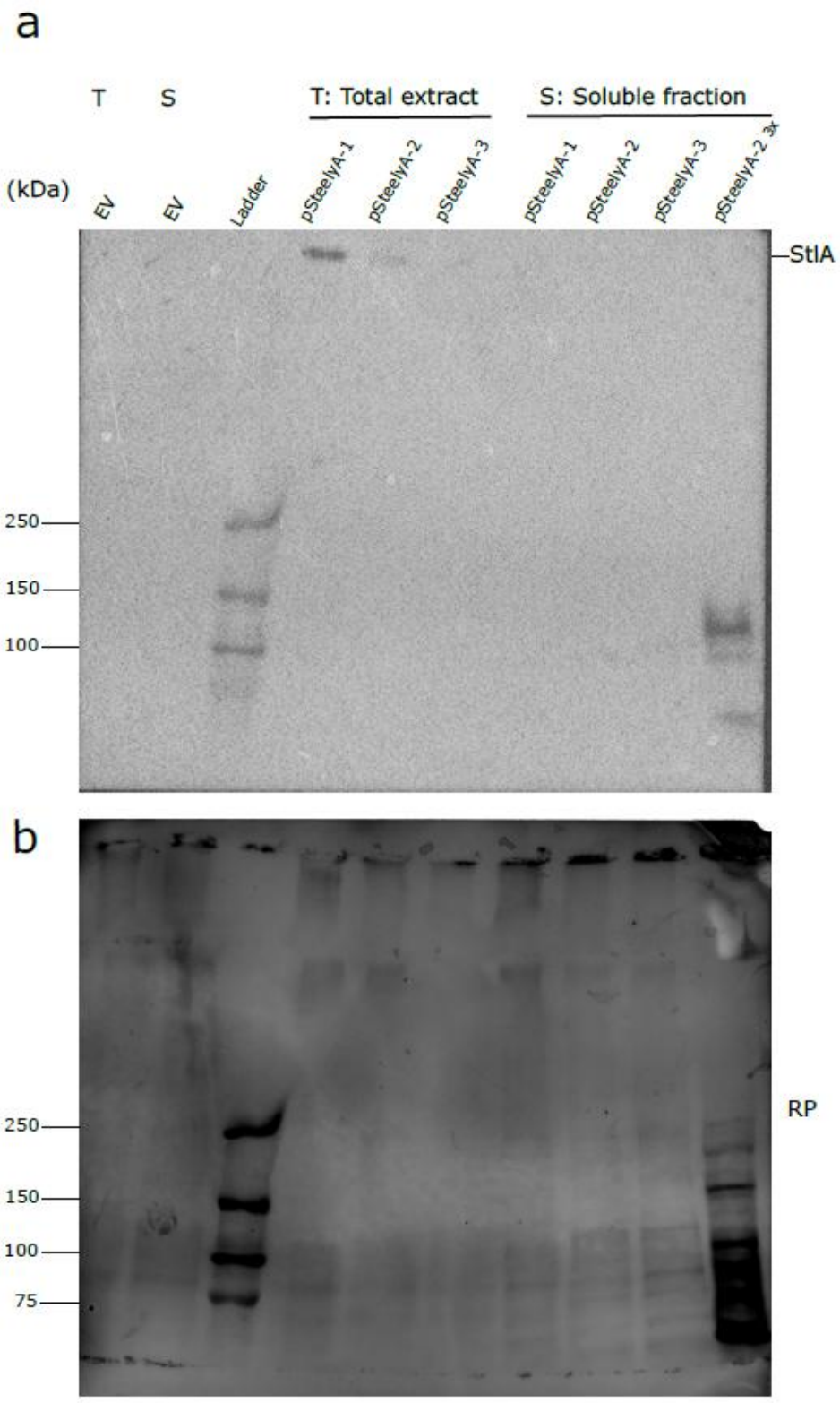


Figure D.1 Western blot Steely A made with a homemade gradient gel.

Bands are observable in the total fraction of pSteelyA-1 and -2 samples high above the 252 kDa ladder.

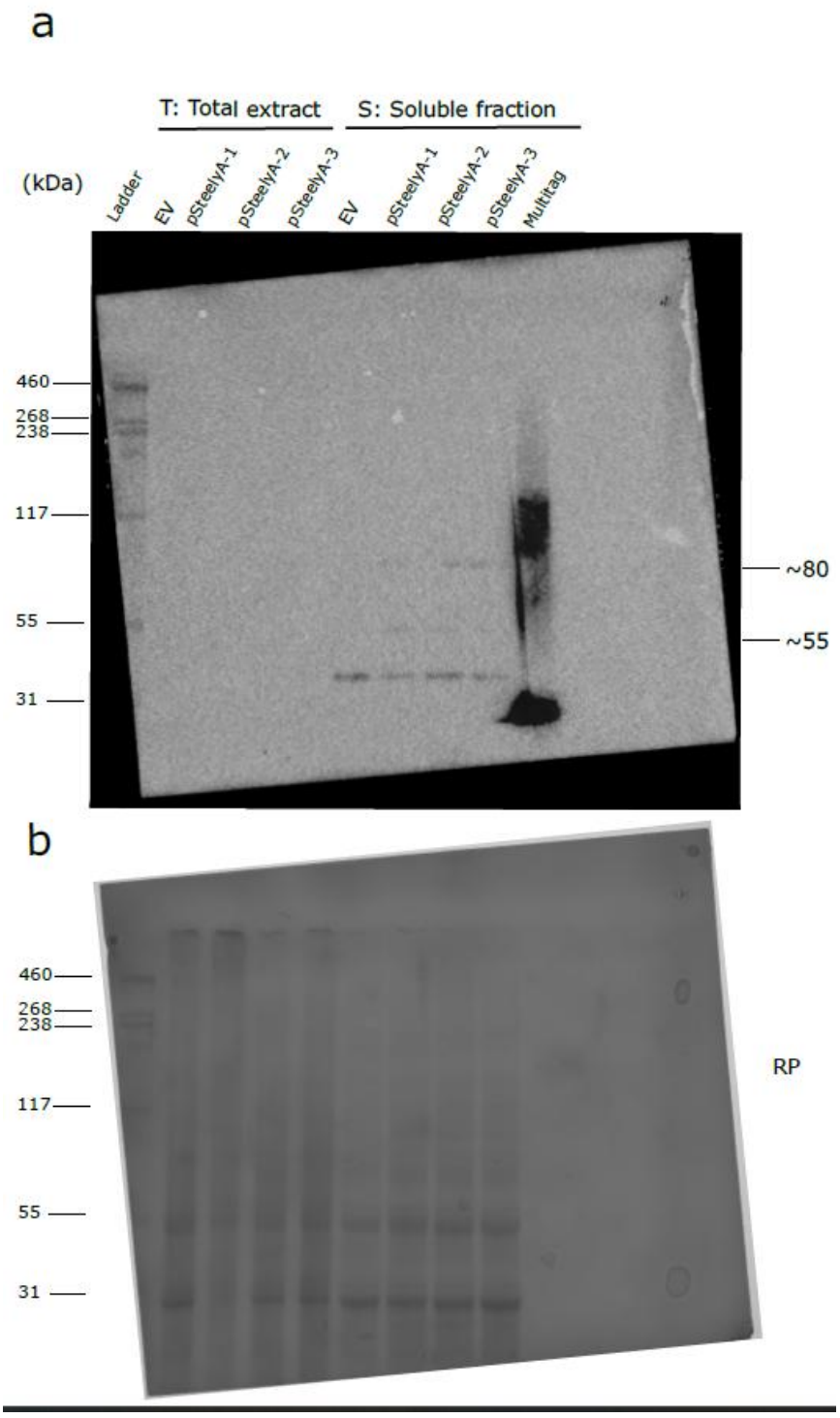


Figure D.2 Western blot Steely A made with a SurePage 4-12% gradient gel.

Bands are observable around 33 kDa in the soluble fraction of all samples. One band is visible in all pSteelyA samples around 55 kDa, and one hardly in pSteelyA-2 and -3 around 80 kDa.

REFERENCES

ABATE, S., LANZAFAME, P., PERATHONER, S. & CENTI, G. 2015. New sustainable model of biorefineries: biofactories and challenges of integrating bio-and solar refineries. *ChemSusChem,* 8**,** 2854-2866.

ABBASI, A. M., KHAN, S. M., AHMAD, M., KHAN, M. A., QUAVE, C. L. & PIERONI, A. 2013. Botanical ethnoveterinary therapies in three districts of the Lesser Himalayas of Pakistan. *Journal of ethnobiology and ethnomedicine,* 9**,** 1-21.

AFRIN, F., CHI, M., EAMENS, A. L., DUCHATEL, R. J., DOUGLAS, A. M., SCHNEIDER, J., GEDYE, C., WOLDU, A. S. & DUN, M. D. 2020. Can hemp help? Low-THC cannabis and non-THC cannabinoids for the treatment of cancer. *Cancers,* 12**,** 1033.

ALASALVAR, C., SALVADÓ, J.-S. & ROS, E. 2020. Bioactives and health benefits of nuts and dried fruits. *Food chemistry,* 314**,** 126192.

ALONSO-CASTRO, A. J., ARANA-ARGÁEZ, V., YÁÑEZ-BARRIENTOS, E., TORRES-ROMERO, J. C., CHABLE-CETZ, R. J., WORBEL, K., DE JESÚS EUAN-CANTO, A., WROBEL, K., GONZÁLEZ-IBARRA, A. & SOLORIO-ALVARADO, C. R. 2021. Pharmacological activities of Asclepias curassavica L.(Apocynaceae) aerial parts. *Journal of Ethnopharmacology,* 281**,** 114554.

AMERI, A. 1999. The effects of cannabinoids on the brain. *Progress in neurobiology,* 58**,** 315-348.

ANDERSON, L. C. 1980. Leaf variation among Cannabis species from a controlled garden. *Botanical Museum Leaflets, Harvard University,* 28**,** 61-69.

ANDRE, C. M., HAUSMAN, J.-F. & GUERRIERO, G. 2016. Cannabis sativa : The Plant of the Thousand and One Molecules.

ANDROVICOVA, R., HORACEK, J., STARK, T., DRAGO, F. & MICALE, V. 2017. Endocannabinoid system in sexual motivational processes: Is it a novel therapeutic horizon? *Pharmacological Research,* 115**,** 200-208.

ANJARD, C., SU, Y. & LOOMIS, W. F. 2011. The polyketide MPBD initiates the SDF-1 signaling cascade that coordinates terminal differentiation in Dictyostelium. *Eukaryotic cell,* 10**,** 956-963.

ANNESLEY, S. J. & FISHER, P. R. 2009. Dictyostelium discoideum—a model for many reasons. *Molecular and cellular biochemistry,* 329**,** 73-91.

APT, K. E., ZASLAVKAIA, L., LIPPMEIER, J. C., LANG, M., KILIAN, O., WETHERBEE, R., GROSSMAN, A. R. & KROTH, P. G. 2002. In vivo characterization of diatom multipartite plastid targeting signals. *J Cell Sci,* 115**,** 4061-9.

ARIAS, C. A. D., OLIVEIRA, C. F. M. D., MOLINO, J. V. D., FERREIRA-CAMARGO, L. S., MATSUDO, M. C. & CARVALHO, J. C. M. D. 2022. Production of recombinant biopharmaceuticals in Chlamydomonas reinhardtii. *International Journal of Plant Biology,* 14**,** 39-52.

ARMBRUST, E. V. 2009. The life of diatoms in the world's oceans. *Nature,* 459**,** 185-192.

ATHANASAKOGLOU, A., GRYPIOTI, E., MICHAILIDOU, S., IGNEA, C., MAKRIS, A. M., KALANTIDIS, K., MASSÉ, G., ARGIRIOU, A., VERRET, F. & KAMPRANIS, S. C. 2019. Isoprenoid biosynthesis in the diatom Haslea ostrearia. *New Phytologist,* 222**,** 230-243.

ATTYGALLE, A., SIEGEL, B., VOSTROWSKY, O., BESTMANN, H. & MASCHWITZ, U. 1989. Chemical composition and function of metapleural gland secretion of the ant, Crematogaster deformis Smith (Hymenoptera: Myrmicinae). *Journal of chemical ecology,* 15**,** 317-328.

AUBRY, L. & FIRTEL, R. 1999. Integration of signaling networks that regulate Dictyostelium differentiation. *Annual review of cell and developmental biology,* 15**,** 469-517.

AUSTIN, M. B., SAITO, T., BOWMAN, M. E., HAYDOCK, S., KATO, A., MOORE, B. S., KAY, R. R. & NOEL, J. P. 2006. Biosynthesis of Dictyostelium discoideum differentiation-inducing factor by a hybrid type I fatty acid–type III polyketide synthase. *Nature chemical biology,* 2**,** 494-502.

AVESANI, L., MERLIN, M., GECCHELE, E., CAPALDI, S., BROZZETTI, A., FALORNI, A. & PEZZOTTI, M. 2014. Comparative analysis of different biofactories for the production of a major diabetes autoantigen. *Transgenic Research,* 23**,** 281-291.

AWWAD, F., FANTINO, E. I., HÉNEAULT, M., DIAZ-GARZA, A. M., MERINDOL, N., CUSTEAU, A., GÉLINAS, S.-E., MEDDEB-MOUELHI, F., LI, J. & LEMAY, J.-F. 2023a. Bioengineering of the marine diatom Phaeodactylum tricornutum with Cannabis genes enables the production of the cannabinoid precursor, olivetolic acid. *International Journal of Molecular Sciences,* 24**,** 16624.

AWWAD, F., FANTINO, E. I., HENEAULT, M., DIAZ-GARZA, A. M., MERINDOL, N., CUSTEAU, A., GELINAS, S. E., MEDDEB-MOUELHI, F., LI, J., LEMAY, J. F., KARAS, B. J. & DESGAGNE-PENIX, I. 2023b. Bioengineering of the Marine Diatom Phaeodactylum tricornutum with Cannabis Genes Enables the Production of the Cannabinoid Precursor, Olivetolic Acid. *Int J Mol Sci,* 24.

AYAKAR, S. R., PAWAR, S. V., HALLAM, S. J., HOSSAIN, S., YADAV, V. G. & SRIVASTAVA, S. K. 2024. Metabolic engineering of E. coli for the biosynthesis of cannabinoid. Google Patents.

