METABOLIC EVOLUTION IN *GALDIERIA SULPHURARIA*

By

CHAD M. TERNES

Bachelor of Science in Botany
Oklahoma State University
Stillwater, Oklahoma
2009

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
May, 2015

METABOLIC EVOLUTION IN *GALDIERIA*

SUPHURARIA

Dissertation Approved:

Dr. Gerald Schoenknecht

Dissertation Adviser

Dr. David Meinke

Dr. Andrew Doust

Dr. Patricia Canaan

Name: CHAD M. TERNES

Date of Degree: MAY, 2015

Title of Study: METABOLIC EVOLUTION IN *GALDIERIA SULPHURARIA*

Major Field: PLANT SCIENCE

Abstract: The thermoacidophilic, unicellular, red alga *Galdieria sulphuraria* possesses characteristics, including salt and heavy metal tolerance, unsurpassed by any other alga. Like most plastid bearing eukaryotes, *G. sulphuraria* can grow photoautotrophically. Additionally, it can also grow solely as a heterotroph, which results in the cessation of photosynthetic pigment biosynthesis. The ability to grow heterotrophically is likely correlated with *G. sulphuraria*'s broad capacity for carbon metabolism, which rivals that of fungi. Annotation of the metabolic pathways encoded by the genome of *G. sulphuraria* revealed several pathways that are uncharacteristic for plants and algae, even red algae. Phylogenetic analyses of the enzymes underlying the metabolic pathways suggest multiple instances of horizontal gene transfer, in addition to endosymbiotic gene transfer and conservation through ancestry. Although some metabolic pathways as a whole appear to be retained through ancestry, genes encoding individual enzymes within a pathway were substituted by genes that were acquired horizontally from other domains of life. Thus, metabolic pathways in *G. sulphuraria* appear to be composed of a 'metabolic patchwork', underscored by a mosaic of genes resulting from multiple evolutionary processes. Substitution of genes encoding pathway enzymes also extends to metabolic pathways in other eukaryotic organisms. Specifically, *de novo* NAD^+ biosynthesis in eukaryotes, including those possessing a plastid, has been subjected to numerous gene transfer events, some of which were responsible for the establishment of novel metabolic pathways in plastid-bearing eukaryotes. Another characteristic of *G. sulphuraria* is observed when cultivating the alga in a liquid medium. Under light limiting conditions, *G. sulphuraria* excretes porphyrins, which fluoresce when illuminated with near-UV wavelengths of light. Examination of the absorption and emission (fluorescence) spectra of the porphyrin mixture led to the hypothesis of spectral shifting, whereby near-UV light that is unusable for photosynthesis is converted into light readily absorbed by phycocyanin, a photosynthetic pigment synthesized by *G. sulphuraria*. The calculated relative fluorescence quantum yield, i.e., the efficiency at which absorbed photons are re-emitted as fluoresced light, of the excreted porphyrins was lower than to be expected for counteracting light limiting conditions, possibly indicating a different biological function for porphyrin excretion in *G. sulphuraria*.

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Figure Page

CHAPTER I

INTRODUCTION

Galdieria sulphuraria is a thermoacidophilic, red microalga which thrives in environments, such as hot sulfur springs, volcanic calderas, and strip mines, that are usually deemed uninhabitable by eukaryotes (Weber, et al. 2004). It can survive temperatures up to 56°C with pH values ranging from 0-4 (Gross 2000) and displays heavy metal (Yoshimura, et al. 1999; Nagasaka, et al. 2004) and salt tolerance (Barbier, et al. 2005). *G. sulphuraria* possesses an extensive metabolic flexibility as shown by the ability to grow photoautotrophically, heterotrophically, or mixotrophically (a combination of the prior two) on over 50 different carbon sources (Weber, et al. 2004) (Fig. 1).

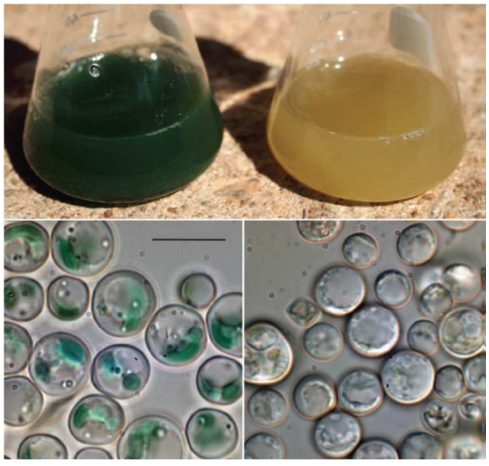


Fig. 1: Phototrophic (left) and heterotrophic (right) *G. sulphuraria* cells taken from Schönknecht, et al. (2013). **Left**: Phototrophic culture of *G. sulphuraria* cells (top) and light microscopic image (bottom; bar represents 10 µm) grown under continuous light with no supplemental carbon source. **Right**: Heterotrophic culture of *G. sulphuraria* cells grown in darkness supplemented with 200 mM glucose (top) and light microscopic image (bottom).

Red algae provide insight into the process of early eukaryotic evolution. One of the earliest fossil records for multicellular eukaryotes is a red alga, *Bangiomorpha pubescens*, estimated to be 1.2 billion years old (Butterfield 2000). Not only does the *B. pubescens* fossil provide a broader perspective on the evolution of multicellular eukaryotes as a whole, it specifically has an impact when considering evolution of the supergroup Archaeplastida (Plantae), which is comprised of Viridiplantae (land plants and green algae), Rhodophyta (red algae), and Glaucophyta (glaucophytes). More specifically, *G. sulphuraria* is a member of Cyanidiophyceae, an early diverging family of red algae. In the context of the *B. pubescens* fossil, it is possible that Cyanidiophyceae evolved well before the estimate 1.2 billion years ago (Yoon, et al. 2002). Furthermore, all Cyanidiophyceae species are unicellular, suggesting that Cyanidiophyceae possibly split off from other red algal lineages prior to the development of multicellularity in red algae. The primary endosymbiosis event, whereby a heterotrophic, protist-like eukaryote engulfed a cyanobacterium ultimately giving rise to modern day chloroplasts, occurred in the common ancestor of Archaeplastida (Gould, et al. 2008; Archibald 2009). The initial establishment of plastids in Archaeplastida provided the foundation for the radiation of plastids throughout other eukaryotic lineages. A substantial number of other plastid bearing eukaryotes, such as brown algae, diatoms, haptophytes, and cryptophytes, possess a plastid of red algal origin that results from secondary endosymbiosis events (McFadden 2014). Several secondary endosymbiosis events entailed the engulfment of a red alga by a once-heterotrophic eukaryote, leading to the establishment of (red algal) plastids in other eukaryotic lineages. Modern day eukaryotes possessing a plastid of red algal origin encompass a wealth of biodiversity on Earth. Thus, understanding the biology and evolution of red algae is essential to understanding the biology and evolution of other photosynthetic eukaryotes that possess plastids of red algal origin.

The genome of *G. sulphuraria* has been published (Schönknecht, et al. 2013) and sheds light on the diversity of metabolic pathways in red algae. In addition, the metabolic pathways present in *G. sulphuraria* and other Cyanidiophyceae algae provide a glimpse into early eukaryotic metabolism and

its evolution. Another species within Cyanidiophyceae, *Cyanidioschyzon merolae*, has also been sequenced (Matsuzaki, et al. 2004) in addition to other red algae, *Porphyridium purpureum* (Bhattacharya, et al. 2013), *Chondrus crispus* (Collén, et al. 2013), and *Pyropia yezoensis* (Nakamura, et al. 2013), making direct comparisons of metabolic pathways possible within Cyanidiophyceae, and more broadly Rhodophyta. While *G. sulphuraria* shares many metabolic pathways in common with other red algae, it also possesses pathways and enzymes that make it unique among red algae.

Coupling phylogenetic analyses with the annotation of the metabolic pathways, the evolutionary history of different pathways present in *G. sulphuraria* was elucidated. For some pathways, simply determining the distribution among different eukaryotic lineages provided insight into the origin of metabolic pathways in eukaryotes, and even photosynthetic eukaryotes in some instances. The diversity of pathways encoded by the genome of *G. sulphuraria* seems to be a 'metabolic patchwork', with some pathways appearing to be conserved through ancestry, and others being acquired from prokaryotes either via endosymbiosis or horizontal gene transfer. Furthermore, some genes encoding enzymes for metabolic pathways, those which have been conserved throughout eukaryotic evolution, have been substituted by genes of prokaryotic origin. Consequently, some metabolic pathways in *G. sulphuraria*, otherwise attributed to ancestry, contain enzymes encoded by genes that have not been conserved through ancestry.

AN 'ANCIENT' PATHWAY FOR *DE NOVO* NAD^+ BIOSYNTHESIS

A distinguishing metabolic pathway in *G. sulphuraria* and other red algae, in contrast with Viridiplantae and Glaucophyta, is the route by which the amino acid L-tryptophan is degraded. Tryptophan degradation in *G. sulphuraria* proceeds through the intermediate metabolite L-kynurenine, similar to that of Opisthokonta (animals and fungi), in what has been deemed the

kynurenine pathway (Fig. 2). Not only is the kynurenine pathway responsible for the degradation of tryptophan, but it is also coupled with *de novo* biosynthesis of the essential cofactor nicotinamide adenine dinucleotide, or NAD^+. NAD^+ is essential for a multitude of cellular processes and is arguably the most ubiquitous electron carrier in the cell, being associated with glycolysis, the TCA cycle, and the electron transport chain for respiration on the inner membrane of the mitochondria. The phosphorylated form of NAD^+, $NADP^+$, is indispensable for biosynthetic pathways, such as the oxidative pentose phosphate pathway that regenerates NADPH and synthesizes carbohydrate intermediates for other biosynthetic pathways (Voet and Voet 2004). In the light-dependent reactions of photosynthesis, $NADP^+$ is reduced to form NADPH, which is used to help fuel the Calvin-Benson cycle to synthesize carbohydrates from CO_2 (Taiz and Zeiger 2010).

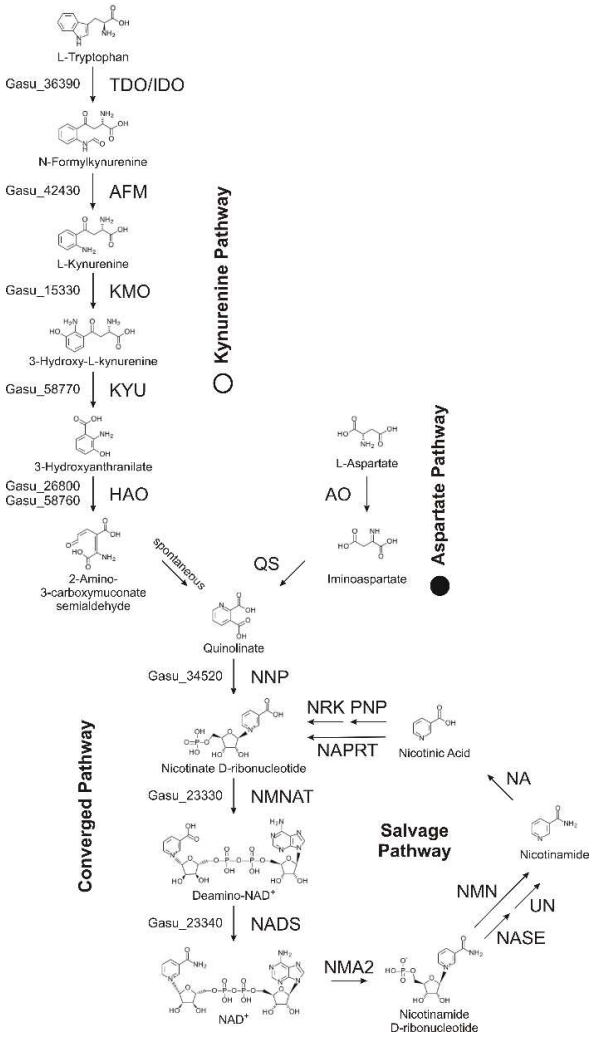


Fig. 2: The two different pathways for *de novo* NAD^+ biosynthesis starting with either tryptophan for the kynurenine pathway or aspartate for the aspartate pathway are shown. Both pathways converge at quinolinate. The salvage pathway recycles NAD^+ components (see text), and pathways are dependent upon lineage, e.g. one or two enzymes required to synthesize particular intermediates. Structural formulas for intermediates and all substrate and enzyme names according to KEGG (Kanehisa, et al. 2014). Abbreviations: AO, aspartate oxidase; QS, quinolinate synthase; TDO/IDO, tryptophan - / indoleamine 2,3-dioxygenase;

AFM, arylformamidase; KMO, kynurenine 3-monooxygenase; KYU, kynureninase; HAO, 3-hydroxyanthranilate 3,4-dioxygenase; NNP, nicotinate-nucleotide adenylyltransferase; NADS, NAD^+ synthase; NMA2, nicotinamide-nucleotide adenylyltransferase 2; NMN, nicotinamide mononucleotide nucleosidase; NASE, 5'-nucleotidase; UN, uridine nucleosidase; NA, nicotinamidase; PNP, purine-nucleotide phosphorylase; NRK, nicotinamide/nicotinate riboside kinase; NAPRT, nicotinate phorphoribosyltransferase.

Based on sequenced eukaryotic genomes, the distribution of which *de novo* NAD^+ biosynthesis pathway is present in different eukaryotic lineages was determined. This resulted in an unexpected pattern for photosynthetic eukaryotes, with red algae using the kynurenine pathway, while Viridiplantae, diatoms, and brown algae (Phaeophyta) appear to use the aspartate pathway for *de novo* NAD^+ biosynthesis (Fig. 2). Phylogenetic analyses (conducted here) indicate that the evolution of *de novo* NAD^+ biosynthesis was driven by several instances of gene transfer, both endosymbiotic gene transfer and horizontal gene transfer from bacteria or archaea into eukaryotic genomes.

G. SULPHURARIA EXCRETES FLUORESCENT PIGMENTS

When grown to high cell densities in liquid medium, *G. sulphuraria* excretes fluorescent pigments into the culture medium. Once the cells are removed, the medium retains a reddish-pink coloration (Fig. 3). The fluorescent pigments displayed interesting spectral properties, whereby near-UV radiation was absorbed and re-emitted as orange light, similar to the spectral characteristics of porphyrins. In essence, these pigments potentially act as 'light converters', transforming near-UV radiation that cannot be used by photosynthesis into 'photosynthetically active radiation' (PAR), radiation that can be used for photosynthesis.

Fig. 3: Stationary growth phase *G. sulphuraria* liquid culture supplemented with 50 mM glucose. Cells were allowed to settle for several hours. **Left**: Culture medium under fluorescent lighting showing a pink coloration. **Right**: culture medium illuminated with blue LED lighting showing fluorescence.

Based on the observation that *G. sulphuraria* excretes fluorescent pigments into the culture medium under increased cell density, and considering the spectroscopic properties of the pigments, it seems reasonable that the fluorescent pigments could serve as a mechanism to counteract conditions when light becomes limiting, either as a result of high cell density or low light environments. This mechanism could ultimately allow cells to utilize light energy otherwise not photosynthetically available by absorbing near-UV radiation and re-emitting photosynthetically active orange light readily used by chlorophylls *a* and *b* and phycocyanin. In order to determine the effectiveness of 'spectral shifting' by the fluorescent pigments excreted by *G. sulphuraria*, the pigments were fully characterized spectroscopically as they existed in the culture medium. A complete, emission (fluorescence) spectroscopy characterization determined the effectiveness of spectral shifting of these pigments; the single, most important measurement was the fluorescence quantum yield (or efficiency). The fluorescence quantum yield describes how many photons are re-emitted in relation to how many photons are absorbed, e.g. an efficiency of 20% would mean for every five photons that are absorbed, one photon is re-emitted.

Photosynthesis is the limiting factor for both terrestrial crops and algal biomass. Photosynthesis in land plants and algae is driven by photosynthetically active radiation (PAR)

between 400 and 700 nm. Only about 50% of the solar radiation that reaches a plant's surface is used to drive photosynthesis (Monteith 1969; Meek, et al. 1984). There is a clear correlation between intercepted light and biomass production (Monteith 1994), and expanding the solar spectrum effective for photosynthesis could significantly increase productivity. Yet, oxygenic photosynthesis is largely limited to the 400-700 nm wavelength range (Chen and Blankenship 2011), and only anoxygenic, photosynthetic bacteria use a considerably wider range of wavelengths (300-1000 nm) (Scheer 2003; Kiang, et al. 2007). It has been experimentally shown that spectral shifting, i.e. the conversion of UV light (<400 nm) into PAR, enhances algal growth (Prokop, et al. 1984). Thus, the fluorescent pigments excreted by *G. sulphuraria* could potentially be used for biotechnology applications, such as algal biofuel production, to increase biomass production in cell cultures that experience low light conditions, as a result of high cell density.

CHAPTER II

REVIEW OF LITERATURE

EVOLUTION SHAPES METABOLIC PATHWAYS AND PROCESSES

Evolution has played a direct role in shaping the metabolic pathways present in photosynthetic organisms. While global metabolism in the cell was undoubtedly established in the ancestor of all eukaryotes, different lineages of photosynthetic eukaryotes have undergone their own evolution, which further shaped metabolism in a lineage-specific manner. The result is a common core of centralized metabolic pathways, i.e. glycolysis, TCA cycle, etc., with differences existing amongst different plastid-bearing lineages, especially when considering compartmentalization, the specific organellar location of a particular metabolic pathway (Sweetlove and Fernie 2013). The primary endosymbiosis event, whereby a cyanobacterial ancestor was engulfed by a protist-like, heterotrophic eukaryote (Gould, et al. 2008), significantly changed eukaryotic evolution. Consequently, thousands of genes were transferred from the cyanobacterium to the nucleus of the eukaryotic host in a process known as endosymbiotic gene transfer (Timmis, et al. 2004). As expected, a large portion of the transferred genes encoded proteins that function enzymatically in metabolic pathways and were ultimately integrated into the existing metabolism of the eukaryotic host. Some introduced pathways simply served as alternative routes to synthesize the same metabolites, such as terpenoid backbone biosynthesis (Lange 2000); however, some pathways were entirely novel due to their association with

photosynthesis, e.g. the Calvin-Benson cycle, which is responsible for converting CO_2 into sugars (Martin and Schnarrenberger 1997). Therefore, the merging of cyanobacterial and eukaryotic genomes provided a new set of genes encoding metabolic pathways that could further be shaped by evolution in eukaryotes.

Within the past several years, it has become increasingly apparent that another evolutionary mechanism plays an important role in eukaryotic evolution. The vertical inheritance of genetic material, DNA passed from parent to offspring, was originally thought to be the only form of transferring genetic information between eukaryotes besides endosymbiotic gene transfer. However, horizontal gene transfer, a common mechanism in bacteria, also occurs in eukaryotes. Horizontal gene transfer is the direct transfer of genetic material from one organism to another as opposed to inheritance via reproduction, whereby genetic material is passed vertically from parent to offspring.

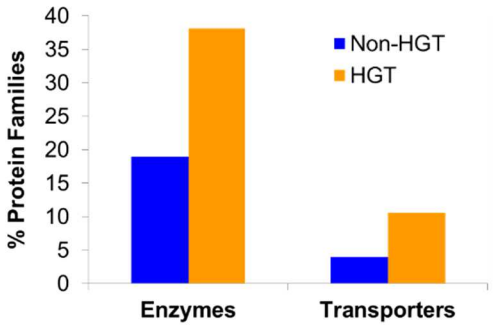


Fig. 4: Comparison between *G. sulphuraria* protein families (both enzymes and transporter proteins) classified as horizontal gene transfer candidates and protein families not considered candidates (Schönknecht, unpublished). Protein families within *G. sulphuraria* were divided into two categories, HGT and Non-HGT, based on protein phylogenies, and the number of introns present and GC% of the respective genes. Then each category was divided into enzymes and transporters. Figure produced by G. Schönknecht.

An overwhelming number of examples, including those where genes encode metabolic pathway enzymes, have shown horizontal gene transfer to occur in eukaryotes, whether it be prokaryote to

eukaryote, or eukaryote to eukaryote gene transfer (Oborník and Green 2005; Kondrashov, et al. 2006; Richards, et al. 2006; Keeling and Palmer 2008; Whitaker, et al. 2009; Qiu, Price, et al. 2013; Qiu, Yoon, et al. 2013; Schönknecht, et al. 2013; Schönknecht, et al. 2014; Ternes and Schönknecht 2014). Furthermore, it has been reported that entire metabolic pathways that were subsequently lost in a lineage have been reacquired via horizontal gene transfer (Nedelcu, et al. 2009; Pombert, et al. 2012). In the specific case of *G. sulphuraria*, numerous protein families (77 total) result either entirely or partly from horizontal gene transfer (Fig. 4). Protein families that are not considered horizontal gene transfer candidates do not necessarily result from eukaryotic ancestry, but may result from endosymbiotic gene transfer, which is not considered as horizontal gene transfer in Figure 4 (Schönknecht, unpublished).

ENDOSYMBIOTIC PLASTID EVOLUTION FUELED RADIATION OF PHOTOSYNTHESIS

Oxygenic photosynthesis first evolved in the ancestor of cyanobacteria, but later spread to eukaryotes through endosymbiosis (Archibald 2009; Dagan, et al. 2013). The earliest photosynthetic, eukaryotic ancestor eventually gave rise to the supergroup Archaeplastida, one of the largest lineages of photosynthetic eukaryotes (Fig. 5). Initially, there was little evidence supporting Archaeplastida as being monophyletic (all groups evolved from a single, common ancestor); however, the surge in sequenced genomes and subsequent phylogenomic studies are providing ever-growing evidence to support the monophyly hypothesis (Bhattacharya, et al. 2004; Keeling 2010; Dorrell and Smith 2011; Burki, Okamoto, et al. 2012). The exact order of divergence within Archaeplastida is unknown. Many scientists studying Archaeplastida evolution propose the glaucophytes diverged first, yet there is little supporting evidence and still remains an active area of debate to this day. The nuclear genome of *C. paradoxa*, the most studied

glaucophyte, was recently published; however, phylogenomic analyses failed to solidify glaucophytes as the basal lineage (Price, et al. 2012).

Following the primary endosymbiosis event, additional endosymbioses resulted in the radiation of other major lineages of photosynthetic eukaryotes. Secondary endosymbiosis events, whereby an ancient Archaeplastida was engulfed by a heterotrophic, protist-like eukaryote, extended photosynthetic ability to other eukaryotic lineages, some of which account for a large portion of phytoplankton biomass in the world's oceans (Archibald 2009; Keeling 2013). There is debate surrounding the origin of the ancient Archaeplastida that was engulfed; yet, the general consensus agrees on a red algal endosymbiont. Of the currently described species regarded as (unicellular) protists, which includes unicellular algae, more than 50% contain secondary plastids (plastids resulting from secondary endosymbiosis) derived from a red alga (Keeling 2010), emphasizing the need to further our understanding of red algal biology and evolution.

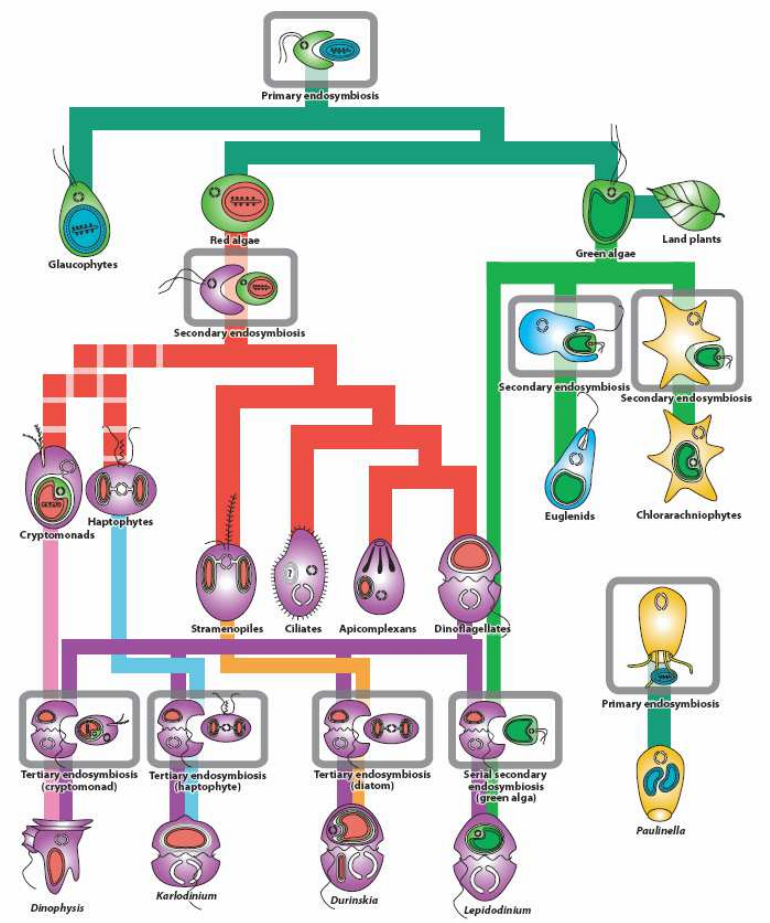


Fig. 5: Schematic of plastid evolution taken from Keeling (2013). From the top, the primary endosymbiosis event established the three major lineages of Archaeplastida (glaucophytes, red algae, green algae and land plants), designated by dark green lines. Plastids present in other linages including cryptomonads (cryptophytes), haptophytes, stramenopiles, ciliates, apicomplexans, and dinoflagellates, which are of red algal origin as a result of secondary endosymbiosis, are indicated by red lines. These individual lineages were then involved in tertiary endosymbiosis with a dinoflagellate host as indicated by pink, cyan, orange, and purple lines depicting the plastid origin in different genera of dinoflagellates. The red lines leading to cryptomonads and haptophytes are dashed due to the uncertainty of evolutionary relationships. While Burki, Okamoto, et al. (2012) proposed explicit relationships (see text), there is still debate regarding where they reside in the tree of life. Lineages containing plastids of green algal origin via secondary endosymbiosis (euglenids, chlorarachniophytes, and *Lepidodinium* [dinoflagellate])

are indicated by light green lines. *Paulinella* appears to have undergone its own independent primary endosymbiosis involving cyanobacteria.

A major eukaryotic lineage with the ability to photosynthesize is the SAR supergroup (Archibald 2009; Burki, et al. 2010), composed of Stramenopiles (brown algae, diatoms, and Oomycota [water molds or downy mildew]), Alveolata (dinoflagellates, apicomplexans, and cilitates), and Rhizaria (chloroarachniophytes) (Fig. 5). The plastids present in stramenopiles and alveolates are the result of a secondary endosymbiosis event whereby a red algal ancestor was engulfed by a heterotrophic, eukaryotic ancestor (Gould, et al. 2008). The plastids present in the chloroarachniophyte *Bigelowiella natans* are also a result of a secondary endosymbiosis event that occurred independently from endosymbiosis events in stramenopiles and alveolates within the SAR supergroup. However, the plastid progenitor in Rhizaria is of green algal origin (Archibald, et al. 2003).

Other algal groups also possess secondarily derived plastids, either of red or green algal origin. Haptophytes, marine phytoplankton that significantly impact global carbon fixation, have been shown to be related to the SAR supergroup. Together, haptophytes and SAR form one, large monophyletic group (Burki, Okamoto, et al. 2012; Read, et al. 2013). Haptophyte plastids are of red algae origin, yet experienced their own endosymbiosis event independently of the groups within SAR (Burki, Okamoto, et al. 2012). Cryptomonads also possess a red algal derived plastid that was acquired independently via secondary endosymbiosis. The position in the tree of life where cryptomonads reside is unknown; however, Burki, Okamoto, et al. (2012) proposed that they are more closely related to Archaeplastida than any other eukaryotic lineage. The number of secondary endosymbiosis events remains unknown and is still a subject of debate. Conversely, the eukaryotic supergroups that possess secondary plastids are not monophyletic, but are para- or polyphyletic (do not directly share a common ancestor) and are a result of their own

independent evolution. Plastid evolution is an important and complex facet of algal biology. Many characteristics of plastid evolution, such as the presence or absence of metabolic pathways (taken together with phylogenetics) and photosynthetic pigments, provide more information regarding the potential origin of plastids present in a particular algal group. The aforementioned characteristics also shed light on early eukaryotic biology. Unraveling the fine details of early eukaryotic evolution will aid in resolving the deepest regions in the tree of life.

THE SWEET LIFE OF *G. SULPHURARIA*

G. sulphuraria is able to grow on more than 50 different carbon sources, a feat unrivaled by any other alga (Oesterhelt, et al. 1999; Weber, et al. 2004). The high degree of metabolic diversity is directly attributed to the enzymes (and transporters [see below]) encoded by the *G. sulphuraria* genome. For example, several different types of sugar kinases not only play a role in sugar sensing (Oesterhelt and Gross 2002), but directly phosphorylate sugars that can subsequently be funneled into various metabolic pathways. These kinases include **glucokinase** (Gasu_04750, EC 2.7.1.2), **galactokinase** (Gasu_09360, EC 2.7.1.6), **fructokinase** (Gasu_01610 & Gasu_04420, EC 2.7.1.4), **glycerol kinase** (Gasu_03170 & Gasu_60010, EC 2.7.1.30), **xylulokinase** (Gasu_46540, EC 2.7.1.17), and **ribokinase** (Gasu_33210, EC 2.7.1.15) (Barbier 2005). Despite the lack of phylogenetic analyses for the majority of these kinases, they are more similar to prokaryotic enzymes rather than sugar kinases from plants (Heilmann, et al. 1997; Barbier, et al. 2005). *G. sulphuraria* is also able to metabolize sugar alcohols by oxidation catalyzed by dehydrogenases. Following oxidation, the resulting sugar can then be phosphorylated by kinases, such as those mentioned above. A

polyol dehydrogenase, **xylitol dehydrogenase** (Gasu_25510, EC 1.1.1.9), was detected in *G. sulphuraria* using ESTs and biochemical assays (Stein, et al. 1997; Barbier 2005).

Mannose, more specifically mannose-6-phosphate, is a sugar that becomes toxic at increased levels in higher plants (Knudson 1917; Stenlid 1954; Morgan and Street 1959). The lack of substrate specificity by plant hexokinases leads to an accumulation of mannose-6-phosphate. Hexokinase can phosphorylate mannose, in addition to glucose, fructose, and other sugars (Herold and Lewis 1977). When mannose-6-phosphate accumulates, phosphate levels decrease, hindering many processes in higher plants such as ATP synthesis, germination, photosynthesis, and starch synthesis. Elevated mannose-6-phosphate levels also inhibit phosphoglucomutase and glucose phosphate isomerase (Herold and Lewis 1977; Barbier 2005). In contrast, *G. sulphuraria* is able to grow solely on mannose. The genome of *G. sulphuaria* does not appear to encode a hexokinase, but rather a **glucokinase** (Gasu_04750, EC 2.7.1.2); the phosphorylation of mannose has been shown to be catalyzed by glucokinase (Heilmann, et al. 1997). In *G. sulphuraria*, **mannose-6-phosphate isomerase** (Gasu_38450, EC 5.3.1.8) converts mannose-6-phosphate to fructose-6-phosphate, a glycolytic intermediate. Even though some higher plants possess this isomerase and do not exhibit toxicity symptoms, the isomerase in *G. sulphuraria* is highly active, ensuring that mannose toxicity will not occur (Heilmann, et al. 1997). Galactose also presents a problem to most plants where it accumulates as galactose-1-phosphate or UDP-galactose, which inhibits carbon metabolism (Roberts, et al. 1971). The presence of a few key enzymes allows *G. sulphuraria* to metabolize galactose: **galactokinase** (Gasu_09360, EC 2.7.1.6), **(UTP)-galactose-1-phosphate-uridylyltransferase** (Gasu_52840, EC 2.7.7.12), and **UDP-glucose-4-epimerase** (Gasu_12030, EC 5.1.3.2) (Gross and Schnarrenberger 1995b; Prosselkov, et al. 1996; Barbier 2005).

WHAT IS METABOLISM WITHOUT TRANSPORTERS?

Transporters are key components that contribute significantly to *G. sulphuraria*'s ability to utilize a wide spectrum of carbon substrates. There seems to be a correlation between the life style of an organism, heterotrophic and/or mixotrophic, and the number and diversity of membrane transport proteins (Ren and Paulsen 2005). The large number of proteins in *G. sulphruaria* annotated as membrane transport proteins indicates—and is probably essential—to a heterotrophic lifestyle that depends on the uptake of organic compounds (Barbier 2005; Schönknecht, et al. 2013). *G. sulphuraria* encodes numerous 'sugar porters', proteins specialized to transport a particular sugar or carbohydrate, which might have been acquired horizontally from fungi (Schönknecht, et al. 2013). Additionally, some of the transport proteins possess partly overlapping specificities, further contributing to the diversity of substrates that can be taken up (Oesterhelt, et al. 1999). Polyol transport proteins, transporters responsible for importing sugar alcohols, are also present in *G. sulphuraria*. Even though only a few polyol dehydrogenases have been characterized and annotated, numerous sugar alcohols can be taken up by *G. sulphuraria*, including sorbitol, arabitol, xylitol, and dulcitol among others (Oesterhelt, et al. 1999). *G. sulphuraria* can also grow solely on glycerol (Gross and Schnarrenberger 1995a). Five genes have been annotated in the *G. sulphuraria* genome as encoding glycerol transporters, and phylogenetic analysis indicates these transporter proteins were probably acquired via horizontal gene transfer from Proteobacteria (Schönknecht, et al. 2013). Of all the proteins encoded by the *G. sulphuraria* genome, 5.2% are annotated as membrane transporters. As expected and given the metabolic diversity of *G. sulphuraria*, the number of encoded transporters is more than for most eukaryotes, especially when compared to *C. merolae*, the other sequenced Cyanidiophyceae red alga (Fig. 6) (Schönknecht, et al. 2013).

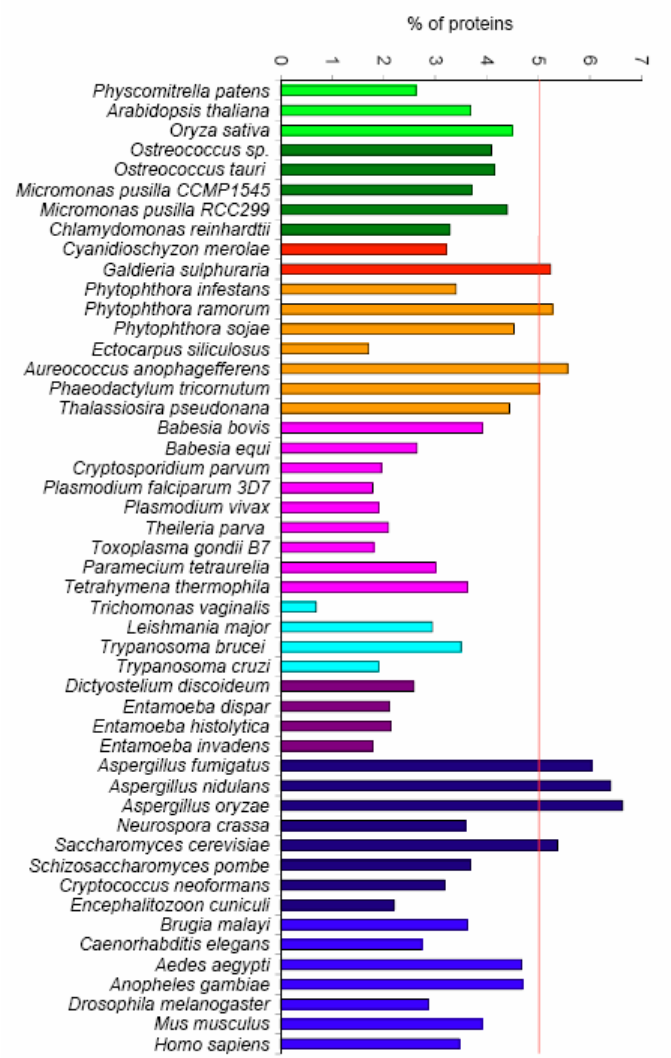


Fig. 6: The percentage of proteins annotated as transporter proteins for 49 various eukaryotes taken from Schönknecht, et al. (2013). The red line indicates the 5% level of which *G. sulphuraria* surpasses as do other organisms such as fungi and two stramenopiles that are capable of heterotrophic growth (Tyler, et al. 2006; Gobler, et al. 2011).

ADAPTATION VIA HORIZONTALLY ACQUIRED TRANSPORTERS

The heavy metal and salt tolerance exhibited by *G. sulphuraria* is made possible due in part to transporters. Arsenic is removed from the cell by first being reduced from arsenate

to arsenite via **arsenate reductase** (Gasu_63270, EC 1.20.4.1). Arsenite is then pumped out of the cell by ArsB, a plasma membrane arsenic pump. Phylogenetic analysis revealed the ArsB to be of prokaryotic origin, namely from thermoacidophilic species, and was probably acquired via horizontal gene transfer. Furthermore, both genes encoding ArsB (Gasu_31570 & Gasu_56050) do not contain any introns, whereas 72.4% of genes in *G. sulphuraria* do contain introns (Schönknecht, et al. 2013). Salt tolerance is also partially achieved by employing a membrane transporter. The monovalent cation/proton antiporter, i.e. Na^+:H^+ antiporter, allows *G. sulphuraria* to exchange sodium ions with protons. Seven different genes encode this type of antiporter. Based on their position in a phylogenetic tree, five of the antiporters are localized either to the plasma membrane or internally, i.e. an organellar membrane. The localization of the other two antiporters are currently unknown; however, the genes encoding them have been acquired via horizontal gene transfer from bacteria (Schönknecht, et al. 2013).

WHERE THERE'S LIFE, THERE'S NAD^+

NAD^+, nicotinamide adenine dinucleotide, is a coenzyme essential for each living cell. NAD^+ functions in redox biochemistry and energetic metabolism as an electron carrier and cofactor of oxidoreductases without being consumed. NAD^+ is consumed in ADP-ribose transfer reactions where it serves as a substrate for the synthesis of the Ca^{2+}-mobilizing second messengers cyclic ADP ribose (cADPR) and nicotinic acid adenine dinucleotide phosphate (NAADP) (Belenky, et al. 2007; Pollak, et al. 2007; Noctor, et al. 2011). These NAD^+-consuming reactions require continuous *de novo* biosynthesis. In addition, so-called salvage pathways recycle components containing a nicotinamide ring (i.e., nicotinic acid, nicotinamide, nicotinamide ribose) (Fig. 2); these components result from NAD^+ cleavage or

are taken up with food (in heterotrophic organisms) (Preiss and Handler 1958; Katoh and Hashimoto 2004). *De novo* NAD^+ biosynthesis in all free-living, eukaryotic organisms is achieved by one of two *de novo* pathways, the aspartate pathway (Griffith, et al. 1975) or the kynurenine pathway starting with tryptophan (Gaertner and Subbayya Shetty 1977). Both pathways converge at the intermediate quinolinate, which in three additional reaction steps is converted into NAD^+ (Fig. 2).

The kynurenine pathway has long been characterized for animals and fungi (Beadle, et al. 1947), but more recently for bacteria (Kurnasov, et al. 2003). In humans, metabolic intermediates of the kynurenine pathway serve an important role in neurophysiology and have implications for neurodegenerative diseases (Schwarcz, et al. 2012). Yeast strains lacking any of the kynurenine pathway enzymes require nicotinic acid for normal growth (Panozzo, et al. 2002). With the sequencing of two red algae, *C. merolae* (Matsuzaki, et al. 2004) and *G. sulphuraria* (Schönknecht, et al. 2013), it became clear that red algae employ the kynurenine pathway as well (Schönknecht, et al. 2013; Ternes and Schönknecht 2014). In contrast, land plants and green algae encode the aspartate pathway, which is commonly referred to as the bacterial pathway (Katoh, et al. 2006; Lin, et al. 2010). The aspartate pathway is localized to the plastid in *Arabidopsis thaliana*, and *A. thaliana* genes encoding aspartate pathway enzymes have been shown to complement *E. coli* knockout mutants (Katoh, et al. 2006). Green algae *Chlamydomonas reinhardtii* mutants lacking any of the five genes converting aspartate into NAD^+ require supplemental nicotinamide for growth (Lin, et al. 2010).

GALDIERIA SPP. ALGAE EXCRETE PORPHYRINS

When supplemented with aminolevulinic acid (ALA), *Cyanidium caldarium* was observed excreting fluorescent pigments into the culture medium (Troxler and Bogorad

1966). In much of the older literature (before sequencing technology rapidly improved), *G. sulphuraria* was commonly referred to as *Cyanidium caldarium*. After the sequencing of the *G. sulphuraria* genome (sequenced in 2004), it became known that what many had thought was *C. caldarium*, was indeed *G. sulphuraria*. Later, another species of Cyanidiophyceae red algae, *Galdieria partita*, was also observed excreting fluorescent pigments during cultivation when supplemented with glucose (Stadnichuk, et al. 1998). Identification of the pigments was achieved by both Troxler and Bogorad (1966) and Stadnichuk, et al. (1998) via chromatography in conjunction with absorption and emission (fluorescence) spectroscopy. Both studies concluded the pigments were porphyrins, cyclic tetrapyrroles that serve as precursor metabolites for other photosynthetic pigments including chlorophyll and cytochromes.

Troxler and Bogorad (1966) did not speculate as to the potential function of porphyrin excretion, nor the mechanism by which it occurred, because porphyrin excretion was only observed when *C. caldarium* cultures were supplemented with δ-aminolevulinic acid (ALA), the first metabolite in the tetrapyrrole biosynthetic pathway. Likewise, porphyrin excretion in *C. caldarium* was also observed in a later study and appeared to be dependent upon incubation with an additional compound, specifically N-methylprotoporphyrin IX (NMP), a phycocyanobilin biosynthesis inhibitor; incubations with both ALA and NMP resulted in porphyrin excretion (Brown, et al. 1982). Growth studies conducted on *G. partita* observed excretion of porphyrins when supplied with exogenous glucose (Stadnichuk, et al. 1998). The excretion of porphyrins by *G. partita* was attributed to an inhibitory effect on chlorophyll *a* and phycocyanobilin biosynthesis caused by D-glucose. Yet, Stadnichuk, et al. (1998) provide no evidence supporting inhibition of pigment biosynthesis by D-glucose. It seems that the conclusions drawn by Stadnichuk, et al. (1998), whereby porphyrin excretion is the result of an inhibitory effect by D-glucose, are merely based on the similar results

presented by Troxler and Bogorad (1966) and Brown, et al. (1982). In mature leaves, a correlation between D-glucose and the inhibition of photosynthetically-driven carbohydrate biosynthesis, owing to source-sink regulation, has been previously observed in terrestrial plants (Sheen 1990; Rolland, et al. 2006). Given that all species within Cyanidiophyceae are unicellular, source-sink regulation hardly seems applicable. Furthermore, Stadnichuk, et al. (1998) made direct comparisons between heterotrophic and photoautotrophic cultures despite the approximately 25-fold difference in cell density. There was no mention of the correlation between porphyrin excretion and overall cell density of the cultures. Thus, further investigation is required surrounding the potential 'trigger' for porphyrin excretion, in addition to the biological significance of the phenomenon.