AZMIR, J., ZAIDUL, I. S. M., RAHMAN, M. M., SHARIF, K., MOHAMED, A., SAHENA, F., JAHURUL, M., GHAFOOR, K., NORULAINI, N. & OMAR, A. 2013. Techniques for extraction of bioactive compounds from plant materials: A review. *Journal of food engineering,* 117**,** 426-436.

BACKER, R., SCHWINGHAMER, T., ROSENBAUM, P., MCCARTY, V., EICHHORN BILODEAU, S., LYU, D., AHMED, M. B., ROBINSON, G., LEFSRUD, M. & WILKINS, O. 2019. Closing the yield gap for cannabis: a meta-analysis of factors determining cannabis yield. *Frontiers in plant science,* 10**,** 495.

BAESHEN, M. N., AL-HEJIN, A. M., BORA, R. S., AHMED, M. M., RAMADAN, H. A., SAINI, K. S., BAESHEN, N. A. & REDWAN, E. M. 2015. Production of biopharmaceuticals in E. coli: current scenario and future perspectives.

BAO, M., YANG, W., LI, X., YU, G., GU, W., GAO, S. & WANG, G. 2025. A genetically modified triradiate strain of Phaeodactylum tricornutum demonstrates considerable advantages in a 60-L photobioreactor. *Journal of Applied Phycology***,** 1-8.

BARNES, M. P. 2006. Sativex®: Clinical efficacy and tolerability in the treatment of symptoms of multiple sclerosis and neuropathic pain. *Expert opinion on pharmacotherapy,* 7**,** 607-615.

BARTOLOMEI, B., GENTILE, G., ROSSO, C., FILIPPINI, G. & PRATO, M. 2021. Turning the light on phenols: new opportunities in organic synthesis. *Chemistry–A European Journal,* 27**,** 16062-16070.

BATTISTA, N., DI TOMMASO, M., BARI, M. & MACCARRONE, M. 2012. The endocannabinoid system: an overview. *Frontiers in behavioral neuroscience,* 6**,** 9.

BELUNS, S., GAIDUKOVS, S., PLATNIEKS, O., GRASE, L., GAIDUKOVA, G. & THAKUR, V. K. 2023. Sustainable hemp-based bioplastics with tunable properties via reversible thermal crosslinking of cellulose. *International Journal of Biological Macromolecules,* 242**,** 125055.

BENNER, S. A. & SISMOUR, A. M. 2005. Synthetic biology. *Nature reviews genetics,* 6**,** 533-543.

BHUYAN, D. J., ALSHERBINY, M. A., PERERA, S., LOW, M., BASU, A., DEVI, O. A., BAROOAH, M. S., LI, C. G. & PAPOUTSIS, K. 2019. The odyssey of bioactive compounds in avocado (Persea americana) and their health benefits. *Antioxidants,* 8**,** 426.

BIE, B., WU, J., FOSS, J. F. & NAGUIB, M. 2018. An overview of the cannabinoid type 2 receptor system and its therapeutic potential. *Current Opinion in Anesthesiology,* 31**,** 407-414.
BIOCYC. 2024. *hexanoyl-CoA* [Online]. Available: https://biocyc.org/compound?orgid=PHAEO&id=HEXANOYL-COA#RXNS [Accessed 23-10 2024].
BIORENDER 2023. BioRender.
BIRCH, A. 1967. Biosynthesis of polyketides and related compounds. *Science,* 156**,** 202-206.
BITWELL, C., INDRA, S. S., LUKE, C. & KAKOMA, M. K. 2023. A review of modern and conventional extraction techniques and their applications for extracting phytochemicals from plants. *Scientific African,* 19**,** e01585.
BLATT-JANMAAT, K. & QU, Y. 2021a. The biochemistry of phytocannabinoids and metabolic engineering of their production in heterologous systems. *International Journal of Molecular Sciences,* 22**,** 2454.
BLATT-JANMAAT, K. & QU, Y. 2021b. The Biochemistry of Phytocannabinoids and Metabolic Engineering of Their Production in Heterologous Systems.
BLOUNT, Z. D. 2015. The unexhausted potential of E. coli. *elife,* 4**,** e05826.
BOGART, J. W., CABEZAS, M. D., VÖGELI, B., WONG, D. A., KARIM, A. S. & JEWETT, M. C. 2021. Cell-free exploration of the natural product chemical space. *ChemBioChem,* 22**,** 84-91.
BOHLIN, K. H. 1897. *Zur morphologie und biologie einzelliger algen*.
BONINI, S. A., PREMOLI, M., TAMBARO, S., KUMAR, A., MACCARINELLI, G., MEMO, M. & MASTINU, A. 2018. Cannabis sativa: A comprehensive ethnopharmacological review of a medicinal plant with a long history. *Journal of ethnopharmacology,* 227**,** 300-315.
BOURDIN, B., ADENIER, H. & PERRIN, Y. 2007. Carnitine is associated with fatty acid metabolism in plants. *Plant Physiology and Biochemistry,* 45**,** 926-931.
BRANCO-VIEIRA, M., SAN MARTIN, S., AGURTO, C., FREITAS, M. A., MARTINS, A. A., MATA, T. M. & CAETANO, N. S. 2020. Biotechnological potential of Phaeodactylum tricornutum for biorefinery processes. *Fuel,* 268**,** 117357.
BROUSSEAU, V. D., WU, B.-S., MACPHERSON, S., MORELLO, V. & LEFSRUD, M. 2021. Cannabinoids and terpenes: how production of photo-protectants can be manipulated to enhance Cannabis sativa L. phytochemistry. *Frontiers in Plant Science,* 12**,** 620021.
BRUGGEMAN, F. J. & WESTERHOFF, H. V. 2007. The nature of systems biology. *TRENDS in Microbiology,* 15**,** 45-50.
BUNTRU, M., HAHNENGRESS, N., CROON, A. & SCHILLBERG, S. 2022. Plant-derived cell-free biofactories for the production of secondary metabolites. *Frontiers in Plant Science,* 12**,** 794999.
BUTLER, T., KAPOORE, R. V. & VAIDYANATHAN, S. 2020. Phaeodactylum tricornutum: a diatom cell factory. *Trends in biotechnology,* 38**,** 606-622.
BUTLER, T. O., PADMAPERUMA, G., LIZZUL, A. M., MCDONALD, J. & VAIDYANATHAN, S. 2022. Towards a Phaeodactylum tricornutum biorefinery in an outdoor UK environment. *Bioresource Technology,* 344**,** 126320.
CALERO, P. & NIKEL, P. I. 2019. Chasing bacterial chassis for metabolic engineering: a perspective review from classical to non-traditional microorganisms. *Microbial Biotechnology,* 12**,** 98-124.
CALLAWAY, J. 2004. Hempseed as a nutritional resource: An overview. *Euphytica,* 140**,** 65-72.
CALLAWAY, J. C. & PATE, D. W. 2009. Hempseed oil. *Gourmet and health-promoting specialty oils.* Elsevier.

CAPORGNO, M. P., OLKIEWICZ, M., TORRAS, C., SALVADÓ, J., CLAVERO, E. & BENGOA, C. 2016. Effect of pre-treatments on the production of biofuels from Phaeodactylum tricornutum. *Journal of Environmental Management,* 177**,** 240-246.
CASTILLO-ARELLANO, J., CANSECO-ALBA, A., CUTLER, S. J. & LEON, F. 2023. The Polypharmacological Effects of Cannabidiol. *Molecules,* 28.
CHAPLIN, M. F. & BUCKE, C. 1990. *Enzyme technology*, CUP Archive.
CHEUNG, P. C. 2010. The nutritional and health benefits of mushrooms. *Nutrition Bulletin,* 35**,** 292-299.
CHIURCHIÙ, E. S., S.; ALLEGRINI, P.; CICERI, D.; BALLINI, R.; PALMIERI A. 2021. A Novel and Practical Continuous Flow Chemical Synthesis of Cannabidiol (CBD) and its CBDV and CBDB Analogues. *European Journal of Organic Chemistry*.
CLAASSENS, N. J., BURGENER, S., VÖGELI, B., ERB, T. J. & BAR-EVEN, A. 2019. A critical comparison of cellular and cell-free bioproduction systems. *Current opinion in biotechnology,* 60**,** 221-229.
CLOMBURG, J. M., CRUMBLEY, A. M. & GONZALEZ, R. 2017. Industrial biomanufacturing: the future of chemical production. *Science,* 355**,** aag0804.
CLOMBURG, J. M. & GONZALEZ, R. 2010. Biofuel production in Escherichia coli: the role of metabolic engineering and synthetic biology. *Applied microbiology and biotechnology,* 86**,** 419-434.
CONNEELY, L. J., MAULEON, R., MIEOG, J., BARKLA, B. J. & KRETZSCHMAR, T. 2021. Characterization of the Cannabis sativa glandular trichome proteome. *PLoS One,* 16**,** e0242633.
COX, J. S. W., P. 1996. A Novel Mechanism for Regulating Activity of a Transcription Factor That Controls the Unfolded Protein Response.
CRIMMINS, E. M. 2015. Lifespan and healthspan: past, present, and promise. *The Gerontologist,* 55**,** 901-911.
CRINI, G., LICHTFOUSE, E., CHANET, G. & MORIN-CRINI, N. 2020. Traditional and new applications of hemp. *Sustainable agriculture reviews 42: hemp production and applications***,** 37-87.
CRISTINO, L., BISOGNO, T. & DI MARZO, V. 2020a. Cannabinoids and the expanded endocannabinoid system in neurological disorders. *Nat Rev Neurol,* 16**,** 9-29.
CRISTINO, L., BISOGNO, T. & DI MARZO, V. 2020b. Cannabinoids and the expanded endocannabinoid system in neurological disorders. *Nature Reviews Neurology,* 16**,** 9-29.
CROCQ, M.-A. 2020a. History of cannabis and the endocannabinoid system. *Dialogues in clinical neuroscience,* 22**,** 223-228.
CROCQ, M. A. 2020b. History of cannabis and the endocannabinoid system. *Dialogues Clin Neurosci,* 22**,** 223-228.
CUI, Y., THOMAS-HALL, S. R. & SCHENK, P. M. 2019. Phaeodactylum tricornutum microalgae as a rich source of omega-3 oil: Progress in lipid induction techniques towards industry adoption. *Food chemistry,* 297**,** 124937.
CUI, Y., THOMAS-HALL, S. R., CHUA, E. T. & SCHENK, P. M. 2021. Development of high-level omega-3 eicosapentaenoic acid (EPA) production from Phaeodactylum tricornutum. *Journal of Phycology,* 57**,** 258-268.
CUI, Y. T.-H., S.R.; CHUA, E.T.; SCHENK, P.M. 2021. Development of high-level omega-3 eicosapentaenoic acid (EPA) production from Phaeodactylum tricornutum. *Journal of Phycology*.
D'ADAMO, S., SCHIANO DI VISCONTE, G., LOWE, G., SZAUB-NEWTON, J., BEACHAM, T., LANDELS, A., ALLEN, M. J., SPICER, A. & MATTHIJS, M. 2019. Engineering the unicellular alga Phaeodactylum tricornutum for high-value plant triterpenoid production. *Plant biotechnology journal,* 17**,** 75-87.

DAI, L. N., T.; LUO, R.; ZHANG, L.; ZHANG, S.; KANG, Y.; CHI, J.; FENG, X.; SHI, J.; TIAN, Y.; GAO, B.; LI, Z. 2024. Improvement of cannabidiolic acid synthetase activity through molecular docking and site-directed mutagenesis. *Industrial Crops and Products*.

DATWYLER, S. L. & WEIBLEN, G. D. 2006. Genetic variation in hemp and marijuana (Cannabis sativa L.) according to amplified fragment length polymorphisms. *Journal of Forensic Sciences,* 51**,** 371-375.

DAVIS, C. C. & CHOISY, P. 2024. Medicinal plants meet modern biodiversity science. *Current Biology,* 34**,** R158-R173.

DE A. LEITE, J., DE OLIVEIRA, M. V. L., CONTI, R., DE, S. B. W., ROSA, T. R., FILGUEIRAS, P. R., LACERDA, V., JR., ROMAO, W. & NETO, A. C. 2018. Extraction and isolation of cannabinoids from marijuana seizures and characterization by (1)H NMR allied to chemometric tools. *Sci Justice,* 58**,** 355-365.

DEARDEN, J. & FORBES, W. 1959. Light absorption studies: Part XIV. The ultraviolet absorption spectra of phenols. *Canadian Journal of Chemistry,* 37**,** 1294-1304.

DEGENHARDT, F., STEHLE, F. & KAYSER, O. 2017. The biosynthesis of cannabinoids. *Handbook of Cannabis and related pathologies.* Elsevier.

DEL MONDO, A., SANSONE, C. & BRUNET, C. 2022. Insights into the biosynthesis pathway of phenolic compounds in microalgae. *Computational and Structural Biotechnology Journal,* 20**,** 1901-1913.

DEMAIN, A. L. & SANCHEZ, S. 2009. Microbial drug discovery: 80 years of progress. *The Journal of antibiotics,* 62**,** 5-16.

DEMIRBAS, A. 2009. Progress and recent trends in biodiesel fuels. *Energy conversion and management,* 50**,** 14-34.

DESAULNIERS BROUSSEAU, V., WU, B. S., MACPHERSON, S., MORELLO, V. & LEFSRUD, M. 2021. Cannabinoids and Terpenes: How Production of Photo-Protectants Can Be Manipulated to Enhance Cannabis sativa L. Phytochemistry. *Front Plant Sci,* 12**,** 620021.

DESBOROUGH, M. J. & KEELING, D. M. 2017. The aspirin story–from willow to wonder drug. *British journal of haematology,* 177**,** 674-683.

DEVANE, W. A., HANUŠ, L., BREUER, A., PERTWEE, R. G., STEVENSON, L. A., GRIFFIN, G., GIBSON, D., MANDELBAUM, A., ETINGER, A. & MECHOULAM, R. 1992. Isolation and structure of a brain constituent that binds to the cannabinoid receptor. *Science,* 258**,** 1946-1949.