CHAPTER III

METHODOLOGY

ENZYME AND METABOLIC PATHWAY ANNOTATION

A preliminary protein annotation spreadsheet assembled by bioinformatics scripts and pipelines was initially provided by Wei-Hua Chen to begin enzyme and metabolic pathway annotation. The spreadsheet included all relevant information about the protein sequences, i.e. GeneID number (Gasu_#), amino acid sequence and length, in addition to the best bi-directional BLASTP (Altschul, et al. 1997) hit at NCBI (Pruitt, et al. 2007) using the non-redundant database, the *C. merolae* genome BLAST database (http://merolae.biol.s.u-tokyo.ac.jp), and The Arabidopsis Information Resource (Lamesch, et al. 2011). Gene Ontology (The Gene Ontology, et al. 2000) and Mercator (Lohse, et al. 2014) BLAST results were also provided for all protein sequences. Both annotation programs classify protein sequences by assigning a 'functional term' to a particular protein sequence, e.g. amino acid biosynthesis, lipid metabolism, etc., that can be used for organizing and annotating protein sequences. Along with the preliminary spreadsheet mentioned above, another spreadsheet was provided containing BLASTP results generated from the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto 2000). All predicted protein sequences (6,623 in total) from *G. sulphuraria* were submitted to BLASTP at KEGG, and the best BLASTP hit (BLASTing against all organisms in the KEGG database) for each predicted protein sequence was recorded. Furthermore, in addition to the 'BLAST against

all'approach, all predicted protein sequences from *G. sulphuraria* were BLASTed against different, organismal databases at KEGG (plants and fungi, archaea, bacteria, and cyanobacteria) due to the unique characteristics of *G. sulphuraria*, i.e. high frequency of horizontal gene transfer and high degree of metabolic diversity. All best bi-directional BLASTP hits were recorded for each predicted protein sequence.

The Kyoto Automatic Annotation Server (KAAS) (Moriya, et al. 2007) was used to generate a set of preliminary metabolic pathway maps based on the pathway map templates used by KEGG. KAAS assigns enzyme classes (EC), which are organized by the type of reactions, to protein sequences based on 'best BLAST hit information'. For example, a submitted query protein sequence is BLASTed against the protein sequences in the KEGG database, and based on the degree of 'best hit information', the query protein sequence is assigned to an enzyme class associated with a particular enzymatic reaction (see Moriya, et al. (2007) for details). Five different sets of KEGG metabolic pathway maps were generated in total for *G. sulphuraria*. The first set of pathway maps were based on the 'BLAST against all' approach mentioned above; no specific organismal database at KEGG was selected to serve as a 'reference' for generating pathway maps. The remaining four sets of pathway maps were generated from using the specific organismal databases mentioned above to serve as a 'reference', i.e. plants and fungi, archaea, bacteria, and cyanobacteria. All annotation data generated from KAAS, such as GeneID number (Gasu_#), EC number, and predicted enzyme annotation (enzyme class assignment via KAAS) was compiled in a spreadsheet to directly compare the different KAAS protein annotations using the 'BLAST against all' approach and the four different 'reference' databases, e.g. KAAS results could potentially provide five different annotations for a given predicted protein sequence. The resulting five different sets of metabolic pathway maps were manually inspected and consolidated to construct a final set of pathway maps using the KEGG Search & Color module (91 individual pathway maps in total). If the KAAS annotations for a predicted protein sequence was unanimous

for at least three of the five KAAS annotation assignments, the protein sequence was assigned that particular annotation after performing an additional BLASTP search at NCBI using the non-redundant database and conserved domain searches using the Conserved Domain Database (CDD) (Marchler-Bauer, et al. 2011), InterProScan (Zdobnov and Apweiler 2001), and HMMer (Finn, et al. 2011) to ensure congruency.

For instances where no protein sequence was assigned to a specific enzyme class, i.e. there was a 'gap' in a metabolic pathway, reference sequences from other organisms were used to BLAST against the *G. sulphuaria* genome (http://genomics.msu.edu/cgi-bin/galdieria/blast.cgi) to manually search for potential candidates. Reference sequences were retrieved from databases including NCBI RefSeq (Pruitt, et al. 2007), TAIR (Lamesch, et al. 2011), and BRENDA (Scheer, et al. 2011). Subsequent protein annotations were made with the aid of BLAST searches from NCBI and Uniprot (Consortium 2014) and conserved domain searches using the Conserved Domain Database (CDD) (Marchler-Bauer, et al. 2011), InterProScan (Zdobnov and Apweiler 2001), and HMMer (Finn, et al. 2011). In several instances the KAAS generated false positives for a given enzyme. False positives usually resulted from using bacteria as the 'reference' organismal database, whereas the other three 'reference' databases (fungi and plant, archaea, and cyanobacteria) usually did not make an assignment. Putative false positives were not immediately discredited due to the potential of being a horizontally acquired gene. Protein sequences were subjected to the battery of BLAST and conserved domain searches mentioned above. If BLAST and domain searches were congruent with the KAAS annotation using bacteria as the 'reference' organismal database, the protein was considered to be a horizontal gene transfer candidate and was further investigated phylogenetically and statistically by Gerald Schönknecht.

Numerous studies, including EST datasets, have previously reported characterized and assayed enzymes from *G. sulphuraria*. Table A1 (appendix) shows a compilation of the literature

describing enzymes in *G. sulphuraria* in comparison with enzymes that have been annotated in the genome of *G. sulphuraria*.

PHYLOGENETIC TREE CONSTRUCTION AND TOPOLOGY TESTING (*DE NOVO* NAD^+ BIOSYNTHESIS)

Homologous protein sequences were identified by BLAST (Altschul, et al. 1997) searches at NCBI (Pruitt, et al. 2007), KEGG (Kanehisa, et al. 2014), http://genome.jgi.doe.gov/, http://www.broadinstitute.org, http://cyanophora.rutgers.edu, and collected using MEGA 5 (Tamura, et al. 2011). Individual BLAST searches were carried out separately for different clades for broad phylogenomic sampling, e.g. an animal sequence was used as a query to BLAST against all animals. Both incomplete and/or highly redundant sequences were removed. No sequences were allowed to share greater than 90% amino acid identity in attempt to remove highly similar (or redundant) sequences, i.e. multiple isoforms. The NCBI GeneID number for each sequence is displayed in the phylogenetic tree. Multiple sequence alignments were generated with T-COFFEE (Notredame, et al. 2000) using the 'accurate' mode. The 'accurate' mode BLASTs all sequences (to be aligned) at NCBI to generate a sequence profile. The sequence profiles are used to search the database (PDB) (Berman, et al. 2000) looking for structural templates. The structural templates are then aligned, which are used with homology extension for generating the multiple sequence alignment (Di Tommaso, et al. 2011; Taly, et al. 2011). T-COFFEE 'scores' each position within the alignment based on the degree of conservation with '9' being the most conserved and '1' being the least conserved. High scoring portions of multiple sequence alignments (T-COFFEE score 5 to 9) were extracted in order to retain the most highly conserved regions of the (amino acid) multiple sequence alignment and also remove regions containing lengthy gaps (Talavera and Castresana 2007). The scenario in which a different score was used is when extraction of high scoring blocks (5 to 9) resulted in an overall alignment length

of less than 200 amino acids. In these cases (nicotinate-nucleotide adenylyltransferase and tryptophan 2,3-dioxygenase alignments) a score of 3 to 9 was used instead. Deviations from the aforementioned 5 to 9 high scoring blocks extraction are also mentioned in the figure legends. The best model of protein evolution estimated to fit the multiple sequence alignment best was generated by ProtTest (Abascal, et al. 2005). The best model, in most cases LG+I+G (deviations mentioned in figure legend), was used to generate phylogenetic trees with PhyML 3.0 (Guindon and Gascuel 2003) estimating branch support values by non-parametric bootstrap with 200 replicates. The LG model of protein evolution is not implemented in the MPI version of MrBayes v3.1.2 (Altekar, et al. 2004), and the best model implemented for all but one multiple sequence alignment (specified in figure legend) was WAG+I+G. MrBayes was run with 12 chains at a temperature of 0.05 for 5,000,000 generations sampling every 100th generation. The first 25% of samples were considered 'burn in', thus being ignored when calculating parameters and consensus tree. Trace files from Bayesian MCMC runs were inspected with Tracer 1.6 (http://beast.bio.ed.ac.uk/Tracer) to ensure that likelihoods beyond the 25% cutoff had converged and stabilized, and also showed an approximately Gaussian distribution. The phylogenetic trees displayed are unrooted Bayesian trees with larger clades collapsed. Support values >50% are shown above the branches. Bootstrap support values as calculated by PhyML 3.0 are shown as percentages below the branches; only support values >50% are given.

Alternative evolutionary scenarios, with specific clades constrained as monophyletic, were tested using likelihood values. All constrained and unconstrained Bayesian trees were constructed using MrBayes v3.1.2, and the harmonic mean (H) and Akaike's information criterion through Markov chain Monte Carlo (AICM) (Baele, et al. 2012) of log-likelihood values from the stationary phase of Bayesian analyses were calculated by Tracer 1.6. Differences between harmonic means (ΔH) and differences between AICM values (ΔAICM) usually showed a ratio close to $\Delta H \approx \Delta$AICM/-2, as expected. Constraint trees were regarded significantly different from

the unconstrained tree when $\Delta H > 3$ SD (SD, standard deviation of log-likelihood values as calculated by Tracer 1.6, ranging from 10.6 to 16; using the larger SD between the constrained and unconstrained tree), and $-\Delta AICM > 6$ SD. Constrained and unconstrained maximum likelihood trees were generated using RAxML 8.0 (Stamatakis 2006) using the same model of protein evolution as for PhyML 3.0 mentioned above. However, a defined number of bootstrap replicates was not specified. Rather, RAxML 8.0 was allowed to determine the adequate number of bootstrap replicates, i.e. bootstopping (-# autoMRE command), based on convergence. Once support values stabilized, or converged, the current number of bootstrap replicates was considered sufficient. The best scoring (lowest log-likelihood) maximum likelihood tree from the bootstrap analysis was used for further statistical analysis in both cases of unconstrained and constrained trees. RAxML 8.0 was also used to calculate site-likelihood values to use as input for CONSEL (Shimodaira and Hasegawa 2001) to calculate *p*-values of the Kishino-Hasegawa (KH) test and the approximate unbiased (AU) test (Shimodaira 2002). At *p*-values <5%, log-likelihood values of constraint trees were regarded as significantly lower.

G. SULPHURARIA LIQUID CULTURES

Phototrophic *G. sulpharia* cells were grown in liquid medium according to Gross and Schnarrenberger (1995a). Liquid culture medium was supplemented with D-glucose, yielding final concentrations of 50 mM, 100 mM and 200 mM. Culture flasks were wrapped in aluminum foil, incubated at 37°C, and shaken continuously at 200 rpm. Biomass (absorbance at 800 nm), porphyrin concentration (absorbance at 400 nm), and glucose concentration were measured every day and every three days after stationary growth was reached. Due to the relatively long (initial) lag phase of *G. sulphuraria* cells, cultures were not successfully grown continuously in the exponential growth phase, as is routinely done for algal cultures in general. Glucose concentrations were measured using the dinitrosalicylic acid assay (Miller 1959).

SPECTROSCOPY & RELATIVE FLUORESCENCE QUANTUM YIELD

The method for determining the relative fluorescence quantum yield (Φ_F) of the porphyrin mixture excreted by *G. sulphuraria* was adapted from Würth, et al. (2013) and Horiba (2014). Experimental determination of the *absolute* fluorescence quantum yield is complex and requires special equipment. In practice, one usually determines the *relative* fluorescence quantum yield (Crosby and Demas 1971; Würth, et al. 2013). The relative fluorescence quantum yield was determined by comparing the porphyrin mixture to well-characterized fluorescence standards, i.e. Coumarin 102 and Coumarin 153, whose fluorescence quantum yields are known (Rurack and Spieles 2011). Using the equation below (Eq. 1), the relative fluorescence quantum yield (Φ_F) for the porphyrin mixture was determined.

$$\Phi_{F(X)} = \Phi_{F(ST)} \frac{A_{ST} \cdot F_X \cdot \eta_X^2}{A_X \cdot F_{ST} \cdot \eta_{ST}^2}, \qquad \text{Eq. (1)}$$

where A is the absorbance at the excitation wavelength, F is the integrated area under the corrected emission curve (expressed in number of photons), and η is the refractive index of the solvents used; subscripts ST and X refer to the standard and to the substance of interest, respectively. For the equation above, it is critical to ensure that the absorbance values for both the standard and the unknown sample are equal at the desired wavelength. Significant differences in absorbance values between the standard and the unknown sample at the desired wavelength requires considerably more calculations in addition to Eq. 1 above (Würth, et al. 2013). The wavelength at which absorbance spectra overlap, i.e. absorbance values for both the standard and the unknown sample are equal at a given wavelength, is referred to as a 'cross point' (Fig. 7). By using the wavelength at which the cross point (of the absorption spectra) resides as the excitation wavelength, it can be assumed that both samples are absorbing the same amount of light at that particular wavelength. Thus, the amount of fluoresced light is a direct result of the molecule's fluorescence quantum yield, i.e. if both samples are excited with the same number of photons

(equal absorbance values), but one sample has a larger fluorescence intensity, that is a relative indicator of the fluorescence quantum yield.

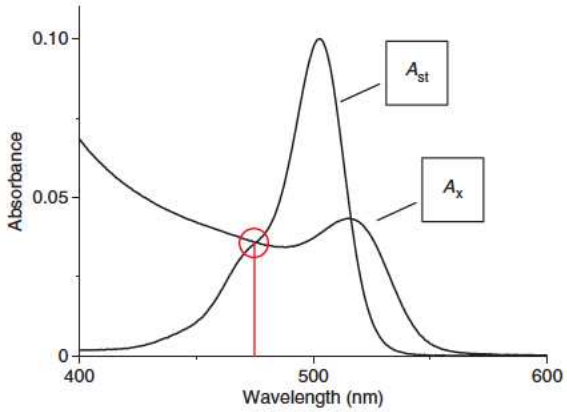
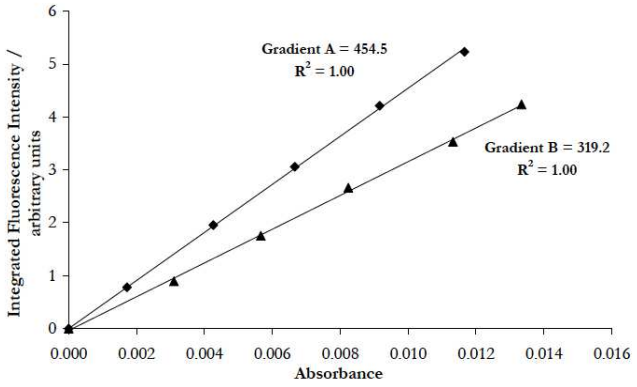


Fig. 7: Using overlapping absorption spectra to determine the excitation wavelength for emission spectra taken from Würth, et al. (2013). The position at which the two absorption spectra overlap (indicated by red circle) signifies a cross point. The wavelength at which the cross point resides indicates the best wavelength to be used as the excitation wavelength for collecting emission spectra (red line below red circle).

Fig. 8: Linear regression showing integrated fluorescence as a function of absorbance [$F=f(A)$] taken from Horiba (2014). The slope values for Gradient A and Gradient B are determined via linear regression, which are proportional to the fluorescence quantum yield, and are used as the *Grad* values in Eq. 2 and Eq. 3 below.



Emission (or fluorescence) spectra were then collected for all of the concentration dilutions for both the standard and the unknown sample using the cross point wavelength as the excitation wavelength. The resulting emission spectra were integrated individually (summation of the area under the emission curve), to determine the number of photons that were fluoresced by the standard and the unknown sample. A series of both absorbance and emission spectra were recorded at five different concentrations for the standard and the unknown sample to plot $F=f(A)$

(integrated fluorescence as a function of absorbance), where the slope was determined via linear regression ($Grad = F/A$) (Fig. 8). By using identical solvents with identical refractive indices ($\eta_X^2 = \eta_{ST}^2$, when possible) for both, the standard and the unknown sample, Eq. 1 can be simplified to

$$\Phi_{F(X)} / \Phi_{F(ST)} = Grad_X / Grad_{ST} , \qquad \text{Eq. (2)}$$

where *Grad* is the slope value as determined by linear regression. Using the slope values generated via linear and the published fluorescence quantum yield value for the standard, determination of the relative fluorescence quantum yield for the unknown sample was achieved.

Absorption spectra were collected with a Shimadzu UV-1800 UV-Vis spectrophotometer using Hellma 10 mm 104B-QS cuvettes. Fluorescence standards Coumarin 102 (CAS 41267-76-9) and Coumarin 153 (CAS 53518-18-6) were purchased from Sigma-Aldrich, and were chosen as the fluorescence standards because they both exhibited similar absorption spectra when compared to the absorption spectra of porphyrin mixture excreted by *G. sulphuraria* (see Findings, Fig. 41). All solutions were diluted to concentrations with low absorbance values ($A < 0.1$) at the excitation wavelength to minimize inner filter effects. Two major components contribute to inner filter effects: reabsorption of fluoresced or emitted light and insufficient excitation of the sample, both of which can yield lower, underestimated values for the total amount of fluoresced light (Dhami, et al. 1995; Lakowicz 2009). Reabsorption, whereby the sample reabsorbs fluoresced light, was assumed to be a nonissue due to the large Stoke's shift (distance between maximum absorption peak and maximum emission peak) exhibited by both fluorescence standards and the porphyrin mixture. Overlap between the absorption and emission spectra was insignificant for both fluorescence standards and the porphyrin mixture. Another component contributing to inner filter effects is high sample concentrations. High sample concentrations (where $A > 0.1$) can lead to unevenly distributed excitation light throughout the

sample, whereby not all of the molecules are equally excited, again resulting in underestimated values of total fluoresced light (Lakowicz 2009).

The maximum absorption peak for the porphyrin mixture was 400 nm, the wavelength at which light was strongly absorbed. Therefore, the excitation wavelength used for determining the fluorescence quantum yield needed to be as close to 400 nm as possible. Preliminary absorption spectra were collected in order to determine which concentrations of both fluorescence standards and the porphyrin mixture would yield a cross point wavelength closest to 400 nm. The cross point (and excitation) wavelength chosen was 404 nm. All of the absorbance spectra for both fluorescence standards and the porphyrin mixture crossed at 404 nm. Five different concentration dilutions, all with $A < 0.1$ and absorption spectra crossing at 404 nm at each respective concentration, were made for both fluorescence standards and the porphyrin mixture.

Emission (fluorescence) spectra were collected using a Horiba Fluoromax-3 fluorescence spectrometer with Hellma 10 mm 111 cuvettes using an excitation wavelength of 404, the cross point wavelength of the absorption spectra. As both fluorescence standards, Coumarin 102 and Coumarin 153, were dissolved in ethanol, a baseline ethanol emission spectrum was subtracted from emission spectra for both Coumarin standards. A baseline water emission spectrum was subtracted from the porphyrin mixture emission spectrum as well. In addition, emission spectra were 'corrected' for all samples. The emission correction factors used to produce 'corrected emission spectra' were previously generated and integrated into the DataMax software by the Fluoromax-3 manufacturer (Horiba) in order to remove any type of bias introduced by the fluorescence spectrometer. This bias includes any (manufacturer specific) settings that may be used by the fluorescence spectrometer that could result in small changes in the emission spectrum simply as a result of which machine was used (Würth, et al. 2013). IGOR Pro 6 (www.wavemetrics.com) was used for integrating the area under emission spectra to determine the total number of photons emitted, or fluoresced. To determine the slope values [$Grad = F(A)$],

when integrated fluorescence was plotted as a function of absorbance [$F=f(A)$], a linear regression was performed using GraphPad Prism 6 (www.graphpad.com).

Both fluorescence standards, Coumarin 102 and Coumarin 153, were cross-calibrated with one another. This step was crucial in order to determine if the published values for the fluorescence standards could be reproduced before attempting to determine the fluorescence quantum yield of the porphyrin mixture. As mentioned above, both Coumarin 102 and Coumarin 153 were dissolved in ethanol; therefore, Eq. 2 was used to calculate the relative fluorescence quantum yield for the fluorescence standards. However, the porphyrin mixture was dissolved in water. Thus, Eq. 3, which still accounts for solvents with differing refractive indices, was used to calculate the relative fluorescence quantum yield of the porphyrin mixture excreted by *G. sulphuraria*.

$$\Phi_{F(X)} = \Phi_{F(ST)} \frac{Grad_X \cdot \eta_X^2}{Grad_{ST} \cdot \eta_{ST}^2} \qquad \text{Eq. (3)}$$

CHAPTER IV

FINDINGS

In total, 91 KEGG metabolic pathway maps were annotated and constructed for the *G. sulphuraria* genome; however, only 11 out of the 91 are presented here. Core metabolic pathways, many that contain enzymes unique to *G. sulphuraria*, are shown below. Furthermore, metabolic pathway maps associated with carbon metabolism are shown to emphasize the diversity of carbon metabolism in *G. sulphuraria*, in addition to pathway maps depicting metabolic pathways unique to *G. sulphuraria*. Not all pathways possess enzymes encoded by genes acquired via horizontal gene transfer, as is common in *G. sulphuraria*, but display a metabolic diversity otherwise not characteristic for land plants and algae.

CORE METABOLIC PATHWAYS IN *G. SULPHURARIA*

As expected, *G. sulphuraria* possesses the core carbon metabolic pathways common throughout all eukaryotes. Glycolysis and a complete citric acid (TCA) cycle including the pyruvate dehydrogenase complex were annotated (Figs. 9 & 10). Similar to animals, photosynthetic eukaryotes require the ability to synthesize glucose from organic acids, such as lactate and pyruvate, in a process known as gluconeogenesis (Rylott, et al. 2003). The gluconeogenesis pathway is very similar to glycolysis, but occurs in reverse for carbohydrate biosynthesis. Both glycolysis and gluconeogenesis employ the same enzymes, with a few exceptions where the enzymatic reactions for glycolysis are thermodynamically unfavorable for

gluconeogenesis (Voet and Voet 2004). These irreversible enzymatic reactions of glycolysis are replaced in the gluconeogenic pathway with enzymes favoring the reverse reaction. *G. sulphuraria*, in addition to land plants and algae, share two enzymes in common with animals for gluconeogenesis, **phosphoenol-pyruvate carboxykinase** (Gasu_48040, EC 4.1.1.49), which synthesizes phosphoenolpyruvate from oxaloacetate, and **fructose-1,6-bisphosphatase** (4 genes, EC 3.1.3.11), which generates fructose-6-phosphate (Fig. 9). Plants and most algae, including *G. sulphuraria* and red algae, do not encode the animal-characterized gluconeogenic enzyme pyruvate carboxylase (EC 6.4.1.1) (Fig. 10), which synthesizes oxaloacetate. Although oxaloacetate is a TCA cycle intermediate, it also serves an intermediate in gluconeogenesis in animals. Instead, *G. sulphuraria* encodes a **pyruvate, water dikinase** (Gasu_42070, EC 2.7.9.2) (Fig. 12), which yields phosphoenolpyruvate (PEP) directly from pyruvate, bypassing oxaloacetate formation. Land plants and algae can also catalyze the formation of phosphoenolpyruvate directly from pyruvate; however, they possess pyruvate, orthophosphate dikinase (EC 2.7.9.1) instead of pyruvate, water dikinase (EC 2.7.9.1). It appears that gluconeogenesis in *G. sulphuraria* does not require the intermediate oxaloacetate, similar to land plants and algae. Bypassing the oxaloacetate intermediate also allows *G. sulphuraria* to operate gluconeogenesis, while sparing oxaloacetate for the TCA cycle. *G. sulphuraria* possesses a **glucose-6-phosphate isomerase** (Gasu_20100 & Gasu_55050, EC 5.3.1.9) (Fig. 9), which yields either α-D-glucose-6-phosphate or β-D-glucose-6-phosphate. It appears that α-D-glucose-6-phosphate and/or β-D-glucose-6-phosphate represent the end product of gluconeogenesis in *G. sulphuraria*, similar land plants and most algae, whereas α-D-glucose is the gluconeogenic end product in animals via glucose-6-phosphatase (EC 3.1.3.9) (Fig. 9).

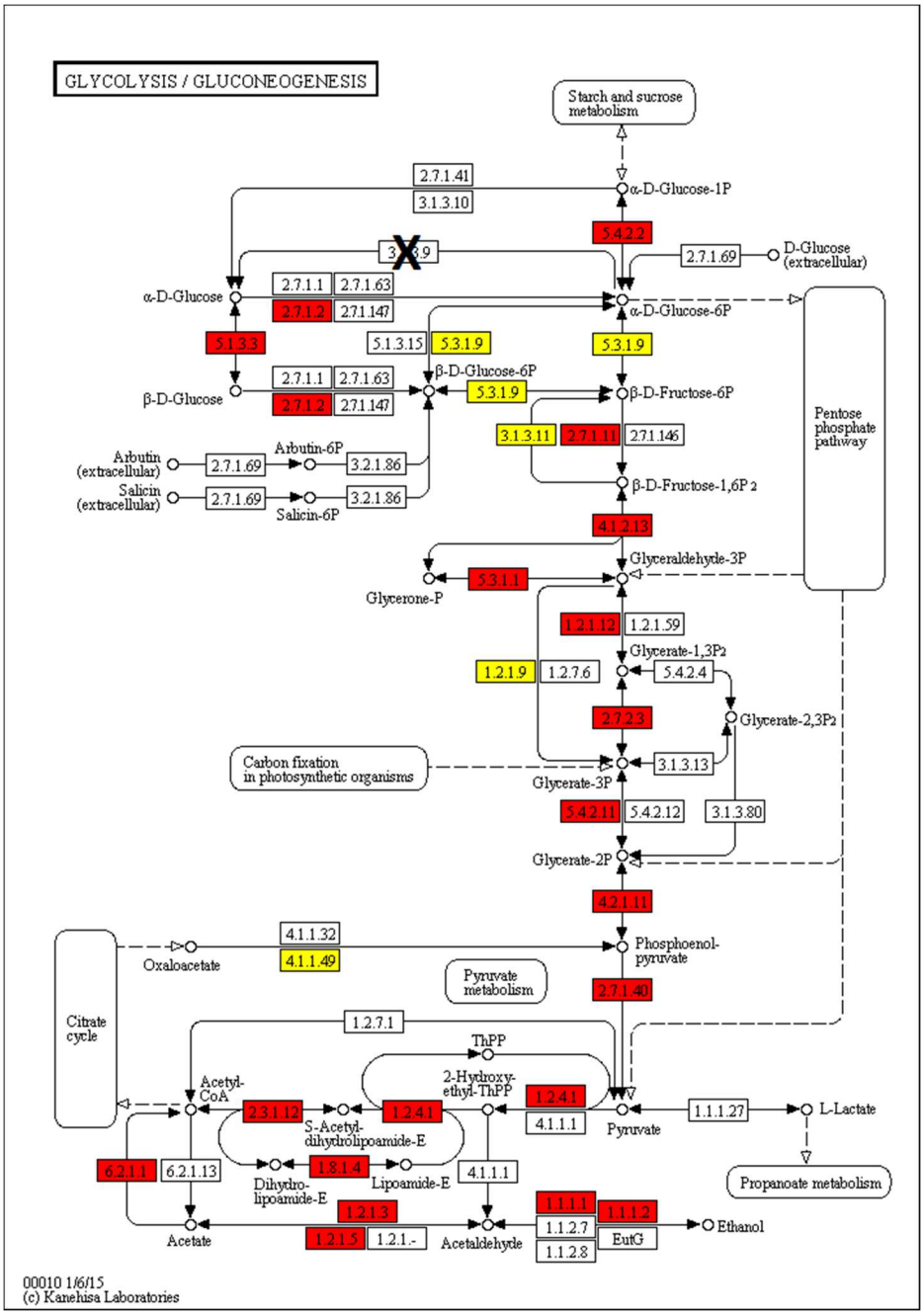


Fig. 9: Glycolysis and gluconeogenesis pathways (KEGG pathway 00010 (Kanehisa, et al. 2014)). Enzymes marked in red and yellow have been annotated in the *G. sulphuraria* genome. Enzymes marked in yellowed have been marked for easy identification and are directly referenced in the main text. The enzyme covered by large **X** (glucose-6-phosphatase, EC 3.1.3.9), commonly annotated in the gluconeogenesis pathway in animals, could not be detected in the *G. sulphuraria* genome.

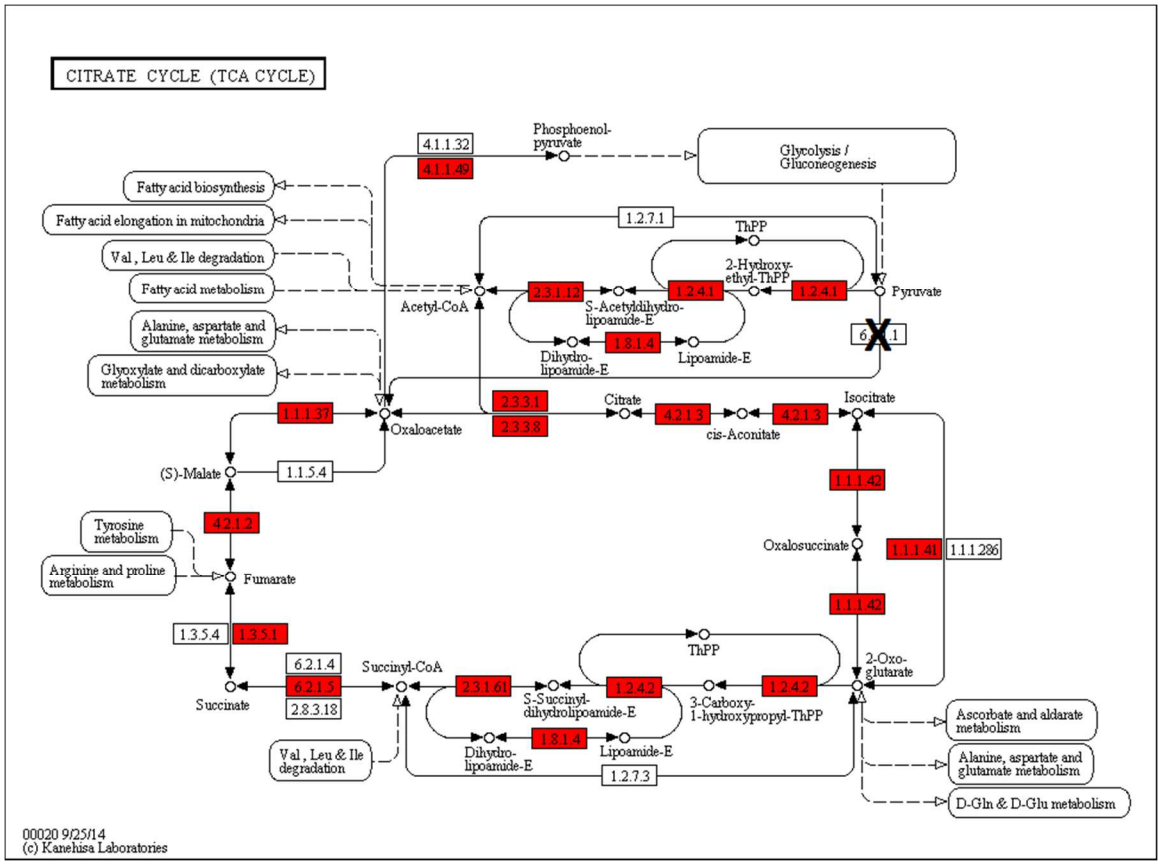


Fig. 10: The TCA cycle (KEGG pathway 00020 (Kanehisa, et al. 2014)). Enzymes marked in red have been annotated in the *G. sulphuraria* genome. Pyruvate carboxylase (EC 6.4.1.1) was surprisingly absent from the *G. sulphuraria* (indicated by large **X**) despite being characteristic for green algae and localized in the plastid of the haptophyte alga *Emiliania huxleyi*, which possesses a plastid of red algal origin (Tsuji, et al. 2012).

Of the alternative glycolytic reactions present in green plants, but absent in most other eukaryotes, **nonphosphorylating glyceraldehyde-3-phosphate dehydrogenase** (Gasu_00840 & Gasu_43210; EC 1.2.1.9) (Fig. 9), which functions in carbohydrate partitioning and storage in heterotrophic cells (Piattoni, et al. 2011), is present in *G. sulphuraria*. Additionally, *G. sulphuraria* encodes a **pyrophosphate-dependent phosphofructokinase** (Gasu_20900, EC 2.7.1.90), which has been proposed to be involved in stress response, such as low oxygen and phosphate availability (Mustroph, et al. 2007) (Fig. 11). Phosphoenolpyruvate phosphatase (EC

3.1.3.60; not shown), which also produces pyruvate from phosphenolpyruvate and functions in land plants to bypass **pyruvate kinase** (4 genes, EC 2.7.1.40) (Fig. 12) under decreased phosphate levels (Duff, et al. 1989), could not be detected in the *G. sulphuraria* genome. Similar to other eukaryotes, fructose 2,6-bisphosphate is synthesized and degraded by a bifunctional enzyme, **6-phosphofructo-2-kinase / fructose-2,6-bisphosphatase** (Gasu_45440 & Gasu_55220, EC 2.7.1.105, EC 3.1.3.46) (Fig. 11). Fructose-2,6-bisphosphatase is a crucial enzyme in plant physiology and metabolism. Fructose-2,6-bisphosphate, the substrate for fructose-2,6,-bisphosphatase, controls carbon partitioning by coordinating glycolysis, gluconeogenesis, and sucrose biosynthesis, ultimately coordinating respiration and carbohydrate (sucrose) biosynthesis (Nielsen, et al. 2004). It seems reasonable to assume that fructose-2,6-bisphosphate in *G. sulphuraria* controls carbon partitioning, similar to land plants.

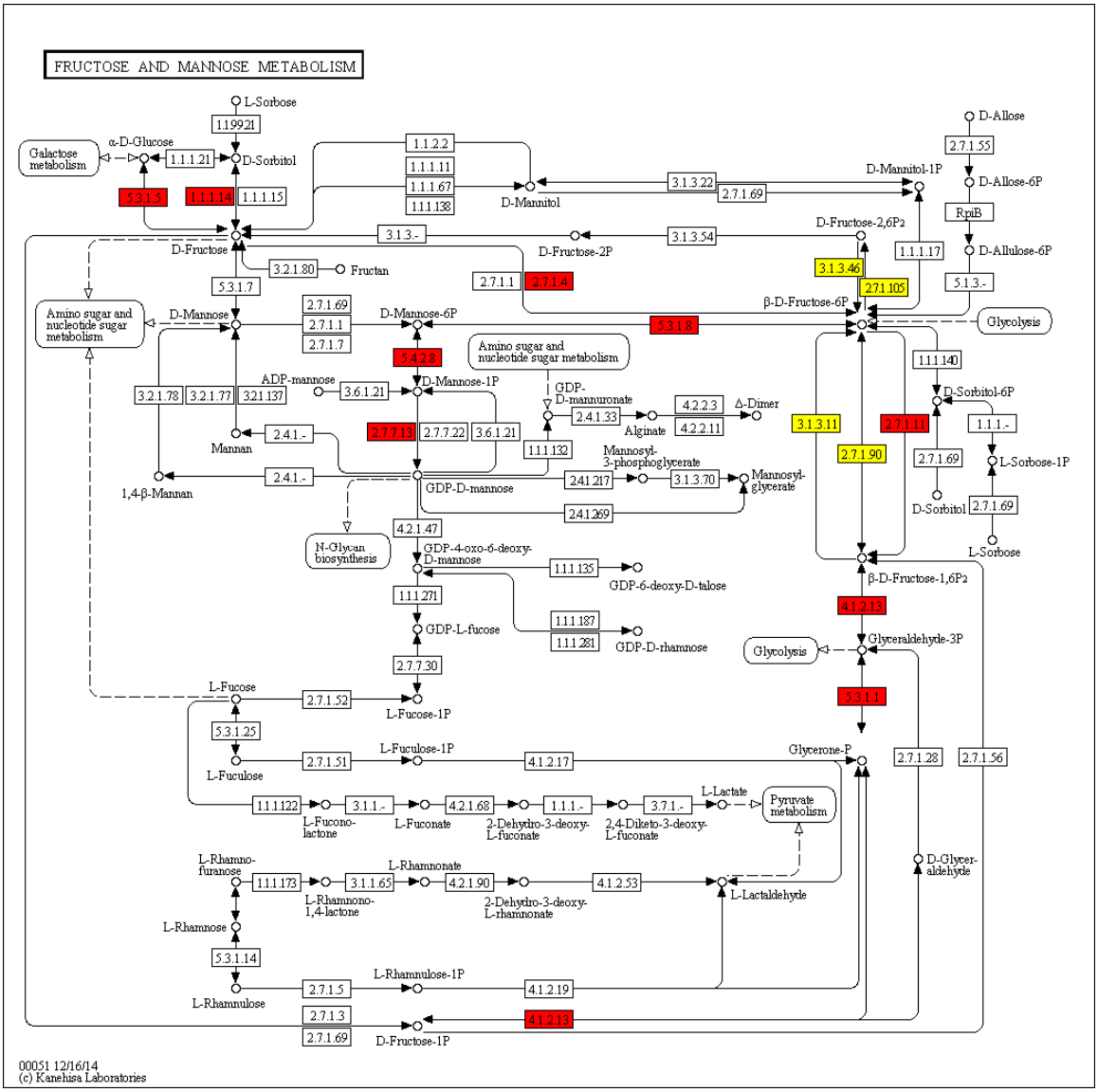


Fig. 11: Fructose and mannose metabolism (KEGG pathway 00051 (Kanehisa, et al. 2014)). Enzymes marked in red and yellow have been annotated in the *G. sulphuraria* genome. Enzymes marked in yellowed have been marked for easy identification and are directly referenced in the main text. Although the pathway map shows many uncolored boxes (enzymes not annotated in the *G. sulphuraria* genome), it does show alternative glycolytic reactions metabolizing phosphorylated β-D-fructose (EC 3.1.3.11 & EC 2.7.1.90) and the bifunctional enzyme 6-phosphofructo-2-kinase / fructose-2,6-bisphosphatase (EC 2.7.1.105, EC 3.1.3.46) which is essential in plant physiology and biochemistry (Nielsen, et al. 2004).

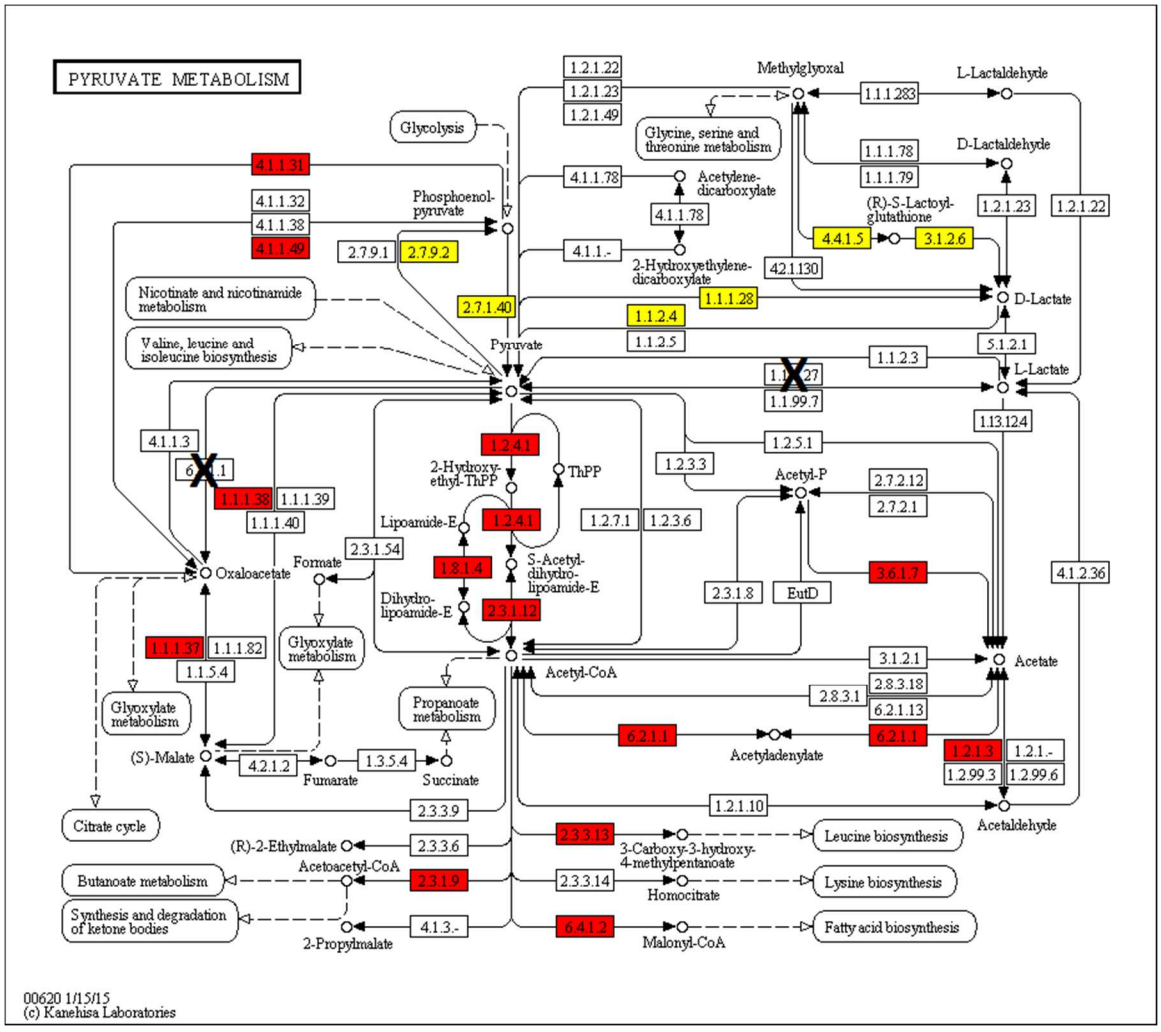


Fig. 12: Pyruvate metabolism (KEGG pathway 00620 (Kanehisa, et al. 2014)). Enzymes marked in red and yellow have been annotated in the *G. sulphuraria* genome. Enzymes marked in yellowed have been marked for easy identification and are directly referenced in the main text. Enzymes covered by large **X** (pyruvate carboxylase, EC 6.4.1.1; L-lactate dehydrogenase, EC 1.1.1.27), was absent from the *G. sulphuraria* genome (see Fig. 10 caption for EC 6.4.1.1).