DEVINSKY, O., CROSS, J. H., LAUX, L., MARSH, E., MILLER, I., NABBOUT, R., SCHEFFER, I. E., THIELE, E. A., WRIGHT, S. & CANNABIDIOL IN DRAVET SYNDROME STUDY, G. 2017. Trial of Cannabidiol for Drug-Resistant Seizures in the Dravet Syndrome. *N Engl J Med,* 376**,** 2011-2020.

DEVINSKY, O., MARSH, E. & FRIEDMAN, D. 2016. Cannabidiol in patients with treatment-resistant epilepsy - Authors' reply. *Lancet Neurol,* 15**,** 545-6.

DHAOUADI, F., AWWAD, F., DIAMOND, A. & DESGAGNÉ-PENIX, I. 2020. Diatoms' breakthroughs in biotechnology: Phaeodactylum tricornutum as a model for producing high-added value molecules. *American Journal of Plant Sciences,* 11**,** 1632.

DIAMOND, A., DIAZ-GARZA, A. M., LI, J., SLATTERY, S. S., MERINDOL, N., FANTINO, E., MEDDEB-MOUELHI, F., KARAS, B. J., BARNAB, S. & DESGAGN, I. 2023a. Instability of extrachromosomal DNA transformed into the diatom Phaeodactylum tricornutum.

DIAMOND, A., DIAZ-GARZA, A. M., LI, J., SLATTERY, S. S., MERINDOL, N., FANTINO, E., MEDDEB-MOUELHI, F., KARAS, B. J., BARNABÉ, S. & DESGAGNÉ-PENIX, I. 2023b. Instability of extrachromosomal DNA transformed into the diatom Phaeodactylum tricornutum. *Algal Research,* 70**,** 102998.

DIAZ-GARZA, A. M., MERINDOL, N., DOS SANTOS, K. C. G., LAVOIE-MARCHAND, F., INGALLS, B. & DESGAGNE-PENIX, I. 2024. No two clones are alike: characterization of

heterologous subpopulations in a transgenic cell line of the model diatom Phaeodactylum tricornutum. *Microb Cell Fact,* 23**,** 286.
DJAOUDENE, O., ROMANO, A., BRADAI, Y., ZEBIRI, F., OUCHENE, A., YOUSFI, Y., AMRANE-ABIDER, M., SAHRAOUI-REMINI, Y. & MADANI, K. 2023. A global overview of dietary supplements: regulation, market trends, usage during the COVID-19 pandemic, and health effects. Nutrients. 2023; 15 (15): 3320.
DONG, M., LI, J., YANG, D., LI, M. & WEI, J. 2023. Biosynthesis and pharmacological activities of flavonoids, Triterpene Saponins and polysaccharides derived from Astragalus membranaceus. *Molecules,* 28**,** 5018.
DUARTE, B., FEIJÃO, E., CRUZ DE CARVALHO, R., DUARTE, I. A., MARQUES, A. P., MAIA, M., HERTZOG, J., MATOS, A. R., CABRITA, M. T. & CAÇADOR, I. 2022. Untargeted metabolomics reveals antidepressant effects in a marine photosynthetic organism: The diatom Phaeodactylum tricornutum as a case study. *Biology,* 11**,** 1770.
DUMONTIER, R., LOUTELIER-BOURHIS, C., WALET-BALIEU, M.-L., BUREL, C., MARECK, A., AFONSO, C., LEROUGE, P. & BARDOR, M. 2021. Identification of N-glycan oligomannoside isomers in the diatom Phaeodactylum tricornutum. *Carbohydrate Polymers,* 259**,** 117660.
DUNN, J. D., BOSMANI, C., BARISCH, C., RAYKOV, L., LEFRANÇOIS, L. H., CARDENAL-MUÑOZ, E., LÓPEZ-JIMÉNEZ, A. T. & SOLDATI, T. 2018. Eat prey, live: Dictyostelium discoideum as a model for cell-autonomous defenses. *Frontiers in immunology,* 8**,** 1906.
EICHHORN BILODEAU, S., WU, B.-S., RUFYIKIRI, A.-S., MACPHERSON, S. & LEFSRUD, M. 2019. An update on plant photobiology and implications for cannabis production. *Frontiers in plant science,* 10**,** 296.
EINHAUS, A., BAIER, T. & KRUSE, O. 2024. Molecular design of microalgae as sustainable cell factories. *Trends Biotechnol,* 42**,** 728-738.
ELSHOBARY, M. E., ABO-SHANAB, W. A., ENDE, S. S., ALQURAISHI, M. & EL-SHENODY, R. A. 2025. Optimizing Phaeodactylum tricornutum cultivation: integrated strategies for enhancing biomass, lipid, and fucoxanthin production. *Biotechnology for Biofuels and Bioproducts,* 18**,** 7.
ELSO, C. M., ROBERTS, L. J., SMYTH, G. K., THOMSON, R. J., BALDWIN, T. M., FOOTE, S. J. & HANDMAN, E. 2004. Leishmaniasis host response loci (lmr1-3) modify disease severity through a Th1/Th2-independent pathway. *Genes Immun,* 5**,** 93-100.
ERDENE-OCHIR, E., SHIN, B.-K., KWON, B. & JUNG, C. 2019a. Identification and characterisation of the novel endogenous promoter HASP1 and its signal peptide from Phaeodactylum tricornutum. *Scientific Reports.* Springer US.
ERDENE-OCHIR, E., SHIN, B.-K., KWON, B., JUNG, C. & PAN, C.-H. 2019b. Identification and characterisation of the novel endogenous promoter HASP1 and its signal peptide from Phaeodactylum tricornutum. *Scientific reports,* 9**,** 9941.
ERDENE-OCHIR, E., SHIN, B. K., HUDA, M. N., KIM, D. H., LEE, E. H., SONG, D. G., KIM, Y. M., KIM, S. M. & PAN, C. H. 2016. Cloning of a novel endogenous promoter for foreign gene expression in Phaeodactylum tricornutum. *Applied Biological Chemistry.* Springer Netherlands.
ESCHERICH, T. 1886. *Die darmbakterien des säuglings und ihre beziehungen zur physiologie der Verdauung*, Enke.
ESPE, S. 2018. MalaCards: the human disease database. *Journal of the Medical Library Association: JMLA,* 106**,** 140.
EVERYCURE. 2024. *The Problem* [Online]. Available: https://everycure.org/the-problem/#:~:text=Today%2C%20less%20than%2022%25%20of,and%20many%20forms%20of%20cancer. [Accessed 13/02 2025].
FABRIS, M., GEORGE, J., KUZHIUMPARAMBIL, U., LAWSON, C. A., JARAMILLO-MADRID, A. C., ABBRIANO, R. M., VICKERS, C. E. & RALPH, P. 2020a. Extrachromosomal Genetic

Engineering of the Marine Diatom Phaeodactylum tricornutum Enables the Heterologous Production of Monoterpenoids. *ACS Synthetic Biology.*

FABRIS, M., GEORGE, J., KUZHIUMPARAMBIL, U., LAWSON, C. A., JARAMILLO-MADRID, A. C., ABBRIANO, R. M., VICKERS, C. E. & RALPH, P. 2020b. Extrachromosomal genetic engineering of the marine diatom Phaeodactylum tricornutum enables the heterologous production of monoterpenoids. *ACS synthetic biology,* 9**,** 598-612.

FAHNERT, B., LILIE, H. & NEUBAUER, P. 2004. Inclusion bodies: formation and utilisation. *Physiological Stress Responses in Bioprocesses: -/-***,** 93-142.

FAJARDO, A. R., CERDÁN, L. E., MEDINA, A. R., FERNÁNDEZ, F. G. A., MORENO, P. A. G. & GRIMA, E. M. 2007. Lipid extraction from the microalga Phaeodactylum tricornutum. *European Journal of Lipid Science and Technology,* 109**,** 120-126.

FANG, H., LI, D., KANG, J., JIANG, P., SUN, J. & ZHANG, D. 2018. Metabolic engineering of Escherichia coli for de novo biosynthesis of vitamin B12. *Nature communications,* 9**,** 4917.

FANTINO, E., AWWAD, F., MERINDOL, N., GARZA, A. M. D., GÉLINAS, S.-E., ROBLES, G. C. G., CUSTEAU, A., MEDDEB-MOUELHI, F. & DESGAGNÉ-PENIX, I. 2024a. Bioengineering Phaeodactylum tricornutum, a marine diatom, for cannabinoid biosynthesis. *Algal Research,* 77**,** 103379.

FANTINO, E., MESSAABI, A., MÉRINDOL, N., AWWAD, F., SENE, N., GÉLINAS, S.-E., CUSTEAU, A., RHÉAUME, K.-L., MEDDEB-MOUELHI, F. & DESGAGNÉ-PENIX, I. 2024b. Extrachromosomal expression of functional Cannabis sativa cannabidiolic acid synthase in Phaedodactylum tricornutum. *Algal Research***,** 103889.

FANTINO, E. A., F.; MERINDOL, N.; DIAZ GARZA, A.M.; GELINAS, S.E; GAJON ROBLES, G.C.; CUSTEAU, A.; MEDDEB-MOUELHI, F.; DESGAGNE-PENIX, I. 2024. Bioengineering Phaeodactylum tricornutum, a marine diatom, for cannabinoid biosynthesis. *Algal Research*.

FARAG, S. & KAYSER, O. 2017. The cannabis plant: Botanical aspects. *Handbook of cannabis and related pathologies.* Elsevier.

FELLERMEIER, M., EISENREICH, W., BACHER, A. & ZENK, M. H. 2001. Biosynthesis of cannabinoids: incorporation experiments with 13C-labeled glucoses. *European Journal of Biochemistry,* 268**,** 1596-1604.

FELLERMEIER, M. & ZENK, M. H. 1998. Prenylation of olivetolate by a hemp transferase yields cannabigerolic acid, the precursor of tetrahydrocannabinol. *FEBS letters,* 427**,** 283-285.

FERNANDES, T. & CORDEIRO, N. 2021. Microalgae as sustainable biofactories to produce high-value lipids: biodiversity, exploitation, and biotechnological applications. *Marine drugs,* 19**,** 573.

FERRARA, M. S. 2021. Peak-experience and the entheogenic use of cannabis in world religions. *Journal of Psychedelic Studies,* 4**,** 179-191.

FIELD, C. B., BEHRENFELD, M. J., RANDERSON, J. T. & FALKOWSKI, P. 1998. Primary production of the biosphere: integrating terrestrial and oceanic components. *science,* 281**,** 237-240.

FISHER, P. R. 2001. Genetic analysis of phototaxis in Dictyostelium. *Comprehensive Series in Photosciences.* Elsevier.

FLEMING, A. 1944. The discovery of penicillin. *British Medical Bulletin,* 2**,** 4-5.

FLOCKHART, D. T., PICHANCOURT, J. B., NORRIS, D. R. & MARTIN, T. G. 2015. Unravelling the annual cycle in a migratory animal: breeding-season habitat loss drives population declines of monarch butterflies. *Journal of Animal Ecology,* 84**,** 155-165.

FORTUNATO, A., STRASSMANN, J., SANTORELLI, L. & QUELLER, D. 2003. Co-occurrence in nature of different clones of the social amoeba, Dictyostelium discoideum. *Molecular ecology,* 12**,** 1031-1038.

GAERTNER, E. E. 1979. The history and use of milkweed (Asclepias syriaca L.). *Economic Botany,* 33**,** 119-123.

GAGNE, S. J., STOUT, J. M., LIU, E., BOUBAKIR, Z., CLARK, S. M. & PAGE, J. E. 2012. Identification of olivetolic acid cyclase from Cannabis sativa reveals a unique catalytic route to plant polyketides. *Proceedings of the National Academy of Sciences,* 109**,** 12811-12816.

GALAS, L., BUREL, C., SCHAPMAN, D., ROPITAUX, M., BERNARD, S., BÉNARD, M. & BARDOR, M. 2021. Comparative structural and functional analyses of the fusiform, oval, and triradiate morphotypes of Phaeodactylum tricornutum Pt3 strain. *Frontiers in Plant Science,* 12**,** 638181.

GALLWITZ, M., ENOKSSON, M., THORPE, M. & HELLMAN, L. 2012. The extended cleavage specificity of human thrombin. *PloS one,* 7**,** e31756.

GAMMATANTRAWET, N., NGUYỄN, C. T., SUSAWAENGSUP, C., RAMLI, A. N. M., TONGKOOM, K., CHATSUNGNOEN, T., DANGTUNGEE, R. & BHUYAR, P. 2024. Phytochemistry of medicinal herbs belongs to asclepiadaceae family for therapeutic applications: A critical review. *Molecular Biotechnology***,** 1-25.

GAO, J., JIANG, L. & LIAN, J. 2021. Development of synthetic biology tools to engineer Pichia pastoris as a chassis for the production of natural products. *Synthetic and systems biotechnology,* 6**,** 110-119.

GARRETT, E. R. & HUNT, C. A. 1974. Physicochemical properties, solubility, and protein binding of Δ9-tetrahydrocannabinol. *Journal of Pharmaceutical Sciences,* 63**,** 1056-1064.

GEISSLER, M., VOLK, J., STEHLE, F., KAYSER, O. & WARZECHA, H. 2018. Subcellular localization defines modification and production of D 9 -tetrahydrocannabinolic acid synthase in transiently transformed Nicotiana benthamiana. *Biotechnology Letters.* Springer Netherlands.

GEORGE, J., KAHLKE, T., ABBRIANO, R. M., KUZHIUMPARAMBIL, U., RALPH, P. J. & FABRIS, M. 2020a. Metabolic Engineering Strategies in Diatoms Reveal Unique Phenotypes and Genetic Configurations With Implications for Algal Genetics and Synthetic Biology. *Frontiers in Bioengineering and Biotechnology.*

GEORGE, J., KAHLKE, T., ABBRIANO, R. M., KUZHIUMPARAMBIL, U., RALPH, P. J. & FABRIS, M. 2020b. Metabolic engineering strategies in diatoms reveal unique phenotypes and genetic configurations with implications for algal genetics and synthetic biology. *Frontiers in Bioengineering and Biotechnology,* 8**,** 513.

GERTSCH, J., PERTWEE, R. G. & DI MARZO, V. 2010. Phytocannabinoids beyond the Cannabis plant–do they exist? *British journal of pharmacology,* 160**,** 523-529.