Methylglyoxal is a ubiquitous, cytotoxic byproduct most commonly associated with glycolysis. The non-enzymatic removal of phosphate from two glycolysis intermediates, dihydroxyacetone phosphate and glyceraldehyde phosphate, results in the production of methylglyoxal, which is enhanced under salt and heavy metal stress (Engqvist, et al. 2009), of

which *G. sulphuraria* is commonly exposed. The most common pathway used for detoxification utilizes two enzymes: **lactoylglutathione lyase** (glyoxalase I; Gasu_09680 & Gasu_49880, EC 4.4.1.5) and **hydroxyacylglutathione hydrolase** (glyoxalase II; Gasu_21860, Gasu_08440, & Gasu_57940, EC 3.1.2.6) (Fig. 12). The enzymatic product of hydroxyacylglutathione hydrolase is D-lactate [(*R*)-lactate], which has been shown to be an accumulated fermentation product and does not appear to be further metabolized in *G. sulphuraria* under anaerobic conditions (Lafraie and Betz 1985). Methylglyoxal also accumulates in *E. coli* when grown using glycerol as a sole carbon source under anaerobic conditions (Zhu, et al. 2001). The *G. sulphuaria* genome encodes for **D-lactate dehydrogenase** (Gasu_43520, EC 1.1.1.28) and **D-lactate dehydrogenase [cytochrome]** (Gasu_04450 & Gasu_38460, EC 1.1.2.4), which both catalyze the reversible formation of pyruvate from D-lactate for further methylglyoxal detoxification, and also pyruvate fermentation under anaerobiosis (Fig. 12). Gasu_00450 and Gasu_38460 have 29% and 45% amino acid identity, respectively, to the D-lactate dehydrogenase [cytochrome] characterized in *Arabidopsis thaliana* (At5g06580), which has a strong preference for D-lactate, significantly more so than for L-lactate (Engqvist, et al. 2009). Interestingly, *G. sulphuraria* does not encode an L-lactate dehydrogenase (EC 1.1.1.27) (Fig. 12) indicating *G. sulphuraria* neither produces L-lactate, nor can readily degrade it in contrast with land plants and *C. merolae*, another Cyanidiophyceae red alga. Thus, both anaerobic fermentation and methylglyoxal detoxification in G. *sulphuraria* proceed through the intermediate D-lactate, in contrast to plants and *C. merolae*, which seem to generate L-lactate during anaerobiosis.

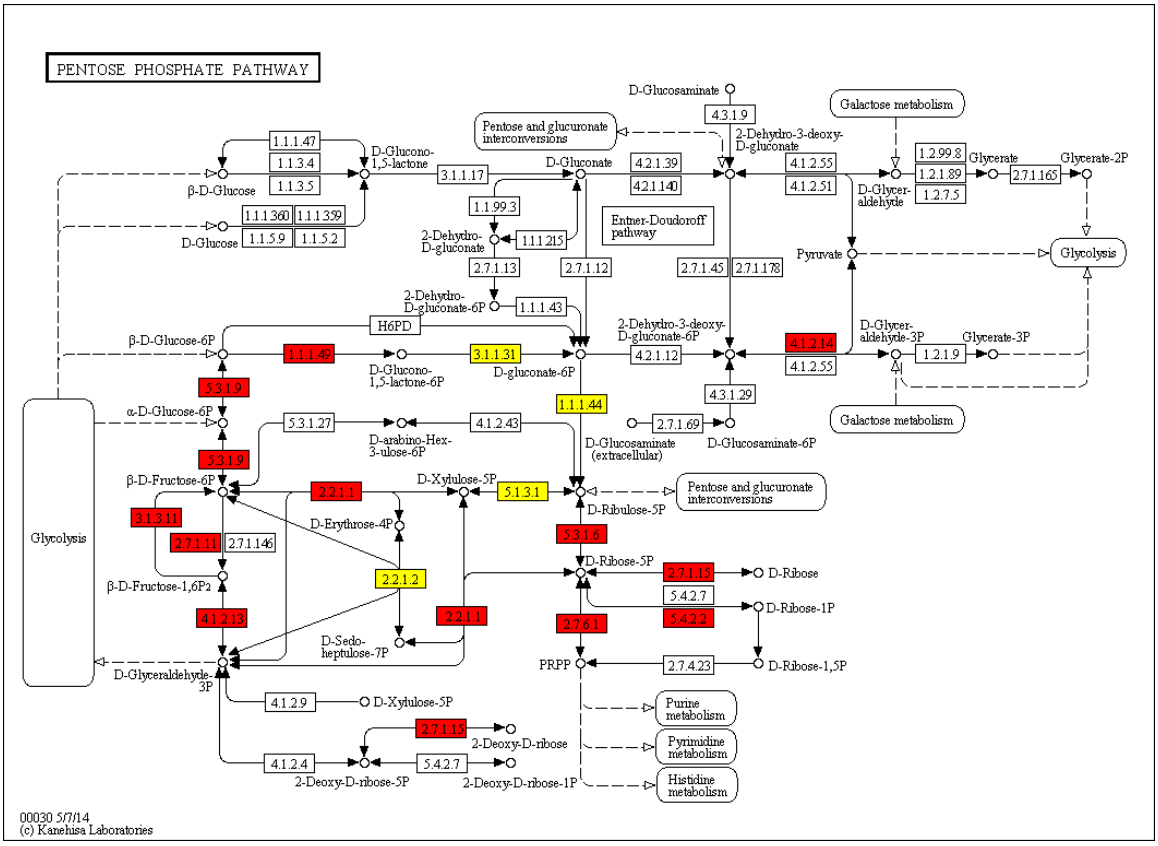


Fig. 13: Oxidative pentose phosphate pathway (KEGG pathway 00030 (Kanehisa, et al. 2014)). Enzymes marked in red and yellow have been annotated in the *G. sulphuraria* genome. Enzymes marked in yellow, **transaldolase** (Gasu_24580, EC 2.2.1.2), **phosphogluconate dehydrogenase** (Gasu_02940, EC 1.1.1.44), **ribulose-phosphate 3-epimerase** (Gasu_18200, EC 5.1.3.1), **6-phosphogluconolactonase** (Gasu_43900, EC 3.1.1.21), are encoded by a single gene, possibly indicating one (cellular) location for the pentose phosphate pathway. Enzymes marked in red are encoded by two or more genes in the *G. sulphuraria* genome.

As expected, the oxidative pentose phosphate pathway (OPPP) is complete in *G. sulphuraria* (Fig. 13). The OPPP serves many functions, such as regenerating NADPH, a reducing agent commonly employed in biosynthetic processes, being one of the most important. Additionally, the OPPP generates (four carbon) erythrose-4-phosphate and (five carbon) ribose-5-phosphate sugar intermediates used as precursors for aromatic amino acid biosynthesis and nucleic acid

biosynthesis, respectively. Several enzymes are encoded by a single gene (Fig. 13, yellow boxes), indicating this pathway probably exists either in the cytosol or in the plastid in *G. sulphuraria*. In contrast, land plants and green algae operate the oxidative pentose phosphate pathway in the cytosol, as well as the plastid (Kruger and von Schaewen 2003). The Calvin-Benson cycle, responsible for converting CO_2 into carbon molecules readily metabolized by the cell, is also present in its entirety (not shown). The pathways shown above represent some of the core metabolic pathways in *G. sulphuraria* involved in carbon metabolism. No comment can be made regarding enzymes resulting from horizontal gene transfer as no evolutionary trees were constructed. Yet, several examples are given that illustrate the metabolic diversity of *G. sulphuraria*, even in core metabolic pathways that are common to eukaryotes as a whole.

The genome appears to be very close to complete given that for each enzymatic step in the aforementioned metabolic pathways, (at least) one protein has been annotated. The glyoxylate cycle appears (after exhaustive searching) to be incomplete. The two enzymes unique to this cycle, isocitrate lyase (EC 4.1.3.1) and malate synthase (EC 2.3.3.9), are absent from the genome and also appear absent from other sequenced red algae based on BLAST searches (Fig. 14). This suggests that substrates ultimately degraded to acetyl-CoA in *G. sulphuraria* (and red algae) are directly incorporated into the citric acid cycle, forgoing carbohydrate biosynthesis.

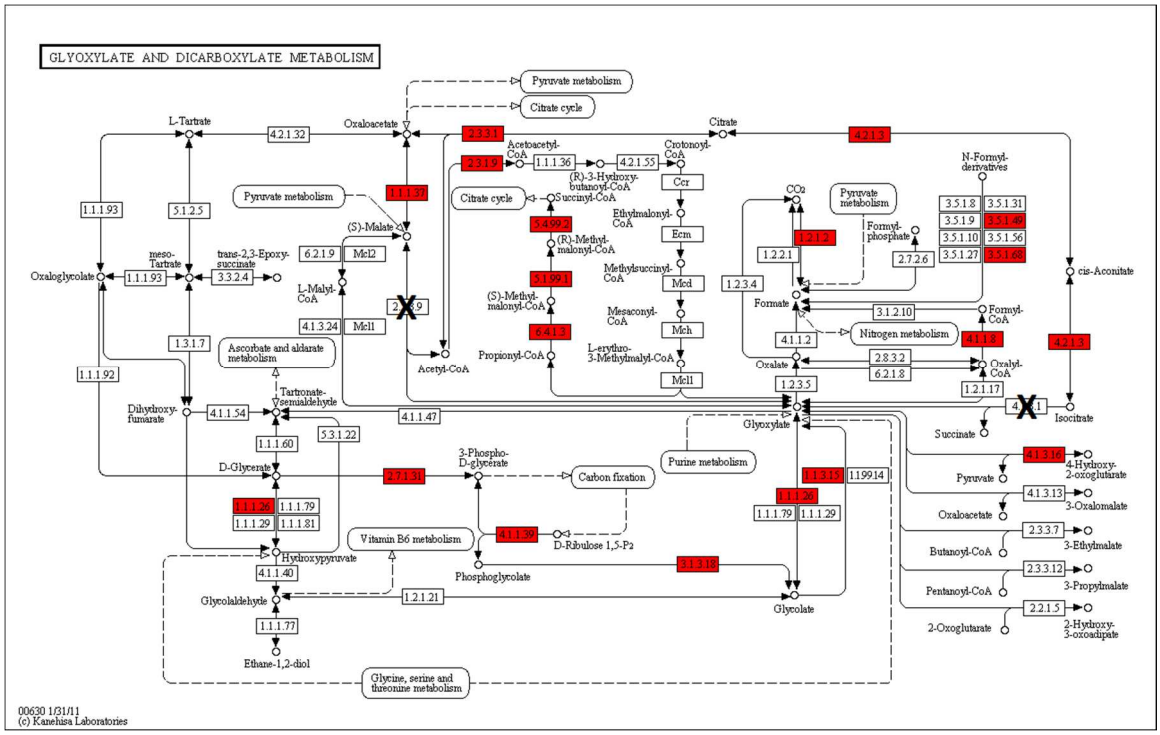


Fig. 14: Glyoxylate and dicarboxylate metabolism (KEGG pathway 00630 (Kanehisa, et al. 2014)). Enzymes marked in red have been annotated in the *G. sulphuraria* genome. Enzymes covered by a large **X**, isocitrate lyase (EC 4.1.3.1) and malate synthase (EC 2.3.3.9), could not be detected in the genome resulting in a missing glyoxylate cycle, in contrast with many photosynthetic, eukaryotic lineages including land plants, some green algae, and stramenopiles.

TRANSPORT SUGARS IN *G. SULPHURARIA*

Green plants and algae have long been characterized as using sucrose as the main transport sugar, whereas red algae are known to produce the soluble heterosides floridoside (2-α-D-galactosylglycerol) and isofloridoside (1-α-D-galactosylglycerol), which also double as compatible solutes, compounds that accumulate to regulate osmotic potential, under salt stress (Kirst 1980; Pade, et al. 2014). Likewise, sucrose is also implicated with abiotic stress responses, such as aiding in osmoregulation. Sucrose can be broken down into glucose and fructose, both of

which can also act as signaling molecules (Ruan 2014). Metabolic profiling of *G. sulphuraria* extracts indicated the presence of both sucrose and floridoside (Horst 2006). Unlike all other red algae sequenced thus far, *G. sulphuaria* synthesizes sucrose and possesses the enzymes that synthesize sucrose from UDP-glucose and fructose-6-phosphate, namely **sucrose-phosphate synthase** (Gasu_03810, EC 2.4.1.14) and **sucrose-phosphatase** (Gasu_32680, EC 3.1.3.24) (Fig. 15). The sucrose-phosphate synthase (Gasu_03810) contains an N-terminal glucosyl-transferase domain and a C-terminal phosphohydrolase domain, analogous to bacterial and land plant sucrose-phosphate synthases (Salerno and Curatti 2003). In an evolutionary tree, the sequence from *G. sulphuraria* (Gasu_03810) forms a separate branch and is neither closely related to land plant, nor bacterial sucrose-phosphate synthases (Schönknecht, unpublished). With the sequences currently available, it is not possible to discern whether the sucrose-phosphate synthase (Gasu_03810) was retained from the last common ancestor of land plants and red algae (i.e. the ancestor of Archaeplastidae), or was acquired via horizontal gene transfer from bacteria.

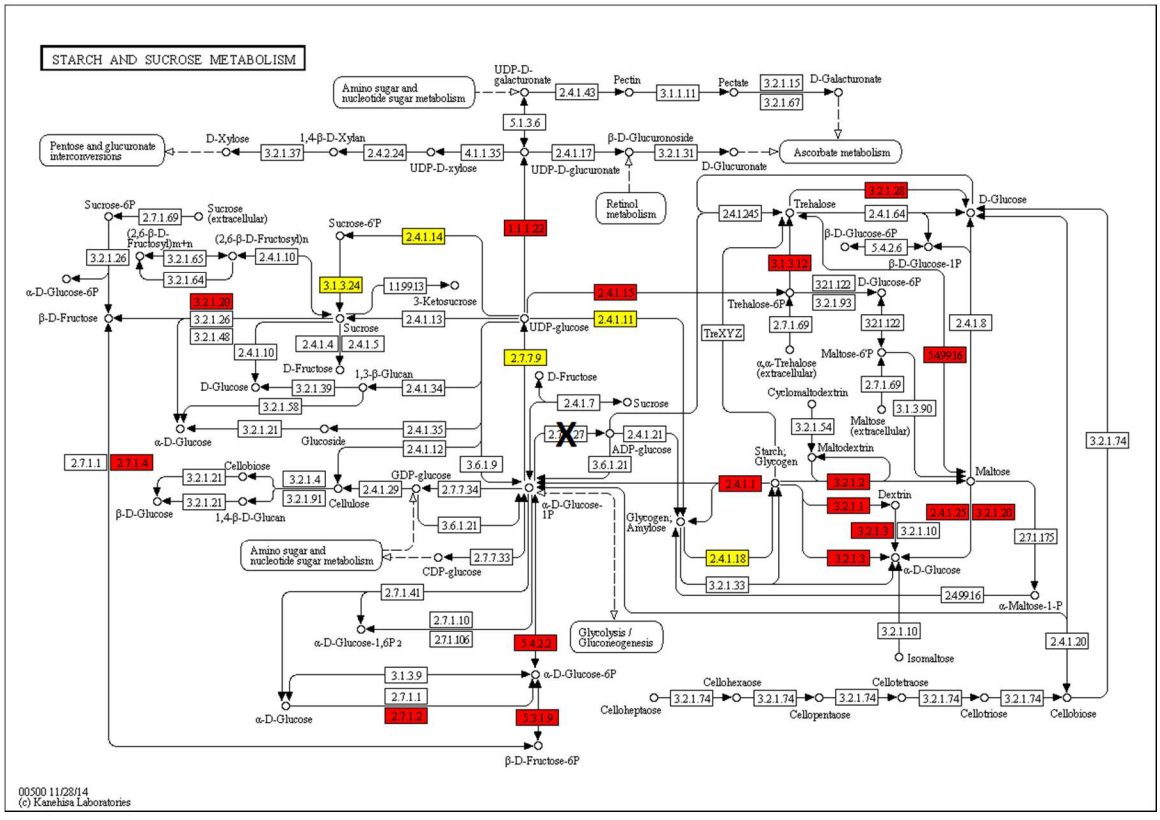


Fig. 15: Starch and sucrose metabolism (KEGG pathway 00500 (Kanehisa, et al. 2014)). Enzymes marked in red and yellow have been annotated in the *G. sulphuraria* genome. Enzymes have been marked in yellow for easy identification are directly mentioned in the main text. The enzyme marked by the large **X**, ADP-glucose pyrophosphorylase (EC 2.7.7.27), could not be detected in the genome of *G. sulphuraria* indicating glucose polymers are synthesized similarly to animals, via UDP-glucose.

G. SULPHURARIA SYNTHESIZES GLYCOGEN

Archaeplastida store carbohydrates as polyglucans, which are polysaccharides composed of D-glucose subunits linked by glycosidic bonds (Ball, et al. 2011). Green plants and algae (Viridiplantae) usually store polyglucans in the form of starch inside chloroplasts. Red algae synthesize floridean starch in the cytosol, i.e. cytosolic starch. *G. sulphuraria* synthesizes glycogen, a highly branched storage glucan usually observed in heterotrophic eukaryotes and bacteria (Hirabaru, et al. 2010; Ball, et al. 2011). In contrast to Viridiplantae, where starch biosynthesis uses the ADP-glucose-based pathway, *G. sulphuraria* and all other eukaryotes, including red algae, use the UDP-glucose-based pathway (Ball, et al. 2011). *G. sulphuraria* encodes a **UDP-glucose pyrophosphorylase** (Gasu_00869, EC 2.7.7.9, UTP-glucose-1-phosphate uridylyltransferase), but does not appear to encode ADP-glucose pyrophosphorylase (EC 2.7.7.27) (Fig. 15). This corroborates earlier observations in red algae whereby polyglucan synthesis proceeds via a UDP-glucose-selective α-glucan synthase (Viola, et al. 2001). Red algal UDP-glucose pyrophosphorylases are related to cytosolic UDP-glucose pyrophosphorylases from Viridiplantae, which utilize UDP-glucose for sucrose biosynthesis, and other eukaryotes (Patron and Keeling 2005). It appears the UDP-G biosynthesis pathway in eukaryotes may originate from ancestry and was retained for cytosolic sucrose biosynthesis in Viridiplantae. On the other hand,

the ADP-G biosynthesis pathway operates solely in the chloroplast of Viridiplantae species, further supporting acquisition via endosymbiosis from a cyanobacterium.

The **glycogen synthase** (Gasu_14030, EC2.4.1.11) in *G. sulphuraria* is 1699 amino acids long, and only the C-terminal 700 amino acids contain the glycogen synthase domain. The N-terminal half shows similarity with **1,4-α- glucan branching enzyme** (Gasu_50970, EC 2.4.1.18, *GlgB* domain) (Fig. 16) that catalyzes the formation of α-1,6-linked side branches (Baecker, et al. 1986).

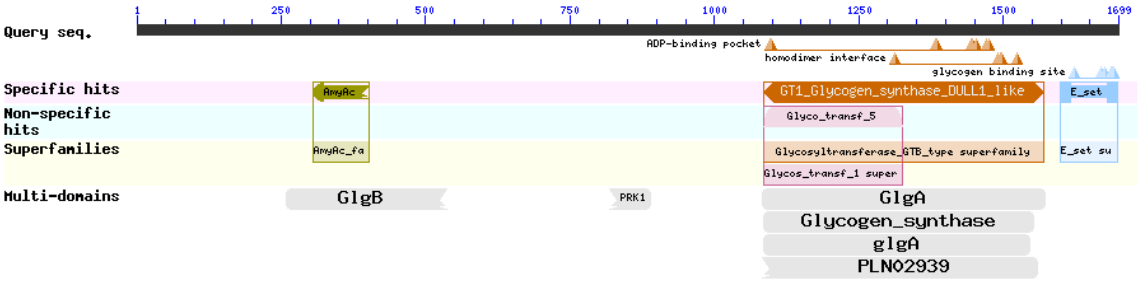


Fig. 16: Conserved Domains Database (CDD) (Marchler-Bauer, et al. 2011) search results indicate Gasu_14030 is a bifunctional protein consisting of both, a glycogen synthase (*GlgA*) and a 1,4-α-glucan branching domains. Glycogen synthase is responsible for the elongation of glucose polymers synthesizing α-1,4 glycosidic bonds, whereas 1,4-α-glucan branching enzyme generates short chains of glucose polymers by cleaving α-1,4 linkages and catalyzing α-1,6 branching linkages leading to the highly branched architecture of glycogen.

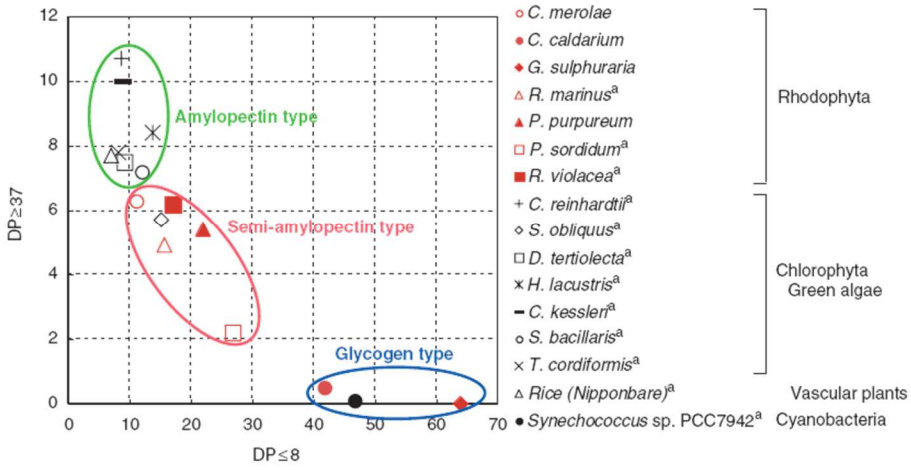


Fig. 17: Degree of polyglucan branching taken from Hirabaru, et al. (2010). Degree of polymerization is abbreviated as DP, which refers to the (linear) length of each chain. Moving from the bottom, right to the top, left of the graph signifies moving from short, highly branched polyglucans to long, linear polylglucans with little branching.

G. sulphuraria synthesizes glycogen-type polysaccharides that possess a very high degree of branching when compared to several other species of red algae (Fig. 17) (Shimonaga, et al. 2008; Hirabaru, et al. 2010). The *G. sulphuraria* genome encodes only one dedicated glycogen branching enzyme (Gasu_50970). However, as mentioned above, the glycogen synthase (Gasu_14030) also contains a glycogen branching enzyme (*GlgB*) domain. Assuming the predicted protein sequence of Gasu_14030 is not the result of a genome assembly error, glycogen synthase and 1,4-α-glucan branching enzyme appear to form a bifunctional, fusion enzyme in *G. sulphuraria*. Based on the large degree of polyglucan branching, the branching domain of the glycogen synthase in *G. sulphuraria* might significantly aid in achieving the abundance of α-1,6 linkages. Using Gasu_14030, the glycogen synthase, as a query sequence, BLASTP searches using the non-redundant database at NCBI resulted in only a handful of hits with similar (amino acid) length. Two species of red algae, *C. merolae* and *C. crispus*, and four alveolates all possess glycogen synthases of comparable length to that of *G. sulphuraria*. However, none of the

sequences from the six organisms mentioned above, contain an N-terminal 1,4-α-glucan branching domain (*GlgB*), making the glycogen synthase from *G. sulphuraria* (Gasu_14030) the only one of its kind to date.

GLYCEROL: A LINK BETWEEN CARBOHYDRATES AND LIPIDS

G. sulphuraria can grow on glycerol as a sole carbon source (Gross and Schnarrenberger 1995a). Multiple pathways exist for metabolizing glycerol (Fig. 18), two of which convert glycerol into glycerone phosphate, a glycolysis intermediate also known as dihydroxyacetone-phosphate. A third pathway for metabolizing glycerol ultimately produces glycerate-3-phosphate, another glycolytic intermediate. All three routes serve as a link between lipid metabolism and carbohydrate metabolism.

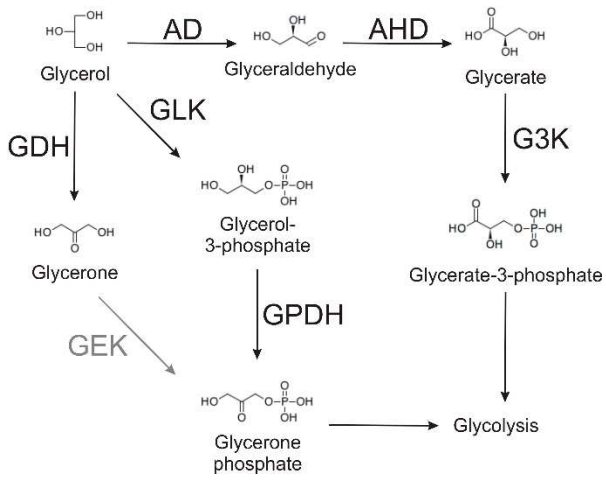


Fig. 18: Pathways for metabolizing glycerol. Enzymes abbreviations given in black have been annotated in the *G. sulphuraria* genome. Glycerone kinase, enzyme abbreviation given in gray (GEK), has not been annotated in *G. sulphuaria*. Abbreviations are as follows: GDH, glycerol dehydrogenase (NAD^+); GEK, glycerone kinase; GLK, glycerol kinase; GPDH, glycerol-3-phosphate dehydrogenase (NAD^+); AD, alcohol dehydrogenase; AHD, aldehyde dehydrogenase;

G3K, glycerate 3-kinase. Structural formulas for intermediates and all substrate and enzyme names according to KEGG (Kanehisa, et al. 2014).

As in most eukaryotes, *G. sulphuraria* can convert glycerol into glycerone phosphate by **glycerol kinase** (Gasu_03170 & Gasu_60010, EC 2.7.1.30) and glycerol-3-phosphate dehydrogenase (Gasu_03180, Gasu_57960, & Gasu_57980; EC 1.1.1.6), i.e. phosphorylation followed by oxidation (GLK and GPDH respectively, Fig. 18). *G. sulphuraria* is also likely able to convert glycerol into glycerone phosphate by **glycerol dehydrogenase** (Gasu_03180, Gasu_57960, & Gasu_57980, EC 1.1.1.6) and glycerone kinase (EC 2.7.1.29), i.e. oxidation followed by phosphorylation (GDH and GEK respectively, Fig. 18). Even though the *G. sulphuraria* genome does not appear to encode a glycerone kinase (GEK, Fig. 18), it is known that proteins from different families can act as a glycerone kinase (Erni, et al. 2006); glycerol kinases (EC 2.7.1.30) have been shown to possess glycerone kinase activity (Hayashi and Lin 1967; Janson and Cleland 1974). Moreover, glycerol dehydrogenase (Gasu_03180) and glycerol kinase (Gasu_03170) are encoded by neighboring, unidirectional genes separated by just 5 bp, pointing to the possibility of coordinated transcription (Schönknecht, unpublished). A third pathway is also employed by *G. sulphuraria* whereby glycerol is converted to glycerate-3-phosphate, another glycolytic intermediate. Glycerol undergoes two rounds of oxidation (AD and AHD in Fig. 18), by **alcohol dehydrogenase** (Gasu_53790, EC 1.1.1.2) and **aldehyde dehydrogenase** (6 genes, EC 1.2.1.3) generating glycerate. Glycerate-3-phosphate is then generated by **glycerate 3-kinase** (Gasu_01520, EC 2.7.1.31) (G3K, Fig. 18). Results from phylogenetic analyses suggest that the diversity of glycerol metabolizing pathways, including the ability to take up glycerol into the cell, was facilitated by horizontal gene transfer from bacteria. All three of the glycerol-3-phosphate dehydrogenases encoded by *G. sulphuraria* (Gasu_03180, Gasu_57960, & Gasu_57980) appear to originate from bacteria followed by gene duplication (Schönknecht, et al. 2013). Furthermore, all five of the genes encoding glycerol uptake

facilitators (transporter proteins) also originate from bacteria via horizontal gene transfer (Schönknecht, et al. 2013).

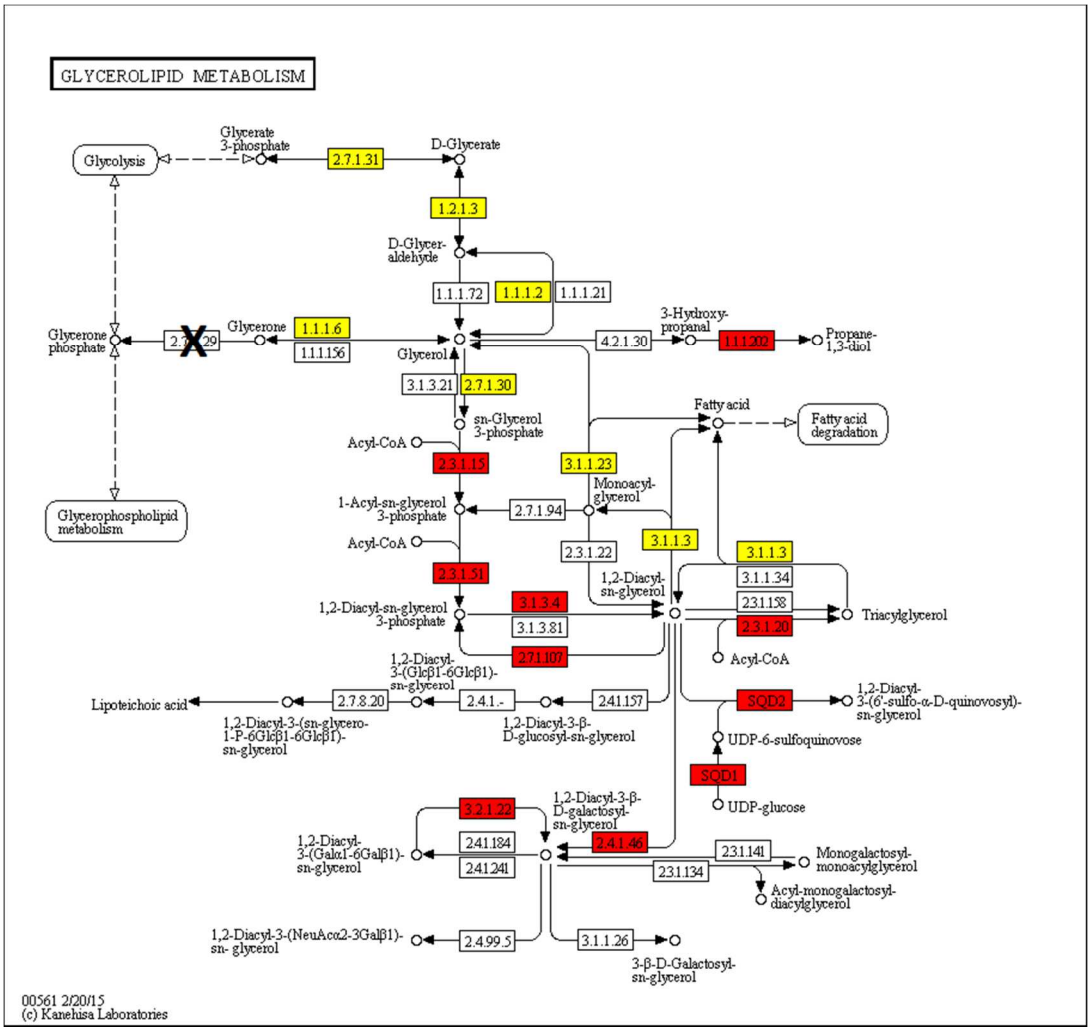


Fig. 19: Glycerolipid metabolism (KEGG pathway 00561 (Kanehisa, et al. 2014)). Enzynmes marked in red and yellow have been annotated in the genome of *G. sulphuraria*. Enzymes directly mentioned in the main text have been marked in yellow for easy identification. As depicted in Fig. 18, glycerone kinase (EC 2.7.1.29) could not be detected in the *G. sulphuraria* genome as indicated by a large **X**.

Even though lipids are not a major storage molecule in *G. sulphuraria* (Graziani, et al. 2013), the metabolic pathways for the synthesis and degradation of lipids are present, as

expected. Triacylglycerides (a type of lipid [ester] composed of glycerol and three fatty acids) can be directly degraded to recover the glycerol residue and the associated fatty acids. **Triacylglycerol lipase** (Gasu_07310 & Gasu_17090, EC 3.1.1.3) and **acylglycerol lipase** (Gasu_37420, EC 3.1.1.23) cleave the glycerol residue, which can then be converted into glycolytic intermediates, and fatty acids (Fig. 19). The remaining fatty acids, which represent the energy-rich component of triacyglycerols, are degraded through successive rounds of oxidation producing acetyl-CoA (and succinyl-CoA from odd-chain fatty acids; see below).

DISTINGUISHING METABOLIC PATHWAYS IN *G. SULPHURARIA*

G. sulphuraria possesses metabolic pathways that not only distinguish it from land plants and green algae, but also from other red algae. The distribution of metabolic pathways present in different eukaryotic lineages is a direct result of evolution, and in some cases, the presence or absence of a pathway has become an identifying characteristic for particular eukaryotic lineages. In the context of lineage-specific, characteristic pathways, *G. sulphuraria* appears to be an exception. *G. sulphuraria* not only encodes pathways that are characteristic for red algae and other photosynthetic eukaryotes, including Viridiplantae, but also encodes pathways uncharacteristic for photosynthetic eukaryotes all together (Fig. 20).

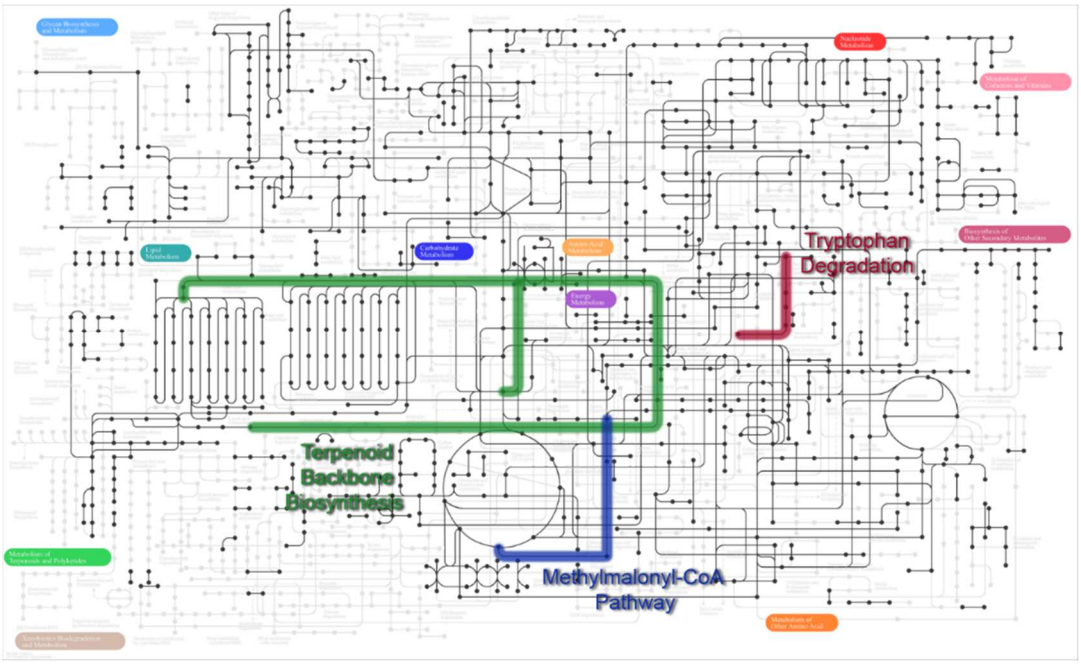


Fig. 20: Global metabolism of *G. sulphuraria* with distinguishing metabolic pathways highlighted and labeled; modified from KEGG (Kanehisa, et al. 2014). Black lines represent enzymes annotated in the genome of *G. sulphuraria*, and black dots represent the corresponding metabolites associated with each enzyme.

METHYLMALONYL-COA PATHWAY

Fatty acids are crucial components of cellular biology. The long, hydrophobic chains are indispensable for lipid membranes and sustain the hydrophobic region spanning the intersection between both sides of a lipid bilayer. Not only are fatty acids considered an import energy source for heterotrophic eukaryotes, they are also the major component of seed oils, a lipid-based energy reserve for germinating seeds. The end products resulting from fatty acid oxidation (or degradation) is dependent upon the type of fatty acid. Oxidation of even-chain fatty acids only results in the end product acetyl-CoA, which is easily incorporated into the TCA cycle. The complete oxidation of odd-chain fatty acids (β-oxidation) results in two end products: acetyl-CoA and propionyl-CoA. Propionyl-CoA must be metabolized by utilizing another pathway. The

three-reaction step methylmalonyl-CoA pathway allows for the degradation of propionyl-CoA, in addition to the degradation of the branched amino acids valine and isoleucine. Valine and isoleucine must first be converted into propionyl-CoA before entering the methylmalonyl-CoA pathway. The final metabolite of the methylmalonyl-CoA pathway is succinyl-CoA, which can also be fed into the TCA cycle (Fig. 21). In plants and seeds, fatty acid β-oxidation occurs in peroxisomes (Engqvist, et al. 2009) and glyoxysomes respectively, but the exact route by which propionyl-CoA is metabolized is not currently understood (Lucas, et al. 2007). Nonetheless, the methylmalonyl-CoA pathway has neither been annotated, nor characterized for any Viridiplantae species. Methylmalonyl-CoA pathway enzymes are most commonly associated with prokaryotes and animals. Yet, other photosynthetic eukaryotes, namely stramenopiles, appear to encode the methylmalonyl-CoA pathway as well.

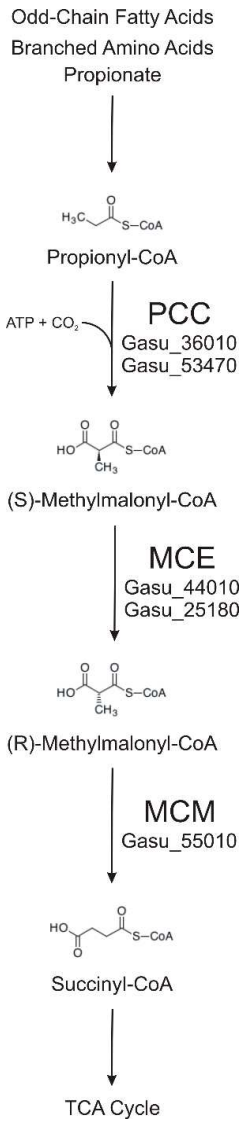


Fig. 21: Enzymes of the methylmalonyl-CoA pathway annotated in *G. sulphuraria* are listed for each enzyme as genome identifiers (Gasu_#). Abbreviations: PCC, propionyl-CoA carboxylase; MCE, methylmalonyl-CoA epimerase; MCM, methylmalonyl-CoA mutase.

Currently, only two red algae appear to encode the methylmalonyl-CoA pathway: *G. sulphuraria* and the unicellular, red alga *Porphyridium purpureum* (Bhattacharya, et al. 2013). Phylogenetic analyses indicate the methylmalonyl-CoA pathway enzymes in *G. sulphuraria* have been conserved through ancestry, and not from horizontal gene transfer, supporting the

idea that the methylmalonyl-CoA pathway underwent multiple losses throughout major eukaryotic lineages, such as Viridiplantae and fungi. Evolutionary trees for methylmalonyl-CoA pathway enzymes showed similar results for all *G. sulphuraria* protein sequences; all enzymes appear to originate from ancestry (Figs. 22, 23, & 24). Sequences from the red alga *P. purpureum*, which possesses the methylmalonyl-CoA pathway, were not included due to sequences being unavailable at the time of the phylogenetic analysis.

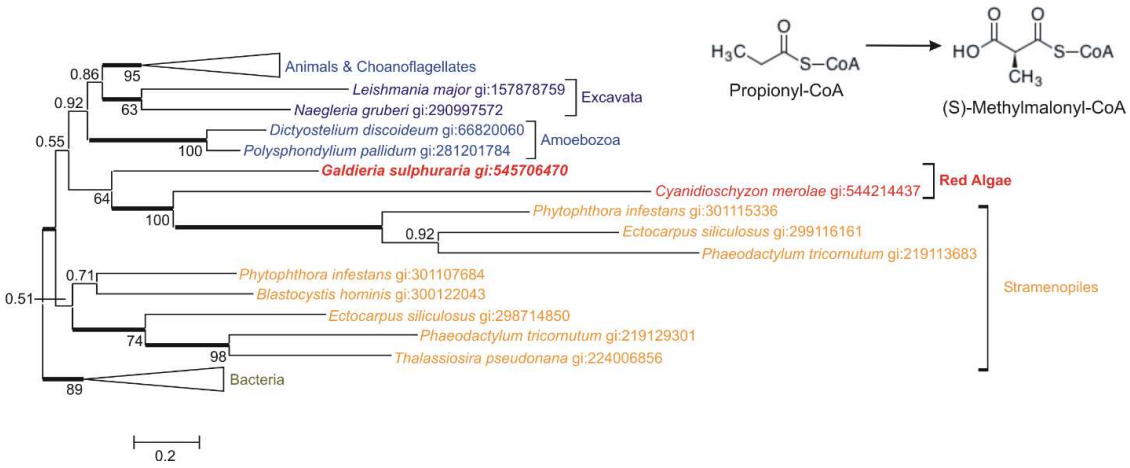


Fig. 22: An evolutionary tree for the alpha chain of propionyl-CoA carboxylase shows the expected separation of bacterial and eukaryotic sequences. The unrooted Bayesian tree was generated using the WAG+I+G model of protein evolution with posterior probabilities above the branches and PhyML bootstrap values (LG+I+G; 200 replicates) displayed as percentages below the branches. Thickened horizontal lines represent 1.0 Bayesian posterior probability. Only support values >50% are shown. Larger clades have been collapsed for presentation and size is indicative of the number of taxa within the clade. Scale bar represents 0.2 substitutions per site.

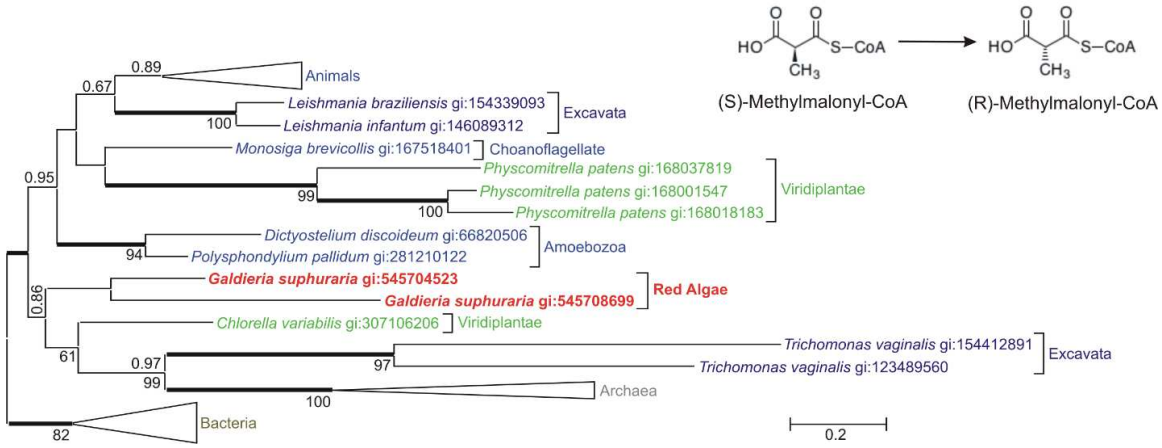


Fig. 23: An evolutionary tree for methylmalonyl-CoA epimerase shows disbanded eukaryotic lineages, i.e. Unikonta, Excavata, and Viridiplantae. The unrooted Bayesian tree was generated using the WAG+I+G model of protein evolution with posterior probabilities above the branches and PhyML bootstrap values (LG+I+G) displayed as percentages below the branches. Thickened horizontal lines represent 1.0 Bayesian posterior probability. Only support values >50% are shown. Larger clades have been collapsed for presentation and size is indicative of the number of taxa within the clade. Scale bar represents 0.2 substitutions per site.

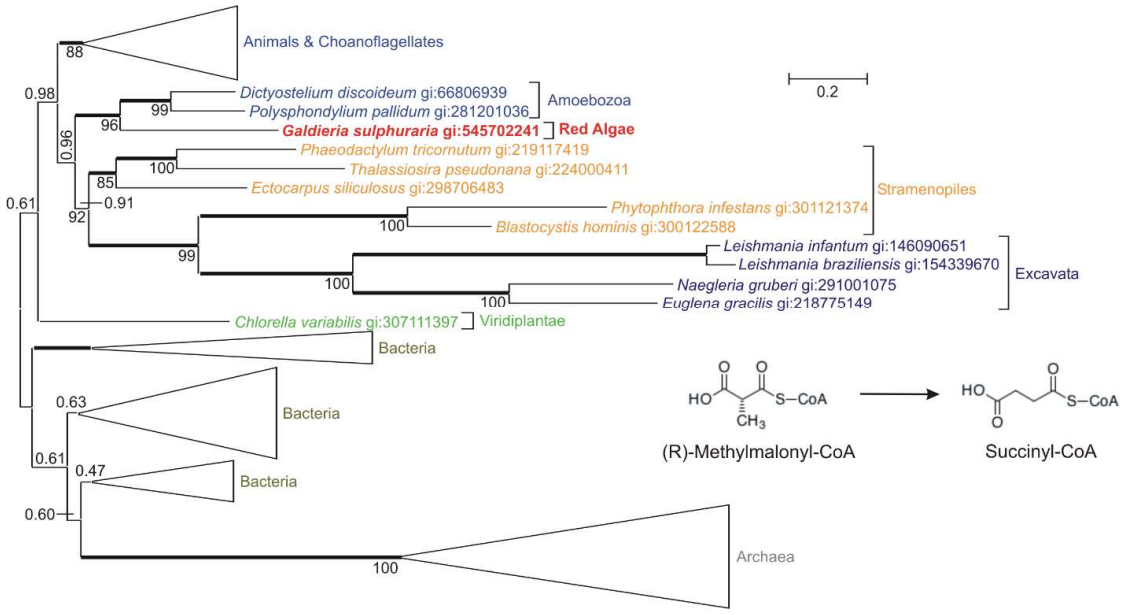


Fig. 24: An evolutionary tree for methylmalonyl-CoA mutase shows the expected separation of bacterial and eukaryotic sequences. The unrooted Bayesian tree was generated using the WAG+I+G+F model of protein evolution with posterior probabilities above the branches and PhyML bootstrap values (LG+I+G; 200 replicates) displayed as percentages below the branches. Thickened horizontal lines represent 1.0 Bayesian posterior probability. Only support values >50% are shown. Larger clades have been collapsed for presentation and size is indicative of the number of taxa within the clade. Scale bar represents 0.2 substitutions per site.

The final enzyme in the methylmalonyl-CoA pathway, methylmalonyl-CoA mutase (EC 5.4.99.2) (MCM in Fig. 25), requires vitamin B_{12} (Ado-cobalamin), an essential cofactor for the reaction mechanism (Cannata, et al. 1965). Bacteria are the only known organisms able to synthesize vitamin B_{12} (Roth, et al. 1996). However, algae have been shown to obtain vitamin B_{12} from bacteria through a symbiotic relationship (Croft, et al. 2005). *G. sulphuraria* can account for more than 98% of the eukaryotic biomass in its specific niches, such as volcanic calderas (Weber, et al. 2004), habitats usually deemed suitable only for prokaryotes. Furthermore, *G. sulphuraria* has been shown to coexist alongside prokaryotes within the same biofilm in extreme environments (Hoeft, et al. 2010). Thus, it seems reasonable to assume *G. sulphuraria* acquires vitamin B_{12} from neighboring bacteria.

Propionyl-CoA degradation is also achieved through another metabolic route, which fungi and some bacteria utilize. The coupling of the methylcitrate pathway and the glyoxylate cycle allows propionyl-CoA to be converted into succinate, another intermediate metabolite of the TCA cycle. As mentioned above, *G. sulphuraria* neither possesses a glyoxylate cycle, nor are the genes encoding the methylcitrate pathway present in the genome. Therefore, the presence of the methylmalonyl-CoA pathway, in the absence of the glyoxylate cycle, provides a route to

metabolize odd-chain fatty acid end products, namely propionyl-CoA, and the branched amino acids, isoleucine and valine.

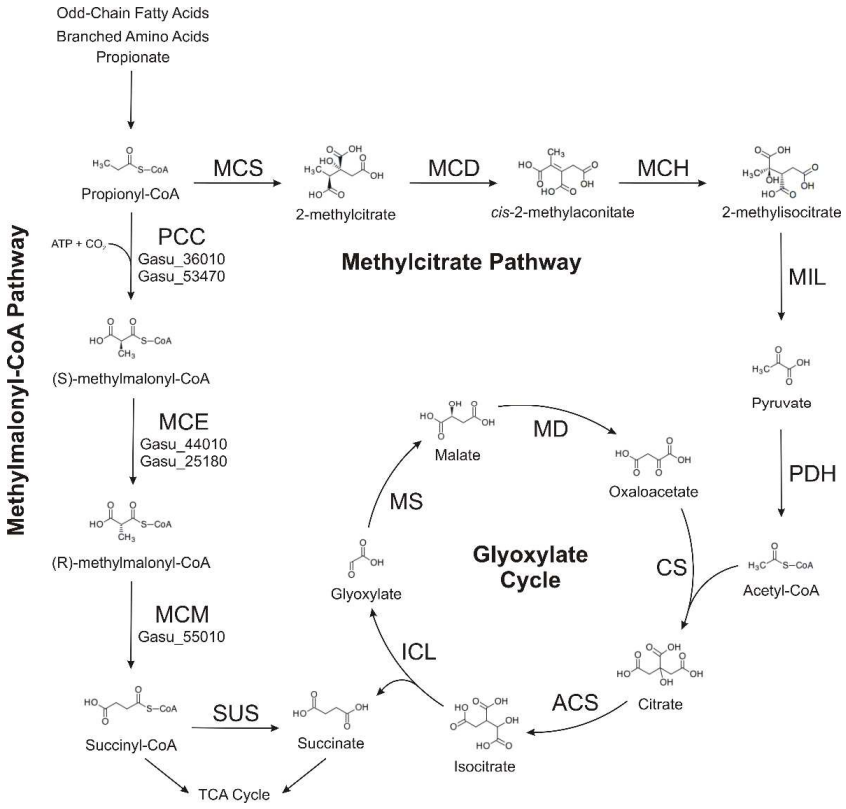


Fig. 25: Two different routes for the degradation of propionyl-CoA. The 3-step methylmalonyl-CoA pathway is employed by *G. sulphuraria* and other organisms including bacteria and animals. The methylcitrate pathway and the glyoxylate cycle are utilized in conjunction by fungi and some bacteria. Abbreviations are: PCC, propionyl-CoA carboxylase; MCE, methylmalonyl-CoA epimerase; MCM, methylmalonyl-CoA mutase; SUS, succinyl-CoA synthetase; MCS, 2-methylcitrate synthase; MCD, 2-methylcitrate dehydratase; MCH, (2*R*,3*S*)-2-methylcitrate hydro-lyase; MIL, 2-methylisocitrate lyase; PDH, pyruvate dehydrogenase complex; CS, citrate synthase; ACS, aconitase; ICL, isocitrate lyase; MS, malate synthase; MD, malate dehydrogenase. Intermediate molecular structures according to KEGG are depicted (Kanehisa, et al. 2014).

TERPENOID BIOSYNTHESIS

Terpenoids, also known as isoprenoids, are ubiquitous throughout the tree of life and serve crucial physiological and biochemical roles in plants. Carotenoids, a broad class of terpenoids, function as accessory pigments for photosynthesis and are also involved in oxidative stress response (Owen and Peñuelas 2005). Another terpenoid, gibberellic acid, is an essential plant hormone in land plants (Lange 2000; Owen and Peñuelas 2005). Terpenoids represent one of the largest classes of biological metabolites and the largest class of natural plant products. Terpenoid backbone biosynthesis can be accomplished by two different pathways: the cytosol-localized mevalonate (MVA) pathway, which uses acetyl-CoA as a precursor, and the plastid-localized methylerythritol phosphate (MEP) pathway, which utilizes glyceraldehyde-3-phosphate and pyruvate as precursors (Fig. 26).

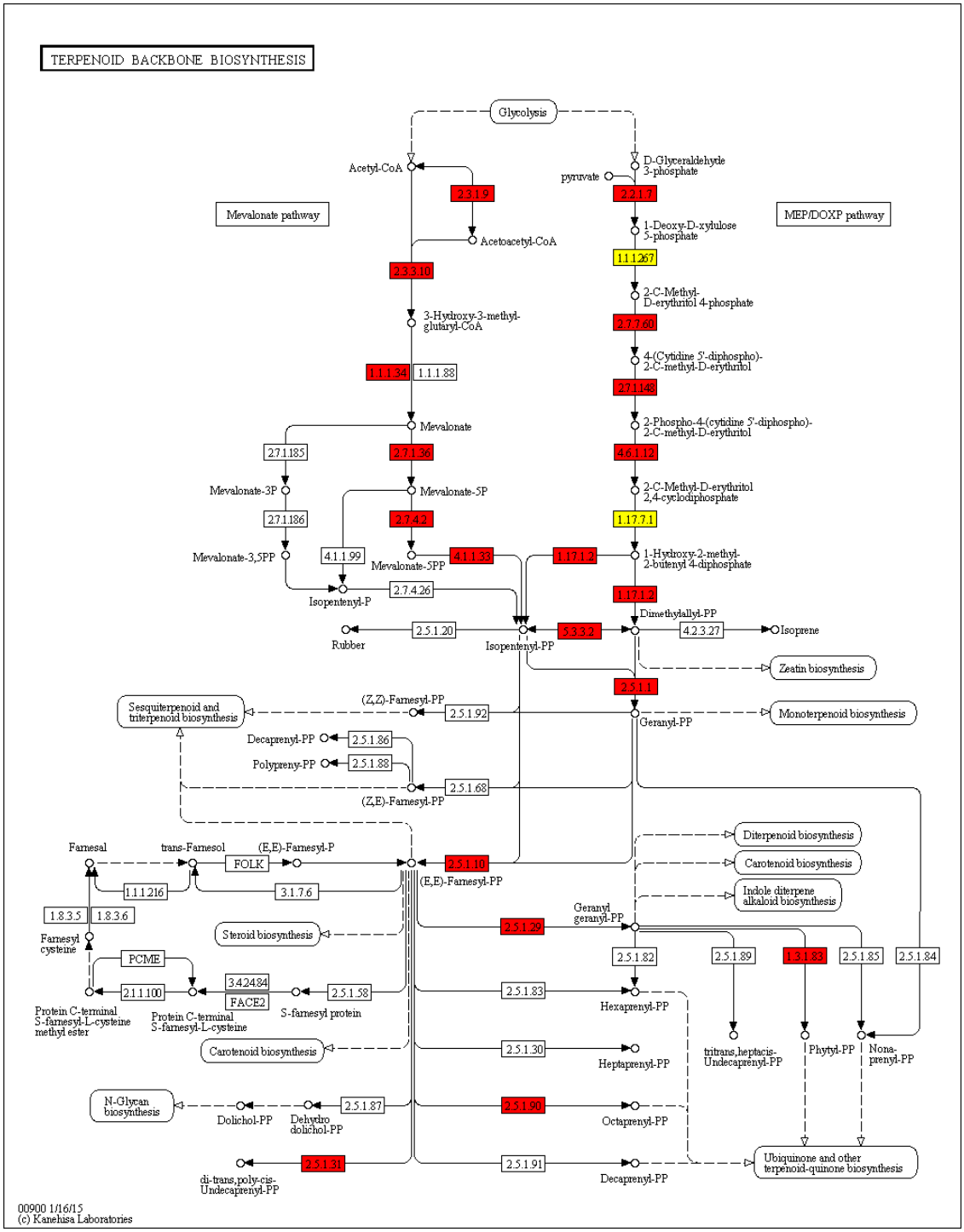


Fig. 26: Terpenoid backbone biosynthesis (KEGG pathway 00900 (Kanehisa, et al. 2014)). Enzymes marked in red and yellow have been annotated in the genome of *G. sulphuraria*. Enzymes have been marked in yellow for easy identification and are directly referenced in the main text.

The mevalonate pathway is most commonly associated with heterotrophic eukaryotes and archaea, and has even been coined as the ancestral, eukaryotic pathway (Lombard and Moreira 2011). However, land plants encode both, the mevalonate and methylerythritol phosphate pathways, while green algae only encode the methylerythritol phosphate pathway. The presence of the bacterial-associated methylerythritol phosphate pathway is believed to have been acquired during the primary endosymbiosis event, further supported by its plastid localization (Lange 2000). Interestingly, *G. sulphuraria* encodes both, the mevalonate and methylerythritol phosphate pathway. Currently, *G. sulphuraria* is the only red alga that possesses both pathways in contrast to other red algae, such as *C. merolae* and *C. crispus*, which only possess the methylerythritol phosphate pathway. No phylogenetic analyses have been conducted specifically investigating the evolutionary history of the enzymes involved in terpenoid backbone biosynthesis in *G. sulphuraria*. Best BLASTP hits from enzymes of both the mevalonate and methylerythritol phosphate pathways resided in expected lineages. Enzymes from the ancestral, eukaryotic mevalonate pathway had best BLASTP hits with other mevalonate pathway encoding eukaryotes, whereas most enzymes from the endosymbioint-derived methylerythritol phosphate pathway had best BLASTP hits in other photosynthetic eukaryotes. Unexpectedly, two enzymes from the methyleythritol phosphate pathway, **1-deoxy-D-xylulose-5-phosphate reducoisomerase** (Gasu_57430, EC 1.1.1.267) and **(*E*)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase** (Gasu_27800, EC 1.17.7.1) (Fig. 26), had best BLASTP hits in cyanobacteria. While best BLAST hits cannot provide conclusive information regarding evolutionary histories, the best hits were surprising considering red algae share a more recent common ancestor with other plastid-bearing eukaryotes than with cyanobacteria, even in the context of genes with cyanobacterial origin.

It is unknown why *G. sulphuraria* is the only red alga that encodes both pathways, and why the mevalonate pathway has been lost in other red algae. Transporters could have played a key role in the establishment of the one pathway over another (Lohr, et al. 2012). If a plastid-localized

transporter was established relatively soon after the acquisition of the plastid, access to plastid-produced terpenoid intermediates could have facilitated the loss of the ancestral mevalonate pathway (Lohr, et al. 2012). Given the unique evolutionary history of not only *G. sulphuraria*, but Cyanidiophyceae as a whole, it is difficult to definitively say which evolutionary scenarios could have accounted for the present day distribution of terpenoid biosynthesis pathways in red algae. Nonetheless, common habitats for *G. sulphuraria* and other Cyanidiophyceae algae are often harsh, and present numerous, potential abiotic stressors that terpenoids may help circumvent as it does in the green alga *Dunaliella salina* (Lohr, et al. 2012). Thus, the presence of both the mevalonate and methylerythritol phosphate pathway may be explained by an abiotic stressor context.

LINEAGE SPECIFIC DISTRIBUTION OF *DE NOVO* NAD^+ BIOSYNTHESIS PATHWAYS

The two *de novo* NAD^+ biosynthesis pathways seem to be mutually exclusive in eukaryotes (Fig. 27). Thus far no eukaryotic organism has been discovered that contains both pathways. Earlier reports (Katoh and Hashimoto 2004) of the existence of enzymes for the kynurenine pathway in rice were not corroborated by the bioinformatics and phylogenetic analyses presented here. Some parasitic eukaryotes, such as *Plasmodium* spp., acquire nicotinic acid and/or nicotinamide from their hosts. These parasitic eukaryotes lack the ability to biosynthesize NAD^+ *de novo* and only operate salvage pathways (Plata, et al. 2010). In Eukaryota, the kynurenine pathway is observed in most lineages, with the exception of some photosynthetic lineages (or lineages closely related to photosynthetic ones), which utilize the aspartate pathway for quinolinate biosynthesis (Fig. 27). The phylogenetic distribution of the two different *de novo* NAD^+ biosynthesis pathways indicates that the kynurenine pathway was the ancestral pathway in

the last common ancestor of all eukaryotes. The primary endosymbiosis event, which gave rise to photosynthesis in some eukaryotic lineages, was more than likely responsible for the substitution of the kynurenine pathway by the aspartate pathway. Since the aspartate pathway is only observed in plastid-bearing eukaryotes, or in lineages which probably had a plastid-bearing ancestor, it seems reasonable to assume that the aspartate pathway was introduced into eukaryotes by endosymbiotic gene transfer from a cyanobacterial genome when photosynthesis was initially established in eukaryotes.

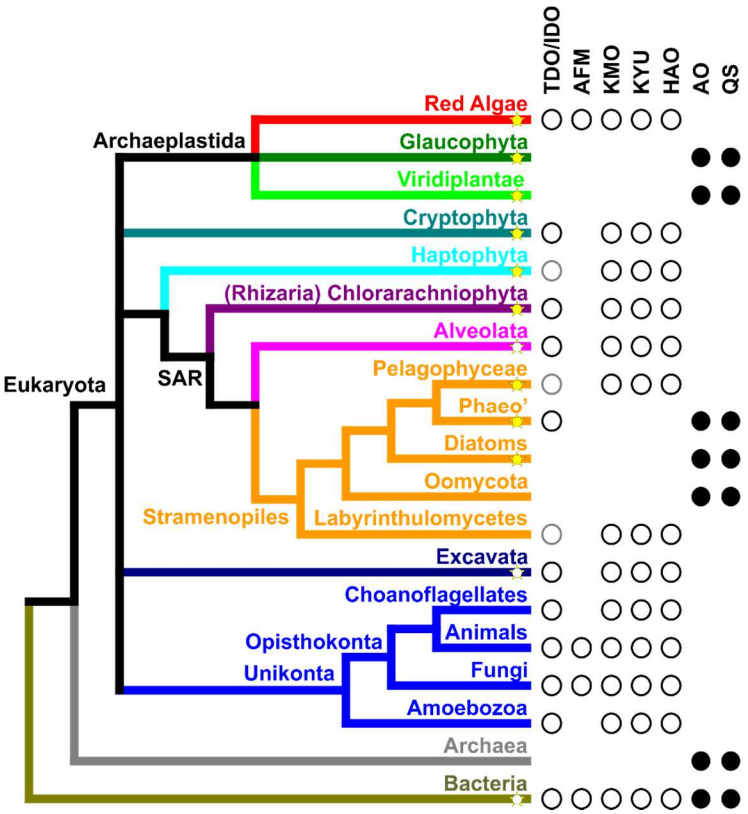


Fig. 27: Phylogenetic distribution of *de novo* NAD^+ biosynthesis pathways in eukaryotes shows a patchy distribution. Shown is the organismal phylogeny (Keeling and Palmer 2008; Burki, Okamoto, et al. 2012) for the major groups of life with emphasis on plastid bearing eukaryotes (star symbols: yellow, most species photoautotroph; white, some species photoautotroph; SAR, Stramenopiles, Alveolates, and Rhizaria; Phaeo', Phaeophyceae or brown algae). For each clade the presence of enzymes catalyzing the aspartate (closed symbols; AO, aspartate oxidase; QS,

quinolinate synthase) or the kynurenine pathway (open symbols; TDO/IDO, tryptophan - / indoleamine 2,3-dioxygenase; AFM, arylformamidase; KMO, kynurenine 3-monooxygenase; KYU, kynureninase; HAO, 3-hydroxyanthranilate 3,4-dioxygenase) are indicated. Grey open symbols indicate that the IDO identified in this work, based on sequence similarity, might not actually catalyze the first reaction of the kynurenine pathway (see text). Most bacteria use the aspartate pathway, while some have the kynurenine pathway (see text). Very few bacteria, if any, seem to use both pathways in parallel.

Within Archaeplastida, the presence of the aspartate pathway (and absence of the kynurenine pathway) is well established in Viridiplantae. In addition, the two aspartate pathway enzymes in *Arabidopsis thaliana* are orthologous to bacterial enzymes (Katoh, et al. 2006; Noctor, et al. 2011). Analysis of the glaucophyte *Cyanophora paradoxa* genome (Price, et al. 2012) indicates the presence of the aspartate pathway (see below). However, the third major lineage within Archaeplastida, red algae (Rhodophyta), solely encodes the kynurenine pathway for *de novo* NAD^+ biosynthesis. All five red algal genomes, *Cyanidioschyzon merolae* (Matsuzaki, et al. 2004), *Galdieria sulphuraria* (Schönknecht, et al. 2013), *Pyropia yezoensis* (Nakamura, et al. 2013), *Chondrus crispus* (Collén, et al. 2013), and *Porphyridium purpureum* (Bhattacharya, et al. 2013) appear to encode the kynurenine pathway, whereas the aspartate pathway appears to be absent.

Interestingly, diatoms and brown algae, which possess secondary plastids of red algal origin, and Oomycota (water molds or downy mildew) appear to use the aspartate pathway for quinolinate biosynthesis, and not the conserved kynurenine pathway observed in red algae and some other Stramenopile groups (Labyrinthulomycetes and Pelagophyceae) (Fig. 27). This raises the question of how diatoms, brown algae, and Oomycota might have acquired the aspartate pathway? Thousands of genes of potential green algal origin have been detected in diatoms, resulting in the hypothesis that Stramenopiles have undergone a cryptic secondary endosymbiosis

with a green alga, prior to acquiring the red algal endosymbiont (Moustafa, et al. 2009). However, this hypothesis has been met with much criticism (Woehle, et al. 2011; Burki, Flegontov, et al. 2012; Deschamps and Moreira 2012). The basal Stramenopile group, Labyrinthulomycetes, (which lack plastids) encodes enzymes for the kynurenine pathway, indicating this lineage might have split off from other Stramenopiles before the aspartate pathway was acquired. Surprisingly, the genome of the pelagophyte *A. anophagefferens* (Gobler, et al. 2011) lacks genes for the aspartate pathway, but encodes for the kynurenine pathway, in contrast to other photosynthetic Stramenopiles (Fig. 27), which can probably not be explained by conservation of the kynurenine pathway (see below). The presence of the kynurenine pathway in Cryptophyta and Haptophyta, which also acquired plastids independently by secondary endosymbiosis with a red alga, can probably be best explained by conservation of the kynurenine pathway in the host (Fig. 27).

EVOLUTIONARY HISTORY OF ASPARTATE PATHWAY ENZYMES

The aspartate pathway consists of two reaction steps whereby aspartate is converted into quinolinate, the precursor metabolite at which both *de novo* NAD^+ biosynthesis pathways converge (Fig. 2). **L-aspartate oxidase** (EC 1.4.3.16) catalyzes the first step; L-aspartate is reduced by electrons from molecular oxygen to form iminoasparate. Surprisingly, phylogenetic analysis revealed that aspartate oxidases found in three (photosynthetic) eukaryotic lineages do not form a monophyletic group. An evolutionary tree of aspartate oxidases (Fig. 28) shows the enzyme from *Cyanophora paradoxa* as monophyletic with cyanobacterial aspartate oxidases. This was expected for a gene that was acquired via endosymbiotic gene transfer from a cyanobacterial endosymbiont. The protein sequence from *C. paradoxa* resides deep within a cyanobacterial branch next to a sequence from *Thermosynechococcus elongates* BP-1, close to the position where plastids were presumed to branch from the cyanobacterial tree according to

Criscuolo and Gribaldo (2011). However, aspartate oxidases from Viridiplantae and Stramenopiles are neither closely related to cyanobacterial sequences, as expected specifically for Viridiplantae, nor show similar evolutionary origins. Instead, sequences from Viridiplantae form a monophyletic group with Bacteroidetes, whereas Stramenopile oxidases are monophyletic with spirochaetes and salt-loving, aquatic archaea.

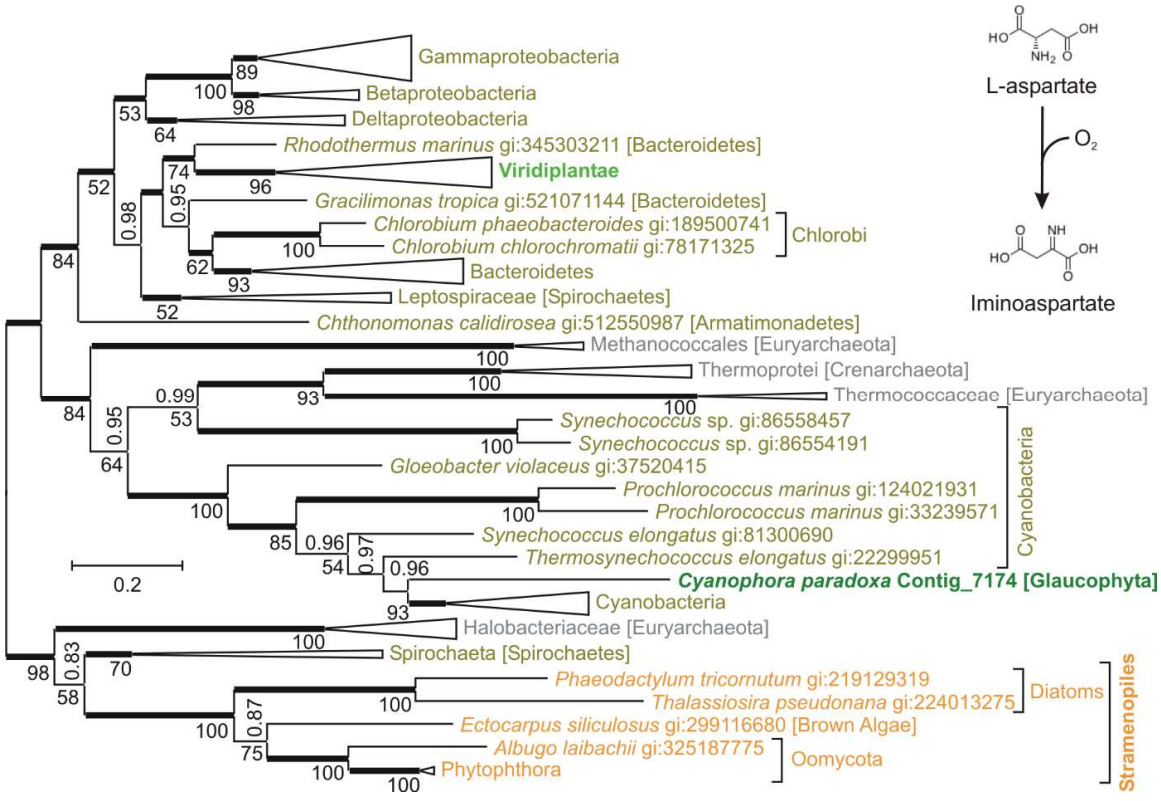


Fig. 28: Evolutionary tree for L-aspartate oxidase indicates separate origins of eukaryotic sequences. The unrooted Bayesian tree shows posterior probabilities above the branches and PhyML bootstrap values displayed as percentages below the branches. Thickened horizontal lines represent 1.0 Bayesian posterior probability. Larger clades have been collapsed for presentation and size is indicative of the number of taxa within the clade. Scale bar represents 0.2 substitutions per site.

The most surprising result of the aspartate oxidase phylogeny is that no eukaryotic lineage is monophyletic with another. All three of the eukaryotic lineages reside within different

parts of the phylogeny ruling out a direct, vertical inheritance of aspartate oxidase among these eukaryotic lineages. Instead, all three eukaryotic lineages appear to originate from three different lineages within Bacteria or Archaea. Alternative evolutionary trees with topological constraints enforcing monophyly of any two of the three (including all three) eukaryotic, photosynthetic lineages had extremely low probabilities (Table 1). Additional evolutionary scenarios were also tested (Table 1) including a monophyletic group consisting of Halobacteriaceae and Stramenopiles, and a monophyletic group consisting of Viridiplantae and Proteobacteria. Even though the Halobacteriaceae sit immediately basal to Spirochaeta and Stramenopiles with good support, a monophyletic group consisting only of Halobacteriaceae and Stramenopiles cannot be ruled out statistically. Thus, results are inconclusive regarding which group (Spirochaeta or Halobacteriaceae) actually forms the sister-relationship with the Stramenopiles. Monophyly of Viridiplantae and proteobacteria was tested based on results from the phylogenetic analysis of quinolinate synthase (see below). Despite the Bayesian posterior probability of 1.0 for some clades deep within bacteria containing Viridiplantae, bootstrap support for some clades was low (<50%). Given the latter, it was not surprising that monophyly of Viridiplantae and proteobacteria also cannot be ruled out statistically under maximum likelihood. Another relationship observed in the quinolinate synthase tree (see below) was tested where proteobacteria and Stramenopiles encoding the aspartate pathway were forced to be monophyletic; that scenario can be ruled out statistically.

Aspartate Oxidase Monophyletic Constraints	Bayesian				Maximum Likelihood			
	H	Δ*H*	AICM	ΔAICM	log *L*	Δlog *L*	AU	KH
Eukaryota	-51141	323.0	102430	-634.2	-50703	314.4	3E-06	0
Viridiplantae & Cyanophora	-51048	230.2	102249	-453.0	-50612	224.1	2E-05	0
Viridiplantae & Stramenopiles (asp)	-50952	134.9	102047	-251.3	-50507	118.4	4E-06	0.0001
Stramenopiles (asp) & Cyanophora	-50979	161.5	102109	-312.4	-50547	158.3	7E-06	9.0E-06
Bacteria	-51153	335.6	102460	-663.8	-50717	328.6	3E-254	0
Archaea	-51148	330.2	102452	-655.6	-50434	46.2	*0.054*	*0.058*
Viridiplantae & Proteobacteria	-50943	125.9	102043	-246.4	-50397	8.2	*0.32*	*0.33*
Stramenopiles (asp) & Proteobacteria	-51020	202.4	102207	-411.0	-50482	93.7	4.0E-05	0
Stramenopiles (asp) & Halobacteria	-50827	*9.1*	101802	*-5.8*	-50391	2.8	*0.274*	*0.29*

Table 1: Bayesian and Maximum Likelihood analyses of constraint trees for aspartate oxidase. From left to right, clades for which protein sequences were forced to be monophyletic, harmonic mean (*H*) of log-likelihood values from stationary phase of Bayesian analysis as calculated by Tracer (http://beast.bio.ed.ac.uk/Tracer), difference between harmonic means of unconstrained and constrained tree (Δ*H*), Akaike's information criterion through Markov chain Monte Carlo (AICM) as calculated by Tracer, difference in AICM between unconstrained and constrained tree (ΔAICM), log-likelihood (log *L*) of best maximum likelihood tree, difference between log-likelihoods (Δlog *L*) of unconstrained and constrained tree, *p*-values from approximately unbiased (AU) test and from Kishino-Hasegawa (KH) test as calculated by CONSEL (Shimodaira and Hasegawa 2001; Shimodaira 2002). Δ*H*, ΔAICM, and *p*-values indicating no significant difference in italics and reduced font size.

The second reaction step of the aspartate pathway is catalyzed by **quinolinate synthase** (EC 2.5.1.72). Iminoasparate reacts with dihydroxyacetone phosphate to generate quinolinate. An evolutionary tree of quinolinate synthases (Fig. 29) shows the sequence from *C. paradoxa* residing deep within a cyanobacterial branch, in agreement with an origin by endosymbiotic gene transfer. The gene for quinolinate synthase is encoded by the cyanelle (chloroplast) genome of *C. paradoxa*, providing further support for endosymbiotic origin. Quinolinate synthases from Stramenopiles and Viridiplantae are not closely related to cyanobacterial sequences, but form a separate monophyletic group, together with a synthase from the proteobacterium *Plesiocystis pacifica*. This separate monophyletic group, as well as the descent of *C. paradoxa* quinolinate

synthase from cyanobacterial sequences, is well supported by both Bayesian posterior probabilities and bootstrap support values. Alternative trees with topological constraints enforcing monophyly of all three photosynthetic eukaryotic clades had extremely low probabilities (Table 2), ruling out a direct, vertical inheritance. The monophyly of *C. paradoxa* and Viridiplante was tested to determine if vertical inheritance was a possible scenario, being they both reside within Archaeplastida and utilize the aspartate pathway, but statistical results ruled out this relationship. Relationships that were observed in the aspartate oxidase phylogeny were also tested to assess if both genes for the aspartate pathway were possibly acquired in the same event that established aspartate oxidases. Alternative trees (constructed irrespectively) with monophyletic Viridiplantae and Bacteriodetes, monophyletic Stramenopiles and Halobacteriaceae, and monophyletic Stramenopiles and Spirochaetes were insignificant, supporting different origins of both aspartate pathway enzymes in Viridiplantae and Stramenopiles. Evolutionary relationships observed in the aspartate oxidase phylogeny above were also tested for quinolinate synthase. These included monophyly of Viridiplantae and Bacteriodetes, Halobacteriaceae and Stramenopiles encoding the aspartate pathway, and Spirochaetes and Stramenopiles encoding the aspartate pathway. All three evolutionary scenarios could be rejected statistically (Table 2).

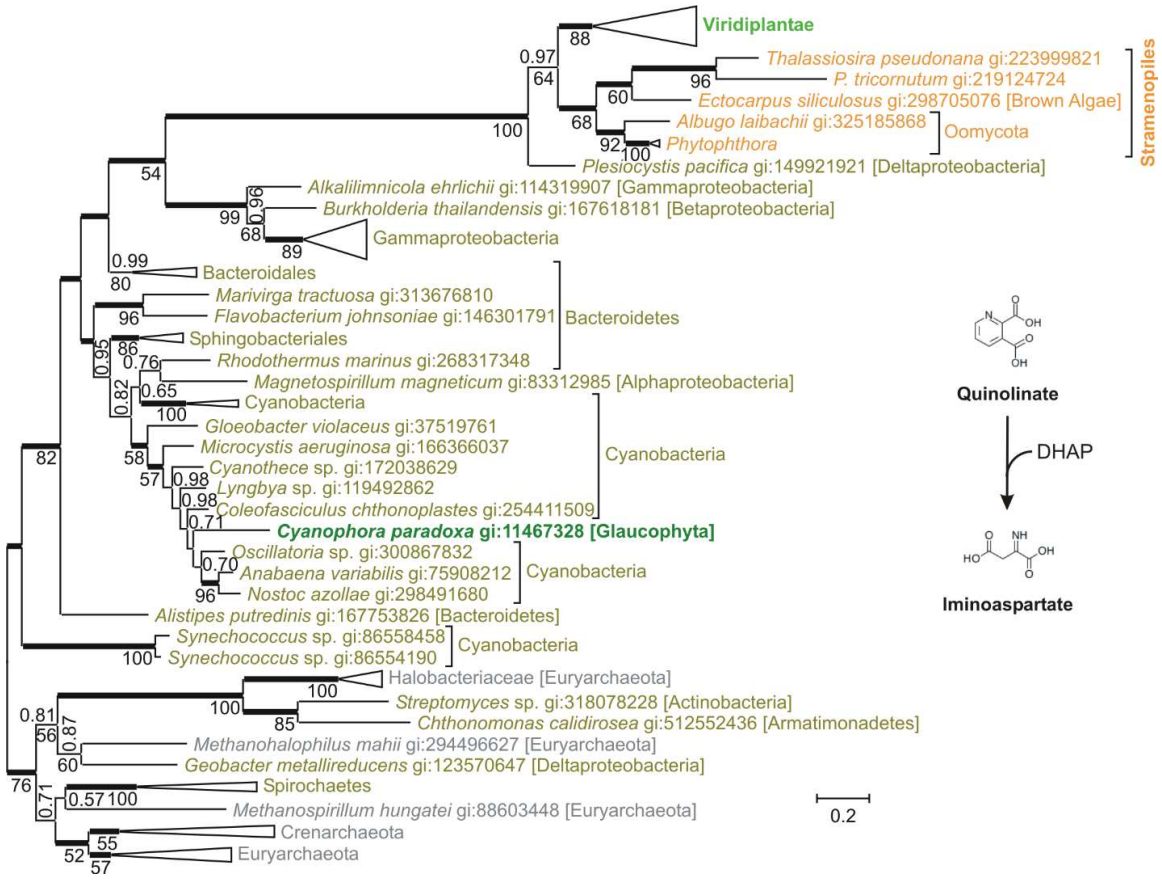


Fig. 29: Evolutionary tree for quinolinate synthase indicates separate origins of eukaryotic sequences. The unrooted Bayesian tree shows posterior probabilities above the branches and PhyML bootstrap values below the branches. Thickened horizontal lines represent 1.0 Bayesian posterior probability. Larger clades have been collapsed for presentation and size is indicative of the number of taxa within the clade. Scale bar represents 0.2 substitutions per site.

It appears that aspartate oxidase and quinolinate synthase have different evolutionary origins in different photosynthetic, eukaryotic clades. To explain this somewhat complex evolutionary pattern, several separate gene transfer events must be postulated. The high similarity of both enzymes of the aspartate pathway between the glaucophyte *C. paradoxa* and cyanobacteria seems to support acquisition of the aspartate pathway by photosynthetic eukaryotes via endosymbiotic gene transfer from the cyanobacterial endosymbiont. When the three lineages of Archaeplastida (namely Glaucophyta, red algae, and green plants and algae) split, Glaucophyta

seem to have kept the cyanobacterial genes encoding aspartate oxidase and quinolinate synthase, while the ancestor of Viridiplantae appears to have substituted the cyanobacterial genes via horizontal gene transfer with genes from non-photosynthetic bacteria. In the lineage giving rise to red algae, the aspartate pathway was not established, and the kynurenine pathway remained active. The monophyly of quinolinate synthases from Viridiplantae and Stramenopiles indicates horizontal gene transfer between green plants and algae and Stramenopiles. The close similarity of Stramenopile aspartate oxidases to prokaryotic aspartate oxidases could be explained by another horizontal gene transfer from a spirochaete or a halobacterium into an ancestor of Stramenopiles. In higher plants it has been demonstrated that aspartate oxidase and quinolinate synthase are located in plastids, in agreement with their endosymbiotic origin (Katoh, et al. 2006). Somewhat unexpected, nicotinate-nucleotide pyrophosphorylase (see below) of higher plants is targeted to plastids as well (Katoh, et al. 2006).

Quinolinate Synthase Monophyletic Constraints	Bayesian				Maximum Likelihood			
	H	Δ*H*	AICM	ΔAICM	log *L*	Δlog *L*	AU	KH
Eukaryota	-29507	343.8	59159	-684.4	-29176	332.1	1.0E-92	0
Viridiplantae & Cyanophora	-29552	388.7	59274	-799.8	-29219	375.8	9.0E-121	0
Bacteria	-29842	678.7	59838	-1363.1	-29465	621.4	3.0E-29	0
Archaea	-29282	118.9	58714	-239.5	-28965	121.4	1.0E-06	0
Viridiplantae & Bacteriodetes	-29577	414.2	59312	-837.6	-29248	404.2	4.0E-116	0
Stramenopiles & Halobacteriace	-29407	244.0	58975	-500.6	-29084	240.1	2.0E-14	0
Stramenopiles & Spirochaetes	-29419	256.2	59000	-524.9	-29089	245.3	3.0E-12	0

Table 2: Bayesian and Maximum Likelihood analyses of constraint trees for quinolinate synthase. From left to right, clades for which protein sequences were forced to be monophyletic, harmonic mean (*H*) of log-likelihood values from stationary phase of Bayesian analysis as calculated by Tracer (http://beast.bio.ed.ac.uk/Tracer), difference between harmonic means of unconstrained and constrained tree (Δ*H*), Akaike's information criterion through Markov chain Monte Carlo (AICM) as calculated by Tracer, difference in AICM between unconstrained and constrained tree (ΔAICM), log-likelihood (log *L*) of best maximum likelihood tree, difference between log-

likelihoods (Δlog *L*) of unconstrained and constrained tree, *p*-values from approximately unbiased (AU) test and from Kishino-Hasegawa (KH) test as calculated by CONSEL (Shimodaira and Hasegawa 2001; Shimodaira 2002).