GHOSH, R., CHHABRA, A., PHATALE, P. A., SAMRAT, S. K., SHARMA, J., GOSAIN, A., MOHANTY, D., SARAN, S. & GOKHALE, R. S. 2008. Dissecting the functional role of polyketide synthases in Dictyostelium discoideum: biosynthesis of the differentiation regulating factor 4-methyl-5-pentylbenzene-1, 3-diol. *Journal of Biological Chemistry,* 283**,** 11348-11354.

GOEDHEER, J. & SWART, J. 1975. Decay of delayed light with the diatom Phaeodactylum tricornutum. *Plant Science Letters,* 4**,** 335-341.

GORZYNIK-DEBICKA, M., PRZYCHODZEN, P., CAPPELLO, F., KUBAN-JANKOWSKA, A., MARINO GAMMAZZA, A., KNAP, N., WOZNIAK, M. & GORSKA-PONIKOWSKA, M. 2018. Potential health benefits of olive oil and plant polyphenols. *International journal of molecular sciences,* 19**,** 686.

GOVINDARAJAN, R. K., MISHRA, A. K., CHO, K.-H., KIM, K.-H., YOON, K. M. & BAEK, K.-H. 2023. Biosynthesis of phytocannabinoids and structural insights: a review. *Metabolites,* 13**,** 442.

GREWAL, P. S., MODAVI, C., RUSS, Z. N., HARRIS, N. C. & DUEBER, J. E. 2018. Bioproduction of a betalain color palette in Saccharomyces cerevisiae. *Metabolic Engineering,* 45**,** 180-188.

GRUBER, A. & OBORNÍK, M. 2024. Evolution of plastids and mitochondria in diatoms. *Diatom Photosynthesis: From Primary Production to High-Value Molecules***,** 81-111.

GU, W., KAVANAGH, J. M. & MCCLURE, D. D. 2022a. A scalable model for EPA and fatty acid production by Phaeodactylum tricornutum. *Frontiers in Bioengineering and Biotechnology,* 10**,** 1011570.

GU, W., KAVANAGH, J. M. & MCCLURE, D. D. 2022b. Towards a sustainable supply of omega-3 fatty acids: Screening microalgae for scalable production of eicosapentaenoic acid (EPA). *Algal Research,* 61**,** 102564.

GÜLCK, T., BOOTH, J. K., CARVALHO, KHAKIMOV, B., CROCOLL, C., MOTAWIA, M. S., MØLLER, B. L., BOHLMANN, J. & GALLAGE, N. J. 2020. Synthetic Biology of Cannabinoids and Cannabinoid Glucosides in Nicotiana benthamiana and Saccharomyces cerevisiae. *Journal of Natural Products.*

GÜLCK, T. & MØLLER, B. L. 2020. Phytocannabinoids: origins and biosynthesis. *Trends in plant science,* 25**,** 985-1004.

GÜNTHER, M., REIMER, C., HERBST, R., KUFS, J. E., RAUTSCHEK, J., UEBERSCHAAR, N., ZHANG, S., PESCHEL, G., REIMER, L. & REGESTEIN, L. 2022. Yellow polyketide pigment suppresses premature hatching in social amoeba. *Proceedings of the National Academy of Sciences,* 119**,** e2116122119.

GUPTA, A. K., JAIN, A., ROY, P. & SINGH, R. 2020. Pharmacological evaluation of Cannabis indica for their aphrodisiac potential. *Int. J. Ayurvedic Med,* 11**,** 399.

GUSTAVSSON, A., NORTON, N., FAST, T., FRÖLICH, L., GEORGES, J., HOLZAPFEL, D., KIRABALI, T., KROLAK-SALMON, P., ROSSINI, P. M. & FERRETTI, M. T. 2023. Global estimates on the number of persons across the Alzheimer's disease continuum. *Alzheimer's & Dementia,* 19**,** 658-670.

HACKER, J. & BLUM-OEHLER, G. 2007. In appreciation of theodor escherich. *Nature Reviews Microbiology,* 5**,** 902-902.

HALL, S., MCDERMOTT, C., ANOOPKUMAR-DUKIE, S., MCFARLAND, A. J., FORBES, A., PERKINS, A. V., DAVEY, A. K., CHESS-WILLIAMS, R., KIEFEL, M. J. & ARORA, D. 2016. Cellular effects of pyocyanin, a secreted virulence factor of Pseudomonas aeruginosa. *Toxins,* 8**,** 236.

HALPIN-MCCORMICK, A., MAAZ, T. M., KANTAR, M. B., BARTON, K. E., MASALIA, R. R., BATORA, N., LAW, K. & KUNTZ, E. J. 2024. Species Distribution of Cannabis sativa: Past, Present and future. *bioRxiv***,** 2024.06. 11.598429.

HAMILTON, M. L., HASLAM, R. P., NAPIER, J. A. & SAYANOVA, O. 2014. Metabolic engineering of Phaeodactylum tricornutum for the enhanced accumulation of omega-3 long chain polyunsaturated fatty acids. *Metabolic engineering,* 22**,** 3-9.

HAMMOND, D., GOODMAN, S., WADSWORTH, E., FREEMAN, T. P., KILMER, B., SCHAUER, G., PACULA, R. L. & HALL, W. 2022. Trends in the use of cannabis products in Canada and the USA, 2018–2020: Findings from the International Cannabis Policy Study. *International Journal of Drug Policy,* 105**,** 103716.

HARA, K. Y., ARAKI, M., OKAI, N., WAKAI, S., HASUNUMA, T. & KONDO, A. 2014. Development of bio-based fine chemical production through synthetic bioengineering. *Microbial cell factories,* 13**,** 1-19.

HASHIESH, H. M., JHA, N. K., THOMAS, S. D., SADEK, B., KUMAR, S., ADEGHATE, E. & OJHA, S. 2023. Targeting the CB (2) receptor to delay progression of cardiovascular diseases. *Medicinal Usage of Cannabis and Cannabinoids.* Elsevier.

HAZEKAMP, A., CHOI, Y. H. & VERPOORTE, R. 2004. Quantitative analysis of cannabinoids from Cannabis sativa using 1H-NMR. *Chemical and pharmaceutical bulletin,* 52**,** 718-721.
HEMPEL, F., LAU, J., KLINGL, A. & MAIER, U. G. 2011. Algae as protein factories: expression of a human antibody and the respective antigen in the diatom Phaeodactylum tricornutum. *PloS one,* 6**,** e28424.
HENDERSON, A. 2020. *Characterising Chinese Hamster Ovary Cell Line Stability in Bioproduction of Therapeutic Proteins.* UCL (University College London).
HERKENHAM, M., LYNN, A. B., LITTLE, M. D., JOHNSON, M. R., MELVIN, L. S., DE COSTA, B. R. & RICE, K. C. 1990. Cannabinoid receptor localization in brain. *Proceedings of the national Academy of sciences,* 87**,** 1932-1936.
HERRING, A. C., KOH, W. S. & KAMINSKI, N. E. 1998. Inhibition of the cyclic AMP signaling cascade and nuclear factor binding to CRE and κB elements by cannabinol, a minimally CNS-active cannabinoid. *Biochemical pharmacology,* 55**,** 1013-1023.
HEYDARIZADEH, P., BOUREBA, W., ZAHEDI, M., HUANG, B., MOREAU, B., LUKOMSKA, E., COUZINET-MOSSION, A., WIELGOSZ-COLLIN, G., MARTIN-JÉZÉQUEL, V. & BOUGARAN, G. 2017. Response of CO2-starved diatom Phaeodactylum tricornutum to light intensity transition. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 372**,** 20160396.
HILDEBRAND, M., DAVIS, A. K., SMITH, S. R., TRALLER, J. C. & ABBRIANO, R. 2012. The place of diatoms in the biofuels industry. *Biofuels,* 3**,** 221-240.
HILGER, D., MASUREEL, M. & KOBILKA, B. K. 2018. Structure and dynamics of GPCR signaling complexes. *Nature structural & molecular biology,* 25**,** 4-12.
HILLIG, K. W. 2005. Genetic evidence for speciation in Cannabis (Cannabaceae). *Genetic Resources and Crop Evolution,* 52**,** 161-180.
HOFFMANN, A. A. & HERCUS, M. J. 2000. Environmental stress as an evolutionary force. *Bioscience,* 50**,** 217-226.
HOLLISTER, L. E. 1986. Health aspects of cannabis. *Pharmacological reviews,* 38**,** 1-20.
HOU, Y. & WU, G. 2018. Nutritionally essential amino acids. *Advances in Nutrition,* 9**,** 849-851.
HOWLETT, A., QUALY, J. M. & KHACHATRIAN, L. L. 1986. Involvement of Gi in the inhibition of adenylate cyclase by cannabimimetic drugs. *Molecular pharmacology,* 29**,** 307-313.
HOWLETT, A. C. & ABOOD, M. E. 2017. CB1 and CB2 receptor pharmacology. *Advances in Pharmacology,* 80**,** 169-206.
HOWLETT, A. C., BARTH, F., BONNER, T., CABRAL, G., CASELLAS, P., DEVANE, W., FELDER, C., HERKENHAM, M., MACKIE, K. & MARTIN, B. 2002. International Union of Pharmacology. XXVII. Classification of cannabinoid receptors. *Pharmacological reviews,* 54**,** 161-202.
HU, J. M., W.; SU, Y.; QIAN, C.; FU, W. 2023. Emerging technologies for advancing microalgal photosynthesis and metabolism toward sustainable production. *Frontiers*.
HUANG, B., GUO, J., YI, B., YU, X., SUN, L. & CHEN, W. 2008. Heterologous production of secondary metabolites as pharmaceuticals in Saccharomyces cerevisiae. *Biotechnology letters,* 30**,** 1121-1137.
HUG, L. A., BAKER, B. J., ANANTHARAMAN, K., BROWN, C. T., PROBST, A. J., CASTELLE, C. J., BUTTERFIELD, C. N., HERNSDORF, A. W., AMANO, Y. & ISE, K. 2016. A new view of the tree of life. *Nature microbiology,* 1**,** 1-6.
HUNT, J. M. 2010. Shiga toxin–producing Escherichia coli (STEC). *Clinics in laboratory medicine,* 30**,** 21-45.
INSIGHT, F. B. 2024. *Cannabis Market Size, Share & COVID-19 Impact Analysis, By Type (Flowers/Buds and Concentrates), By Application (Medical, Recreational (Edibles and Topicals), and Industrial Hemp) By Component (THC-Dominant, Balanced THC & CBD,*

and CBD Dominant), and Regional Forecast, 2023-2030 [Online]. Available: https://www.fortunebusinessinsights.com/industry-reports/cannabis-marijuana-market-100219 [Accessed 02-10-2024 2024].

JABŁOŃSKA, J., AUGUSTYNIAK, A., DUBROWSKA, K. & RAKOCZY, R. 2023. The two faces of pyocyanin-why and how to steer its production? *World Journal of Microbiology and Biotechnology,* 39**,** 103.

JARAMILLO-MADRID, A. C., ABBRIANO, R., ASHWORTH, J., FABRIS, M., PERNICE, M. & RALPH, P. J. 2020. Overexpression of key sterol pathway enzymes in two model marine diatoms alters sterol profiles in phaeodactylum tricornutum. *Pharmaceuticals.*

JENNY, R. J., MANN, K. G. & LUNDBLAD, R. L. 2003. A critical review of the methods for cleavage of fusion proteins with thrombin and factor Xa. *Protein Expr Purif,* 31**,** 1-11.

KANEHISA, M., FURUMICHI, M., SATO, Y., KAWASHIMA, M. & ISHIGURO-WATANABE, M. 2023. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic acids research,* 51**,** D587-D592.

KARAS, B., DINER, R., LEFEBVRE, S., MCQUAID, J., PHILLIPS, A., NODDINGS, C., BRUNSON, J., VALAS, R., DEERINCK, T. & JABLANOVIC, J. 2015a. Designer diatom episomes delivered by bacterial conjugation. Nat Commun 6: 6925.

KARAS, B. J., DINER, R. E., LEFEBVRE, S. C., MCQUAID, J., PHILLIPS, A. P., NODDINGS, C. M., BRUNSON, J. K., VALAS, R. E., DEERINCK, T. J. & JABLANOVIC, J. 2015b. Designer diatom episomes delivered by bacterial conjugation. *Nature communications,* 6**,** 6925.

KARAS, B. J., DINER, R. E., LEFEBVRE, S. C., MCQUAID, J., PHILLIPS, A. P. R., NODDINGS, C. M., BRUNSON, J. K., VALAS, R. E., DEERINCK, T. J., JABLANOVIC, J., GILLARD, J. T. F., BEERI, K., ELLISMAN, M. H., GLASS, J. I., HUTCHISON, C. A., SMITH, H. O., VENTER, J. C., ALLEN, A. E., DUPONT, C. L. & WEYMAN, P. D. 2015c. Designer diatom episomes delivered by bacterial conjugation. *Nature Communications.* Nature Publishing Group.

KARAS, B. J., DINER, R. E., LEFEBVRE, S. C., MCQUAID, J., PHILLIPS, A. P. R., NODDINGS, C. M., BRUNSON, J. K., VALAS, R. E., DEERINCK, T. J., JABLANOVIC, J., GILLARD, J. T. F., BEERI, K., ELLISMAN, M. H., GLASS, J. I., HUTCHISON III, C. A., SMITH, H. O., VENTER, J. C., ALLEN, A. E., DUPONT, C. L. & WEYMAN, P. D. 2015d. Designer diatom episomes delivered by bacterial conjugation. *Nature Communications,* 6**,** 6925.

KARAS, B. J., MOLPARIA, B., JABLANOVIC, J., HERMANN, W. J., LIN, Y.-C., DUPONT, C. L., TAGWERKER, C., YONEMOTO, I. T., NOSKOV, V. N., CHUANG, R.-Y., ALLEN, A. E., GLASS, J. I., HUTCHISON, C. A., SMITH, H. O., VENTER, J. C. & WEYMAN, P. D. 2013. Assembly of eukaryotic algal chromosomes in yeast. *Journal of Biological Engineering,* 7**,** 30.

KASSAW, T. K., PATON, A. J. & PEERS, G. 2022. Episome-based gene expression modulation platform in the model diatom Phaeodactylum tricornutum. *ACS Synthetic Biology,* 11**,** 191-204.