EVOLUTIONARY HISTORY OF THE CONVERGED PATHWAY ENZYMES

Nicotinate-nucleotide pyrophosphorylase (EC 2.4.2.19) catalyzes the reaction of quinolinate with 5-phosphoribosyl diphosphate to synthesize nicotinate D-ribonucleotide plus pyrophosphate and CO_2 (Fig. 2). Because the aspartate pathway and kynurenine pathway converge at quinolinate, all free-living eukaryotes contain nicotinate-nucleotide pyrophosphorylase. The evolutionary tree for this enzyme shows eukaryotic sequences forming one monophyletic branch, with the exception of sequences from green plants and algae (Viridiplantae) and *Bigelowiella natans*, which reside together deep within the bacterial branch (Fig. 30). Nicotinate-nucleotide pyrophosphorylases in most eukaryotic lineages appear to descend from the last common ancestor of all eukaryotes. In contrast, nicotinate-nucleotide pyrophosphorylases from Viridiplantae form a sister group to sequences from Bacteroidetes, indicating the nicotinate-nucleotide pyrophosphorylase gene of eukaryotic origin was substituted by horizontal gene transfer from a bacterium in an early ancestor of Viridiplantae. The sequence from *C. paradoxa* was omitted from an evolutionary analysis because it is incomplete. The best BLASTP hit for the incomplete sequence from *C. paradoxa* using the non-redundant database from NCBI was with a nicotinate-nucleotide pyrophosphorylase sequence from *Capsaspora owczarzaki* (167 score, 67% coverage), and all top 100 BLASTP hits were with sequences from non-photosynthetic eukaryotes, indicating possible descent of nicotinate-nucleotide pyrophosphorylase in *C. paradoxa* from the last common ancestor of eukaryotes. Alternative trees enforcing monophyly of all eukaryotic sequences had low probabilities (Table 3). The clade comprised of *B. natans*, Viridiplantae, and Bacteriodetes is not statistically concrete. Alternative

trees (constructed irrespectively) with monophyletic Viridiplantae and *B. natans*, and monophyletic Viridiplantae and cyanobacteria cannot be ruled out. Despite the uncertain topology regarding Viridiplantae and *B. natans***,** the pyrophorphorylases present in these organisms is not of eukaryotic origin. The nicotinate-nucleotide pyrophosphorylase gene from tobacco has been shown to functionally complement *E. coli* cells lacking the orthologous enzyme (Sinclair, et al. 2000), emphasizing the similarity between higher plant and bacterial orthologs.

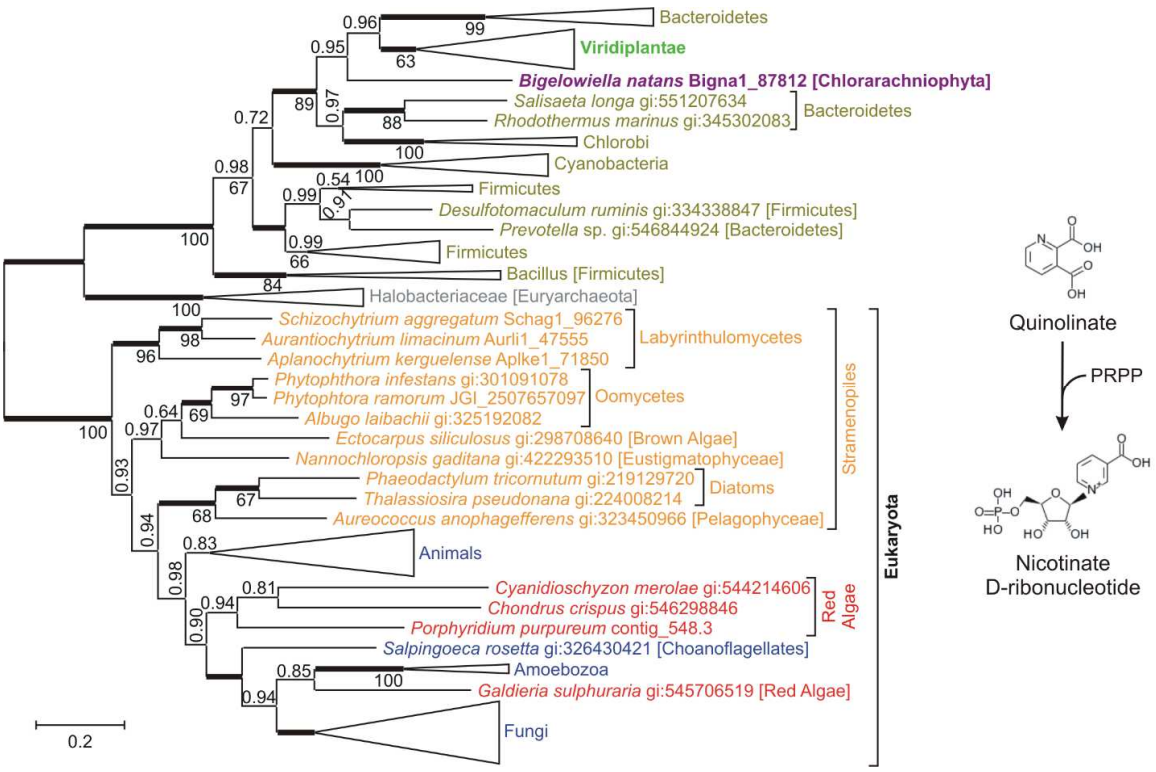


Fig. 30: An evolutionary tree for nicotinate-nucleotide pyrophosphorylase indicates that some photosynthetic eukaryotes acquired genes via transfers. The unrooted Bayesian tree shows posterior probabilities above the branches and PhyML bootstrap values below the branches. Thickened horizontal lines represent 1.0 Bayesian posterior probability. Larger clades have been collapsed for presentation and size is indicative of the number of taxa within the clade. Scale bar represents 0.2 substitutions per site. Tree produced by G. Schönknecht.

The monophyletic relationship between Viridiplantae and Stramenopiles (possessing the aspartate pathway) observed in the quinolinate synthase phylogeny can be ruled out statistically (Table 3). Some discrepancies are present in the eukaryotic clade, i.e. a disbanded Opisthokonta (red algal sequences reside within the clade resulting in paraphyletic opisthokonts). However, forcing Opisthokonta to be monophyletic resulted in a topology that was not significantly different from the topology presented here. Furthermore, enforcing other major eukaryotic lineages to be monophyletic (i.e. Stramenopiles, Unikonta, red algae, etc.) resulted in topologies also not significantly different from the topology presented here. Lastly, monophyly of Stramenopiles and Halobacteriaceae cannot be statistically ruled out indicating a possible gene transfer between the ancestors of the two lineages. The similarity of nicotinate-nucleotide pyrophosphorylase from *B. natans* and green plants can probably be explained by endosymbiotic gene transfer. The chlorarachniophyte *B. natans* is known to have acquired its plastid via secondary endosymbiosis with an ancient green alga (Archibald, et al. 2003). Endosymbiotic gene transfer from an ancestral green algal endosymbiont into the genome of *B. natans* probably resulted in the substitution of the eukaryotic pyrophosphorylase gene.

Nicotinate-Nucleotide Pyrophosphorylase Monophyletic Constraints	Bayesian				Maximum Likelihood			
	H	Δ*H*	AICM	ΔAICM	log *L*	Δlog *L*	AU	KH
Eukaryota	-38680	61.6	77583	-103.6	-38283	59.9	0.019	0.02
Viridiplantae & Bigelowiella	-38630	*11.5*	77485	*-5.5*	-38224	1.1	*0.474*	*0.476*
Viridiplantae & Cyanobacteria	-38654	*36.1*	77555	*-75.7*	-38250	27.8	*0.099*	*0.092*
Viridiplantae & Red Algae	-38851	232.9	77943	-463.2	-38443	219.9	4.0E-06	0
Viridiplantae & Stramenopiles (asp)	-38857	238.4	77958	-478.1	-38429	206.4	5.0E-15	0
Bigelowiella & Stramenopiles [SAR]	-38783	165.0	77803	-322.9	-38381	158.5	7.0E-08	0
Bacteria	-38642	*23.5*	77519	*-39.3*	-38253	29.9	*0.077*	*0.072*
Opisthokonta	-38641	*22.9*	77524	*-43.9*	-38235	11.8	*0.173*	*0.172*
Unikonta	-38637	*19.3*	77496	*-16.8*	-38235	11.8	*0.172*	*0.174*
Red Algae	-38624	*5.7*	77478	*1.2*	-38231	8.5	*0.232*	*0.223*
Stramenopiles	-38633	*15.0*	77478	*2.1*	-38226	3.4	*0.417*	*0.424*
Stramenopiles (asp) & Halobacteriaceae	-38686	67.9	77623	-143.1	-38273	49.9	*0.057*	*0.06*

Table 3: Bayesian and Maximum Likelihood analyses of constraint trees for nicotinate-nucleotide pyrophosphorylase. From left to right, clades for which protein sequences were forced to be

monophyletic, harmonic mean (*H*) of log-likelihood values from stationary phase of Bayesian analysis as calculated by Tracer (http://beast.bio.ed.ac.uk/Tracer), difference between harmonic means of unconstrained and constrained tree (Δ*H*), Akaike's information criterion through Markov chain Monte Carlo (AICM) as calculated by Tracer, difference in AICM between unconstrained and constrained tree (ΔAICM), log-likelihood (log *L*) of best maximum likelihood tree, difference between log-likelihoods (Δlog *L*) of unconstrained and constrained tree, *p*-values from approximately unbiased (AU) test and from Kishino-Hasegawa (KH) test as calculated by CONSEL (Shimodaira and Hasegawa 2001; Shimodaira 2002). Δ*H*, ΔAICM, and *p*-values indicating no significant difference in italics and reduced font size.

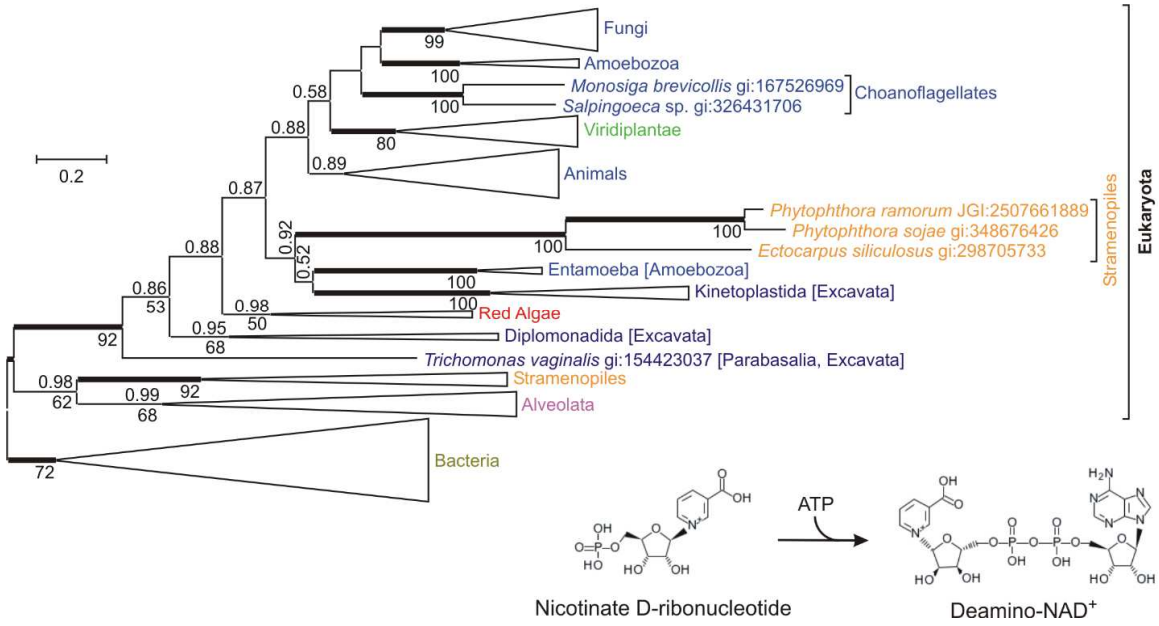


Fig. 31: An evolutionary tree for nicotinate-nucleotide adenylyltransferase shows the expected separation of bacterial and eukaryotic sequences. The unrooted Bayesian tree was generated using the VT+G model of protein evolution with posterior probabilities above the branches and PhyML bootstrap values (LG+G) displayed as percentages below the branches. A score of 3 to 9 instead of 5 to 9 was used to extract multiple sequence alignments. Thickened horizontal lines represent 1.0 Bayesian posterior probability. Larger clades have been collapsed for presentation and size is indicative of the number of taxa within the clade. Scale bar represents 0.2 substitutions per site.

Nicotinate-nucleotide adenylyltransferase catalyzes the conversion of ATP plus β-nicotinate D-ribonucleotide into deamido-NAD^+ plus pyrophosphate (EC 2.7.7.18). This enzyme also catalyzes the conversion of ATP plus nicotinamide D-ribonucleotide into NAD^+ plus diphosphate (nicotinamide-nucleotide adenylyltransferase; EC 2.7.7.1). In an evolutionary tree of nicotinate-nucleotide adenylyltransferases (Fig. 31), eukaryotic sequences form one monophyletic branch, and bacterial sequences form another branch. The topology of the eukaryotic branch does not reflect established eukaryotic phylogeny (compare to Fig. 27). Enforcing monophyly of established eukaryotic lineages (with the exception of the Stramenopiles and Amoebozoa) results in alternative trees with almost identical log-likelihood values (Table 4).

Nicotinate-Nucleotide Adenylyltransferase	Bayesian				Maximum Likelihood			
Monophyletic Constraints	*H*	Δ*H*	AICM	ΔAICM	log *L*	Δlog *L*	AU	KH
Amoebozoa	-41091	*15.5*	82552	*-24.9*	-40058	20.2	0.016	0.030
Excavata	-41101	*25.9*	82577	*-50.6*	-40051	13.0	*0.310*	*0.307*
Stramenopiles	-41135	59.1	82601	*-74.5*	-40086	48.2	0.004	0.006
Unikonta	-41089	*13.1*	82537	*-10.7*	-40053	15.3	*0.191*	*0.196*
Viridiplantae & Red Algae	-41087	*11.4*	82550	*-23.3*	-40048	10.5	*0.367*	*0.350*

Table 4: Bayesian and Maximum Likelihood analyses of constraint trees for nicotinate-nucleotide adenylyltransferase. From left to right, clades for which protein sequences were forced to be monophyletic, harmonic mean (*H*) of log-likelihood values from stationary phase of Bayesian analysis as calculated by Tracer (http://beast.bio.ed.ac.uk/Tracer), difference between harmonic means of unconstrained and constrained tree (Δ*H*), Akaike's information criterion through Markov chain Monte Carlo (AICM) as calculated by Tracer, difference in AICM between unconstrained and constrained tree (ΔAICM), log-likelihood (log *L*) of best maximum likelihood tree, difference between log-likelihoods (Δlog *L*) of unconstrained and constrained tree, *p*-values from approximately unbiased (AU) test and from Kishino-Hasegawa (KH) test as calculated by CONSEL (Shimodaira and Hasegawa 2001; Shimodaira 2002). Δ*H*, ΔAICM, and *p*-values indicating no significant difference in italics and reduced font size.

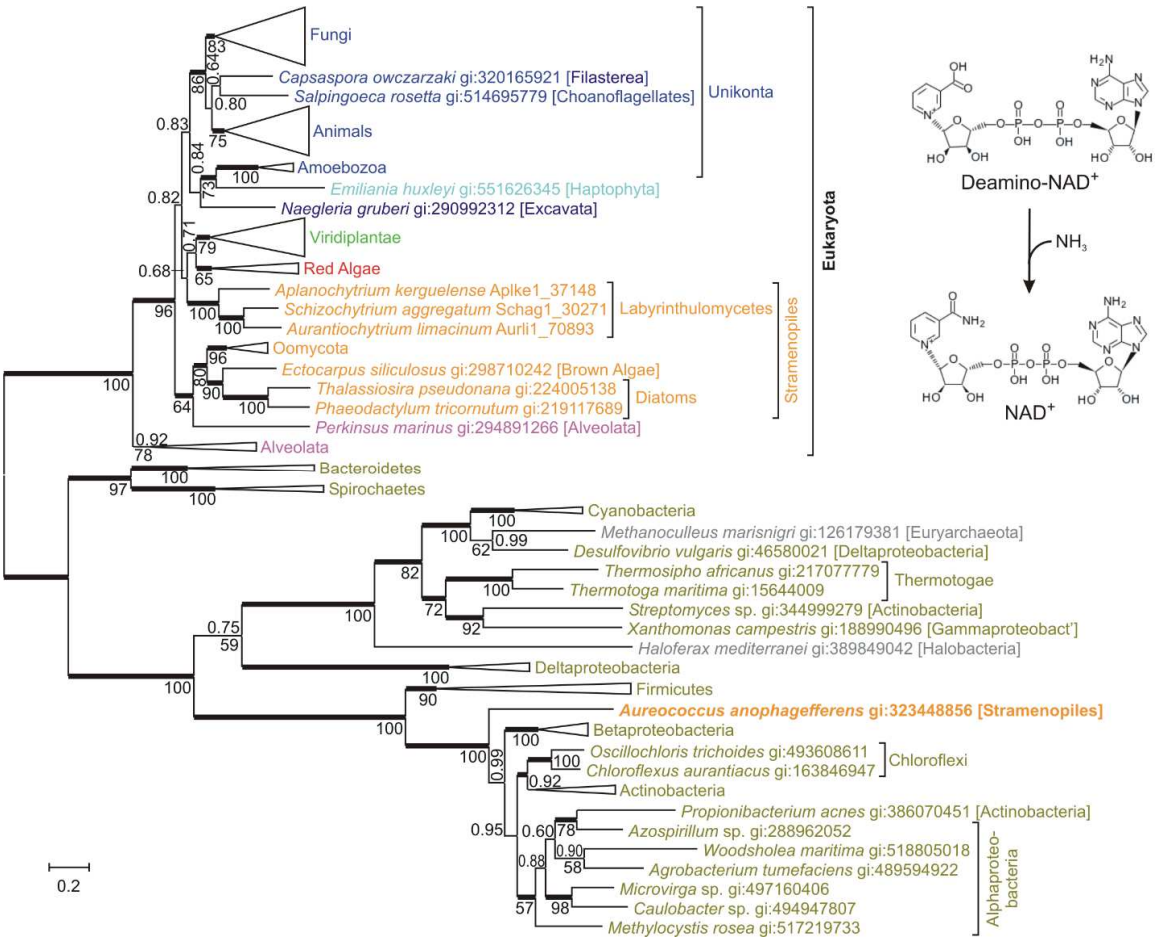


Fig. 32: An evolutionary tree for NAD+ synthases shows the expected separation between eukaryotes and prokaryotes with the exception of *A. anophagefferens*, which possibly acquired a gene via horizontal gene transfer from a bacterium. The unrooted Bayesian tree shows posterior probabilities above the branches and PhyML bootstrap values (LG+I+G+F) below the branches. Thickened horizontal lines represent 1.0 Bayesian posterior probability. Larger clades have been collapsed for presentation and size is indicative of the number of taxa within the clade. Scale bar represents 0.2 substitutions per site.

The final step in *de novo* NAD^+ biosynthesis is catalyzed by **NAD^+ synthase** (EC 6.3.5.1), catalyzing the conversion of deamino-NAD^+ into NAD^+ using glutamine as an NH_3 donor (Fig. 2). An evolutionary tree of NAD^+ synthases shows all eukaryotic sequences, with exception of the sequence from *A. anophagefferens*, as a monophyletic group (Fig. 32). The

NAD^+ synthase sequence from *A. anophagefferens* resides within a branch composed entirely of bacterial protein sequences, suggesting horizontal gene transfer from a bacterium. An alternative tree enforcing monophyly of all eukaryotic sequences had extremely low probability (Table 5). Enforcing monophyly of established eukaryotic lineages that are disbanded in the unconstrained phylogeny (Unikonta, stramenopiles [minus *A. anophagefferens*], and Alveolata) resulted in trees that were not significantly different from the tree presented here.

NAD^+ Synthase Monophyletic Constraints	Bayesian				Maximum Likelihood			
	H	Δ*H*	AICM	ΔAICM	log *L*	Δlog *L*	AU	KH
Eukaryota	-101181	222.2	202620	-457.7	-100249	226.8	0	0
Bacteria	-101692	733.3	203653	-1490.4	-100755	733.3	0	0
Unikonta	-100973	*13.9*	202195	*-32.0*	-100040	18.3	*0.209*	*0.190*
Stramenopiles (all - A.a.)	-101135	176.1	202510	-347.6	-100040	17.9	*0.204*	*0.203*
Alveolata	-101000	40.8	202213	*-50.1*	-100046	24.1	*0.143*	*0.138*

Table 5: Bayesian and Maximum Likelihood analyses of constraint trees for NAD^+ synthase. From left to right, clades for which protein sequences were forced to be monophyletic, harmonic mean (*H*) of log-likelihood values from stationary phase of Bayesian analysis as calculated by Tracer (http://beast.bio.ed.ac.uk/Tracer), difference between harmonic means of unconstrained and constrained tree (Δ*H*), Akaike's information criterion through Markov chain Monte Carlo (AICM) as calculated by Tracer, difference in AICM between unconstrained and constrained tree (ΔAICM), log-likelihood (log *L*) of best maximum likelihood tree, difference between log-likelihoods (Δlog *L*) of unconstrained and constrained tree, *p*-values from approximately unbiased (AU) test and from Kishino-Hasegawa (KH) test as calculated by CONSEL (Shimodaira and Hasegawa 2001; Shimodaira 2002). Δ*H*, ΔAICM, and *p*-values indicating no significant difference in italics and reduced font size.

EVOLUTIONARY HISTORY OF KYNURENINE PATHWAY ENZYMES

Given the numerous gene transfer events for the aspartate pathway, phylogenetic analyses were also conducted for the kynurenine pathway to elucidate possible gene transfer events. The first reaction in the kynurenine pathway produces N-formylkynurenine from tryptophan and molecular oxygen. This reaction is catalyzed by either the highly specific tryptophan 2,3-dioxygenase (EC 1.13.11.11) or the less specific indoleamine 2,3-dioxygenase (EC 1.13.11.52). An evolutionary tree of **tryptophan 2,3-dioxygenase** does not show the expected separation of eukaryotic and bacterial protein sequences (Fig. 33), and alternative trees enforcing monophyly of all eukaryotic or all bacterial sequences have low probability (Table 6). The sequence from *Naegleria gruberi* is the only eukaryotic protein sequence in a prokaryotic branch, probably indicating horizontal gene transfer of a bacterial tryptophan 2,3-dioxygenase gene into (an ancestor of) *N. gruberi*.

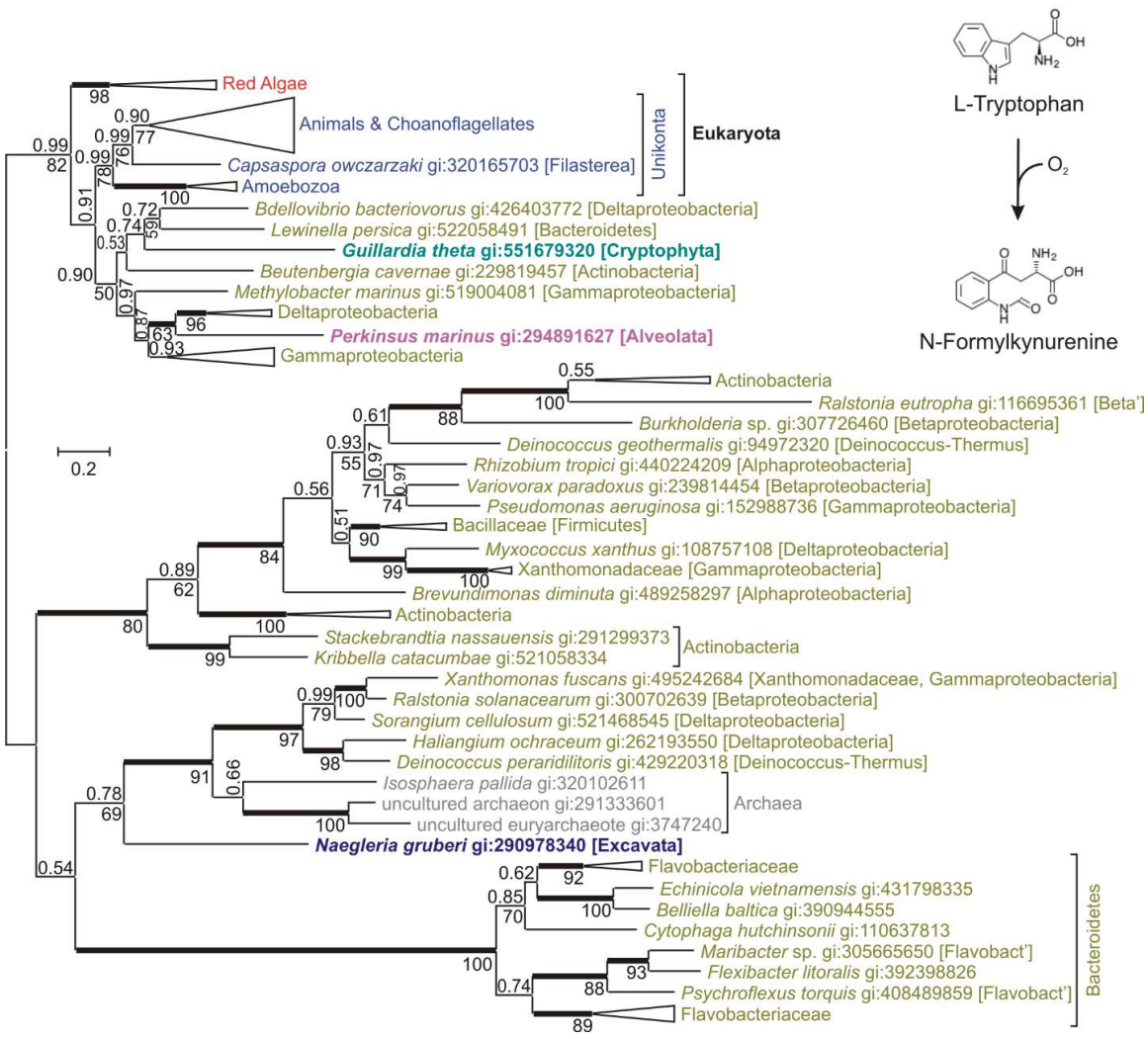


Fig. 33: An evolutionary tree for tryprophan 2,3-dioxygenase indicates that *N. gruberi* acquired a gene via horizontal gene transfer. The unrooted Bayesian tree shows posterior probabilities above the branches and PhyML bootstrap values below the branches. A score of 3 to 9 instead of 5 to 9 was used to extract multiple sequence alignments. Thickened horizontal lines represent 1.0 Bayesian posterior probability. Larger clades have been collapsed for presentation and size is indicative of the number of taxa within the clade. Scale bar represents 0.2 substitutions per site.

Monophyly of unikonts and red algae was tested in order to determine if they could be treated as one monophyletic group when testing other topologies. Enforcing monophyly of 1) unikonts, red algae, and *Perkinsus marinus* and 2) red algae, unikonts, and *Guillardia theta*

(irrespective of each other) resulted in topologies that could not be rejected statistically. Furthermore, an alternative tree forcing *G. theta* and *P. marinus* to be monophyletic was also significant. Horizontal gene transfer between eukaryotes is a possible scenario. However, this opens the door for more possibilities such as eukaryote→eukaryote→bacteria or bacteria→eukaryote→eukaryote gene transfer. The tryptophan 2,3-dioxygenases from *P. marinus* and *G. theta* might originate from horizontal gene transfer from bacteria, or bacteria might have acquired eukaryotic genes, followed by gene transfers between bacteria. The evolutionary analysis of tryptophan 2,3-dioxygenases was limited by relatively short multiple sequence alignments (250 to 350 amino acids) and limited sequence availability (some eukaryotic clades only contain indoleamine 2,3-dioxygenase, such as fungi mentioned below).

Tryptophan Dioxygenase Monophyletic Constraints	Bayesian				Maximum Likelihood			
	H	Δ*H*	AICM	ΔAICM	log *L*	Δlog *L*	AU	KH
Eukaryotes (all)	-40116	68.6	80399	-149.6	-39838	73.6	0.013	0.011
Unikonts & Red Algae	-40038	*-8.7*	80267	*-17.2*	-39769	5.1	*0.370*	*0.357*
Unikonts & Red Algae & Naegleria	-40095	48.2	80375	-125.2	-39817	52.6	0.001	0.001
Unikonts & Red Algae & Perkinsus	-40057	*10.5*	80300	*-50.2*	-39791	26.6	*0.077*	*0.081*
Unikonts & Red Algae & Guillardia	-40044	*-2.7*	80257	*-7.1*	-39772	8.4	*0.283*	*0.284*
Perkinsus & Guillardia	-40051	*4.0*	80282	*-31.8*	-39781	16.7	*0.054*	*0.054*
Bacteria	-40206	158.9	80596	-346.0	-39927	162.6	6.00E-18	0
Gammaproteobacteria	-40772	725.5	81733	-1483.1	-40492	728.2	9.00E-70	0
Deltaproteobacteria	-40771	724.5	81690	-1440.3	-40469	704.8	2.00E-05	0

Table 6: Bayesian and Maximum Likelihood analyses of constraint trees for tryptophan 2,3-dioxygenase. From left to right, clades for which protein sequences were forced to be monophyletic, harmonic mean (*H*) of log-likelihood values from stationary phase of Bayesian analysis as calculated by Tracer (http://beast.bio.ed.ac.uk/Tracer), difference between harmonic means of unconstrained and constrained tree (Δ*H*), Akaike's information criterion through Markov chain Monte Carlo (AICM) as calculated by Tracer, difference in AICM between unconstrained and constrained tree (ΔAICM), log-likelihood (log *L*) of best maximum likelihood tree, difference between log-likelihoods (Δlog *L*) of unconstrained and constrained tree, *p*-values from approximately unbiased (AU) test and from Kishino-Hasegawa (KH) test as calculated by

CONSEL (Shimodaira and Hasegawa 2001; Shimodaira 2002). Δ*H*, ΔAICM, and *p*-values indicating no significant difference in italics and reduced font size.

An evolutionary tree of **indoleamine 2,3-dioxygenase** (Fig. 34) has three main branches, all three containing sequences from both Eukaryota and Bacteria. Alternative trees enforcing monophyly of all eukaryotic or all bacterial sequences have low probability (Table 7). However, only enzymes in one of the three main branches seem to catalyze the reaction of tryptophan and O_2 with an affinity, and at a rate expected for effective tryptophan degradation. This is the branch containing sequences from mammals, fungi, and the bacteria *Gemmatimonas aurantiaca* and *Streptomyces scabiei* (Yuasa, et al. 2011). Enzymes from the other two branches that were biochemically characterized show very low reaction rates, questioning a function in tryptophan degradation. These are fungal indoleamine 2,3-dioxygenases γ (Yuasa and Ball 2012) and recombinant indoleamine 2,3-dioxygenases from *Erythrobacter litoralis* (Alphaproeobacteria) and *Neptuniibacter caesariensis* (Gammaproteobacteria) (Yuasa, et al. 2011). On the other hand, indoleamine 2,3-dioxygenases from the stramenopiles *Aplanochytrium kerguelense*, *Aurantiochytrium limacinum*, *Schizochytrium aggregatum* (Labyrinthulomycetes), *A. anophagefferens* (Pelagophyceae), and the haptophyte *Emiliania huxleyi* located in these two branches were the only candidates identified in BLAST searches that seem suitable (based on multiple sequence alignments) to catalyze the reaction with tryptophan. It seems possible that for some Stramenopiles and Haptophyta, the first enzyme of the kynurenine pathway still awaits identification (indicated by grey open circles in Fig. 27).

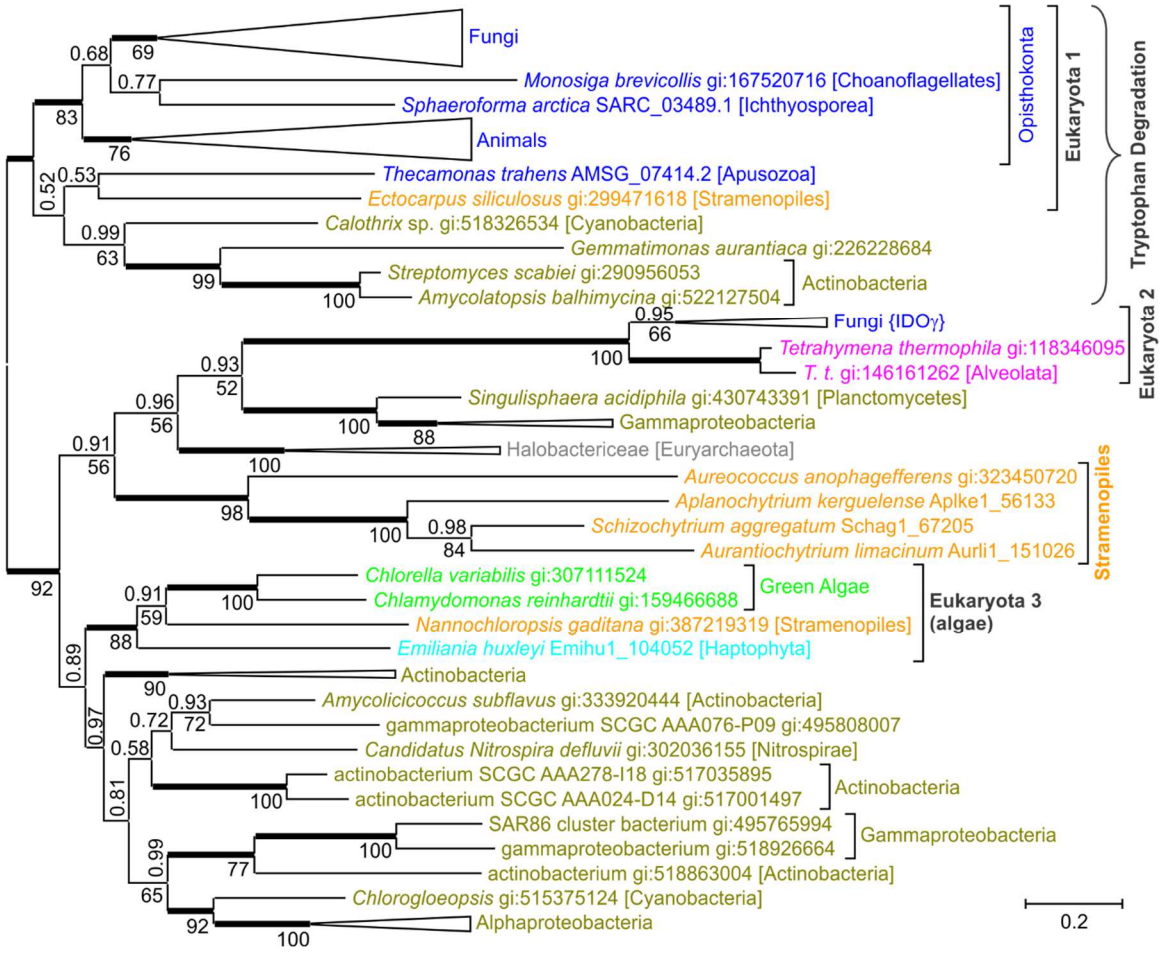


Fig. 34: An evolutionary tree for indoleamine-2,3-dioxygenase shows a complex mixture of bacterial and eukaryotic sequences. The unrooted Bayesian tree shows posterior probabilities above the branches and PhyML bootstrap values below the branches. Thickened horizontal lines represent 1.0 Bayesian posterior probability. Curly brackets refer to enzymatic activity; only enzymes from the upper branch show reaction rates as expected for effective tryptophan degradation, while fungal indoleamine-2,3-dioxygenase γ (IDOγ) and some bacterial enzymes in the lower two branches have very low reaction rates (Yuasa and Ball 2011; Yuasa and Ball 2012). Larger clades have been collapsed for presentation and size is indicative of the number of taxa within the clade. Scale bar represents 0.2 substitutions per site. Tree produced by G. Schönknecht.

An alternative tree enforcing the stramenopiles clade (three species of Labyrinthulomycetes and *A. anophagefferens*) to be monophyletic with *Ectocarpus siliculosus*

could not be rejected statistically (Table 7). The evolutionary tree containing the former constraint showed the Stamenopiles group being 'pulled up' to where *E. siliculosus* resides. Also, enforcing the 'Stramenopiles' and 'Eukaryota 2' clades to be monophyletic cannot be rejected (Table 7). While the unconstrained topology (Fig. 34) shows a rather distant relationship between 'Stramenopiles' and *E. siliculosus*, the exact position where the 'Stramenopiles' clade resides is uncertain, and could have implications regarding enzyme kinetics. Further biochemical investigation would be required to determine if the dioxygenases in the 'Stramenopiles' clade do in fact catalyze the formation of N-formylkynurenine from L-tryptophan.

Indoleamine Dioxygenase Monophyletic Constraints	Bayesian				Maximum Likelihood			
	H	Δ*H*	AICM	ΔAICM	log *L*	Δlog *L*	AU	KH
Eukaryota	-44187	82.5	88534	-166.0	-43890	78.8	9.0E-05	0.001
Bacteria	-44198	93.2	88541	-173.3	-43913	101.9	6.0E-08	0
Stramenopiles & Eukaryota 2	-44112	*7.9*	88374	*-6.3*	-43813	1.8	*0.408*	*0.414*
Stramenopiles & Eukaryota 2 & Eukaryota 1	-44150	45.2	88459	-91.2	-43851	40.0	0.026	0.030
Stramenopiles & Eukaryota 2 & Eukaryota 3	-44135	*30.9*	88422	*-54.2*	-43836	24.7	0.036	0.037
Eukaryota 1 & Eukaryota 3	-44171	66.7	88497	-129.0	-43873	62.0	0.0001	0.001
Ectocarpus & Eukaryota 3 (algae)	-44146	41.2	88451	-83.0	-43853	41.8	0.001	0.001
Ectocarpus & Stramenopiles	-44120	*15.6*	88398	*-30.2*	-43828	16.9	*0.134*	*0.136*
Stramenopiles (from all branches)	-44164	59.4	88490	-122.0	-43871	59.7	0.0001	0
Fungi	-44424	319.0	89004	-636.1	-44118	307.2	4.0E-09	0

Table 7: Bayesian and Maximum Likelihood analyses of constraint trees for indoleamine 2,3-dioxygenase. From left to right, clades for which protein sequences were forced to be monophyletic, harmonic mean (*H*) of log-likelihood values from stationary phase of Bayesian analysis as calculated by Tracer (http://beast.bio.ed.ac.uk/Tracer), difference between harmonic means of unconstrained and constrained tree (Δ*H*), Akaike's information criterion through Markov chain Monte Carlo (AICM) as calculated by Tracer, difference in AICM between unconstrained and constrained tree (ΔAICM), log-likelihood (log *L*) of best maximum likelihood tree, difference between log-likelihoods (Δlog *L*) of unconstrained and constrained tree, *p*-values from approximately unbiased (AU) test and from Kishino-Hasegawa (KH) test as calculated by CONSEL (Shimodaira and Hasegawa 2001; Shimodaira 2002). Δ*H*, ΔAICM, and *p*-values indicating no significant difference in italics and reduced font size.

In the second reaction step of the kynurenine pathway, N-formylkynurenine is hydrolyzed into L-kynurenine and formate. This reaction can be catalyzed by two different, nonorthologous enzymes. In many eukaryotes an α/β hydrolase fold enzyme with an esterase/lipase domain functions as **arylformamidase** (EC 3.5.1.9) (Pabarcus and Casida 2002; Wogulis, et al. 2008), whereas a cyclase functions as an arylformamidase in many bacteria (Kurnasov, et al. 2003). Eukaryotic arylformamidases show little conservation, with just 22% amino acid similarity (11% identity) between yeast and mouse (Wogulis, et al. 2008). Outside Opisthokonta very few eukaryotic arylformamidases have been annotated, and BLAST searches with opisthokont sequences do not result in meaningful hits outside Opisthokonta, preventing an phylogenetic analysis of eukaryotic arylformamidases. In the genome of the red alga *G. sulphuraria*, a cyclase with homology to bacterial arylformamidases has been annotated, and a phylogenetic analysis indicates that *G. sulphuraria* acquired this enzyme via horizontal gene transfer from Actinobacteria (Schönknecht, et al. 2013).

Kynurenine-3-monooxygenase (EC 1.14.13.9) catalyzes the third reaction step of the kynurenine pathway, the hydroxylation of L-kynurenine into 3-hydroxy-L-kynurenine. In an evolutionary tree of this enzyme, sequences from most eukaryotic clades form one major branch (Fig. 35). However, the other major branch, composed mostly of bacterial sequences, has several eukaryotic sequences interspersed, indicating possible instances of horizontal gene transfer. Alternative evolutionary trees enforcing monophyly of eukaryotic clades have low probability (Table 8). Earlier phylogenetic analyses of kynurenine-3-monooxygenases also place the sequence from *Dictyostelium discoideum* (Amoebozoa) in a branch with bacterial sequences (Lima, et al. 2009), in good agreement with results shown here. Phylogenomic analyses for the amoebozoans *D. discoideum* (Eichinger, et al. 2005; Andersson 2011) and *Entamoeba histolytica* (Loftus, et al. 2005) identified several instances of horizontal gene transfer from bacterial genomes, and it has been suggested that the phagotrophic life style of amoebozoa promoted DNA

uptake (Ford Doolittle 1998). Interestingly, none of the three phylogenomic studies with amoebozoa seem to have identified kynurenine-3-monooxygenase as a candidate for horizontal gene transfer. This might be caused by several interspersed eukaryotic sequences rather than just one, making it harder to detect a specific horizontal gene transfer event.