KATSAROS, C. & HEIMANN, K. 2012. *Advances in Algal Cell Biology*, De Gruyter.

KATSUYAMA, Y. & OHNISHI, Y. 2012. Type III polyketide synthases in microorganisms. *Methods in enzymology.* Elsevier.

KEARSEY, L. J., PRANDI, N., KARUPPIAH, V., YAN, C., LEYS, D., TOOGOOD, H., TAKANO, E. & SCRUTTON, N. S. 2019. Structure of the Cannabis sativa olivetol-producing enzyme reveals cyclization plasticity in type III polyketide synthases. *FEBS Journal.*

KEARSEY, L. J., PRANDI, N., KARUPPIAH, V., YAN, C., LEYS, D., TOOGOOD, H., TAKANO, E. & SCRUTTON, N. S. 2020. Structure of the Cannabis sativa olivetol-producing enzyme reveals cyclization plasticity in type III polyketide synthases. *The FEBS Journal,* 287**,** 1511-1524.

KEARSEY, L. J., YAN, C., PRANDI, N., TOOGOOD, H. S., TAKANO, E. & SCRUTTON, N. S. 2023. Biosynthesis of cannabigerol and cannabigerolic acid: the gateways to further cannabinoid production. *Synthetic Biology,* 8**,** ysad010.

KELLY, P. S., MIGUEZ, A. A., ALVES, C. & BARRON, N. 2018. From media to mitochondria–rewiring cellular energy metabolism of Chinese hamster ovary cells for the enhanced production of biopharmaceuticals. *Current opinion in chemical engineering,* 22**,** 71-80.

KHANNA-CHOPRA, R., SEMWAL, V. K., LAKRA, N. & PAREEK, A. 2019. Proline–A key regulator conferring plant tolerance to salinity and drought. *Plant tolerance to environmental stress.* CRC Press.

KILIAN, O. & KROTH, P. G. 2005. Identification and characterization of a new conserved motif within the presequence of proteins targeted into complex diatom plastids. *The Plant journal : for cell and molecular biology.*

KOPPOLU, V. & VASIGALA, V. K. 2016. Role of Escherichia coli in biofuel production. *Microbiology insights,* 9**,** MBI. S10878.

KOSALKOVÁ, K. B., C.; SÁNCHEZ-OREJAS, I.C.; CUETO, L.; GARCÍA-ESTRADA, C. 2023. Biotechnological Fungal Platforms for the Production of Biosynthetic Cannabinoids. *Journal of Fungi*.

KUMAR, G., SHEKH, A., JAKHU, S., SHARMA, Y., KAPOOR, R. & SHARMA, T. R. 2020. Bioengineering of microalgae: recent advances, perspectives, and regulatory challenges for industrial application. *Frontiers in Bioengineering and Biotechnology,* 8**,** 914.

KUMAR, M., DAHUJA, A., TIWARI, S., PUNIA, S., TAK, Y., AMAROWICZ, R., BHOITE, A. G., SINGH, S., JOSHI, S. & PANESAR, P. S. 2021. Recent trends in extraction of plant bioactives using green technologies: A review. *Food Chemistry,* 353**,** 129431.

LABAN, A. 2021. Cannabinoid production in algae. Google Patents.

LAI, Z., TSUGAWA, H., WOHLGEMUTH, G., MEHTA, S., MUELLER, M., ZHENG, Y., OGIWARA, A., MEISSEN, J., SHOWALTER, M., TAKEUCHI, K., KIND, T., BEAL, P., ARITA, M. & FIEHN, O. 2018. Identifying metabolites by integrating metabolome databases with mass spectrometry cheminformatics. *Nature Methods,* 15**,** 53-56.

LAJIS, A. F. B. 2020. Biomanufacturing process for the production of bacteriocins from Bacillaceae family. *Bioresources and Bioprocessing,* 7**,** 1-26.

LAM, M. K. & LEE, K. T. 2012. Microalgae biofuels: a critical review of issues, problems and the way forward. *Biotechnology advances,* 30**,** 673-690.

LANGE, K., SCHMID, A. & JULSING, M. K. 2015. Δ9-Tetrahydrocannabinolic acid synthase production in Pichia pastoris enables chemical synthesis of cannabinoids. *Journal of Biotechnology,* 211**,** 68-76.

LAUTERBACH, A. & OBER, G. 2000. Iodine and iodine compounds. *Kirk-Othmer Encyclopedia of Chemical Technology*.

LAZARJANI, M. P., YOUNG, O., KEBEDE, L. & SEYFODDIN, A. 2021. Processing and extraction methods of medicinal cannabis: A narrative review. *Journal of cannabis research,* 3**,** 1-15.

LE BOISSELIER, R., ALEXANDRE, J., LELONG-BOULOUARD, V. & DEBRUYNE, D. 2017. Focus on cannabinoids and synthetic cannabinoids. *Clinical Pharmacology & Therapeutics,* 101**,** 220-229.

LEBEAU, J., VENKATACHALAM, M., FOUILLAUD, M., PETIT, T., VINALE, F., DUFOSSÉ, L. & CARO, Y. 2017. Production and new extraction method of polyketide red pigments produced by ascomycetous fungi from terrestrial and marine habitats. *Journal of Fungi,* 3**,** 34.

LEWIS, M. M., YANG, Y., WASILEWSKI, E., CLARKE, H. A. & KOTRA, L. P. 2017. Chemical profiling of medical cannabis extracts. *ACS omega,* 2**,** 6091-6103.

LI, J., ZHANG, L. & LIU, W. 2018. Cell-free synthetic biology for in vitro biosynthesis of pharmaceutical natural products. *Synthetic and Systems Biotechnology,* 3**,** 83-89.
LI, Y., DENG, L., WALKER, E. J. L., KARAS, B. J. & MOCK, T. 2025. Genetic engineering in diatoms: advances and prospects. *The Plant Journal,* 121**,** e70102.
LIM, K. J. H., LIM, Y. P., HARTONO, Y. D., GO, M. K., FAN, H. & YEW, W. S. 2021. Biosynthesis of nature-inspired unnatural cannabinoids. *Molecules.*
LINGG, N., ZHANG, P., SONG, Z. & BARDOR, M. 2012. The sweet tooth of biopharmaceuticals: importance of recombinant protein glycosylation analysis. *Biotechnol J,* 7**,** 1462-72.
LIU, X., BIMEREW, M., MA, Y., MÜLLER, H., OVADIS, M., EBERL, L., BERG, G. & CHERNIN, L. 2007. Quorum-sensing signaling is required for production of the antibiotic pyrrolnitrin in a rhizospheric biocontrol strain of Serratia plymuthica. *FEMS microbiology letters,* 270**,** 299-305.
LIU, X., HEMPEL, F., STORK, S., BOLTE, K., MOOG, D., HEIMERL, T., MAIER, U. G. & ZAUNER, S. 2016. Addressing various compartments of the diatom model organism Phaeodactylum tricornutum via sub-cellular marker proteins. *Algal Research.* Elsevier B.V.
LLAVERO PASQUINA, M. 2020. *Molecular biology of B vitamin metabolism genes and their regulation in Phaeodactylum tricornutum.*
LOZANO TEROL, G., GALLEGO-JARA, J., SOLA MARTÍNEZ, R. A., CÁNOVAS DÍAZ, M. & DE DIEGO PUENTE, T. 2019. Engineering protein production by rationally choosing a carbon and nitrogen source using E. coli BL21 acetate metabolism knockout strains. *Microbial cell factories,* 18**,** 1-19.
LU, H.-C. & MACKIE, K. 2021. Review of the endocannabinoid system. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging,* 6**,** 607-615.
LUCA, S. V., BRAUMANN, L., GERIGK, M., FRANK, O. & MINCEVA, M. 2021. Separation of minor cannabinoids from hemp extract with trapping multiple dual mode liquid-liquid chromatography. *Journal of Chromatography A,* 1658**,** 462608.
LUND, E. K. 2013. Health benefits of seafood; is it just the fatty acids? *Food chemistry,* 140**,** 413-420.
LUO, X., REITER, M. A., D'ESPAUX, L., WONG, J., DENBY, C. M., LECHNER, A., ZHANG, Y., GRZYBOWSKI, A. T., HARTH, S. & LIN, W. 2019a. Complete biosynthesis of cannabinoids and their unnatural analogues in yeast. *Nature,* 567**,** 123-126.
LUO, X., REITER, M. A., D'ESPAUX, L., WONG, J., DENBY, C. M., LECHNER, A., ZHANG, Y., GRZYBOWSKI, A. T., HARTH, S., LIN, W., LEE, H., YU, C., SHIN, J., DENG, K., BENITES, V. T., WANG, G., BAIDOO, E. E. K., CHEN, Y., DEV, I., PETZOLD, C. J. & KEASLING, J. D. 2019b. Complete biosynthesis of cannabinoids and their unnatural analogues in yeast. *Nature.* Springer US.
MALFITANO, A. M., BASU, S., MARESZ, K., BIFULCO, M. & DITTEL, B. N. What we know and do not know about the cannabinoid receptor 2 (CB2). Seminars in immunology, 2014. Elsevier, 369-379.
MANN, N. J. 2018. A brief history of meat in the human diet and current health implications. *Meat science,* 144**,** 169-179.
MAREY, M. A., ABOZAHRA, R., EL-NIKHELY, N. A., KAMAL, M. F., ABDELHAMID, S. M. & EL-KHOLY, M. A. 2024. Transforming microbial pigment into therapeutic revelation: extraction and characterization of pyocyanin from Pseudomonas aeruginosa and its therapeutic potential as an antibacterial and anticancer agent. *Microbial Cell Factories,* 23**,** 174.
MATHIEU-RIVET, E., KIEFER-MEYER, M.-C., VANIER, G., OVIDE, C., BUREL, C., LEROUGE, P. & BARDOR, M. 2014. Protein N-glycosylation in eukaryotic microalgae and its impact on the production of nuclear expressed biopharmaceuticals. *Frontiers in Plant Science,* 5.

MAXWELL, K. L., MITTERMAIER, A. K., FORMAN-KAY, J. D. & DAVIDSON, A. R. 1999. A simple in vivo assay for increased protein solubility. *Protein Science,* 8**,** 1908-1911.
MCCOURT, R. 2016. Archaeplastida: diversification of red algae and the green plant lineage. *Encyclopedia of evolutionary biology.* Elsevier Inc.
MCFADDEN, G. I. 2001. Primary and secondary endosymbiosis and the origin of plastids. *Journal of Phycology,* 37**,** 951-959.
MCPARTLAND, J. M. 2017. Cannabis sativa and Cannabis indica versus "Sativa" and "Indica". *Cannabis sativa L.-botany and biotechnology***,** 101-121.
MCPARTLAND, J. M. & SMALL, E. 2020. A classification of endangered high-THC cannabis (Cannabis sativa subsp. indica) domesticates and their wild relatives. *PhytoKeys,* 144**,** 81.
MENDEZ-PEREZ, D., ALONSO-GUTIERREZ, J., HU, Q., MOLINAS, M., BAIDOO, E. E., WANG, G., CHAN, L. J., ADAMS, P. D., PETZOLD, C. J. & KEASLING, J. D. 2017. Production of jet fuel precursor monoterpenoids from engineered Escherichia coli. *Biotechnology and bioengineering,* 114**,** 1703-1712.
MESSAABI, A., MERINDOL, N., BOHNENBLUST, L., FANTINO, E., MEDDEB-MOUELHI, F. & DESGAGNE-PENIX, I. 2024a. In vivo thrombin activity in the diatom Phaeodactylum tricornutum: biotechnological insights. *Appl Microbiol Biotechnol,* 108**,** 481.
MESSAABI, A., MERINDOL, N., BOHNENBLUST, L., FANTINO, E., MEDDEB-MOUELHI, F. & DESGAGNÉ-PENIX, I. 2024b. In vivo thrombin activity in the diatom Phaeodactylum tricornutum: biotechnological insights. *Applied Microbiology and Biotechnology,* 108**,** 481.
MICALE, V., MAZZOLA, C. & DRAGO, F. 2007. Endocannabinoids and neurodegenerative diseases. *Pharmacological Research,* 56**,** 382-392.
MILLS, B., YEPES, A. & NUGENT, K. 2015. Synthetic cannabinoids. *The American journal of the medical sciences,* 350**,** 59-62.
MITAL, S., CHRISTIE, G. & DIKICIOGLU, D. 2021. Recombinant expression of insoluble enzymes in Escherichia coli: a systematic review of experimental design and its manufacturing implications. *Microbial cell factories,* 20**,** 1-20.
MOJZEŠ, P., GAO, L., ISMAGULOVA, T., PILÁTOVÁ, J., MOUDŘÍKOVÁ, Š., GORELOVA, O., SOLOVCHENKO, A., NEDBAL, L. & SALIH, A. 2020. Guanine, a high-capacity and rapid-turnover nitrogen reserve in microalgal cells. *Proceedings of the National Academy of Sciences,* 117**,** 32722-32730.
MONA, M. O. N. A. 2025. *Databank* [Online]. Available: https://mona.fiehnlab.ucdavis.edu/ [Accessed 2024].
MOOKERJEE, S., CAMPBELL, A. J., WILTSHIRE, Z. D. & CHEN, K. J. 2018. Method and cell line for production of phytocannabinoids and phytocannabinoid analogues in yeast. Google Patents.
MOOKERJEE, S., CAMPBELL, A. J., WILTSHIRE, Z. D. & CHEN, K. J. 2022. Method and cell line for production of phytocannabinoids and phytocannabinoid analogues in yeast. Google Patents.
MUHTASEB, M. & MUHTASEB, A. 2022. Caduceus and Asclepius: A Tale of two Rods. *Eye,* 36**,** 2226-2227.
MUSIO, S., MÜSSIG, J. & AMADUCCI, S. 2018. Optimizing hemp fiber production for high performance composite applications. *Frontiers in Plant Science,* 9**,** 1702.
MUTHURAJ, P. G., KRISHNAMOORTHY, C., ANDERSON-BERRY, A., HANSON, C. & NATARAJAN, S. K. 2022. Novel Therapeutic Nutrients Molecules That Protect against Zika Virus Infection with a Special Note on Palmitoleate. *Nutrients,* 15**,** 124.
NARITA, T. B., KOIDE, K., MORITA, N. & SAITO, T. 2011. Dictyostelium hybrid polyketide synthase, SteelyA, produces 4-methyl-5-pentylbenzene-1, 3-diol and induces spore maturation. *FEMS microbiology letters,* 319**,** 82-87.