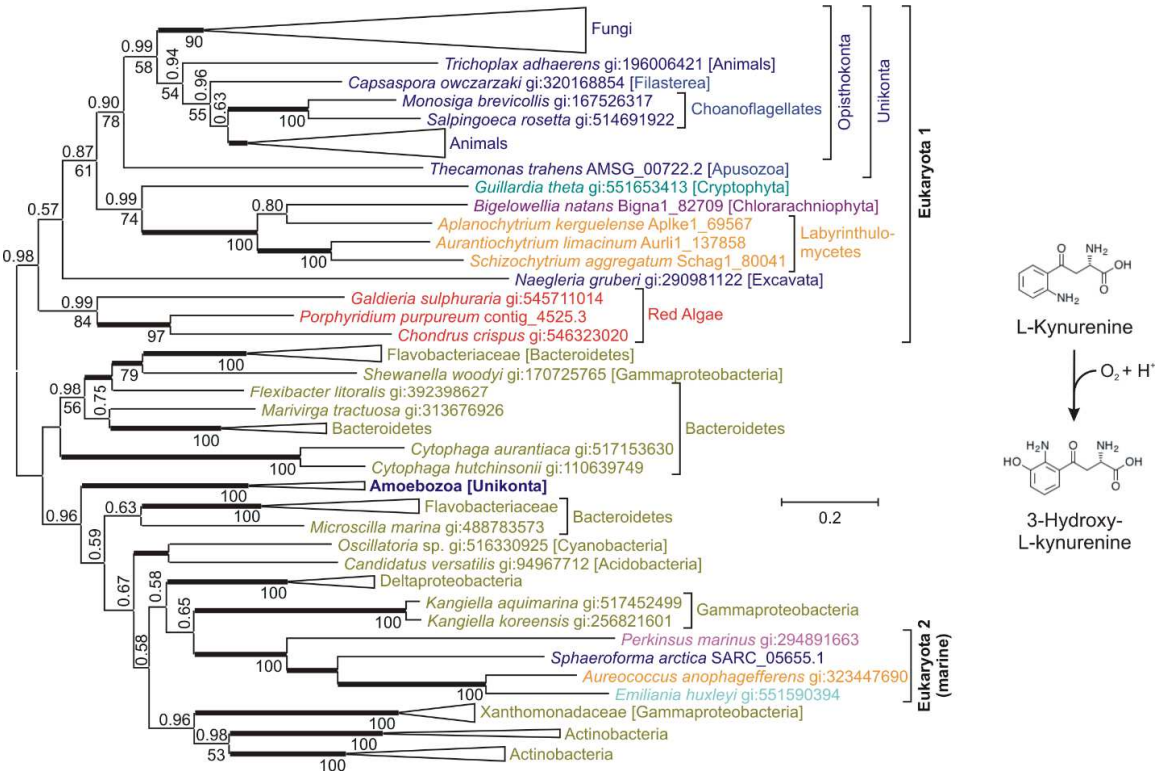


Fig. 35: An evolutionary tree for kynurenine-3-monooxygenase indicates that some unicellular eukaryotes acquired genes via horizontal transfers. The unrooted Bayesian tree shows posterior probabilities above the branches and PhyML bootstrap values below the branches. Thickened horizontal lines represent 1.0 Bayesian posterior probability. Larger clades have been collapsed for presentation and size is indicative of the number of taxa within the clade. Scale bar represents 0.2 substitutions per site.

Somewhat surprising is a branch with sequences from four eukaryotes from different lineages embedded in the bacterial branch ('Eukaryota 2 (marine)' in Fig. 35). *A.*

anophagefferens and *P. marinus* both belong to the SAR (Stramenopiles, Alveolata, Rhizaria) supergroup, which may have a common ancestor with the haptophyte *E. huxleyi* (Fig. 35) (Burki, Okamoto, et al. 2012; Read, et al. 2013). However, sequences from other Stramenopiles (Labyrinthulomycetes) and from *B. natans* (Rhiziaria) are not monophyletic with the sequences from *A. anophagefferens*, *P. marinus*, and *E. huxleyi* (Table 8), indicating that the latter three probably acquired their kynurenine-3-monooxygenase via horizontal gene transfer. Moreover, the sequence from *Sphaeroforma artica* (Ichthyosporea, Opisthokonta) is not monophyletic with other opisthokont sequences (Fig. 35; Table 8), and was probably acquired via horizontal gene transfer as well. While phylogenetically distant, *A. anophagefferens*, *P. marinus*, *E. huxleyi*, and *S. artica* are marine organisms, all four occurring in the northern Atlantic Ocean. The sequences next to this eukaryotic branch are from *Kangiella koreensis* and *Kangiella aquimarina*, marine bacteria from the family of Alcanivoracaceae, with several species also occurring in the northern Atlantic Ocean. The co-occurrence of homologous proteins in distantly related organisms within the same environments has been observed before in studies of horizontal gene transfer, both, in prokaryotes (Beiko, et al. 2005), and in eukaryotes (Andersson 2011). Enforcing Amobozoa and 'Eukaryota 2 (marine)' to be monophyletic could not be rejected statistically, indicating that these two possible horizontal gene transfer events happened independently.

Kynurenine Monooxygenase Monophyletic Constraints	Bayesian				Maximum Likelihood			
	H	Δ*H*	AICM	ΔAICM	log *L*	Δlog *L*	AU	KH
Eukaryotes	-58159	41.0	116502	-94.6	-57736	45.3	0.015	0.02
Unikonta without *S. artica*	-58183	64.7	116546	-139.3	-57755	64.2	0.032	0.04
Eukaryota 1 & Eukaryota 2 (marine)	-58138	*20.1*	116470	*-62.7*	-57723	32.4	0.003	0.013
Opisthokonts	-58230	112.4	116659	-252.1	-57819	128.0	2.0E-06	0
SAR & Haptophyta	-58250	131.8	116712	-304.5	-57830	139.1	3.0E-79	0
Stramenopiles	-58438	319.7	117095	-687.5	-58009	318.5	2.0E-38	0
Amoebozoa & Eukaryota 2 (marine)	-58138	*20.5*	116443	*-35.5*	-57704	13.3	*0.096*	*0.107*

Table 8: Bayesian and Maximum Likelihood analyses of constraint trees for kynurenine 3-monooxygenase. From left to right, clades for which protein sequences were forced to be

monophyletic, harmonic mean (*H*) of log-likelihood values from stationary phase of Bayesian analysis as calculated by Tracer (http://beast.bio.ed.ac.uk/Tracer), difference between harmonic means of unconstrained and constrained tree (Δ*H*), Akaike's information criterion through Markov chain Monte Carlo (AICM) as calculated by Tracer, difference in AICM between unconstrained and constrained tree (ΔAICM), log-likelihood (log *L*) of best maximum likelihood tree, difference between log-likelihoods (Δlog *L*) of unconstrained and constrained tree, *p*-values from approximately unbiased (AU) test and from Kishino-Hasegawa (KH) test as calculated by CONSEL (Shimodaira and Hasegawa 2001; Shimodaira 2002). Δ*H*, ΔAICM, and *p*-values indicating no significant difference in italics and reduced font size.

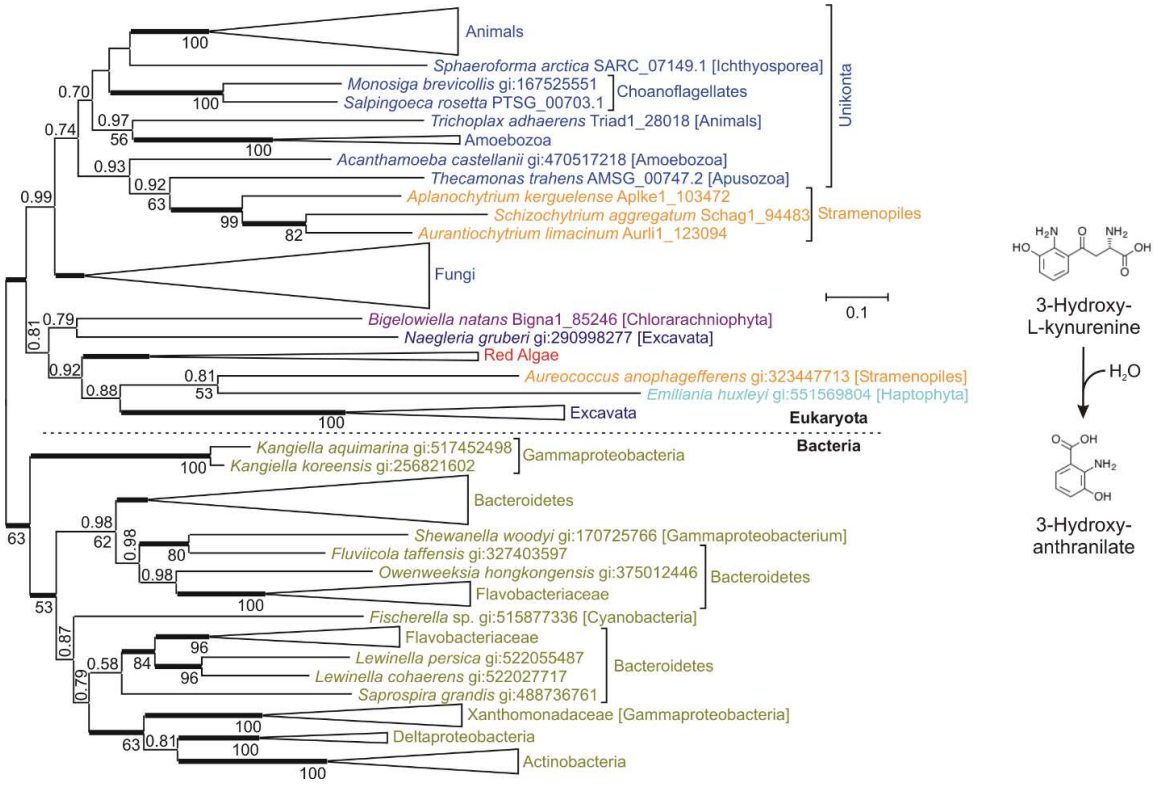


Fig. 36: An evolutionary tree of kynureninases shows the expected separation of eukaryotic and bacterial sequences. The unrooted Bayesian tree shows posterior probabilities above the branches and PhyML bootstrap values below the branches. Thickened horizontal lines represent 1.0 Bayesian posterior probability. Larger clades have been collapsed for presentation and size is

indicative of the number of taxa within the clade. Scale bar represents 0.1 substitutions per site. Tree produced by G. Schönknecht.

Kynureninase (EC 3.7.1.3) catalyzes the fourth step of the kynurenine pathway, the hydrolysis of 3-hydroxy-L-kynurenine into 3-hydroxyanthranilate and alanine, and **3-hydroxyanthranilate 3,4-dioxygenase** (EC:1.13.11.6) catalyzes the fifth step, the decyclization of 3-hydroxyanthranilate into 2-amino-3-carboxymuconate semialdehyde (Fig. 2). For both enzymes evolutionary trees have two major branches, one with eukaryotic sequences, and the other with bacterial sequences (Figs. 36 & 37). Earlier evolutionary analyses of kynureninases and 3-hydroxyanthranilate 3,4-dioxygenases were limited to Unikonta, but had a wider coverage of bacterial sequences (Lima, et al. 2009), and resulted in comparable tree topologies. The kynureninase phylogeny shows several eukaryotic lineages that are disbanded, i.e. Unikonta, stramenopiles, Excavata, and SAR. Enforcing any of these major groups to be monophyletic (irrespective of one another) could not be rejected (Table 9). Similar to the kynureninase phylogeny, the evolutionary tree for 3-hydroxyanthranilate 3,4-dioxygenase also contains major eukaryotic groups that are disbanded (i.e., red algae and SAR). Enforcing these groups to be monophyletic could not be rejected. However, enforcing monophyly among Unikonta or Opisthokonta had extremely low probability, supporting a paraphyletic relationship between fungi and animals in contrast to established organismal evolution (Fig. 27).

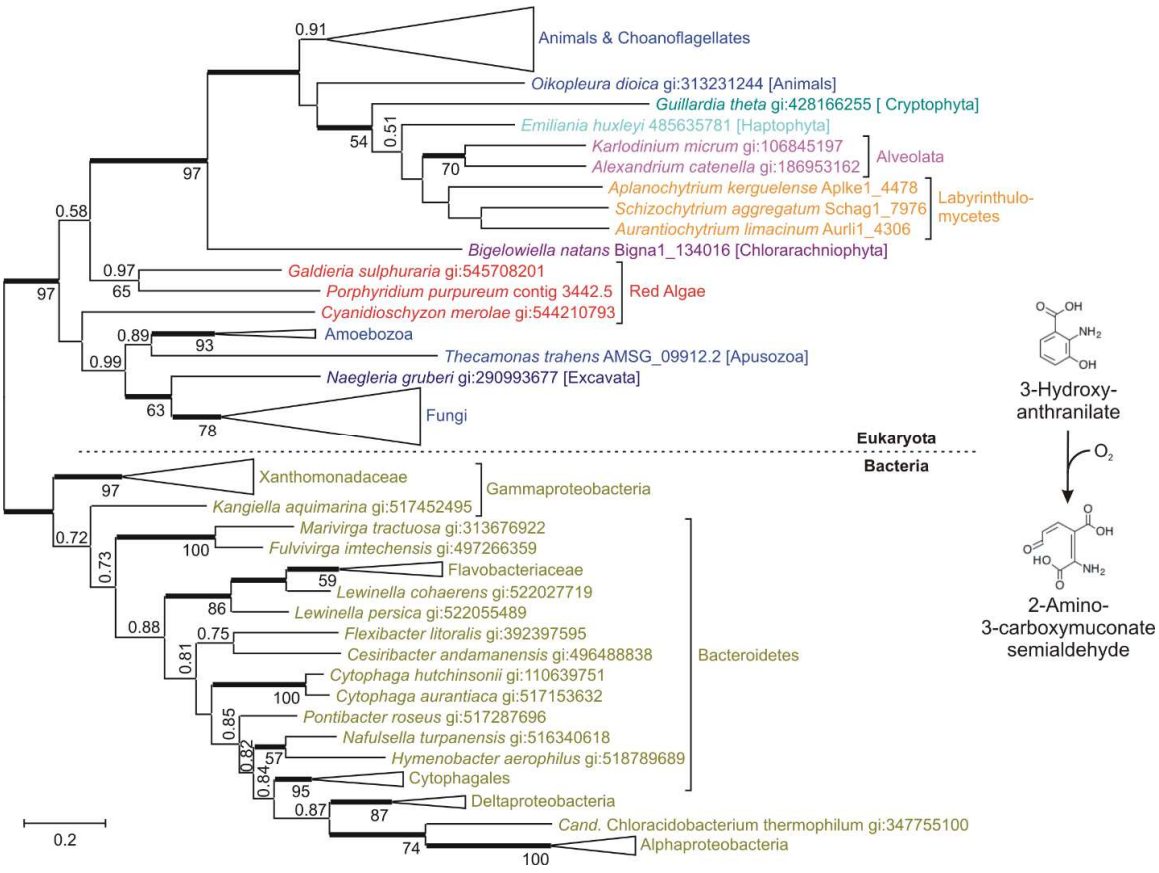


Fig. 37: An evolutionary tree of 3-hydroxyanthranilate-3,4-dioxygenases shows the expected separation of eukaryotic and bacterial sequences. The unrooted Bayesian tree shows posterior probabilities above the branches and PhyML bootstrap values below the branches. Thickened horizontal lines represent 1.0 Bayesian posterior probability. Larger clades have been collapsed for presentation and size is indicative of the number of taxa within the clade. Scale bar represents 0.2 substitutions per site. The sequence from *Aureococcus anophagefferens* was omitted from an evolutionary analysis because it is incomplete.

Kynureninase Monophyletic Constraints	Bayesian				Maximum Likelihood			
	H	Δ*H*	AICM	ΔAICM	log *L*	Δlog *L*	AU	KH
Unikonta	-73137	*11.5*	146577	*-38.6*	-72570	9.6	*0.155*	*0.160*
Stramenopiles	-73152	*27.2*	146574	*-35.3*	-72588	27.7	*0.103*	*0.102*
Excavata	-73131	*5.4*	146546	*-7.4*	-72564	3.0	*0.446*	*0.434*
SAR	-73152	*27.2*	146574	*-35.3*	-72596	34.9	*0.070*	*0.072*
3-Hydroxyanthranilate-3,4-Dioxygenase Monophyletic Constraints	**Bayesian**				**Maximum Likelihood**			
	H	Δ*H*	AICM	ΔAICM	log *L*	Δlog *L*	AU	KH
Unikonta	-25839	111.2	51894	-189.5	-25578	104.6	4.00E-08	0.0003
Opisthokonta	-25843	115.1	51926	-221.4	-25584	110.8	3.00E-44	0
Red Algae	-25737	*8.8*	51713	*-8.2*	-25475	1.4	*0.161*	*0.226*
SAR	-25755	*26.6*	51742	*-36.9*	-25497	23.7	0.046	*0.051*

Table 9: Bayesian and Maximum Likelihood analyses of constraint trees for kynureninase and 3-hydroxyanthranilate-3,4-dioxygenase. From left to right, clades for which protein sequences were forced to be monophyletic, harmonic mean (*H*) of log-likelihood values from stationary phase of Bayesian analysis as calculated by Tracer (http://beast.bio.ed.ac.uk/Tracer), difference between harmonic means of unconstrained and constrained tree (Δ*H*), Akaike's information criterion through Markov chain Monte Carlo (AICM) as calculated by Tracer, difference in AICM between unconstrained and constrained tree (ΔAICM), log-likelihood (log *L*) of best maximum likelihood tree, difference between log-likelihoods (Δlog *L*) of unconstrained and constrained tree, *p*-values from approximately unbiased (AU) test and from Kishino-Hasegawa (KH) test as calculated by CONSEL (Shimodaira and Hasegawa 2001; Shimodaira 2002). Δ*H*, ΔAICM, and *p*-values indicating no significant difference in italics and reduced font size.

THE HIGHLY CONSERVED PATHWAY FOR TETRAPYRROLE BIOSYNTHESIS

Tetrapyrroles, including porphyrins, chlorophylls, and heme, are synthesized though a pathway possessing high conservation among all eukaryotic lineages throughout evolution (Kořený and Oborník 2011). Tetrapyrrole biosynthesis begins with the precursor δ-aminolevulinic acid (ALA) (Figs. 38 & 39); however, two different routes can be used for δ-

aminolevulinic synthesis. Heterotrophic eukaryotes, such as animals and fungi, employ the so-called C4, or Shemin pathway, where δ-aminolevulinic acid is generated through the condensation of glycine and succinyl-CoA. *G. sulphuraria* synthesizes δ-aminolevulinic acid similar to almost all photosynthetic eukaryotes, starting with the amino acid glutamate (Fig. 39).

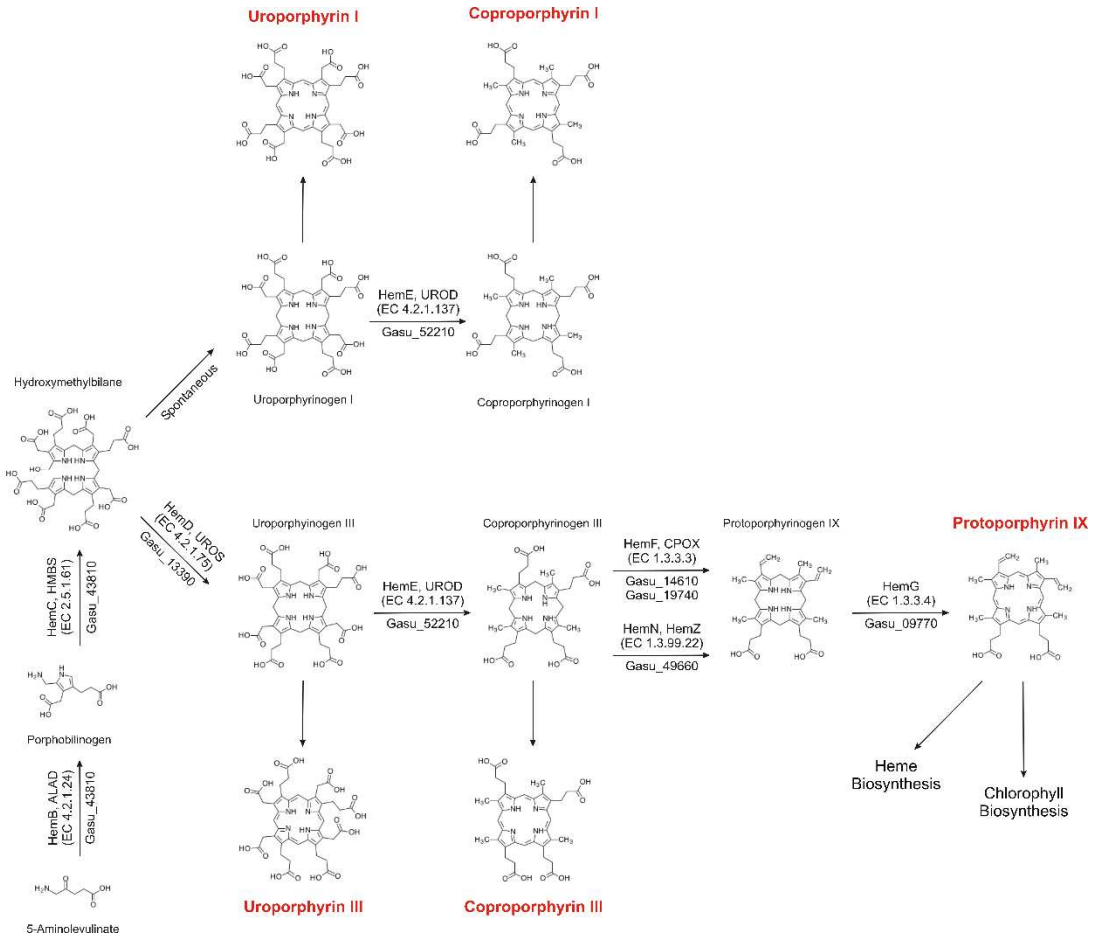


Fig. 38: Porphyrin metabolism modified from KEGG (Kanehisa, et al. 2014). For known enzyme catalyzed reactions, abbreviated enzyme names, EC numbers, and ID's for the *G. sulphuraria* genome (Gasu_#) are given. Porphyrin names are printed in larger font (and red).

The results obtained by Troxler and Bogorad (1966) and Brown, et al. (1982) are somewhat expected when considering the tetrapyrrole biosynthetic pathway. Supplementing cultures with exogenous aminolevulinic acid, the initial precursor metabolite of the pathway

whose synthesis represents the rate-limiting step of overall tetrapyrrole biosynthesis (Papenbrock, et al. 2000; Vasileuskaya, et al. 2005), resulted in porphyrin excretion in *C. caldarium*. The formation of uroporphyrin I and III, and coproporphyrin I and III occurs by a spontaneous reaction (Fig. 38). Therefore, one might expect that 'flooding' the biosynthetic pathway with an excess of the precursor metabolite, aminolevulinic acid, would invoke an increased total amount of tetrapyrroles, especially porphryins which are partially synthesized through spontaneous reactions. Furthermore, incubation with both aminolevulinic acid and *N*-methylprotoporphyrin IX, a strong **ferrochelatase** (Gasu_58560, EC 4.99.1.1; Fig. 39) inhibitor, also resulted in the excretion of porphyrins in *C. caldarium*. Again, this would be expected, as the biosynthetic pathway is 'flooded' with the precursor metabolite, aminolevulinic acid, and one of the two routes, phycobilin synthesis (via heme synthesis) is blocked, resulting in an increase porphyrin biosynthesis.

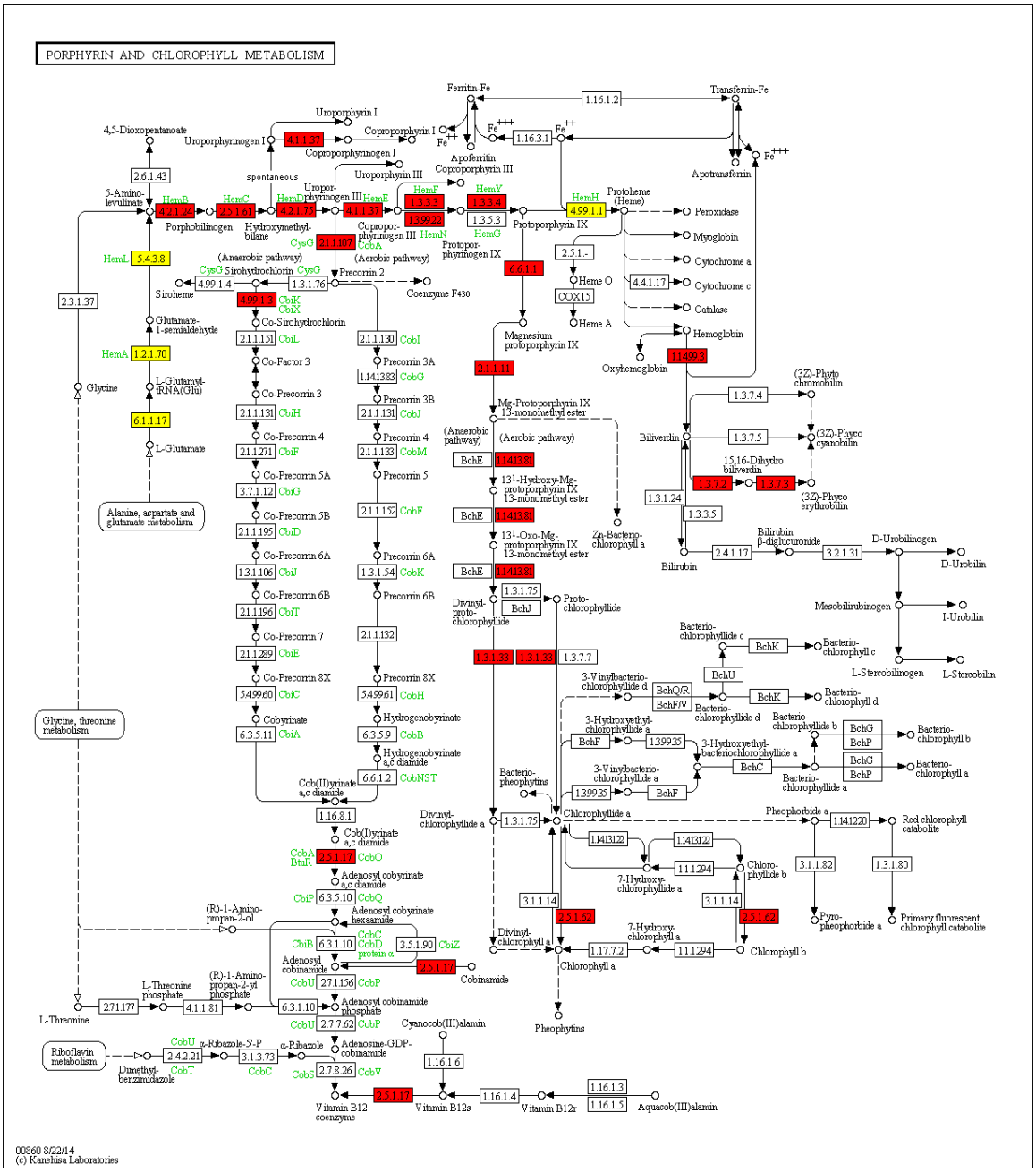


Fig. 39: Porphyrin and chlorophyll metabolism (KEGG pathway 00860 (Kanehisa, et al. 2014)). Enzymes marked in red and yellow have been annotated in the *G. sulphuraria* genome. *G. sulphuraria* synthesizes δ-aminolevulinic acid via the C5 pathway using three enzymes (marked in yellow): **glutamyl-tRNA synthetase** (Gasu_18280 & Gasu_27890, EC 6.1.1.17), **glutamyl-tRNA reductase** (Gasu_39730, EC 1.2.1.70), and **glutamate-1-semialdehyde 2,1-aminomutase** (Gasu_02530, EC 5.4.3.8).

SPECTRAL SHIFTING BY PORPHYRINS EXCRETED FROM *G. SULPHURARIA*

When *G. sulphuraria* cells were transferred to lower light intensities or grown to produce dense cultures causing light-limitation, a pink color and a strong fluorescence of the culture medium were observed. Growth studies showed a correlation between cell density and the excretion of porphyrins (Fig. 40), suggesting conditions exhibiting light limitation may be responsible for eliciting the response. No correlation was observed when comparing porphyrin excretion to supplemental glucose, in contrast to observations and interpretations with *G. partita* (Stadnichuk, et al. 1998).

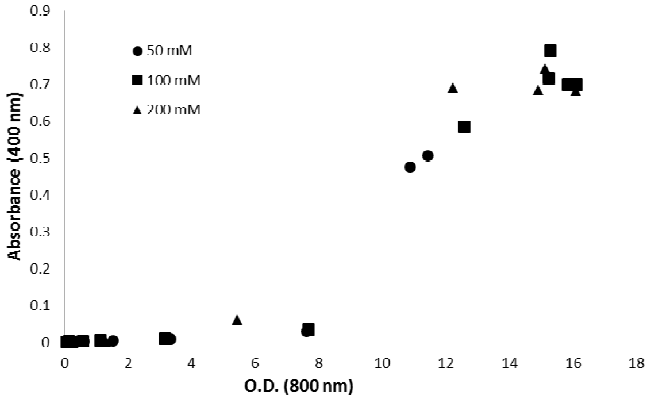


Fig. 40: Concentration of excreted pigments appears to be density dependent in glucose supplemented cultures (glucose concentrations indicted in legend). Pigment concentration increases prior to cultures reaching stationary phase ruling out cell lysis as a major contributing factor. Absorption at 400 nm was used to determine the presence of the porphyrin mixture as 400 nm was the maximum absorption peak in the absorption spectrum. Abbreviation: O.D., optical density.

Separation on a horizontal, native (1.2%) agarose gel resulted in three distinct bands with differing fluorescence intensities (Fig. 41), indicating the presence of three different porphyrins in the culture medium.

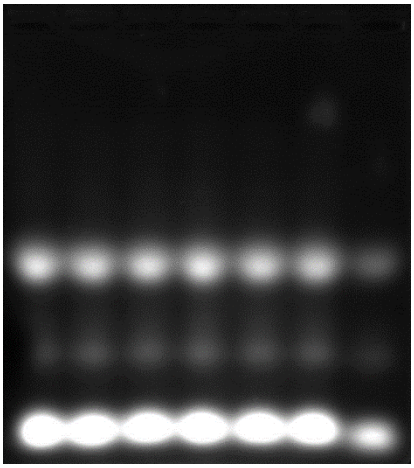


Fig. 41: Horizontal, native agarose (1.2%) gel showing the separation of the three different porphyrins excreted by *G. sulphuraria* when cultivated in liquid medium. All lanes (with exception of the far right lane) were loaded with supernatant from the same liquid culture medium. The lane furthest to the right was loaded with supernatant from a different liquid culture medium. Samples were mixed with glycerol at a 9:1 ratio (sample:glycerol).

THE BASICS: ABSORPTION, EMISSION & EXCITATION

To determine the fluorescence quantum efficiency of the porphyrin mixture excreted by *G. sulphuraria*, three different types of spectra were recorded: absorption, emission, and excitation. A simple absorption spectrum was recorded to determine which wavelengths of light were absorbed by the excreted pigments (Fig. 42, solid line). The wavelength at which the maximum absorption peak resides (400 nm) was then used as the excitation wavelength to measure a fluorescence (or emission) spectrum. Emission spectra requires the selection of a fixed excitation wavelength, i.e. the wavelength of light at which the sample is excited. The resulting emission spectrum is the spectrum of light which is emitted (or fluoresced) by the sample (Fig. 42, dotted line). An excitation spectrum shows the wavelengths of light which can excite the sample to elicit fluorescence. Excitation and absorption spectra are generally similar. The maximum emission peak (591 nm) from the emission spectrum was used as the emission wavelength to measure an excitation spectrum (Fig. 42, dashed line). The emission wavelength is used to 'see' the spectrum of light which excites the sample. The excitation spectrum for the porphyrin mixture showed a broad peak in the near-UV region overlapping with the maximum

absorbance peak. Thus, even wavelengths shorter than 400 nm could be used to generate fluorescence from the excreted porphyrin mixture.

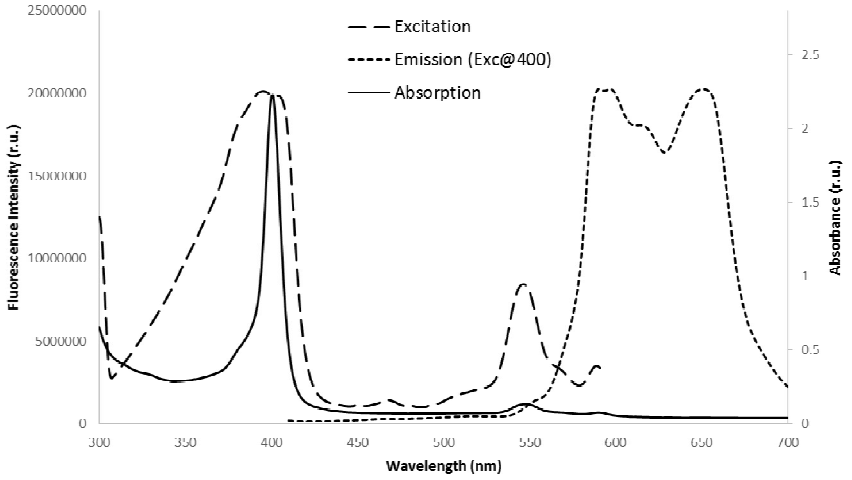


Fig. 42: Three different types of spectra for the porphyrin mixture excreted by *G. sulphuraria*. Culture medium containing porphyrin mixture was at pH 2.

Spectra of the pink culture medium (porphyrin mixture) show a strong excitation peak in the near-UV (320-420 nm), while emission (or fluorescence) spectra revealed two large peaks for yellow and orange light (520 and 660 nm) (Figs. 42 & 43). The excitation peak resides in the near-UV region of the visible spectrum of light, capable of absorbing higher-energy photons in an area of shorter wavelengths when compared to the maximum absorption peaks for both chlorophylls *a* and *b* (Fig. 43). The two large emission peaks in yellow and orange light overlap with the absorption spectra of another photosynthetic pigment, phycocyanin, which *G. sulphuraria* produces in large amounts (Schmidt, et al. 2005). The excitation and emission spectra, taken together with the absorbance spectra of other photosynthetic pigments, namely phycocyanin, gave rise to the hypothesis of spectral shifting. Near-UV light that is not absorbed by photosynthetic pigments is absorbed by the excreted porphyrin mixture, and is then remitted as yellowish-orange light that could be absorbed by phycocyanin (Fig. 43). The excreted porphyrin

mixture could potentially act as light converters, converting light that is unusable photosynthetically into photosynthetically active radiation (PAR) usable by plants and algae.

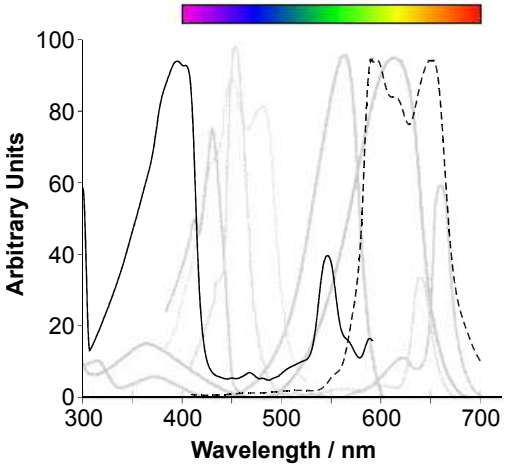


Fig. 43: Excitation spectrum (solid line) and emission spectrum (dashed line) of culture medium (pH 2.0) from light-limited *G. sulphuraria* cultures compared to absorption spectra of different photosynthetic pigments (grey; modified from http://www.biologie.redio.de/pigment.gif). The excitation spectrum was recorded at 600 nm emission; the emission spectrum was recorded with 400 nm excitation. The absorption spectra are for (peaks from left to right) chlorophyll a, chlorophyll b, β-carotene (double peak), phycoerythrin, and phycocyanin. Excitation, emission, and absorption spectra are normalized to have roughly the same amplitude. The color bar on top reflects the different colors visible between 400 nm and 700 nm, the range of wavelengths considered 'photosynthetically active radiation' (PAR).

ABSORPTION AND EMISSION: THE FINER DETAILS

Absorption spectra of the liquid supernatant containing the porphyrin mixture from mixotrophic *G. sulphuraria* cultures supplemented with 50 mM glucose display three distinct absorption peaks at pH 2. The largest peak is very narrow at 400 nm with a small, short-wavelength shoulder. The two smaller peaks are at 550 nm and 590 nm. The absorption spectrum slightly changes at pH 8. However, the only significant change occurs in the region of the two smaller peaks between 550 and 590 nm. The two absorbance peaks observed at pH 2 become four peaks at pH 8 (Fig. 44, bottom middle). At pH 8, the emission spectrum contains two peaks at

608 nm and 674 nm when excited at 400 nm, the wavelength of the maximum absorption peak (Fig. 44). A small difference does exist when comparing emission spectra from pH 2 and pH 8, mainly the wavelength of the emission peaks. At pH 8 the emission curve is shifted to the right towards longer wavelengths. The longer-wavelength emission peak also has a slightly lower fluorescence intensity (Fig. 45). As mentioned above, the region at which the emission curve (at pH 2) resides, is within the absorption spectrum for phycocyanin, a photosynthetic pigment produced in *G. sulphuraria*.

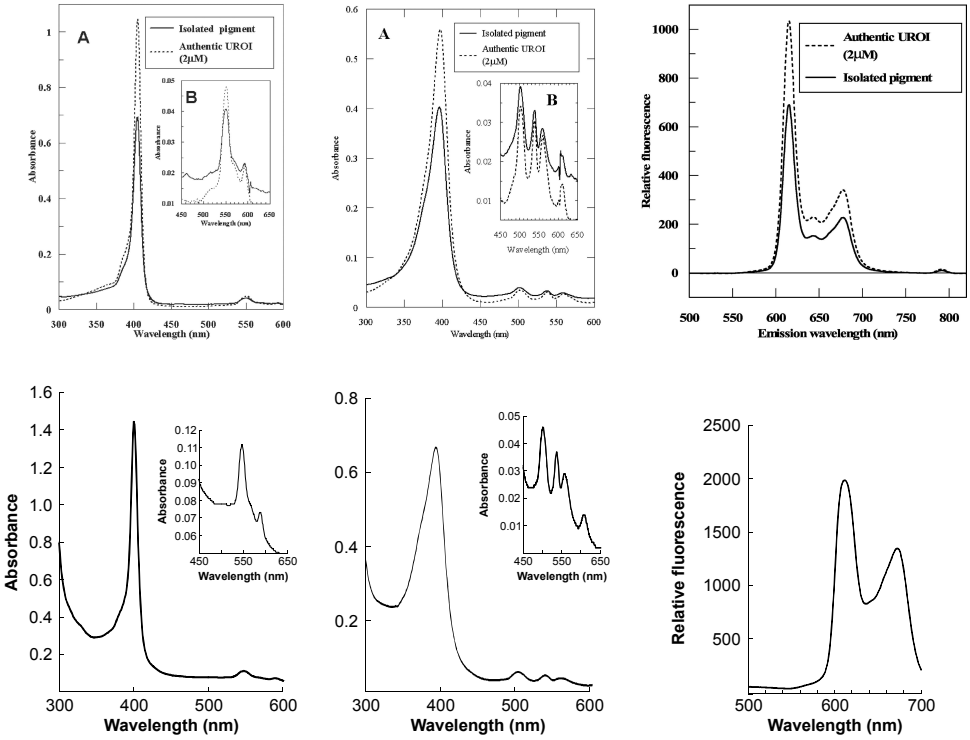


Fig 44: Spectra of uroporphyrin I (top) and spectra of culture medium from light-limited *G. sulphuraria* cultures supplemented with 50 mM glucose (bottom). **Top**: Uroporphyrin I spectra from the authentic substance (dashed line) and the isolated pigment from *E. coli* cultures (solid line) (Akhtar, et al. 2003). **Left**: Absorption spectra at pH 2.0. Insets show minor peaks between 450 and 650 nm at an expanded Y-axis. **Middle**: Absorption spectra at pH 8.0. **Right**: Fluorescence emission spectra at pH 8.0. Figure produced by G. Schönknecht.

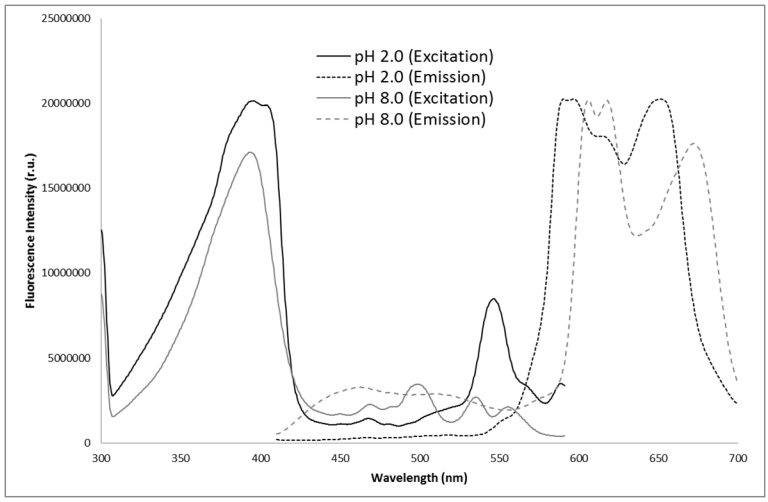


Fig. 45: The effect of pH on excitation and emission spectra. In alkaline conditions the emission spectrum is shifted to longer wavelengths. The maximum peak of the excitation spectrum (400 nm) remains the same at different pH values; however, the 400 nm peak is smaller at pH 8 suggesting a lower energy input is required to elicit a similar magnitude of fluorescence. An excitation wavelength of 400 nm was used for both emission spectra.

The absorption and emission spectra collected for the excreted porphyrin mixture by *G. sulphuraria* show a remarkable similarity to a pigment excreted by an engineered strain of *E. coli*. Akhtar, et al. (2003) identified the pigment excreted by the engineered *E. coli* to be uroporphyrin I, a cyclic tetrapyrrole similar to that of chlorophyll and heme. Comparing the spectra between uroporphyin I excreted by engineered *E. coli* to isolated, authentic uroporphyrin I, showed no difference except for the height of the peaks (Fig. 44, top, (Akhtar, et al. 2003)). At both pH 2 and pH 8, absorption spectra (Fig. 44, top left and middle) and emission spectra (Fig. 44, top right) were practically identical, confirming the excreted pigment from engineered *E. coli* to be uroporphyrin I. The absorption and emission spectra for the excreted porphyrin mixture from *G. sulphuraria* are nearly identical to the spectra presented by Akhtar, et al. (2003), even for both pH 2 and pH 8. Taken together with the results from Troxler and Bogorad (1966) and Stadnichuk, et al. (1998), and comparing absorption and emission spectra between *G. sulphuraria* culture

medium and published uroporphyrin I, the mixture of fluorescent pigments excreted by *G. sulphuraria* appeared to contain, if not mostly composed of, uroporphyin I and/or III.

DETERMINING RELATIVE FLUORESCENCE QUANTUM YIELD

Two characterized fluorescence standards were used in order to determine the **relative fluorescence quantum yield (Φ_F)** of the porphyrin mixture excreted by *G. sulphuraria*. Coumarin 102 and Coumarin 153 both possess an absorption spectrum with the maximum absorption peak near that of the porphyrin mixture at 400 nm (Fig. 46). Coumarin 102 and Coumarin 153 both exhibit a single peak absorption spectrum. Coumarin 102 has a maximum absorption peak at 391 nm and Coumarin 153 at 426 nm.

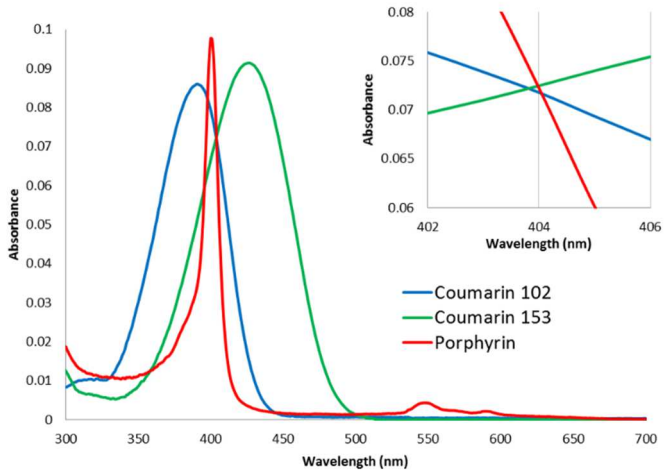


Fig. 46: Absorption spectra for Coumarin 102, Coumarin 153, and porphyrin mixture in culture medium showing maximum absorption peak similarities in the short wavelength region. Inset shows the 404 nm wavelength at which the absorption spectrum of all three samples overlap signifying a cross point.

Preliminary absorption spectra were collected for both Coumarin standards and the porphyrin mixture at differing concentrations. These spectra were collected in order to determine at which wavelengths the cross points resided when using different concentrations. The ideal wavelength for all absorption spectra to cross would be 400 nm, the wavelength for the maximum absorption

peak for the porphyrin mixture. However, generating a cross point for all three absorption spectra at 400 nm was not possible. To minimize inner filter effects, which can lead to underestimated values of the relative Φ_F, the maximum absorption peak for both Coumarin standards and the porphyrin mixture had to have a maximum absorbance value at or below 0.1. Thus, the closest wavelength that could be achieved was 404 nm (Fig. 46). The relative Φ_F was calculated by cross-calibrating the two Coumarin standards first, to ensure published values for relative Φ_F could be reproduced before proceeding with the porphyrin mixture. The initial starting concentrations for Coumarin 102 and Coumarin 153 were 6 μM and 10 μM, respectively. At these concentrations, both Coumarin 102 and Coumarin 153 had an absorbance of less than 0.1 at the maximum absorbance peak, removing any possible inner filter effects. Additionally, these concentrations resulted in absorption spectra that crossed at 404 nm, the (cross point) wavelength used for excitation to collect emission spectra. Each solution was then diluted 20% (four times) with ethanol creating dilutions with 80%, 60%, 40%, and 20% concentrations (for both Coumarin standards) to be used for performing a linear regression (see below). Final dilutions of the concentration series were adjusted slightly with ethanol and/or '100% stock' (6 μM Coumarin 102 and 10 μM Coumarin 153) to ensure all absorption curves overlapped at the cross point wavelength of 404 nm (Fig. 47) as determined above. When exciting at 404 nm for both Coumarin standards at the concentrations given above, the Coumarin standards were assumed to be absorbing the same amount of light (both had equal absorbance values). Thus, the differing magnitudes of emission (or fluorescence) intensity are dependent upon the molecule's Φ_F.

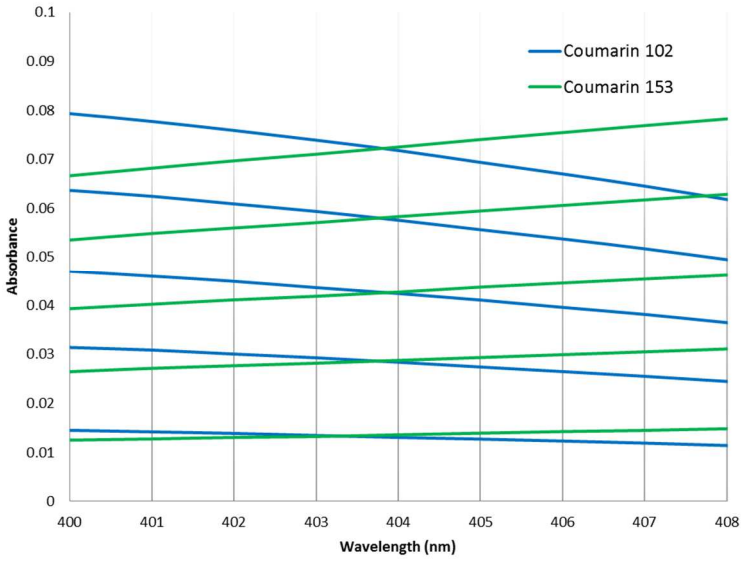


Fig. 47: Absorption spectra of the diluted concentration series of both Coumarin 102 and Coumarin 153. The X-axis resolution has been increased to show the cross points of each concentration dilution. The absorption spectrum for each concentration dilution crosses at or very close to 404 nm, the wavelength that was used for excitation to collect emission spectra.

The cross point wavelength of 404 nm was then used as the excitation wavelength for collecting emission spectra of the Coumarin standards. The first several attempts to measure emission spectra that matched published emission spectra for the Coumarin standards were unsuccessful due to the increased sensitivity of the fluorescence spectrometer. Emission peaks were routinely different than those of published spectra and were the result of saturating the fluorescence detector, which leads to inaccurately quantified fluorescence intensities by the fluorescence spectrometer at various wavelengths. Excitation and emission slit widths were reduced from 5 mm (maximum sensitivity) to 1 mm (minimum sensitivity). Reducing the slit widths resulted in emission spectra that matched published spectra (Rurack and Spieles 2011) and did not exhibit saturation by the fluorescence detector (Fig. 48).

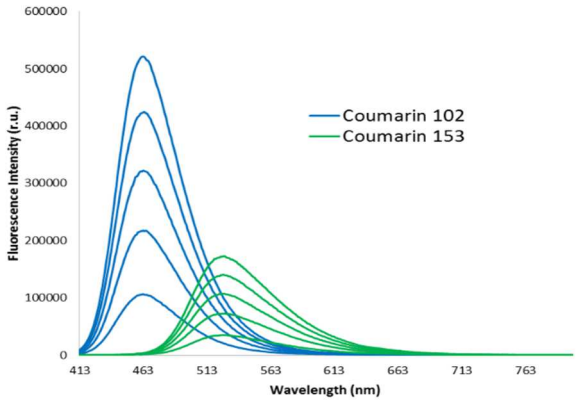


Fig. 48: Raw emission spectra for all five concentration dilutions for both Coumarin 102 and Coumarin 153. The differing magnitudes of fluorescence intensity are a reflection of each Coumarin standard's fluorescence quantum yield as depicted in Fig. 49 below. The cross point wavelength (404 nm) was used as the excitation wavelength for collecting emission spectra. R.U., relative units.

The emission spectra for each concentration dilution of both Coumarin standards were integrated (summation of the area under the emission spectrum), and integrated fluorescence was plotted as a function of absorbance [$F=f(A)$], in order determine the slope via linear regression ($Grad=F/A$) (Fig. 49). As both Coumarin standards were dissolved (and diluted) in ethanol, the refractive indices for all solutions were identical. Thus, Equation 2 (Methodology, pg. 30) was used in order to calculate the relative Φ_F for Coumarin 153, using the published value for the Φ_F of Coumarin 102 (Φ_F =0.764) (Rurack and Spieles 2011) as the known standard. The calculated Φ_F of Coumarin 153 was **0.494**. Coumarin 153 has a published value of 0.544 for the Φ_F (Rurack and Spieles 2011), which gives a difference of 9.2% between published and experimental values (shown here). The calculated experimental value of 0.494 lies within the 10% allowance for the determination of relative Φ_F (Horiba 2014). Given that published values for the relative Φ_F of Coumarin 153 could be reproduced with confidence (within ±10%), determination of the relative Φ_F for the porphyrins excreted by *G. sulphuraria* was attempted next.

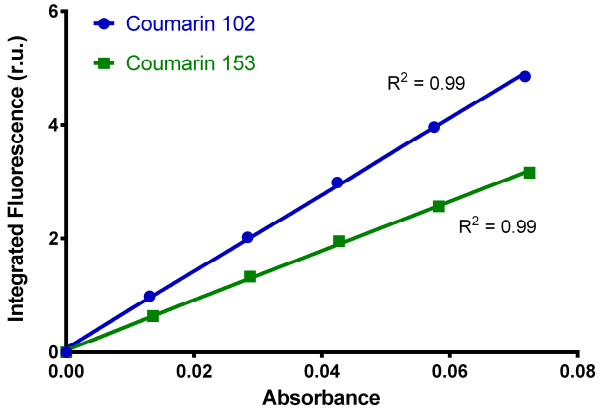


Fig. 49: Linear regression showing integrated fluorescence as a function of absorbance [$F=f(A)$]. The slope values are 67.53 and 43.68 for Coumarin 102 and Coumarin 153 respectively (Table 10, pg. 107), and were used as *Grad* values in Eq. 2 (pg. 30). Integrated fluorescence values were divided by 10^7 for presentation purposes. R.U., relative units.

Similar to the concentration dilutions for the Coumarin standards, the porphyrin mixture was diluted to differing concentrations in order to determine which concentration would yield an absorbance value equal to, or less than, 0.1 at the maximum absorption peak, and additionally, overlap the absorption spectra of the Coumarin standards at the cross point wavelength of 404 nm. The same concentration dilutions used above for cross calibrating the Coumarin standards were used for determining the relative Φ_F of the porphyrin mixture. By all absorption spectra for both Coumarin standards and the porphyrin mixture overlapping at the cross point wavelength (404 nm) for all concentration dilutions, both Coumarins 102 and 153 were used individually to calculate the relative Φ_F of the porphyrin mixture. The initial concentration of the porphyrin mixture used to generate the concentration dilutions was 12% (diluted with water) of the culture medium as it existed, after centrifugation and sterile filtration to remove any cells or cell debris. The 12% porphyrin mixture solution was then used to generate the concentration dilutions by diluting the solution 20% (four times) with water, to create dilutions with 80%, 60%, 40%, and 20% concentrations. The final concentration dilutions of the porphyrin mixture were slightly adjusted with water and/or '100% stock' (12% porphyrin mixture solution) to ensure absorption spectra for each concentration dilution crossed the absorption spectra of the concentration dilutions for the Coumarin standards at 404 nm (Fig. 50).

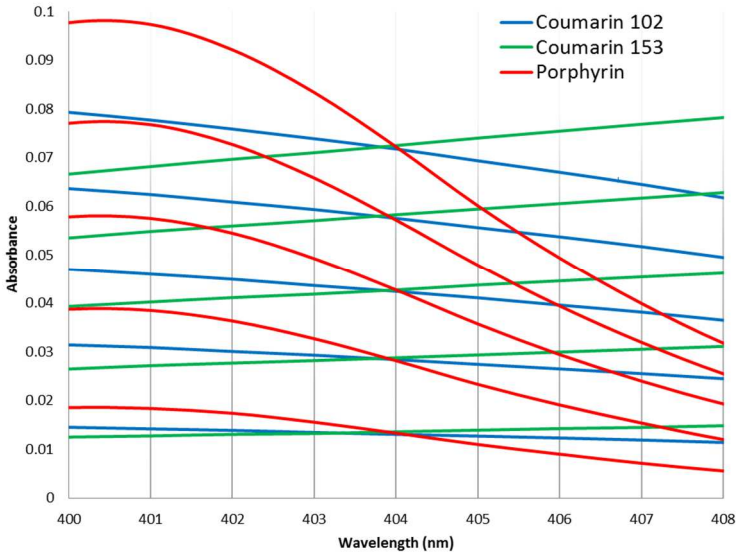


Fig. 50: Absorption spectra of the concentration dilutions for both Coumarin standards and the porphyrin mixture. X-axis resolution has been increased to show cross points and region surrounding cross point wavelength of 404 nm, the wavelength used for excitation when collecting emission spectra.

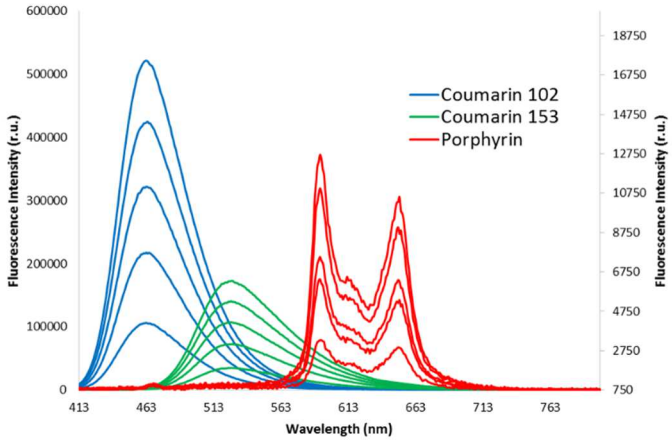


Fig. 51: Raw emission spectra for both Coumarin standards and the porphyrin mixture. Emission spectra for the porphyrin mixture were placed on a secondary Y-axis for presentation purposes due to substantially lower fluorescence intensity for the concentration dilutions. The cross point wavelength (404 nm) was used as the excitation wavelength for collecting emission spectra R.U., relative units.

The emission spectra for each concentration dilution of the porphyrin mixture and the Coumarin standards (Fig. 51) were integrated (summation of the area under the emission spectrum), and integrated fluorescence was plotted as a function of absorbance [$F=f(A)$], in order determine the slope via linear regression ($Grad=F/A$) (Fig. 52). Because the porphyrin mixture was dissolved (and diluted) in water, whereas the Coumarins were dissolved (and diluted) in ethanol, the differing refractive indices of the solutions (water vs. ethanol) had to be taken into consideration. Thus, Equation 3 (Methodology, pg. 32) was used for calculating the relative Φ_F of the porphyrin mixture.

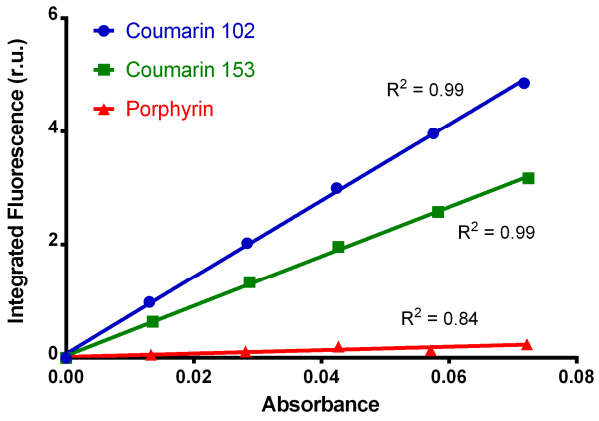


Fig. 52: Linear regression showing integrated fluorescence as a function of absorbance [$F=f(A)$] for both Coumarin standards and the porphyrin mixture. The slope values of the linear regressions are listed in Table 10. Integrated fluorescence values were divided by 10^7 for presentation purposes. R.U., relative units.

Using Coumarin 102 and Coumarin 153 individually as the known fluorescence standards, the porphyrin mixture was calculated to have a **relative Φ_F** of **0.032** and **0.035**, respectively (approximately 9% difference). The calculated relative Φ_F of the porphyrin mixture was lower than to be expected for counteracting light limiting conditions. The large difference in the fluorescence intensities between the Coumarin standards and the porphyrin mixture was also reflected by the large difference in relative Φ_F values. The excretion of porphyrins by *G. sulphuraria* might not be directly associated with environmental lighting conditions, but may serve another biological function that is not easily observed during laboratory cultivation.

	Coumarin 102	Coumarin 153	Porphyrin Mixture
Best-fit values			
Slope	67.53	43.68	2.979
Y-intercept; X=0.0	0.0716	0.0388	0.0167
X-intercept; Y=0.0	-0.0011	-0.0009	-0.0056
95% Confidence Intervals			
Slope	64.73 to 70.32	41.82 to 45.54	1.163 to 4.794
Y-intercept; X=0.0	-0.0494 to 0.1927	-0.0426 to 0.1202	-0.0621 to 0.0954
X-intercept; Y=0.0	-0.003 to 0.0008	-0.0029 to 0.0009	-0.0742 to 0.0143
Goodness of Fit			
R square	0.9991	0.9991	0.8383
Equation	Y = 67.53*X + 0.07162	Y = 43.68*X + 0.03880	Y = 2.979*X + 0.01667

Table 10: Linear regression analyses of all samples (Coumarin standards and porphyrin mixture) where absorption is plotted as a function of integrated fluorescence [$F=f(A)$] (Fig. 52). The slope values given were used for the *Grad* values in Eq. 2 (pg. 30) and Eq. 3 (pg. 32) to calculate the relative fluorescence quantum yield of Coumarin 153 and the porphyrin mixture excreted by *G. sulphuraria* respectively.

CHAPTER V

CONCLUSION

THE METABOLIC DIVERSITY OF *G. SULPHURARIA*

The diversity of metabolic pathways and associated enzymes make *G. sulphuraria* unique not only when compared to other red algae, but also to other algal lineages including green algae. Without question, several evolutionary events have played an important role in shaping metabolic pathways in *G. sulphuraria*, from the primary endosymbiosis event to horizontal gene transfer. Ancestral inheritance was even responsible for the establishment of certain metabolic pathways, such as the methylmalonyl-CoA pathway and tryptophan degradation, which is coupled with *de novo* NAD^+ biosynthesis (via kynurenine pathway). Thus, the metabolic diversity present in *G. sulphuraria* is a direct result of *G. sulphuraria*'s evolutionary history that was also shaped by numerous gene transfers.

Several enzymes implicated in glycogen metabolism in *G. sulphuraria* have been shown to have eukaryotic origins that do not directly result from the primary endosymbiosis event (Patron and Keeling 2005; Deschamps, et al. 2008). While the primary endosymbiosis event unquestionably impacted polyglucan metabolism in several photosynthetic lineages, namely Viridiplantae, polyglucan metabolism in *G. sulphuraria* appears to have mostly retained and still utilizes the eukaryotic, ancestor-established pathways. The architecture of glycogen synthesized by *G. sulphuraria* contains more α-1,6 branching linkages than other red algae, in addition to

Viridiplantae and cyanobacteria (Hirabaru, et al. 2010). Even though synthesizing glycogen may seem peculiar compared to other photosynthetic eukaryotes, utilizing highly-branched glycogen in favor of minimally-branched starch does have its advantages. Glycogen molecules are completely hydrosoluble and like starch (which is not entirely hydrosoluble), have little to no impact on osmoregulation (Ball, et al. 2011). Considering *G. sulphuraria*'s habitat and the mechanisms by which it deals with increased levels of salt (McCoy, et al. 2009; Schönknecht, et al. 2013), osmoregulation is an important facet of *G. sulphuraria*'s physiology. Additionally, due to the highly soluble nature and architecture, carbon reserves stored as glycogen are more readily available compared to starch. When additional energy is required, glucose monomers can be readily mobilized as opposed to starch, which requires a rather lengthy process for mobilization. *G. sulphuraria* synthesizes glycogen via a glycogen synthase (Gasu_14030) that appears to be a bifunctional, fusion enzyme possessing both, a glycogen synthase domain (*GlgA*) and a glycogen branching domain (*GlgB*). To date, no other organism, whether eukaryotic or prokaryotic, appears to possess a glycogen biosynthesis enzyme of similar nature. The underlying evolution of how the glycogen synthase fusion enzyme came to be is unclear; nonetheless, the glycogen synthase in *G. sulphuraria* is unique.

Of all the currently sequenced red algae to date, *G. sulphuraria* is the only alga that can synthesize sucrose. Furthermore, based on phylogenetic analysis, the enzymes responsible for synthesizing sucrose in *G. sulphuraria* do not provide any clue regarding their evolutionary history. The enzymes do not form a sister relationship with either Viridiplantae or bacteria, despite possessing conserved domains similar to both lineages. In land plants, sucrose functions as a transport sugar, as well as being involved in sugar signaling (Ruan 2014). Given that *G. sulphuraria* is unicellular, a transportation role for sucrose in *G. sulphuraria* seems unlikely. In light of the extensive diversity for carbon metabolism exhibited by *G. sulphuraria*, sucrose

metabolizing enzymes provide yet another example of the extensive capacity for carbon metabolism in *G. sulphuraria.*

Both glycerol metabolism and terpenoid backbone biosynthesis appear to result from multiple evolutionary events. Two routes metabolizing glycerol can likely be attributed to ancestral conservation, while the third route, initially catalyzed by glycerol dehydrogenase, was acquired from bacteria via horizontal gene transfer (Schönknecht, et al. 2013). The mevalonate pathway of terpenoid backbone biosynthesis in *G. sulphuraria* appears to be of eukaryotic origin, similar to other eukaryotes possessing the mevalonate pathway. Conversely, the methylerythritol phosphate pathway was probably acquired from the primary endosymbiosis event, which is responsible for the establishment of plastids in eukaryotes. Thus, three different types of evolutionary processes, namely ancestry, endosymbiosis, and horizontal gene transfer, have shaped the diversity of metabolic pathways in *G. sulphuraria*, creating a metabolic patchwork.

THE UNCERTAIN EXTENT OF CARBON METABOLISM IN *G. SULPHURARIA*

While *G. sulphuraria* displays an unparalleled ability to metabolize carbon compounds, the exact extent of carbon metabolism in *G. sulphuraria* is unknown, and might even be underestimated. For example, a D-arabitol dehydrogenase has been described experimentally for *G. sulphuraria* (Stein, et al. 1997); however, the D-arabitol dehydrogenase is not directly supported, nor easily detected in the genome. The D-arabitol dehydrogenase activity can most likely be attributed to another enzyme. Hallborn, et al. (1995) isolated a short chain dehydrogenase from the fungus *Pichia stipitis* that was shown to exhibit D-arabitol dehydrogenase activity, and even showed a preference for D-arabitol. The short chain dehydrogenase/reductase has a 51% amino acid identity to a short chain dehydrogenase/reductase found in *G. sulphuraria* (Gasu_09280). The short chain dehydrogenase/reductase (Gasu_09280)

might be the enzyme exhibiting D-arabitol dehydrogenase activity. The *G. sulphuraria* genome encodes several short chain dehydrogenases/reductases, some of which have been annotated as specific enzymes based on BLAST and conserved domain searches (see Methodology). Given the ambiguous nature of the substrate specificity for just one short chain dehydrogenase/reductase (Stein, et al. 1997), it is difficult to assess the exact capabilities of the short chain dehydrogenases/reductases, and their impact on carbon metabolism.

Like most red algae, *G. sulphuraria* does possess a cell wall. Generally, red algal cell walls are composed of agar, carragennan, lignin (or lignin-like compounds), and (1-3)(1-4)-β-D-glucans (Popper, et al. 2011). Exact biosynthetic pathways for cell wall biosynthesis in red algae still remain unknown. However, several enzymes are known to participate in cell wall biosynthesis, several of which are encoded in the *G. sulphuraria* genome including **fucosyltransferase** (Gasu_23840), **galactosyltranferase** (6 genes), **mannosyltransferase** (15 genes), and **glucosyltransferase** (8 genes). It is unlikely that all of the aforementioned genes encode enzymes directly associated with cell wall biosynthesis, yet it is very likely that many more unannotated enzymes are implicated with synthesizing cell walls. Approximately 38% of the predicted protein sequences encoded by the *G. sulphuraria* genome are still annotated as 'hypothetical protein'. It would not be surprising if several of the enzymes annotated as 'hypothetical protein' participated in cell wall biosynthesis, and even more broadly in carbon metabolism. Thus, without further experimental work it, is difficult to assess the full spectrum of carbon metabolizing enzymes and pathways present in *G. sulphuraria*.

LIFE WITHOUT A GLYOXYLATE CYCLE

As mentioned above, *G. sulphuraria* surprisingly does not possess a glyoxylate cycle. The two enzymes unique to the glyoxylate cycle, isocitrate lyase (EC 4.1.3.1) and malate

synthase (EC 2.3.3.9), have yet to be annotated in any red algal genome to date. The glyoxylate cycle serves a crucial function in higher plants by providing the metabolic intermediates for carbohydrate biosynthesis. In seeds, carbon reserves are typically stored as lipids (fatty acids), polysaccharides, and proteins. Upon germination, fatty acids are broken down and used as precursors via the glyoxylate cycle for anabolic carbohydrate biosynthesis to sustain the growth of the developing plant, until photosynthesis can be established. Given the absence of the glyoxylate cycle, *G. sulphuraria* does not appear to be able to directly convert fatty acids into carbohydrates, and it seems logical that a minimal amount of carbon is allocated to lipid reserves. Lipids account for approximately 10% of storage molecules in *G. sulphuraria*, while carbohydrates (~65%) and proteins (~25%) make up the remainder (Graziani, et al. 2013).

Although the *G. sulphuraria* genome does encode the methylmalonyl-CoA pathway, it does not entirely circumvent the absence of a glyoxylate cycle. While both the methylmalonyl-CoA pathway and glyoxylate cycle (to a somewhat lesser extent) are involved in metabolizing propionyl-CoA, an end product of fatty acid β-oxidation, the methylmalonyl-CoA pathway does not serve the same fundamental role, the direct conversion of fatty acids to carbohydrates, as the glyoxylate cycle. However, the methylmalonyl-CoA pathway in *G. sulphuraria* might assume a supporting role specifically for propionyl-CoA metabolism. In the bacterium *Mycobacterium tuberculosis*, the glyoxylate cycle, together with the methylcitrate pathway, detoxify propioniate (via propionyl-CoA), which becomes lethal at elevated levels (Savvi, et al. 2008). When grown on propionate and supplemented with 3-nitropropionate, a methylcitrate pathway and glyoxylate cycle inhibitor, *M. tuberculosis* was shown to utilize the methylmalonyl-CoA pathway as a bypass for propionate degradation. The use of the methylmalonyl-CoA pathway in *G. sulphuraria* may serve as a comparable approach to that of *M. tuberculosis* for propionate (via propionyl-CoA) degradation. The presence of the methylmalonyl-CoA pathway may bypass the absence of the methylcitrate pathway and the glyoxylate cycle, which would allow *G. sulphuraria* to

potentially utilize propionate as a carbon source without experiencing toxicity. The enzymes underlying the methylmalonyl-CoA pathway in *G. sulphuraria* seem to have been retained through ancestry, suggesting this pathway may have been operating in the earliest of eukaryotes, but was subsequently lost throughout different eukaryotic lineages. Habitats where *G. sulphuraria* is commonly found are characteristically acidic, and may have provided a persistent, selective pressure to retain the methylmalonyl-CoA pathway, in addition to the added benefit of metabolizing end products of fatty acid β-oxidation.

GENE TRANSFERS UNDERSCORE THE EVOLUTION OF *DE NOVO* NAD^+ BIOSYNTHESIS

The different gene transfers contributing to the evolution of *de novo* NAD^+ biosynthesis are summarized In Figure 53. The gene transfer events displayed are somewhat conservative, as only transfer events with good statistical support, both from Bayesian ($\Delta H > 3$ SD and $-\Delta AICM >$ 6 SD) and maximum likelihood analyses (*p*-values < 5%), were included. Possible transfer events of indoleamine dioxygenase homologs (Fig. 34) were not included, because these enzymes may not actually catalyze tryptophan degradation (see above).

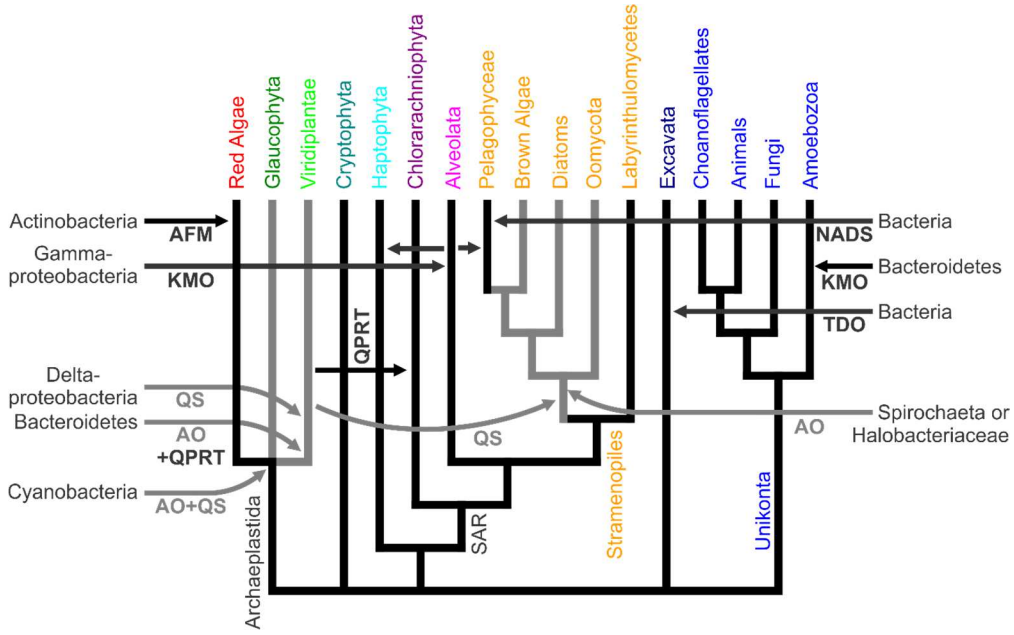


Fig. 53: Gene transfers in the evolution of *de novo* NAD^+ synthesis. Shown is an organismal phylogeny (Keeling and Palmer 2008; Burki, Okamoto, et al. 2012) of the major eukaryotic lineages (SAR, Stramenopiles, Alveolates, and Rhizaria) with emphasis on plastid bearing clades. Horizontal arrows indicate putative horizontal gene transfers from Bacteria or Archaea, with the most likely 'donor clade' indicated. The horizontal gene transfer of a KMO gene from a proteobacterium into *Sphaeroforma arctica* (Ichthyosporea, Opisthokonta) is omitted for clarity. Abbreviated enzymes names are AFM, arylformamidase; AO, aspartate oxidase; KMO, kynurenine 3-monooxygenase; NADS, NAD^+ synthase; QPRT, nicotinate-nucleotide pyrophosphorylase (quinolinate phosphoribosyltransferase); QS, quinolinate synthase; TDO, tryptophan dioxygenase. Black indicates kynurenine (and converged) pathway, grey indicates aspartate pathway.

THE ASPARTATE PATHWAY: EVOLUTIONARY ORIGINS IN EUKARYOTES

To explain the current phylogenetic pattern of aspartate oxidases and quinolinate synthases in eukaryotes, a minimum of five gene transfer events have to be postulated. (1) Both

genes were probably introduced into eukaryotes by endosymbiotic gene transfer from the cyanobacterial endosymbiont into the host genome of an early ancestor of Archaeplastidae. While Glaucophyta kept the genes of cyanobacterial origin, the aspartate pathway was not established, and the kynurenine pathway remained active in red algae. In an early ancestor of Viridiplantae, (2) aspartate oxidase and (3) quinolinate synthase were substituted by horizontal gene transfers from non-photosynthetic bacteria. (4) Gene transfer from an ancient green alga established quinolinate synthase (and possibly aspartate oxidase) in an early ancestor of Stramenopiles. (5) The aspartate oxidase gene observed in Stramenopiles was most likely acquired via horizontal gene transfer from a spirochaete or a halobacterium. This seems to be the most parsimonious explanation of the current evolutionary pattern of aspartate oxidases (Fig. 28) and quinolinate synthases (Fig. 29) in eukaryotes.

The hypothesis that the aspartate pathway was introduced into eukaryotes by endosymbiotic gene transfer during the evolution of plastids is based on the observations that i) the aspartate pathway is only detected in photosynthetic (or, in the case of Oomycota, closely related) eukaryotic clades, ii) in the glaucophyte *C. paradoxa*, aspartate oxidase (Fig. 28) and quinolinate synthase (Fig. 29) seem to be of cyanobacterial origin, and iii) in land plants, both enzymes are targeted to plastids (Katoh, et al. 2006), as many gene products of endosymbiotic origin are. While endosymbiotic acquisition of the aspartate pathway at the base of the Archaeplastidae seems like the most parsimonious explanation of the evolutionary trees presented here, alternative explanations cannot be ruled out. Not all photosynthetic eukaryotes use the aspartate pathway, such as red algae or the pelagophyte *A. anophagefferens*, and not all eukaryotes using the aspartate pathway are photosynthetic, such as Oomycota. Yet, Oomycota form a sister clade to several photosynthetic Stramenopiles, and probably descend from an ancestor with photosynthetic capacity (Cavalier-Smith and Chao 2006; Tyler, et al. 2006). The aspartate oxidase and quinolinate synthase in *C. paradoxa* theoretically might have been acquired

by horizontal, instead of endosymbiotic gene transfer, and with just one glaucophyte genome available, gene transfer from cyanobacteria could be much more recent than indicated in Figure 53. The phylogeny of the plastid ancestor has not been resolved, and due to frequent gene transfer between bacteria the plastid ancestor, may never be fully resolved (Dagan, et al. 2013). Evolutionary trees from single proteins, as in Figures 28 and 29, do not allow for the differentiation between endosymbiotic and horizontal gene transfer from cyanobacteria into *C. paradoxa*. Yet, with both aspartate oxidase (Fig. 28) and quinolinate synthase (Fig. 29) originating from cyanobacteria, and quinolinate synthase being encoded in the cyanelle genome, endosymbiotic acquisition by an ancestor of *C. paradoxa* seems most parsimonious. While many gene products with endosymbiotic ancestry are targeted to plastids (Timmis, et al. 2004), not all plastid-targeted proteins are probably of endosymbiotic origin (Qiu, Price, et al. 2013). Therefore, it is possible that the glaucophyte *C. paradoxa* and Viridiplantae could have acquired the aspartate pathway independently, instead of a single, endosymbiotic gene transfer at the base of the Archaeplastidae, as indicated in Figure 53.

How some stramenopile clades acquired the aspartate pathway is not clear. Stramenopile aspartate oxidases form a sister group to aspartate oxidases from Spirochaeta and Halobacteriaeceae (Fig. 28), whereas quinolinate synthases from Stramenopiles and Viridiplantae are sister groups neighboring Deltaproteobacteria and other Proteobacteria (Fig. 29). It is assumed that aspartate oxidase was acquired by a Stramenopile ancestor via horizontal gene transfer from Sprochaeta or Halobacteriaceae, and that stramenopile quinolinate synthases originate from a gene transfer from Viridiplantae (Fig. 53). However, due to the limited sequences in the evolutionary trees, the direction of potential gene transfers cannot be determined, and theoretically, Viridiplantae might have acquired quinolinate synthase from Stramenopiles. A strong phylogenetic signal connecting Viridiplantae and Stramenopiles has been observed earlier, and gave rise to the hypothesis of a cryptic endosymbiosis between a heterotrophic ancestral

stramenopile and an ancient green alga (Moustafa, et al. 2009). However, both the extent of this phylogenetic signal and how it is to be interpreted are a matter of debate (Woehle, et al. 2011; Burki, Flegontov, et al. 2012; Deschamps and Moreira 2012). Both aspartate oxidase and nicotinate-nucleotide pyrophosphorylase (Fig. 30) in Viridiplantae seem to originate from Bacteroidetes, even though a proteobacterial origin for aspartate oxidase cannot be ruled out (Table 1). Therefore, aspartate oxidase and nicotinate-nucleotide pyrophosphorylase are assumed to have been acquired in a single transfer event, possibly from a bacteroidetes encoding both enzymes in one operon. The quinolinate synthase in Viridiplantae was most likely acquired separately from a deltaproteobacterium.

EVOLUTION OF THE KYNURENINE PATHWAY IN EUKARYOTES

The evolutionary origin of the kynurenine pathway is unclear. The last common ancestor of all eukaryotes probably possessed enzymes for the kynurenine pathway (see above; Fig. 27), whereas archaea and most bacteria use the aspartate pathway. The patchy distribution of genes for the kynurenine pathway among different bacterial lineages gave rise to the hypothesis that some bacteria, especially Xanthomonadales and Flavobacteriales, have acquired these genes via horizontal transfer from eukaryotes (Lima, et al. 2009). Yet, with more and more bacterial genomes being sequenced and annotations improving, the kynurenine pathway has been detected in numerous bacterial lineages. The Kyoto Encyclopedia of Genes and Genomes (Kanehisa, et al. 2014) currently contains 91 bacterial genomes that contain genes encoding enzymes for at least four of the five catalyzed reactions of the kynurenine pathway. These are genomes from Gammaproteobacteria (23), Betaproteobacteria (13), Deltaproteobacteria (5), Alphaproteobacteria (1), Firmicutes (9), Actinobacteria (11), Acidobacteria (1), and Bacteroidetes (28). Evolutionary analyses of the enzymes catalyzing the kynurenine pathway (Figs. 33 through 37) currently do

not support horizontal gene transfer from eukaryotes into Xanthomonadales, Flavobacteriales, or other bacterial lineages. Bacterial lineages are not nested within a clade of eukaryotic lineages showing the expected organismal phylogeny, as it is required to determine the direction of a gene transfer (Stanhope, et al. 2001). This also holds for earlier evolutionary analyses (Lima, et al. 2009), which do not show sequences from Xanthomonadales, Flavobacteriales, or other bacterial lineages embedded within a eukaryotic clade. It remains to be clarified whether the kynurenine pathway evolved in a very early eukaryote and spread to bacterial lineages (Lima, et al. 2009), or evolved in a bacterial lineage that contributed to the genome of the first eukaryotic organisms.

For some enzymes of the kynurenine pathway, horizontal gene transfer from Bacteria or Archaea into eukaryotic genomes seems likely, such as the acquisition of tryprophan-2,3-dioxygenase by the excavate *N. gruberi* from Bacteria (or Archaea; Fig. 33), and the acquisition of bacterial kynurenine-3-monooxygenase by different unicellular eukaryotes (Fig. 35). In addition to a horizontal gene transfer from Bacteroidetes into an early ancestor of Amoebozoa, which has been reported earlier (Lima, et al. 2009), kynurenine-3-monooxygenase has been acquired from Gammaproteobacteria by four unrelated marine eukaryotes, namely *P. marinus* (Alveolata), *S. artica* (Ichthyosporea, Opisthokonta), *A. anophagefferens* (Pelagophyceae, Stramenopiles), and *E. huxleyi* (Haptophyta). In this case it seems likely that a horizontal gene transfer from a marine gammaproteobacterium into a eukaryote, probably *P. marinus*, was followed be gene transfers from one unicellular, marine eukaryote to another (Fig. 53).

IMPACT OF GENE TRANSFER ON METABOLIC EVOLUTION IN EUKARYOTES

Numerous gene transfers shaped *de novo* NAD^+ biosynthesis in eukaryotes, resulting in the acquisition of the aspartate pathway in different photosynthetic clades, and the substitution of several enzymes of the kynurenine and the converged pathways. In Figure 53, which summarizes

these gene transfers, endosymbiotic- and horizontal gene transfers were not differentiated. Due to currently available sequences, these can hardly be discriminated. For the enzymes of the aspartate pathway in *C. paradoxa*, which are of cyanobacterial origin, it seems reasonable to assume they were acquired via endosymbiotic gene transfer. The same holds for nicotinate-nucleotide pyrophosphorylase in *B. natans*, which probably was acquired from green plants during the secondary endosymbiosis giving rise to plastids in Chlorarachniophyta. However, horizontal gene transfer cannot be ruled out in either case as an alternative explanation. On the other hand, the enzymes of the aspartate pathway in Viridiplantae seem to originate from non-photosynthetic bacteria making an acquisition via endosymbiotic gene transfer during plastid evolution unlikely. However, the plastid ancestor, while still a free-living cyanobacterium, might have acquired these two genes encoding the aspartate pathway via horizontal gene transfer from non-photosynthetic bacteria, followed by endosymbiotic gene transfer of these two genes during plastid evolution. Neither has the phylogenetic origin of plastids within cyanobacteria been resolved, nor is the acquisition of hundreds of genes in green plants originating from other prokaryotes well understood (Dagan, et al. 2013).

Regardless whether endosymbiotic or horizontal, the relative large number of gene transfers shaping NAD^+ biosynthesis in eukaryotes is unexpected. *De novo* NAD^+ biosynthesis is essential, and lack of just one enzyme is usually lethal (Panozzo, et al. 2002; Katoh, et al. 2006; Lin, et al. 2010). Therefore, one might expect high conservation of *de novo* NAD^+ biosynthesis. In contrast, phylogenetic analyses show that entire pathways were swapped (kynurenine versus aspartate pathway), and enzymes were substituted by orthologs from bacteria via horizontal gene transfer. Horizontal gene transfer is very common in Bacteria and Archaea (Koonin, et al. 2001; Treangen and Rocha 2011), and there are indications that adaptive evolution of bacterial metabolic networks was largely driven by horizontal gene transfers (Pal, et al. 2005). Systematic screens of eukaryotic genomes for genes originating from horizontal transfer identified many

genes encoding enzymes (Whitaker, et al. 2009; Schönknecht, et al. 2014), and essential metabolic pathways, such as isoprenoid biosynthesis (Lange 2000), heme biosynthesis (Oborník and Green 2005), or the shikimate pathway (Richards, et al. 2006), have been shown to be encoded by a mosaic of genes inherited vertically, and acquired via endosymbiotic or horizontal gene transfers. In addition to *de novo* NAD^+ biosynthesis, salvage biosynthesis of NAD^+ starting with nicotinamide was shaped by horizontal gene transfers as well. Phylogenetic analyses of the two nicotinamide-metabolizing enzymes nicotinamide phosphoribosyltransferase (EC 2.4.2.12) and nicotinamidase (EC 3.5.1.19) show clear deviations from established organismal phylogenies, possibly indicating several instances of horizontal gene transfer (Gazzaniga, et al. 2009). For nicotinamide phosphoribosyltransferase there is evidence for virus- and plasmid-mediated gene transfers (Gazzaniga, et al. 2009). Gene transfers seem to have contributed significantly to the evolution of metabolic pathways in eukaryotes, as shown here for *de novo* NAD^+ biosynthesis.

BIOLOGICAL SIGNIFICANCE OF PORPHYIN EXCRETION IN *G. SULPHURARIA*

Light limited *G. sulphuraria* cultures, whether light limitation is caused by high density or low-light conditions, excrete a mixture of porphyrins into the liquid medium. Our hypothesis proposed that the excreted porphyrins may serve as a 'light converter', by shifting the light spectrum in order to utilize (short) wavelengths of light that were unusable by photosynthetic pigments synthesized in *G. sulphuraria*. The premise of the spectral-shifting hypothesis was based on the absorption and emission spectra of the porphyrins excreted by *G. sulphuraria*. Light was strongly absorbed in the near-UV around 400 nm and re-emitted (fluoresced) between 600- and 700 nm, the region where the photosynthetic pigment phycocyanin shows a high level of absorption. However, given the low calculated relative fluorescence quantum yield of the porphyrin mixture secreted by *G. sulphuraria* (approximately 3%), the effect on photosynthetic

efficiency is likely to be minimal at best. The work presented here is the first time a fluorescence quantum yield has been determined for the Soret band of a porphyrin (400 nm maximum absorption peak). Previous studies on the fluorescence quantum yield of prophyrins solely concentrated on the much smaller Q bands (550 and 590 nm absorption peaks) (Seybold and Gouterman 1969; Figueiredo, et al. 1999; Hu, et al. 2011).