NIELSEN, J. 2013. Production of biopharmaceutical proteins by yeast: advances through metabolic engineering. *Bioengineered,* 4**,** 207-211.
NIKITASHINA, V., STETTIN, D. & POHNERT, G. 2022. Metabolic adaptation of diatoms to hypersalinity. *Phytochemistry,* 201**,** 113267.
NOUEMSSI, S. B., GHRIBI, M., BEAUCHEMIN, R., MEDDEB-MOUELHI, F., GERMAIN, H. & DESGAGNÉ-PENIX, I. 2020. Rapid and efficient colony-PCR for high throughput screening of genetically transformed Chlamydomonas reinhardtii. *Life,* 10**,** 186.
NYYSSÖLÄ, A., KEROVUO, J., KAUKINEN, P., VON WEYMARN, N. & REINIKAINEN, T. 2000. Extreme halophiles synthesize betaine from glycine by methylation. *Journal of Biological Chemistry,* 275**,** 22196-22201.
OBEMBE, O. O., POPOOLA, J. O., LEELAVATHI, S. & REDDY, S. V. 2011. Advances in plant molecular farming. *Biotechnology advances,* 29**,** 210-222.
OETTL, S. K., GERSTMEIER, J., KHAN, S. Y., WIECHMANN, K., BAUER, J., ATANASOV, A. G., MALAINER, C., AWAD, E. M., UHRIN, P. & HEISS, E. H. 2013. Imbricaric acid and perlatolic acid: Multi-targeting anti-inflammatory depsides from Cetrelia monachorum. *PLoS One,* 8**,** e76929.
OHTSUKI, T., FRIESEN, J. B., CHEN, S.-N., MCALPINE, J. B. & PAULI, G. F. 2022. Selective Preparation and High Dynamic-Range Analysis of Cannabinoids in "CBD Oil" and Other Cannabis sativa Preparations. *Journal of natural products.*
OMASA, T., ONITSUKA, M. & KIM, W.-D. 2010. Cell engineering and cultivation of Chinese hamster ovary (CHO) cells. *Current pharmaceutical biotechnology,* 11**,** 233-240.
OVA OZCAN, D. & OVEZ, B. 2020. Evaluation of the interaction of temperature and light intensity on the growth of Phaeodactylum tricornutum: Kinetic modeling and optimization. *Biochemical Engineering Journal.* Elsevier.
OVERKAMP, K. M., BAKKER, B. M., KÖTTER, P., LUTTIK, M. A., VAN DIJKEN, J. P. & PRONK, J. T. 2002. Metabolic engineering of glycerol production in Saccharomyces cerevisiae. *Applied and Environmental Microbiology,* 68**,** 2814-2821.
OVIDE, C., KIEFER-MEYER, M.-C., BÉRARD, C., VERGNE, N., LECROQ, T., PLASSON, C., BUREL, C., BERNARD, S., DRIOUICH, A. & LEROUGE, P. 2018. Comparative in depth RNA sequencing of P. tricornutum's morphotypes reveals specific features of the oval morphotype. *Scientific reports,* 8**,** 14340.
PANG, Y., DUAN, L., SONG, B., CUI, Y., LIU, X. & WANG, T. 2024a. A Review of fucoxanthin biomanufacturing from Phaeodactylum tricornutum. *Bioprocess and Biosystems Engineering***,** 1-22.
PANG, Z., XU, L., VIAU, C., LU, Y., SALAVATI, R., BASU, N. & XIA, J. 2024b. MetaboAnalystR 4.0: a unified LC-MS workflow for global metabolomics. *Nature Communications,* 15.
PARAPOULI, M., VASILEIADIS, A., AFENDRA, A.-S. & HATZILOUKAS, E. 2020. Saccharomyces cerevisiae and its industrial applications. *AIMS microbiology,* 6**,** 1.
PARK, S., HONGU, N. & DAILY III, J. W. 2016. Native American foods: History, culture, and influence on modern diets. *Journal of Ethnic Foods,* 3**,** 171-177.
PASTEUR, L. 1858. Nouveaux faits concernant l'histoire de la fermentation alcoolique. *Comptes Rendus Chimie,* 47**,** 1011-1013.
PATEL, A. K., CHOI, Y. Y. & SIM, S. J. 2020. Emerging prospects of mixotrophic microalgae: Way forward to sustainable bioprocess for environmental remediation and cost-effective biofuels. *Bioresource technology,* 300**,** 122741.
PELLATI, F., BORGONETTI, V., BRIGHENTI, V., BIAGI, M., BENVENUTI, S. & CORSI, L. 2018. Cannabis sativa L. and nonpsychoactive cannabinoids: their chemistry and role against oxidative stress, inflammation, and cancer. *BioMed research international,* 2018**,** 1691428.
PÉREZ-LÓPEZ, P., GONZÁLEZ-GARCÍA, S., ALLEWAERT, C., VERWEEN, A., MURRAY, P., FEIJOO, G. & MOREIRA, M. T. 2014. Environmental evaluation of eicosapentaenoic acid

production by Phaeodactylum tricornutum. *Science of the total environment,* 466**,** 991-1002.
PERTWEE, R. G. 2000. Neuropharmacology and therapeutic potential of cannabinoids. *Addiction biology,* 5**,** 37-46.
PERTWEE, R. G. 2001. Cannabinoid receptors and pain. *Progress in neurobiology,* 63**,** 569-611.
PERTWEE, R. G., HOWLETT, A., ABOOD, M. E., ALEXANDER, S., DI MARZO, V., ELPHICK, M., GREASLEY, P., HANSEN, H., KUNOS, G. & MACKIE, K. 2010. International Union of Basic and Clinical Pharmacology. LXXIX. Cannabinoid receptors and their ligands: beyond CB1 and CB2. *Pharmacological reviews,* 62**,** 588-631.
PETROVSKA, B. B. 2012. Historical review of medicinal plants' usage. *Pharmacognosy reviews,* 6**,** 1.
PHAM, J. V., YILMA, M. A., FELIZ, A., MAJID, M. T., MAFFETONE, N., WALKER, J. R., KIM, E., CHO, H. J., REYNOLDS, J. M. & SONG, M. C. 2019. A review of the microbial production of bioactive natural products and biologics. *Frontiers in microbiology,* 10**,** 1404.
PISANTI, S. & BIFULCO, M. 2019. Medical Cannabis: A plurimillennial history of an evergreen. *Journal of cellular physiology,* 234**,** 8342-8351.
POCHA, C. K. R., CHIA, W. Y., CHEW, K. W., MUNAWAROH, H. S. H. & SHOW, P. L. 2022. Current advances in recovery and biorefinery of fucoxanthin from Phaeodactylum tricornutum. *Algal Research,* 65**,** 102735.
POLLASTRO, F., CAPRIOGLIO, D., DEL PRETE, D., ROGATI, F., MINASSI, A., TAGLIALATELA-SCAFATI, O., MUNOZ, E. & APPENDINO, G. 2018. Cannabichromene. *Natural Product Communications,* 13**,** 1934578X1801300922.
POPKO, J., HERRFURTH, C., FEUSSNER, K., ISCHEBECK, T., IVEN, T., HASLAM, R., HAMILTON, M., SAYANOVA, O., NAPIER, J. & KHOZIN-GOLDBERG, I. 2016. Metabolome analysis reveals betaine lipids as major source for triglyceride formation, and the accumulation of sedoheptulose during nitrogen-starvation of Phaeodactylum tricornutum. *PLoS One,* 11**,** e0164673.
PORTER, A. C. & FELDER, C. C. 2001. The endocannabinoid nervous system: unique opportunities for therapeutic intervention. *Pharmacology & therapeutics,* 90**,** 45-60.
PRINGSHEIM, T., WILTSHIRE, K., DAY, L., DYKEMAN, J., STEEVES, T. & JETTE, N. 2012. The incidence and prevalence of Huntington's disease: a systematic review and meta-analysis. *Movement Disorders,* 27**,** 1083-1091.
PUDNEY, A., GANDINI, C., ECONOMOU, C. K., SMITH, R., GODDARD, P., NAPIER, J. A., SPICER, A. & SAYANOVA, O. 2019. Multifunctionalizing the marine diatom Phaeodactylum tricornutum for sustainable co-production of omega-3 long chain polyunsaturated fatty acids and recombinant phytase. *Scientific reports,* 9**,** 11444.
PUNJA, Z. K., SUTTON, D. B. & KIM, T. 2023. Glandular trichome development, morphology, and maturation are influenced by plant age and genotype in high THC-containing cannabis (Cannabis sativa L.) inflorescences. *Journal of Cannabis Research,* 5**,** 12.
PURI, V., NAGPAL, M., SINGH, I., SINGH, M., DHINGRA, G. A., HUANBUTTA, K., DHEER, D., SHARMA, A. & SANGNIM, T. 2022. A comprehensive review on nutraceuticals: therapy support and formulation challenges. *Nutrients,* 14**,** 4637.
QUAGHEBEUR, K., COOSEMANS, J., TOPPET, S. & COMPERNOLLE, F. 1994. Cannabiorci-and 8-chlorocannabiorcichromenic acid as fungal antagonists from Cylindrocarpon olidum. *Phytochemistry,* 37**,** 159-161.
QUINONES, R., MORENO, S., SMYTHERS, A. L., SULLINS, C., PIJOR, H., BROWN, G., TROUTEN, A., RICHARDS-WAUGH, L. L. & SIDDIG, A. 2022. Quantification of Cannabis in Infused Consumer Products and Their Residues on Skin. *ACS Pharmacol Transl Sci,* 5**,** 642-651.

RAKHA, A., SHEHZAD, A. & KHAN, K. 2024. Plant bioactives: challenges of extraction and processing. Frontiers Media SA.
RAO, J. S. 2024. Phytocannabinoids: a new frontier in Alzheimer's disease management. *Academia Biology,* 2.
RASALA, B. A. & MAYFIELD, S. P. 2015. Photosynthetic biomanufacturing in green algae; production of recombinant proteins for industrial, nutritional, and medical uses. *Photosynthesis research,* 123**,** 227-239.
RATES, S. M. K. 2001. Plants as source of drugs. *Toxicon,* 39**,** 603-613.
RAY, L. & MOORE, B. S. 2016. Recent advances in the biosynthesis of unusual polyketide synthase substrates. *Natural product reports,* 33**,** 150-161.
RAYMAN, M. P. 2012. Selenium and human health. *The Lancet,* 379**,** 1256-1268.
REIMER, C., KUFS, J. E., RAUTSCHEK, J., REGESTEIN, L., VALIANTE, V. & HILLMANN, F. 2022. Engineering the amoeba Dictyostelium discoideum for biosynthesis of a cannabinoid precursor and other polyketides. *Nature biotechnology,* 40**,** 751-758.
REITER, M. A., BRADLEY, T., BÜCHEL, L. A., KELLER, P., HEGEDIS, E., GASSLER, T. & VORHOLT, J. A. 2024. A synthetic methylotrophic Escherichia coli as a chassis for bioproduction from methanol. *Nature catalysis,* 7**,** 560-573.
RESEARCH, G. V. 2023a. *Protein Supplements Market Size, Share & Trends Analysis Report By Product (Protein Powders, Protein Bars), By Distribution Channel (Supermarkets, Online), By Application, By Source, By Region, And Segment Forecasts, 2023 - 2030* [Online]. Grand View Research. Available: https://www.grandviewresearch.com/industry-analysis/protein-supplements-market# [Accessed 14.02 2025].
REYNAUD, E. 2010. Protein misfolding and degenerative diseases. *Nature Education,* 3**,** 28.
RIGANO, M. M., MANNA, C., GIULINI, A., VITALE, A. & CARDI, T. 2009. Plants as biofactories for the production of subunit vaccines against bio-security-related bacteria and viruses. *Vaccine,* 27**,** 3463-3466.
RINNINELLA, E., RAOUL, P., CINTONI, M., FRANCESCHI, F., MIGGIANO, G. A. D., GASBARRINI, A. & MELE, M. C. 2019. What is the healthy gut microbiota composition? A changing ecosystem across age, environment, diet, and diseases. *Microorganisms,* 7**,** 14.
RODRIGUEZ-LOPEZ, P., RUEDA-ROBLES, A., SÁNCHEZ-RODRÍGUEZ, L., BLANCA-HERRERA, R. M., QUIRANTES-PINÉ, R. M., BORRÁS-LINARES, I., SEGURA-CARRETERO, A. & LOZANO-SÁNCHEZ, J. 2022. Analysis and screening of commercialized protein supplements for sports practice. *Foods,* 11**,** 3500.
ROUPAS, P., KEOGH, J., NOAKES, M., MARGETTS, C. & TAYLOR, P. 2012. The role of edible mushrooms in health: Evaluation of the evidence. *Journal of functional foods,* 4**,** 687-709.
RUSSO, E. B. 2007. History of cannabis and its preparations in saga, science, and sobriquet. *Chemistry & biodiversity,* 4**,** 1614-1648.
RUSSO, E. B., GUY, G. W. & ROBSON, P. J. 2007. Cannabis, pain, and sleep: lessons from therapeutic clinical trials of Sativex®, a cannabis-based medicine. *Chemistry & biodiversity,* 4**,** 1729-1743.
RUSSO, M. T., ROGATO, A., JAUBERT, M., KARAS, B. J. & FALCIATORE, A. 2023. Phaeodactylum tricornutum: An established model species for diatom molecular research and an emerging chassis for algal synthetic biology. *Journal of Phycology,* 59**,** 1114-1122.
RYAN ZAKLIN, M. 2019. *Grand Rounds: The Endocannabinoid System & Cannabis Therapeutics: An Integrative Approach to Peripheral Neuropathy* [Online]. Available: https://oshercenter.org/oc-event/grand-rounds-the-endocannabinoid-system-and-cannabis-a-key-to-integrative-wellness/ [Accessed 21.02 2025].
SABIR, J. S., THERIOT, E. C., MANNING, S. R., AL-MALKI, A. L., KHIYAMI, M. A., AL-GHAMDI, A. K., SABIR, M. J., ROMANOVICZ, D. K., HAJRAH, N. H. & EL OMRI, A. 2018.

Phylogenetic analysis and a review of the history of the accidental phytoplankter, Phaeodactylum tricornutum Bohlin (Bacillariophyta). *PloS one,* 13**,** e0196744.
SANDS, L. B., HAIDEN, S. R., MA, Y. & BERKOWITZ, G. A. 2023. Hormonal control of promoter activities of Cannabis sativa prenyltransferase 1 and 4 and salicylic acid mediated regulation of cannabinoid biosynthesis. *Scientific Reports,* 13**,** 8620.
SANGHAMITRA, S., DESHMUKH, S. & NARAYAN, K. P. 2020. Effects of alternate nutrient medium on microalgae biomass and lipid production as a bioenergy source for fuel production. *Materials Today: Proceedings,* 28**,** 659-664.
SARKAR, S., ALVAREZ, V. H. & SALDANA, M. D. 2014. Relevance of ions in pressurized fluid extraction of carbohydrates and phenolics from barley hull. *The Journal of Supercritical Fluids,* 93**,** 27-37.
SAWLER, J., STOUT, J. M., GARDNER, K. M., HUDSON, D., VIDMAR, J., BUTLER, L., PAGE, J. E. & MYLES, S. 2015. The genetic structure of marijuana and hemp. *PloS one,* 10**,** e0133292.
SAYIN, H. U. 2016. Psychoactive plants used during religious rituals. *Neuropathology of drug addictions and substance misuse.* Elsevier.
SCHEAU, C., BADARAU, I. A., MIHAI, L.-G., SCHEAU, A.-E., COSTACHE, D. O., CONSTANTIN, C., CALINA, D., CARUNTU, C., COSTACHE, R. S. & CARUNTU, A. 2020. Cannabinoids in the pathophysiology of skin inflammation. *Molecules,* 25**,** 652.
SCHMIDT, C. A., M.; KAYSER, O. 2024. Engineering cannabinoid production in Saccharomyces cerevisiae. *Biotechnology journal*.
SCHWENDER, H. 2023. siggenes: Multiple Testing using SAM and Efron's Empirical Bayes Approaches. 1.76.0 ed.
SCOTTER, E. L., ABOOD, M. E. & GLASS, M. 2010. The endocannabinoid system as a target for the treatment of neurodegenerative disease. *British journal of pharmacology,* 160**,** 480-498.
SCRANTON, M. A., OSTRAND, J. T., FIELDS, F. J. & MAYFIELD, S. P. 2015. Chlamydomonas as a model for biofuels and bio-products production. *The Plant Journal,* 82**,** 523-531.
SETHI, D., BUTLER, T. O., SHUHAILI, F. & VAIDYANATHAN, S. 2020. Diatoms for Carbon Sequestration and Bio-Based Manufacturing. *Biology (Basel),* 9.
SHAHZAD, A. 2012. Hemp fiber and its composites–a review. *Journal of composite materials,* 46**,** 973-986.
SHARMA, A. K., NYMARK, M., SPARSTAD, T., BONES, A. M. & WINGE, P. 2018. Transgene-free genome editing in marine algae by bacterial conjugation–comparison with biolistic CRISPR/Cas9 transformation. *Scientific reports,* 8**,** 14401.
SHARMA, A. K. & SHARMA, M. K. 2009. Plants as bioreactors: recent developments and emerging opportunities. *Biotechnology advances,* 27**,** 811-832.
SHARMA, N., SIMON, D. P., DIAZ-GARZA, A. M., FANTINO, E., MESSAABI, A., MEDDEB-MOUELHI, F., GERMAIN, H. & DESGAGNÉ-PENIX, I. 2021. Diatoms Biotechnology: Various Industrial Applications for a Greener Tomorrow. *Frontiers in Marine Science,* 8.
SHOYAMA, Y., TAMADA, T., KURIHARA, K., TAKEUCHI, A., TAURA, F., ARAI, S., BLABER, M., SHOYAMA, Y., MORIMOTO, S. & KUROKI, R. 2012. Structure and Function of Δ 1-Tetrahydrocannabinolic Acid (THCA) Synthase , the Enzyme Controlling the Psychoactivity of Cannabis sativa. *Journal of Molecular Biology.* Elsevier Ltd.
SIAUT, M., HEIJDE, M., MANGOGNA, M., MONTSANT, A., COESEL, S., ALLEN, A., MANFREDONIA, A., FALCIATORE, A. & BOWLER, C. 2007. Molecular toolbox for studying diatom biology in Phaeodactylum tricornutum. *Gene,* 406**,** 23-35.
SILVA, C., DA SILVA FILHO, J., PINHEIRO, G., TEIXEIRA, A. & FREIRE, P. 2018. Vibrational and structural properties of L-Alanyl-L-Phenylalanine dipeptide by Raman spectroscopy, infrared and DFT calculations. *Vibrational spectroscopy,* 98**,** 128-133.

SIMOPOULOS, A. P. 2010. The omega-6/omega-3 fatty acid ratio: health implications. *Oléagineux, Corps gras, Lipides,* 17**,** 267-275.
SINGH, A., BHATT, G., GUJRE, N., MITRA, S., SWAMINATHAN, R., LIMAYE, A. & RANGAN, L. 2021. Karanjin. *Phytochemistry,* 183**,** 112641.
SIRIKANTARAMAS, S., MORIMOTO, S., SHOYAMA, Y., ISHIKAWA, Y., WADA, Y., SHOYAMA, Y. & TAURA, F. 2004a. The gene controlling marijuana psychoactivity. Molecular cloning and heterologous expression of Δ1-tetrahydrocannabinolic acid synthase from Cannabis sativa L. *Journal of Biological Chemistry.* Â© 2004 ASBMB. Currently published by Elsevier Inc; originally published by American Society for Biochemistry and Molecular Biology.
SIRIKANTARAMAS, S., MORIMOTO, S., SHOYAMA, Y., ISHIKAWA, Y., WADA, Y., SHOYAMA, Y. & TAURA, F. 2004b. The gene controlling marijuana psychoactivity: molecular cloning and heterologous expression of Δ1-tetrahydrocannabinolic acid synthase from Cannabis sativa L. *Journal of Biological Chemistry,* 279**,** 39767-39774.
SIRIKANTARAMAS, S., TAURA, F., TANAKA, Y., ISHIKAWA, Y. & MORIMOTO, S. 2005. Tetrahydrocannabinolic Acid Synthase , the Enzyme Controlling Marijuana Psychoactivity , is Secreted into the Storage Cavity of the Glandular Trichomes.
SKAPER, S. D. & DI MARZO, V. 2012. Endocannabinoids in nervous system health and disease: the big picture in a nutshell. The Royal Society.
SLATTERY, S. S., DIAMOND, A., WANG, H., THERRIEN, J. A., LANT, J. T., JAZEY, T., LEE, K., KLASSEN, Z., DESGAGNE-PENIX, I., KARAS, B. J. & EDGELL, D. R. 2018. An Expanded Plasmid-Based Genetic Toolbox Enables Cas9 Genome Editing and Stable Maintenance of Synthetic Pathways in Phaeodactylum tricornutum. *ACS synthetic biology.*
SLATTERY, S. S., GIGUERE, D. J., STUCKLESS, E. E., SHRESTHA, A., BRIERE, L. A. K., GALBRAITH, A., REAUME, S., BOYKO, X., SAY, H. H., BROWNE, T. S., FREDERICK, M. I., LANT, J. T., HEINEMANN, I. U., DONOGHUE, P. O., DSOUZA, L., MARTIN, S., HOWARD, P., JEDESZKO, C., ALI, K., STYBA, G., FLATLEY, M., KARAS, B. J., GLOOR, G. B. & EDGELL, D. R. 2022. Phosphate - regulated expression of the SARS - CoV - 2 receptor - binding domain in the diatom Phaeodactylum tricornutum for pandemic diagnostics. *Scientific Reports.* Nature Publishing Group UK.
ŚLEDZIŃSKI, P., NOWAK-TERPIŁOWSKA, A. & ZEYLAND, J. 2020. Cannabinoids in medicine: cancer, immunity, and microbial diseases. *International journal of molecular sciences,* 22**,** 263.
SMALL, E. 2015. Evolution and classification of Cannabis sativa (marijuana, hemp) in relation to human utilization. *The botanical review,* 81**,** 189-294.
SMITH, T. H., SIM-SELLEY, L. J. & SELLEY, D. E. 2010. Cannabinoid CB1 receptor-interacting proteins: novel targets for central nervous system drug discovery? *British journal of pharmacology,* 160**,** 454-466.
SMOLECULE. 2024. *methyl 2-[(4-methyl-2-oxo-2H-chromen-7-yl)oxy]propanoate* [Online]. Available: https://www.smolecule.com/products/s12619316#pos-top [Accessed 10/10/2024 2024].
SODEINDE, O. A. & KINDLE, K. L. 1993. Homologous recombination in the nuclear genome of Chlamydomonas reinhardtii. *Proceedings of the National Academy of Sciences,* 90**,** 9199-9203.
SOINI, J., FALSCHLEHNER, C., LIEDERT, C., BERNHARDT, J., VUORISTO, J. & NEUBAUER, P. 2008. Norvaline is accumulated after a down-shift of oxygen in Escherichia coli W3110. *Microbial cell factories,* 7**,** 1-14.
SOLÁ, R. J. & GRIEBENOW, K. 2009. Effects of glycosylation on the stability of protein pharmaceuticals. *Journal of pharmaceutical sciences,* 98**,** 1223-1245.

SONG, Z., LYE, G. J. & PARKER, B. M. 2020. Morphological and biochemical changes in Phaeodactylum tricornutum triggered by culture media: Implications for industrial exploitation. *Algal Research,* 47**,** 101822.
SOSA-CARRILLO, S., GALEZ, H., NAPOLITANO, S., BERTAUX, F. & BATT, G. 2023. Maximizing protein production by keeping cells at optimal secretory stress levels using real-time control approaches. *Nat Commun,* 14**,** 3028.
SRIVASTAVA, N. & AKHILA, A. 2010. Biosynthesis of andrographolide in Andrographis paniculata. *Phytochemistry,* 71**,** 1298-1304.
STANLEY, P. 2011. Golgi glycosylation. *Cold Spring Harbor perspectives in biology,* 3**,** a005199.
STELLA, N. 2023. THC and CBD: Similarities and differences between siblings. *Neuron,* 111**,** 302-327.
STEWART, M. P., LANGER, R. & JENSEN, K. F. 2018. Intracellular delivery by membrane disruption: mechanisms, strategies, and concepts. *Chemical reviews,* 118**,** 7409-7531.
STOJANOV, S., BERLEC, A. & ŠTRUKELJ, B. 2020. The influence of probiotics on the firmicutes/bacteroidetes ratio in the treatment of obesity and inflammatory bowel disease. *Microorganisms,* 8**,** 1715.
STOLFA, G., SMONSKEY, M. T., BONIFACE, R., HACHMANN, A. B., GULDE, P., JOSHI, A. D., PIERCE, A. P., JACOBIA, S. J. & CAMPBELL, A. 2018. CHO-omics review: The impact of current and emerging technologies on Chinese hamster ovary based bioproduction. *Biotechnology journal,* 13**,** 1700227.
STRIJBIS, K., VAZ, F. M. & DISTEL, B. 2010. Enzymology of the carnitine biosynthesis pathway. *IUBMB life,* 62**,** 357-362.
SZAMECZ, B. K., VARSZEGI, S., NEMETH, A., SZABO, L. & KUMAR, A. 2024. Microorganisms and Methods for the Fermentation of Cannabinoids. Google Patents.
TAHIR, M. N., SHAHBAZI, F., RONDEAU-GAGNÉ, S. & TRANT, J. F. 2021. The biosynthesis of the cannabinoids. *Journal of cannabis research,* 3**,** 1-12.
TAJ, M. K., SAMREEN, Z., LING, J. X., TAJ, I., HASSAN, T. & YUNLIN, W. 2014. Escherichia coli as a model organism. *International Journal of Engineering Research and Science and Technology,* 3**,** 1-8.
TALAPATRA, S. K., TALAPATRA, B., TALAPATRA, S. K. & TALAPATRA, B. 2015. Polyketide pathway. biosynthesis of diverse classes of aromatic compounds. *Chemistry of Plant Natural Products: Stereochemistry, Conformation, Synthesis, Biology, and Medicine***,** 679-715.
TANNEY, C. A., BACKER, R., GEITMANN, A. & SMITH, D. L. 2021. Cannabis glandular trichomes: A cellular metabolite factory. *Frontiers in Plant Science,* 12**,** 721986.
TAURA, F., MORIMOTO, S., SHOYAMA, Y., ACID, C., CANNABIDIOLIC, T. O., TAURA, F., MORIMOTO, S. & SHOYAMA, Y. 1996. Purification and characterization of cannabidiolic-acid synthase from Cannabis sativa L. Biochemical analysis of a novel enzyme that catalyzes the oxidocyclization of cannabigerolic acid to cannabidiolic acid. *Journal of Biological Chemistry.* Â© 1996 ASBMB. Currently published by Elsevier Inc; originally published by American Society for Biochemistry and Molecular Biology.
TEAFORD, M. F. & UNGAR, P. S. 2016. Diet and the evolution of the earliest human ancestors. *Human Evolution Source Book.* Routledge.
TESSON, B., GAILLARD, C. & MARTIN-JEZEQUEL, V. 2009a. Insights into the polymorphism of the diatom Phaeodactylum tricornutum Bohlin.
TESSON, B., GENET, M. J., FERNANDEZ, V., DEGAND, S., ROUXHET, P. G. & MARTIN-JÉZÉQUEL, V. 2009b. Surface chemical composition of diatoms. *ChemBioChem,* 10**,** 2011-2024.
THAKUR, G. A., NIKAS, S. P. & MAKRIYANNIS, A. 2005. CB1 cannabinoid receptor ligands. *Mini reviews in medicinal chemistry,* 5**,** 631-640.

TSUGAWA, H., CAJKA, T., KIND, T., MA, Y., HIGGINS, B., IKEDA, K., KANAZAWA, M., VANDERGHEYNST, J., FIEHN, O. & ARITA, M. 2015. MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nature Methods,* 12**,** 523-526.

TSUGAWA, H., NAKABAYASHI, R., MORI, T., YAMADA, Y., TAKAHASHI, M., RAI, A., SUGIYAMA, R., YAMAMOTO, H., NAKAYA, T., YAMAZAKI, M., KOOKE, R., BAC-MOLENAAR, J. A., OZTOLAN-EROL, N., KEURENTJES, J. J. B., ARITA, M. & SAITO, K. 2019. A cheminformatics approach to characterize metabolomes in stable-isotope-labeled organisms. *Nature Methods,* 16**,** 295-298.

TUNICK, M. H. & NASSER, J. A. 2019. The chemistry of chocolate and pleasure. *Sex, smoke, and spirits: The role of chemistry.* ACS Publications.

TUOMAINEN, M., KÄRKKÄINEN, O., LEPPÄNEN, J., AURIOLA, S., LEHTONEN, M., SAVOLAINEN, M. J., HERMANSEN, K., RISÉRUS, U., ÅKESSON, B. & THORSDOTTIR, I. 2019. Quantitative assessment of betainized compounds and associations with dietary and metabolic biomarkers in the randomized study of the healthy Nordic diet (SYSDIET). *The American journal of clinical nutrition,* 110**,** 1108-1118.

TURNER, C. E. & ELSOHLY, M. A. 1981. Biological activity of cannabichromene, its homologs and isomers. *The Journal of Clinical Pharmacology,* 21**,** 283S-291S.

TWYMAN, R. M., STOGER, E., SCHILLBERG, S., CHRISTOU, P. & FISCHER, R. 2003. Molecular farming in plants: host systems and expression technology. *TRENDS in Biotechnology,* 21**,** 570-578.

VALLIERE, M. A., KORMAN, T. P., ARBING, M. A. & BOWIE, J. U. 2020. A bio-inspired cell-free system for cannabinoid production from inexpensive inputs. *Nature Chemical Biology,* 16**,** 1427-1433.

VALLIERE, M. A., KORMAN, T. P., WOODALL, N. B., KHITROV, G. A., TAYLOR, R. E., BAKER, D. & BOWIE, J. U. 2019. A cell-free platform for the prenylation of natural products and application to cannabinoid production. *Nature Communications.* Springer US.

VAN SICKLE, M. D., DUNCAN, M., KINGSLEY, P. J., MOUIHATE, A., URBANI, P., MACKIE, K., STELLA, N., MAKRIYANNIS, A., PIOMELLI, D. & DAVISON, J. S. 2005. Identification and functional characterization of brainstem cannabinoid CB2 receptors. *Science,* 310**,** 329-332.

VASINCU, A., RUSU, R.-N., ABABEI, D.-C., LARION, M., BILD, W., STANCIU, G. D., SOLCAN, C. & BILD, V. 2022. Endocannabinoid modulation in neurodegenerative diseases: In pursuit of certainty. *Biology,* 11**,** 440.

VASISHT, K., SHARMA, N. & KARAN, M. 2016. Current perspective in the international trade of medicinal plants material: an update. *Current pharmaceutical design,* 22**,** 4288-4336.

WADITEE, R., TANAKA, Y., AOKI, K., HIBINO, T., JIKUYA, H., TAKANO, J., TAKABE, T. & TAKABE, T. 2003. Isolation and Functional Characterization ofN-Methyltransferases That Catalyze Betaine Synthesis from Glycine in a Halotolerant Photosynthetic Organism Aphanothece halophytica. *Journal of Biological Chemistry,* 278**,** 4932-4942.

WALKER, E. J., PAMPUCH, M., DENG, L., LI, Y., TRAN, G., MOCK, T. & KARAS, B. 2024. Spheroplasted cells: a game changer for DNA delivery to diatoms. *bioRxiv***,** 2024.10. 10.617634.

WALKER, T. L., PURTON, S., BECKER, D. K. & COLLET, C. 2005. Microalgae as bioreactors. *Plant cell reports,* 24**,** 629-641.

WALSH, K. B., MCKINNEY, A. E. & HOLMES, A. E. 2021a. Minor Cannabinoids: Biosynthesis, Molecular Pharmacology and Potential Therapeutic Uses. *Frontiers in Pharmacology.*

WALSH, K. B., MCKINNEY, A. E. & HOLMES, A. E. 2021b. Minor cannabinoids: Biosynthesis, molecular pharmacology and potential therapeutic uses. *Frontiers in pharmacology,* 12**,** 777804.

WANG, B., KOVALCHUK, A., LI, D., RODRIGUEZ-JUAREZ, R., ILNYTSKYY, Y., KOVALCHUK, I., KOVALCHUK, O., KOVALCHUK, A., KOVALCHUK, I., KOVALCHUK, O., WANG, B., LI, D., RODRIGUEZ-JUAREZ, R., ILNYTSKYY, Y., KOVALCHUK, I., KOVALCHUK, O. & KOVALCHUK, A. 2020a. In search of preventive strategies: novel high-CBD Cannabis sativa extracts modulate ACE2 expression in COVID-19 gateway tissues. *Aging.*

WANG, J., DING, N., WU, Y., SHI, X., QI, B., LIU, X., WANG, X., LI, J., TU, P. & SHI, S. 2020b. Enzymatic synthesis of 2-hydroxy-4H-quinolizin-4-one scaffolds by integrating coenzyme a ligases and a type III PKS from Huperzia serrata. *RSC advances,* 10**,** 23566-23572.

WANG, L., LI, C. & LUO, K. 2024. Biosynthesis and metabolic engineering of isoflavonoids in model plants and crops: A review. *Frontiers in Plant Science,* 15**,** 1384091.

WANG, T., COLLET, J.-P., SHAPIRO, S. & WARE, M. A. 2008. Adverse effects of medical cannabinoids: a systematic review. *Cmaj,* 178**,** 1669-1678.

WATTS, K. T., LEE, P. C. & SCHMIDT-DANNERT, C. 2006. Biosynthesis of plant-specific stilbene polyketides in metabolically engineered Escherichia coli. *BMC biotechnology,* 6**,** 1-12.

WEAVER, J. C. 1995. Electroporation theory: concepts and mechanisms. *Plant cell electroporation and electrofusion protocols***,** 3-28.

WEBB, J. P., ARNOLD, S. A., BAXTER, S., HALL, S. J., EASTHAM, G. & STEPHENS, G. 2018. Efficient bio-production of citramalate using an engineered Escherichia coli strain. *Microbiology,* 164**,** 133-141.

WEILAND-BRÄUER, N., KISCH, M. J., PINNOW, N., LIESE, A. & SCHMITZ, R. A. 2016. Highly effective inhibition of biofilm formation by the first metagenome-derived AI-2 quenching enzyme. *Frontiers in microbiology,* 7**,** 1098.

WHO. 2024. *Global cancer burden growing, amidst mounting need for services* [Online]. World Health Organization. Available: https://www.who.int/news/item/01-02-2024-global-cancer-burden-growing--amidst-mounting-need-for-services [Accessed 14.02 2025].

WILDING, K. M., SCHINN, S.-M., LONG, E. A. & BUNDY, B. C. 2018. The emerging impact of cell-free chemical biosynthesis. *Current opinion in biotechnology,* 53**,** 115-121.

WILES, D., SHANBHAG, B. K., O'BRIEN, M., DOBLIN, M. S., BACIC, A. & BEDDOE, T. 2022. Heterologous production of Cannabis sativa-derived specialised metabolites of medicinal significance–Insights into engineering strategies. *Phytochemistry,* 203**,** 113380.

WILLIAMSON, E. M. & EVANS, F. J. 2000. Cannabinoids in clinical practice. *Drugs,* 60**,** 1303-1314.

WITTE, C.-P. & HERDE, M. 2020. Nucleotide metabolism in plants. *Plant physiology,* 182**,** 63-78.

XIE, Z., MI, Y., KONG, L., GAO, M., CHEN, S., CHEN, W., MENG, X., SUN, W., CHEN, S. & XU, Z. 2023. Cannabis sativa: origin and history, glandular trichome development, and cannabinoid biosynthesis. *Horticulture Research,* 10**,** uhad150.

XU, C., LI, H., YANG, X., GU, C., MU, H., YUE, Y. & WANG, L. 2016. Cloning and expression analysis of MEP pathway enzyme-encoding genes in Osmanthus fragrans. *Genes,* 7**,** 78.

XU, J., DOLAN, M. C., MEDRANO, G., CRAMER, C. L. & WEATHERS, P. J. 2012a. Green factory: plants as bioproduction platforms for recombinant proteins. *Biotechnology advances,* 30**,** 1171-1184.

XU, L., LIU, T., LIU, L., YAO, X., CHEN, L., FAN, D., ZHAN, S. & WANG, S. 2020. Global variation in prevalence and incidence of amyotrophic lateral sclerosis: a systematic review and meta-analysis. *Journal of neurology,* 267**,** 944-953.

XU, P., VANSIRI, A., BHAN, N. & KOFFAS, M. A. 2012b. ePathBrick: a synthetic biology platform for engineering metabolic pathways in E. coli. *ACS synthetic biology,* 1**,** 256-266.

XUE, J., NIU, Y.-F., HUANG, T., YANG, W.-D., LIU, J.-S. & LI, H.-Y. 2015. Genetic improvement of the microalga Phaeodactylum tricornutum for boosting neutral lipid accumulation. *Metabolic engineering,* 27**,** 1-9.

YANG, D., PARK, S. Y., PARK, Y. S., EUN, H. & LEE, S. Y. 2020. Metabolic engineering of Escherichia coli for natural product biosynthesis. *Trends in Biotechnology,* 38**,** 745-765.

YANG, M.-Q., VAN VELZEN, R., BAKKER, F. T., SATTARIAN, A., LI, D.-Z. & YI, T.-S. 2013. Molecular phylogenetics and character evolution of Cannabaceae. *Taxon,* 62**,** 473-485.

YANG, S., SONG, L., WANG, J., ZHAO, J., TANG, H. & BAO, X. 2024a. Engineering Saccharomyces cerevisiae for efficient production of recombinant proteins. *Engineering Microbiology,* 4**,** 100122.

YANG, X., ZHENG, Z. & WANG, Y. 2024b. Bacillus methanolicus: an emerging chassis for low-carbon biomanufacturing. *Trends in Biotechnology*.

YANG, Y.-X., WANG, J.-X., WANG, Q., LI, H.-L., TAO, M., LUO, Q. & LIU, H. 2018. New chromane and chromene meroterpenoids from flowers of Rhododendron rubiginosum Franch. var. rubiginosum. *Fitoterapia,* 127**,** 396-401.

YOO, C., MAURY, J., GONZALEZ, D. E., KO, J., XING, D., JENKINS, V., DICKERSON, B., LEONARD, M., ESTES, L. & JOHNSON, S. 2024. Effects of Supplementation with a Microalgae Extract from Phaeodactylum tricornutum Containing Fucoxanthin on Cognition and Markers of Health in Older Individuals with Perceptions of Cognitive Decline. *Nutrients,* 16**,** 2999.

ZHANG, P., AN, Q., YI, P., CUI, Y., ZOU, J.-B., YUAN, C.-M., ZHANG, Y., GU, W., HUANG, L.-J. & ZHAO, L.-H. 2022a. Thermlanseedlines A–G, seven thermopsine-based alkaloids with antiviral and insecticidal activities from the seeds of Thermopsis lanceolata R. Br. *Fitoterapia,* 158**,** 105140.

ZHANG, S., ZHANG, L., XU, G., LI, F. & LI, X. 2022b. A review on biodiesel production from microalgae: Influencing parameters and recent advanced technologies. *Frontiers in Microbiology,* 13**,** 970028.

ZHANG, X., TERVO, C. J. & REED, J. L. 2016. Metabolic assessment of E. coli as a Biofactory for commercial products. *Metabolic engineering,* 35**,** 64-74.

ZHONG, Q., XIAO, X., QIU, Y., XU, Z., CHEN, C., CHONG, B., ZHAO, X., HAI, S., LI, S. & AN, Z. 2023. Protein posttranslational modifications in health and diseases: Functions, regulatory mechanisms, and therapeutic implications. *MedComm,* 4**,** e261.

ZHOU, H., YANG, Y., SHANG, W., RAO, Y., CHEN, J., PENG, H., HUANG, J., HU, Z., ZHANG, R. & RAO, X. 2022. Pyocyanin biosynthesis protects Pseudomonas aeruginosa from nonthermal plasma inactivation. *Microbial biotechnology,* 15**,** 1910-1921.

ZIMMER, C. 2009. *Microcosm: E. coli and the New Science of Life*, Vintage.

ZIRPEL, B., DEGENHARDT, F., MARTIN, C., KAYSER, O. & STEHLE, F. 2017a. Engineering yeasts as platform organisms for cannabinoid biosynthesis. *Journal of biotechnology,* 259**,** 204-212.

ZIRPEL, B., DEGENHARDT, F., MARTIN, C., KAYSER, O. & STEHLE, F. 2017b. Engineering yeasts as platform organisms for cannabinoid biosynthesis. *Journal of Biotechnology.* Elsevier.

ZIRPEL, B., DEGENHARDT, F., ZAMMARELLI, C., WIBBERG, D., KALINOWSKI, J., STEHLE, F. & KAYSER, O. 2018a. Optimization of Δ9-tetrahydrocannabinolic acid synthase production in Komagataella phaffii via post-translational bottleneck identification. *Journal of Biotechnology.* Elsevier.

ZIRPEL, B., KAYSER, O. & STEHLE, F. 2018b. Elucidation of structure-function relationship of THCA and CBDA synthase from Cannabis sativa L . *Journal of Biotechnology.* Elsevier.

ZIRPEL, B., STEHLE, F. & KAYSER, O. 2015. Production of Delta9-tetrahydrocannabinolic acid from cannabigerolic acid by whole cells of Pichia (Komagataella) pastoris expressing Delta9-tetrahydrocannabinolic acid synthase from Cannabis sativa L. *Biotechnology letters.*

ZOU, S. & KUMAR, U. 2018. Cannabinoid receptors and the endocannabinoid system: signaling and function in the central nervous system. *International journal of molecular sciences,* 19**,** 833.

ZULU, N. N., ZIENKIEWICZ, K., VOLLHEYDE, K. & FEUSSNER, I. 2018. Current trends to comprehend lipid metabolism in diatoms. *Progress in lipid research,* 70**,** 1-16.