G. sulphuraria only excreted porphyrins when supplemented with glucose. A photoautotrophic culture with high density was achieved only once (using the first approach mentioned below), and did not appear to excrete porphyrins into the liquid medium, i.e. no pink coloration of the culture medium was observed. High densities of phototrophically grown *G. sulphuraria* was not repeatedly achieved due to the less than optimal growth conditions resulting from a lack of available equipment. Phototrophic cultures were grown two different ways. The first was using a Percival growth chamber set to 37°C under continuous, 24 hour (fluorescent) light. An exogenous supply of CO_2 was provided by directly aerating the culture flasks with pure CO_2. This method was largely unsuccessful in achieving high density cultures comparable to densities achieved with mixotrophic/heterotrophic cultures supplemented with glucose. The other approach was similar to that used for mixotrophic/heterotrophic cultures. Phototrophic cultures were grown in an incubated, shaker table under 24 hour continuous light from a 120W flood lamp; these cultures were not supplemented with glucose. However, carbon was assumed to be limiting under the second approach, and could not be mediated by supplying carbonate without affecting the pH of the liquid medium (pH 2). Nonetheless, the suggested glucose inhibition of photosynthetic pigments in *G. partita* (Stadnichuk, et al. 1998) leaves some questions unanswered. Stadnichuk, et al. (1998) never achieved cell densities of photoautotrophic cultures comparable to that of heterotrophic cultures supplemented with glucose. Even though high density photoautotrophic cultures were never continuously achieved here either, there is no way to rule out porphyrin excretion as a result of high cell density. Furthermore, growth studies

conducted here showed no correlation with glucose supplementation, but only with cell density (Fig. 40).

Porphyrins are conjugated tetrapyrroles, most of which contain, and are identified by, their alkyl and/or carboxylic acid embellishments. Porphyrins are ubiquitous among all living organisms, and have even been proposed as a biomarker in the search for extraterrestrial life (Lim 2010). Porphyrin precursors are used as skeletons for many molecules essential for life including, but not limited to, chlorophyll and heme groups found in hemoglobin, cytochromes, and peroxidases (Lim 2010). Porphyrins are synthesized by a metabolic pathway that is highly conserved throughout all life. All of the enzymes for this pathway have been identified in the *G. sulphuraria* genome (Figs. 38 & 39). Thus, the question remains: why does *G. sulphuraria* excrete porphyrins into its environment? Spectral shifting, although seemingly unlikely based on the low relative fluorescence quantum yield, cannot entirely be ruled out. To ultimately address the hypothesis of spectral shifting, growth studies would need to be conducted not only with *G. sulphuraria*, but with other algal species, some of which hold significance to algal biofuel production. The idea of spectral shifting is not entirely novel. Several patents exist that are centered on repurposing light to be used for photosynthesis. A polymeric material containing 2,5-di-(5-tert-butyl-2-benzoxazolyl) thiophene was patented to shield plants in greenhouses from UV radiation by converting UV into visible light, thereby increasing photosynthetically active radiation (PAR) (Treadaway 1980). One patent describes a variety of approaches to add wavelength-shifting pigments to algal cultures to increase PAR, and thus, growth rates (Sayre 2009). Another patent application describes the transformation of algae or cyanobacteria to express "fluorescent proteins, that absorb the harmful UV or near UV wavelengths and emit wavelengths that are photosynthetically more active" (Einbinder, et al. 2010).

Porphyrins can accumulate to levels where they become toxic. High levels of porphyrins have been shown to promote the formation of reactive oxygen species (ROS), which damage cell

components via oxidation (Böhm, et al. 2001). The excretion of porphyrins by *G. sulphuraria* could simply be a mechanism by which the potentially cytotoxic accumulation of porphyrins is removed from the cell. However, endogenous porphyrins could potentially impact the growth of neighboring organisms. Porphyrins are relatively hydrophobic by nature (as dictated by their side chains), and some have been shown to penetrate and become lodged in lipid bilayers (Ricchelli 1995). Once embedded in a membrane, porphyrins can cause oxidative damage compromising membrane integrity. In the specific case of heme, a porphyrin chelated with an iron atom, oxidative damage from embedded heme molecules results in the formation of lipid peroxide, which amplifies membrane permeability, ultimately leading to cell lysis and death (Chiabrando, et al. 2014). Currently, no studies have specifically investigated 'free' porphyrins, those that are not chelated with a metal atom, such as uroporphyrin I and III, and their ability to penetrate lipid bilayers. The ability of porphyrins to become lodged into membranes, subsequently causing oxidative membrane damage, has been exploited by biomedical technology. Photodynamic therapy and photodynamic inhibition, whereby membrane-lodged porphyrins are photoexcited to cause oxidative damage through reactive oxygen species formation, have been extensively studied for applications treating cancerous tumors, in addition to combating antibiotic-resistant bacteria (Dolmans, et al. 2003; Huang, et al. 2010; Pereira, et al. 2014).

G. sulphuraria could potentially use porphyrins as a quorum sensing molecule. Historically, quorum sensing, the mechanism by which intraspecies and interspecies communication is achieved typically in the context of population density, has been characteristic for bacteria. More recently, it has been shown that fungi also employ their own variation of quorum sensing (Sprague and Winans 2006). Even plants possess the ability to detect bacterial quorum sensing molecules (Williams 2007). It seems rather unlikely that *G. sulphuraria* excretes porphyrins to initiate a quorum sensing mechanism, given the ubiquitous nature of porphyrins throughout the tree of life. Porphyrins serve several biological functions, many of which persist

throughout all three domains of life; however, porphyrins could potentially function to alter quorum sensing molecules produced by neighboring prokaryotes. Some plant and algal species produce molecules that mimic or possess the ability to manipulate bacterial quorum sensing molecules. The red alga *Delisea pulchra* produces halogenated furanones, which are structurally similar to *N*-acyl homoserine lactones, a common quorum sensing molecule produced by bacteria, and alter lactone-mediated quorum sensing behavior in bacteria (Bauer and Mathesius 2004). Although no studies exist investigating the interaction between porphyrins and bacterial quorum sensing molecules, it is possible that porphyrins could promote the degradation or manipulation of quorum sensing signals created by neighboring bacteria.

Porphyrin excretion may result from poor regulation of gene expression. This would be surprising considering the biosynthesis of pophyrins is energetically costly from a biological standpoint. If secreted porphyrins aren't serving a direct biological purpose, one would suspect the corresponding genes to be down regulated, in order to conserve cellular resources. Previous studies on *G. sulphuraria* revealed a very slow response time when supplied with alternative carbon sources. Oesterhelt and Gross (2002) observed a 45 day 'lag phase' after switching carbon sources when investigating changes in gene expression of genes involved in carbon metabolism. On the other hand, considering that porphyrin excretion was only observed when cultures were supplemented with high concentrations of glucose, carbon might not be limiting under that particular scenario, providing no incentive to ration exogenous carbon supplies. Additionally, porphyrin excretion may serve as a mechanism—although a seemingly wasteful one—which triggers the switch from a phototrophic lifestyle to that of heterotrophy. Established heterotrophic cells do not appear to produce photosynthetic pigments as indicated by the loss of the green coloration (Fig. 1). The offloading of excess photosynthetic pigment precursors, i.e. porphyrins, could further promote the switch to a heterotrophic lifestyle. Nonetheless, the biological reason why *G. sulphuraria* excretes porphyrins remains unclear. Several speculations have been made

addressing this phenomenon. However, without performing additional experiments, even those aimed to support the possible speculations listed here, it is difficult to definitively determine the underlying cause for porphyrin excretion in *G. sulphuraria.*

REFERENCES

Abascal F, Zardoya R, Posada D. 2005. ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21:2104-2105.

Akhtar MK, Kaderbhai NN, Hopper DJ, Kelly SL, Kaderbhai MA. 2003. Export of a heterologous cytochrome P450 (CYP105D1) in Escherichia coli is associated with periplasmic accumulation of uroporphyrin. Journal of Biological Chemistry 278:45555-45562.

Altekar G, Dwarkadas S, Huelsenbeck JP, Ronquist F. 2004. Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. Bioinformatics 20:407-415.

Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research 25:3389-3402.

Andersson JO. 2011. Evolution of Patchily Distributed Proteins Shared between Eukaryotes and Prokaryotes: Dictyostelium as a Case Study. Journal of Molecular Microbiology and Biotechnology 20:83-95.

Archibald JM. 2009. The Puzzle of Plastid Evolution. Current Biology 19:R81-R88.

Archibald JM, Rogers MB, Toop M, Ishida K, Keeling PJ. 2003. Lateral gene transfer and the evolution of plastid-targeted proteins in the secondary plastid-containing alga Bigelowiella natans. Proceedings of the National Academy of Sciences of the United States of America 100:7678-7683.

Baecker PA, Greenberg E, Preiss J. 1986. Biosynthesis of bacterial glycogen. Primary structure of Escherichia coli 1,4-alpha-D-glucan:1,4-alpha-D-glucan 6-alpha-D-(1, 4-alpha-D-glucano)-transferase as deduced from the nucleotide sequence of the glg B gene. Journal of Biological Chemistry 261:8738-8743.

Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV. 2012. Improving the Accuracy of Demographic and Molecular Clock Model Comparison While Accommodating Phylogenetic Uncertainty. Molecular Biology and Evolution 29:2157-2167.

Ball S, Colleoni C, Cenci U, Raj JN, Tirtiaux C. 2011. The evolution of glycogen and starch metabolism in eukaryotes gives molecular clues to understand the establishment of plastid endosymbiosis. Journal of Experimental Botany 62:1775-1801.

Barbier G, et. al. 2005. Comparative Genomics of Two Closely Related Unicellular Thermo-Acidophilic Red Algae, Galdieria sulphuraria and Cyanidioschyzon merolae, Reveals the Molecular Basis of the Metabolic Flexibility of Galdieria sulphuraria and Significant Differences in Carbohydrate Metabolism of Both Algae. Plant Physiology 137:460-474.

Barbier GG, Zimmermann M, Weber APM editors.; 2005.

Bauer WD, Mathesius U. 2004. Plant responses to bacterial quorum sensing signals. Current Opinion in Plant Biology 7:429-433.

Beadle GW, Mitchell HK, Nyc JF. 1947. KYNURENINE AS AN INTERMEDIATE IN THE FORMATION OF NICOTINIC ACID FROM TRYPTOPHANE BY NEUROSPORA. Proceedings of the National Academy of Sciences of the United States of America 33:155-158.

Beiko RG, Harlow TJ, Ragan MA. 2005. Highways of gene sharing in prokaryotes. Proceedings of the National Academy of Sciences of the United States of America 102:14332-14337.

Belenky P, Bogan KL, Brenner C. 2007. NAD(+) metabolism in health and disease. Trends in Biochemical Sciences 32:12-19.

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. 2000. The Protein Data Bank. Nucleic Acids Research 28:235-242.

Bhattacharya D, Price DC, Chan CX, Qiu H, Rose N, Ball S, Weber APM, Cecilia Arias M, Henrissat B, Coutinho PM, et al. 2013. Genome of the red alga Porphyridium purpureum. Nat Commun 4.

Bhattacharya D, Yoon HS, Hackett JD. 2004. Photosynthetic eukaryotes unite: endosymbiosis connects the dots. Bioessays 26:50-60.

Böhm F, Edge R, Foley S, Lange L, Truscott TG. 2001. Antioxidant inhibition of porphyrin-induced cellular phototoxicity. Journal of Photochemistry and Photobiology B: Biology 65:177-183.

Brown SB, Holroyd JA, Vernon DI, Troxler RF, Smith KM. 1982. The Effect of N-Mtheylprotoporphyrin-IX on the Syntheis of Photosynthetic Pigments in *Cyanidium caldarium* - Further Evidence for the Role of Heme in the Biosynthesis of Plant Bilins. Biochemical Journal 208:487-491.

Burki F, Flegontov P, Oborník M, Cihlář J, Pain A, Lukeš J, Keeling PJ. 2012. Re-evaluating the Green versus Red Signal in Eukaryotes with Secondary Plastid of Red Algal Origin. Genome Biology and Evolution 4:738-747.

Burki F, Kudryavtsev A, Matz MV, Aglyamova GV, Bulman S, Fiers M, Keeling PJ, Pawlowski J. 2010. Evolution of Rhizaria: new insights from phylogenomic analysis of uncultivated protists. BMC Evolutionary Biology 10.

Burki F, Okamoto N, Pombert J-F, Keeling PJ. 2012. The evolutionary history of haptophytes and cryptophytes: phylogenomic evidence for separate origins. Proceedings of the Royal Society B: Biological Sciences.

Butterfield NJ. 2000. Bangiomorpha pubescens n. Gen., n. sp.: Implications for the Evolution of Sex, Multicellularity, and the Mesoproterozoic/Neoproterozoic Radiation of Eukaryotes. Paleobiology 26:386-404.

Cannata JJB, Focesi A, Mazumder R, Warner RC, Ochoa S. 1965. Metabolism of Propionic Acid in Animal Tissues .12. Propities of Mammalian Methylmalonyl Coenzyme-A Mutase. Journal of Biological Chemistry 240:3249-&.

Cavalier-Smith T, Chao E-Y. 2006. Phylogeny and Megasystematics of Phagotrophic Heterokonts (Kingdom Chromista). Journal of Molecular Evolution 62:388-420.

Chen M, Blankenship RE. 2011. Expanding the solar spectrum used by photosynthesis. Trends in Plant Science 16:427-431.

Chiabrando D, Vinchi F, Fiorito V, Mercurio S, Tolosano E. 2014. Heme in pathophysiology: a matter of scavenging, metabolism and trafficking across cell membranes. Frontiers in Pharmacology 5.

Collén J, Porcel B, Carré W, Ball SG, Chaparro C, Tonon T, Barbeyron T, Michel G, Noel B, Valentin K, et al. 2013. Genome structure and metabolic features in the red seaweed Chondrus crispus shed light on evolution of the Archaeplastida. Proceedings of the National Academy of Sciences.

Consortium TU. 2014. Activities at the Universal Protein Resource (UniProt). Nucleic Acids Research 42:D191-D198.

Criscuolo A, Gribaldo S. 2011. Large-Scale Phylogenomic Analyses Indicate a Deep Origin of Primary Plastids within Cyanobacteria. Molecular Biology and Evolution 28:3019-3032.
Croft MT, Lawrence AD, Raux-Deery E, Warren MJ, Smith AG. 2005. Algae acquire vitamin B12 through a symbiotic relationship with bacteria. Nature 438:90-93.
Crosby GA, Demas JN. 1971. Measurement of photoluminescence quantum yields. Review. The Journal of Physical Chemistry 75:991-1024.
Dagan T, Roettger M, Stucken K, Landan G, Koch R, Major P, Gould SB, Goremykin VV, Rippka R, de Marsac NT, et al. 2013. Genomes of Stigonematalean Cyanobacteria (Subsection V) and the Evolution of Oxygenic Photosynthesis from Prokaryotes to Plastids. Genome Biology and Evolution 5:31-44.
Deschamps P, Colleoni C, Nakamura Y, Suzuki E, Putaux J-L, Buléon A, Haebel S, Ritte G, Steup M, Falcón LI, et al. 2008. Metabolic Symbiosis and the Birth of the Plant Kingdom. Molecular Biology and Evolution 25:536-548.
Deschamps P, Moreira D. 2012. Reevaluating the Green Contribution to Diatom Genomes. Genome Biology and Evolution 4:795-800.
Dhami S, Mello AJD, Rumbles G, Bishop SM, Phillips D, Beeby A. 1995. PHTHALOCYANINE FLUORESCENCE AT HIGH CONCENTRATION: DIMERS OR REABSORPTION EFFECT? Photochemistry and Photobiology 61:341-346.
Di Tommaso P, Moretti S, Xenarios I, Orobitg M, Montanyola A, Chang J-M, Taly J-F, Notredame C. 2011. T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. Nucleic Acids Research.
Dolmans DEJGJ, Fukumura D, Jain RK. 2003. Photodynamic therapy for cancer. Nat Rev Cancer 3:380-387.
Dorrell RG, Smith AG. 2011. Do Red and Green Make Brown?: Perspectives on Plastid Acquisitions within Chromalveolates. Eukaryotic Cell 10:856-868.
Duff SMG, Lefebvre DD, Plaxton WC. 1989. Purification and Characterization of a Phosphoenolpyruvate Phosphatase from Brassica nigra Suspension Cells. Plant Physiology 90:734-741.
Eichinger L, Pachebat JA, Glockner G, Rajandream MA, Sucgang R, Berriman M, Song J, Olsen R, Szafranski K, Xu Q, et al. 2005. The genome of the social amoeba Dictyostelium discoideum. Nature 435:43-57.
Einbinder S, Ufaz S, Eisenstadt D, Schatz D, Gressel J. 2010. Use of fluorescent protien in cyanobacteria and algae for improving photosynthesis and preventing cell damage. In: Google Patents.
Engqvist M, Drincovich MF, Flügge U-I, Maurino VG. 2009. Two d-2-Hydroxy-acid Dehydrogenases in Arabidopsis thaliana with Catalytic Capacities to Participate in the Last Reactions of the Methylglyoxal and β-Oxidation Pathways. Journal of Biological Chemistry 284:25026-25037.
Erni B, Siebold C, Christen S, Srinivas A, Oberholzer A, Baumann U. 2006. Small substrate, big surprise: fold, function and phylogeny of dihydroxyacetone kinases. Cellular and Molecular Life Sciences CMLS 63:890-900.
Figueiredo TLC, Johnstone RAW, Sorensen A, Burget D, Jacques P. 1999. Determination of fluorescence yields, singlet lifetimes and singlet oxygen yields of water-insoluble porphyrins and metalloporphyrins in organic solvents and in aqueous media. Photochemistry and Photobiology 69:517-528.
Finn RD, Clements J, Eddy SR. 2011. HMMER web server: interactive sequence similarity searching. Nucleic Acids Research 39:W29-W37.
Ford Doolittle W. 1998. You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. Trends in Genetics 14:307-311.

Gaertner FH, Subbayya Shetty A. 1977. Kynureninase-type enzymes and the evolution of the aerobic tryptophan-to-nicotinamide adenine dinucleotide pathway. Biochimica et Biophysica Acta (BBA) - Enzymology 482:453-460.
Gazzaniga F, Stebbins R, Chang SZ, McPeek MA, Brenner C. 2009. Microbial NAD Metabolism: Lessons from Comparative Genomics. Microbiology and Molecular Biology Reviews : MMBR 73:529-541.
Gobler CJ, Berry DL, Dyhrman ST, Wilhelm SW, Salamov A, Lobanov AV, Zhang Y, Collier JL, Wurch LL, Kustka AB, et al. 2011. Niche of harmful alga Aureococcus anophagefferens revealed through ecogenomics. Proceedings of the National Academy of Sciences of the United States of America 108:4352-4357.
Gould SB, Waller RR, McFadden GI. 2008. Plastid evolution. Annual Review of Plant Biology 59:491-517.
Graziani G, Schiavo S, Nicolai MA, Buono S, Fogliano V, Pinto G, Pollio A. 2013. Microalgae as human food: chemical and nutritional characteristics of the thermo-acidophilic microalga Galdieria sulphuraria. Food & Function 4:144-152.
Griffith GR, Chandler JLR, Gholson RK. 1975. Studies on the de novo Biosynthesis of NAD in Escherichia coli. European Journal of Biochemistry 54:239-245.
Gross W. 2000. Ecophysiology of algae living in highly acidic environments. Hydrobiologia 433:31-37.
Gross W, Schnarrenberger C. 1995a. Heterotrophic Growth of 2 Strains of the Acido-Thermophilic Red Alga *Galdieria sulphuraria*. Plant and Cell Physiology 36:633-638.
Gross W, Schnarrenberger C. 1995b. Purification and characterization of a galactose-1-phosphate: UDP-glucose uridyltransferase from the red alga Galdieria sulphuraria. European Journal of Biochemistry 234:258-263.
Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic Biology 52:696-704.
Hallborn J, Walfridsson M, Penttila M, Keranen S, Hahnhagerdal B. 1995. A SHORT-CHAIN DEHYDROGENASE GENE FROM PICHIA-STIPITIS HAVING D-ARABINITOL DEHYDROGENASE-ACTIVITY. Yeast 11:839-847.
Hayashi S-I, Lin ECC. 1967. Purification and Properties of Glycerol Kinase from Escherichia coli. Journal of Biological Chemistry 242:1030-1035.
Heilmann I, Schnarrenberger C, Gross W. 1997. Mannose metabolizing enzymes from the red alga Galdieria sulphuraria. Phytochemistry 45:903-906.
Herold A, Lewis DH. 1977. Mannose and Green Plants: Occurrence, Physiology and Metabolism, and Use as a Tool to Study the Role of Orthophosphate. New Phytologist 79:1-40.
Hirabaru C, Izumo A, Fujiwara S, Tadokoro Y, Shimonaga T, Konishi M, Yoshida M, Fujita N, Nakamura Y, Kuroiwa T, et al. 2010. The primitive rhodophyte Cyanidioschyzon merolae contains a semiamylopectin-type, but not an amylose-type, -glucan. Plant and Cell Physiology 51:682-693.
Hoeft SE, Kulp TR, Han S, Lanoil B, Oremland RS. 2010. Coupled Arsenotrophy in a Hot Spring Photosynthetic Biofilm at Mono Lake, California. Applied and Environmental Microbiology 76:4633-4639.
A Guide to Recording Fluorescence Quantum Yields [Internet]. 2014. Available from: http://www.horiba.com/fileadmin/uploads/Scientific/Documents/Fluorescence/quantumyieldstrad.pdf
Horst RJ. 2006. Changes in carbohydrate metabolism of Galdieria sulphuraria in response to salt stress. [
Hu Y, Geissinger P, Woehl JC. 2011. Potential of protoporphyrin IX and metal derivatives for single molecule fluorescence studies. Journal of Luminescence 131:477-481.

Huang L, Dai T, Hamblin M. 2010. Antimicrobial Photodynamic Inactivation and Photodynamic Therapy for Infections. In: Gomer CJ, editor. Photodynamic Therapy: Humana Press. p. 155-173.
Janson CA, Cleland WW. 1974. The Kinetic Mechanism of Glycerokinase. Journal of Biological Chemistry 249:2562-2566.
Kanehisa M, Goto S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research 28:27-30.
Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2014. Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Research 42:D199-D205.
Katoh A, Hashimoto T. 2004. Molecular biology of pyridine nucleotide and nicotine biosynthesis. Frontiers in Bioscience 9:1577-1586.
Katoh A, Uenohara K, Akita M, Hashimoto T. 2006. Early steps in the biosynthesis of NAD in Arabidopsis start with aspartate and occur in the plastid. Plant Physiology 141:851-857.
Keeling PJ. 2010. The endosymbiotic origin, diversification and fate of plastids. Philosophical Transactions of the Royal Society B: Biological Sciences 365:729-748.
Keeling PJ. 2013. The Number, Speed, and Impact of Plastid Endosymbioses in Eukaryotic Evolution. Annual Review of Plant Biology 64:583-607.
Keeling PJ, Palmer JD. 2008. Horizontal gene transfer in eukaryotic evolution. Nat Rev Genet 9:605-618.
Kiang NY, Siefert J, Blankenship RE. 2007. Spectral signatures of photosynthesis. I. Review of earth organisms. Astrobiology 7:222-251.
Kirst GO. 1980. Low MW carbohydrates and ions in rhodophyceae: Quantitative measurement of floridoside and digeneaside. Phytochemistry 19:1107-1110.
Knudson L. 1917. The Toxicity of Galactose and Mannose for Green Plants and the Antagonistic Action of other Sugars Toward These. American Journal of Botany 4:430-437.
Kondrashov FA, Koonin EV, Morgunov IG, Finogenova TV, Kondrashova MN. 2006. Evolution of glyoxylate cycle enzymes in Metazoa: evidence of multiple horizontal transfer events and pseudogene formation. Biology Direct 1:31-31.
Koonin EV, Makarova KS, Aravind L. 2001. Horizontal gene transfer in prokaryotes: Quantification and classification. Annual Review of Microbiology 55:709-742.
Kořený L, Oborník M. 2011. Sequence Evidence for the Presence of Two Tetrapyrrole Pathways in Euglena gracilis. Genome Biology and Evolution 3:359-364.
Kruger NJ, von Schaewen A. 2003. The oxidative pentose phosphate pathway: structure and organisation. Current Opinion in Plant Biology 6:236-246.
Kurnasov O, Goral V, Colabroy K, Gerdes S, Anantha S, Osterman A, Begley TP. 2003. NAD biosynthesis: Identification of the tryptophan to quinolinate pathway in bacteria. Chemistry & Biology 10:1195-1204.
Lafraie MA, Betz A. 1985. Anaerobic Fermentation in *Cyanidium caldarium*. Planta 163:38-42.
Lakowicz JR. 2009. Principles of Fluorescence Spectroscopy: Springer London, Limited.
Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M, et al. 2011. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Research.
Lange BM. 2000. Isoprenoid biosynthesis: The evolution of two ancient and distinct pathways across genomes. Proceedings of the National Academy of Sciences 97:13172-13177.
Lim CK. 2010. High-performance liquid chromatography and mass spectrometry of porphyrins, chlorophylls and bilins: World Scientific.
Lima WC, Varani AM, Menck CFM. 2009. NAD Biosynthesis Evolution in Bacteria: Lateral Gene Transfer of Kynurenine Pathway in Xanthomonadales and Flavobacteriales. Molecular Biology and Evolution 26:399-406.

Lin HW, Kwan AL, Dutcher SK. 2010. Synthesizing and Salvaging NAD(+): Lessons Learned from Chlamydomonas reinhardtii. Plos Genetics 6.
Loftus B, Anderson I, Davies R, Alsmark UCM, Samuelson J, Amedeo P, Roncaglia P, Berriman M, Hirt RP, Mann BJ, et al. 2005. The genome of the protist parasite Entamoeba histolytica. Nature 433:865-868.
Lohr M, Schwender J, Polle JEW. 2012. Isoprenoid biosynthesis in eukaryotic phototrophs: A spotlight on algae. Plant Science 185–186:9-22.
Lohse M, Nagel A, Herter T, May P, Schroda M, Zrenner R, Tohge T, Fernie AR, Stitt M, Usadel B. 2014. Mercator: a fast and simple web server for genome scale functional annotation of plant sequence data. Plant, Cell & Environment 37:1250-1258.
Lombard J, Moreira D. 2011. Origins and Early Evolution of the Mevalonate Pathway of Isoprenoid Biosynthesis in the Three Domains of Life. Molecular Biology and Evolution 28:87-99.
Lucas KA, Filley JR, Erb JM, Graybill ER, Hawes JW. 2007. Peroxisomal metabolism of propionic acid and isobutyric acid in plants. Journal of Biological Chemistry 282:24980-24989.
Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, et al. 2011. CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Research 39:D225-D229.
Martin W, Schnarrenberger C. 1997. The evolution of the Calvin cycle from prokaryotic to eukaryotic chromosomes: a case study of functional redundancy in ancient pathways through endosymbiosis. Current Genetics 32:1-18.
Matsuzaki M, Misumi O, Shin-I T, Maruyama S, Takahara M, Miyagishima SY, Mori T, Nishida K, Yagisawa F, Yoshida Y, et al. 2004. Genome sequence of the ultrasmall unicellular red alga Cyanidioschyzon merolae 10D. Nature 428:653-657.
McCoy JG, Bailey LJ, Ng YH, Bingman CA, Wrobel R, Weber APM, Fox BG, Phillips GN. 2009. Discovery of sarcosine dimethylglycine methyltransferase fromGaldieria sulphuraria. Proteins: Structure, Function, and Bioinformatics 74:368-377.
McFadden GI. 2014. Origin and Evolution of Plastids and Photosynthesis in Eukaryotes. Cold Spring Harbor Perspectives in Biology 6.
Meek DW, Hatfield JL, Howell TA, Idso SB, Reginato RJ. 1984. A Generalized Relationship between Photosynthetically Active Radiation and Solar Radiation. Agron. J. 76:939-945.
Miller GL. 1959. Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. Analytical Chemistry 31:426-428.
Monteith JL. 1969. Light Interception and Radiative Exchange in Crop Stands. In: Eastin JD, Haskins FA, Sullivan CY, van Bavel CHM, editors. Physiological Aspects of Crop Yield: American Society of Agronomy, Crop Science Society of America. p. 89-111.
Monteith JL. 1994. Validity of the correlation between intercepted radiation and biomass. Agricultural and Forest Meteorology 68:213-220.
Morgan DR, Street HE. 1959. The Carbohydrate Nutrition of Tomato Roots: VII. Sugars, Sugar Phosphates, and Sugar Alcohols as Respiratory Substrates for Excised Roots. Annals of Botany 23:89-105.
Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. 2007. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Research 35:W182-W185.
Moustafa A, Beszteri B, Maier UG, Bowler C, Valentin K, Bhattacharya D. 2009. Genomic Footprints of a Cryptic Plastid Endosymbiosis in Diatoms. Science 324:1724-1726.
Mustroph A, Sonnewald U, Biemelt S. 2007. Characterisation of the ATP-dependent phosphofructokinase gene family from Arabidopsis thaliana. Febs Letters 581:2401-2410.
Nagasaka S, Nishizawa NK, Mori S, Yoshimura E. 2004. Metal metabolism in the red alga Cyanidium caldarium and its relationship to metal tolerance. Biometals 17:177-181.

Nakamura Y, Sasaki N, Kobayashi M, Ojima N, Yasuike M, Shigenobu Y, Satomi M, Fukuma Y, Shiwaku K, Tsujimoto A, et al. 2013. The First Symbiont-Free Genome Sequence of Marine Red Alga, Susabi-nori (<italic>Pyropia yezoensis</italic>). PLoS ONE 8:e57122.
Nedelcu AM, Blakney AJC, Logue KD. 2009. Functional replacement of a primary metabolic pathway via multiple independent eukaryote-to-eukaryote gene transfers and selective retention. Journal of Evolutionary Biology 22:1882-1894.
Nielsen TH, Rung JH, Villadsen D. 2004. Fructose-2,6-bisphosphate: a traffic signal in plant metabolism. Trends in Plant Science 9:556-563.
Noctor G, Hager J, Li S. 2011. Biosynthesis of NAD and Its Manipulation in Plants. In: Fabrice R, Roland D, editors. Advances in Botanical Research: Academic Press. p. 153-201.
Notredame C, Higgins DG, Heringa J. 2000. T-Coffee: A novel method for fast and accurate multiple sequence alignment. Journal of Molecular Biology 302:205-217.
Oborník M, Green BR. 2005. Mosaic Origin of the Heme Biosynthesis Pathway in Photosynthetic Eukaryotes. Molecular Biology and Evolution 22:2343-2353.
Oesterhelt C, Gross W. 2002. Different Sugar Kinases Are Involved in the Sugar Sensing of Galdieria sulphuraria. Plant Physiology 128:291-299.
Oesterhelt C, Schnarrenberger C, Gross W. 1999. Characterization of a sugar/polyol uptake system in the red alga Galdieria sulphuraria. European Journal of Phycology 34:271-277.
Owen SM, Peñuelas J. 2005. Opportunistic emissions of volatile isoprenoids. Trends in Plant Science 10:420-426.
Pabarcus MK, Casida JE. 2002. Kynurenine formamidase: determination of primary structure and modeling-based prediction of tertiary structure and catalytic triad1. Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology 1596:201-211.
Pade N, Linka N, Ruth W, Weber APM, Hagemann M. 2014. Floridoside and isofloridoside are synthesized by trehalose 6-phosphate synthase-like enzymes in the red alga Galdieria sulphuraria. New Phytologist:n/a-n/a.
Pal C, Papp B, Lercher MJ. 2005. Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. Nature Genetics 37:1372-1375.
Panozzo C, Nawara M, Suski C, Kucharczyka R, Skoneczny M, Bécam A-M, Rytka J, Herbert CJ. 2002. Aerobic and anaerobic NAD+ metabolism in Saccharomyces cerevisiae. Febs Letters 517:97-102.
Papenbrock J, Mock H-P, Tanaka R, Kruse E, Grimm B. 2000. Role of Magnesium Chelatase Activity in the Early Steps of the Tetrapyrrole Biosynthetic Pathway. Plant Physiology 122:1161-1170.
Patron NJ, Keeling PJ. 2005. Common Evolutionary Origin of Starch Biosynthetic Enzymes in Green and Red Algae. Journal of Phycology 41:1131-1141.
Pereira MA, Faustino MAF, Tome JPC, Neves MGPMS, Tome AC, Cavaleiro JAS, Cunha A, Almeida A. 2014. Influence of external bacterial structures on the efficiency of photodynamic inactivation by a cationic porphyrin. Photochemical & Photobiological Sciences 13:680-690.
Piattoni CV, Bustos DM, Guerrero SA, Iglesias AÁ. 2011. Nonphosphorylating Glyceraldehyde-3-Phosphate Dehydrogenase Is Phosphorylated in Wheat Endosperm at Serine-404 by an SNF1-Related Protein Kinase Allosterically Inhibited by Ribose-5-Phosphate. Plant Physiology 156:1337-1350.
Plata G, Hsiao T-L, Olszewski KL, LlinasManuel, Vitkup D. 2010. Reconstruction and flux-balance analysis of the Plasmodium falciparum metabolic network. Mol Syst Biol 6.
Pollak N, Dolle C, Ziegler M. 2007. The power to reduce: pyridine nucleotides - small molecules with a multitude of functions. Biochemical Journal 402:205-218.
Pombert J-F, Selman M, Burki F, Bardell FT, Farinelli L, Solter LF, Whitman DW, Weiss LM, Corradi N, Keeling PJ. 2012. Gain and loss of multiple functionally related, horizontally

transferred genes in the reduced genomes of two microsporidian parasites. Proceedings of the National Academy of Sciences 109:12638-12643.
Popper ZA, Michel G, Hervé C, Domozych DS, Willats WGT, Tuohy MG, Kloareg B, Stengel DB. 2011. Evolution and Diversity of Plant Cell Walls: From Algae to Flowering Plants. Annual Review of Plant Biology 62:567-590.
Preiss J, Handler P. 1958. Biosynthesis of Diphosphopyridine Nucleotide. Journal of Biological Chemistry 233:488-492.
Price DC, Chan CX, Yoon HS, Yang EC, Qiu H, Weber APM, Schwacke R, Gross J, Blouin NA, Lane C, et al. 2012. Cyanophora paradoxa Genome Elucidates Origin of Photosynthesis in Algae and Plants. Science 335:843-847.
Prokop A, Quinn MF, Fekri M, Murad M, Ahmed SA. 1984. Spectral shifting by dyes to enhance algae growth. Biotechnology and Bioengineering 26:1313-1322.
Prosselkov PV, Gross W, Igamberdiev AU, Schnarrenberger C. 1996. Purification and characterization of UDP-D-galactose 4-epimerase from the red alga Galdieria sulphuraria. Physiologia Plantarum 98:753-758.
Pruitt KD, Tatusova T, Maglott DR. 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Research 35:D61-D65.
Qiu H, Price DC, Weber APM, Facchinelli F, Yoon HS, Bhattacharya D. 2013. Assessing the bacterial contribution to the plastid proteome. Trends in Plant Science.
Qiu H, Yoon HS, Bhattacharya D. 2013. Algal endosymbionts as vectors of horizontal gene transfer in photosynthetic eukaryotes. Frontiers in Plant Science 4.
Read BA, Kegel J, Klute MJ, Kuo A, Lefebvre SC, Maumus F, Mayer C, Miller J, Monier A, Salamov A, et al. 2013. Pan genome of the phytoplankton Emiliania underpins its global distribution. Nature 499:209-213.
Ren Q, Paulsen IT. 2005. Comparative Analyses of Fundamental Differences in Membrane Transport Capabilities in Prokaryotes and Eukaryotes. PLoS Comput Biol 1:e27.
Ricchelli F. 1995. Photophysical properties of porphyrins in biological membranes. Journal of Photochemistry and Photobiology B: Biology 29:109-118.
Richards TA, Dacks JB, Campbell SA, Blanchard JL, Foster PG, Mcleod R, Roberts CW. 2006. Evolutionary origins of the eukaryotic shikimate pathway: gene fusions, horizontal gene transfer and endosymbiotic replacements.
Roberts RM, Heishman A, Wicklin C. 1971. Growth Inhibition and Metabolite Pool Levels in Plant Tissues Fed D-Glucosamine and D-Galactose. Plant Physiology 48:36-42.
Rolland F, Baena-Gonzalez E, Sheen J. 2006. SUGAR SENSING AND SIGNALING IN PLANTS: Conserved and Novel Mechanisms. Annual Review of Plant Biology 57:675-709.
Roth JR, Lawrence JG, Bobik TA. 1996. Cobalamin (coenzyme B-12): Synthesis and biological significance. Annual Review of Microbiology 50:137-181.
Ruan Y-L. 2014. Sucrose Metabolism: Gateway to Diverse Carbon Use and Sugar Signaling. Annual Review of Plant Biology 65:33-67.
Rurack K, Spieles M. 2011. Fluorescence Quantum Yields of a Series of Red and Near-Infrared Dyes Emitting at 600−1000 nm. Analytical Chemistry 83:1232-1242.
Rylott EL, Gilday AD, Graham IA. 2003. The Gluconeogenic Enzyme Phosphoenolpyruvate Carboxykinase in Arabidopsis Is Essential for Seedling Establishment. Plant Physiology 131:1834-1842.
Salerno GL, Curatti L. 2003. Origin of sucrose metabolism in higher plants: when, how and why? Trends in Plant Science 8:63-69.
Savvi S, Warner DF, Kana BD, McKinney JD, Mizrahi V, Dawes SS. 2008. Functional Characterization of a Vitamin B12-Dependent Methylmalonyl Pathway in

Mycobacterium tuberculosis: Implications for Propionate Metabolism during Growth on Fatty Acids. Journal of Bacteriology 190:3886-3895.
Sayre RT. 2009. Optimization of biofuel production. In: Google Patents.
Scheer H. 2003. The Pigments. In: Green B, Parson W, editors. Light-Harvesting Antennas in Photosynthesis: Springer Netherlands. p. 29-81.
Scheer M, Grote A, Chang A, Schomburg I, Munaretto C, Rother M, Sohngen C, Stelzer M, Thiele J, Schomburg D. 2011. BRENDA, the enzyme information system in 2011. Nucleic Acids Research 39:D670-D676.
Schmidt RA, Wiebe MG, Eriksen NT. 2005. Heterotrophic high cell-density fed-batch cultures of the phycocyanin-producing red algaGaldieria sulphuraria. Biotechnology and Bioengineering 90:77-84.
Schönknecht G, Chen W-H, Ternes CM, Barbier GG, Shrestha RP, Stanke M, Bräutigam A, Baker BJ, Banfield JF, Garavito RM, et al. 2013. Gene Transfer from Bacteria and Archaea Facilitated Evolution of an Extremophilic Eukaryote. Science 339:1207-1210.
Schönknecht G, Weber APM, Lercher MJ. 2014. Horizontal gene acquisitions by eukaryotes as drivers of adaptive evolution. Bioessays 36:9-20.
Schwarcz R, Bruno JP, Muchowski PJ, Wu H-Q. 2012. Kynurenines in the mammalian brain: when physiology meets pathology. Nat Rev Neurosci 13:465-477.
Seybold PG, Gouterman M. 1969. Porphyrins: XIII: Fluorescence spectra and quantum yields. Journal of Molecular Spectroscopy 31:1-13.
Sheen J. 1990. Metabolic Repression of Transcription in Higher Plants. The Plant Cell Online 2:1027-1038.
Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. Systematic Biology 51:492-508.
Shimodaira H, Hasegawa M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17:1246-1247.
Shimonaga T, Konishi M, Oyama Y, Fujiwara S, Satoh A, Fujita N, Colleoni C, Buleon A, Putaux JL, Ball SG, et al. 2008. Variation in Storage -Glucans of the Porphyridiales (Rhodophyta). Plant and Cell Physiology 49:103-116.
Sinclair SJ, Murphy KJ, Birch CD, Hamill JD. 2000. Molecular characterization of quinolinate phosphoribosyltransferase (QPRTase) in Nicotiana. Plant Molecular Biology 44:603-617.
Sprague GF, Winans SC. 2006. Eukaryotes learn how to count: quorum sensing by yeast. Genes & Development 20:1045-1049.
Stadnichuk IN, Rakhimberdieva MG, Bolychevtseva YV, Yurina NP, Karapetyan NV, Selyakh IO. 1998. Inhibition by glucose of chlorophyll a and phycocyanobilin biosynthesis in the unicellular red alga Galdieria partita at the stage of coproporphyrinogen III formation. Plant Science 136:11-23.
Stamatakis A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688-2690.
Stanhope MJ, Lupas A, Italia MJ, Koretke KK, Volker C, Brown JR. 2001. Phylogenetic analyses do not support horizontal gene transfers from bacteria to vertebrates. Nature 411:940-944.
Stein R, Gross W, Schnarrenberger C. 1997. Characterization of a xylitol dehydrogenase and a D-arabitol dehydrogenase from the thermo- and acidophilic red alga Galdieria sulphuraria. Planta 202:487-493.
Stenlid G. 1954. Toxic Effects of d-Mannose, 2-Desoxy-d-glucose, and d-Glucosamine upon Respiration and Ion Absorption in Wheat Roots. Physiologia Plantarum 7:173-181.
Sweetlove LJ, Fernie AR. 2013. The Spatial Organization of Metabolism Within the Plant Cell. Annual Review of Plant Biology 64:723-746.
Taiz L, Zeiger E. 2010. Plant Physiology: Sinauer Associates.

Talavera G, Castresana J. 2007. Improvement of Phylogenies after Removing Divergent and Ambiguously Aligned Blocks from Protein Sequence Alignments. Systematic Biology 56:564-577.
Taly J-F, Magis C, Bussotti G, Chang J-M, Di Tommaso P, Erb I, Espinosa-Carrasco J, Kemena C, Notredame C. 2011. Using the T-Coffee package to build multiple sequence alignments of protein, RNA, DNA sequences and 3D structures. Nat. Protocols 6:1669-1682.
Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology and Evolution.
Ternes CM, Schönknecht G. 2014. Gene Transfers Shaped the Evolution of De Novo NAD+ Biosynthesis in Eukaryotes. Genome Biology and Evolution 6:2335-2349.
The Gene Ontology C, Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, et al. 2000. Gene Ontology: tool for the unification of biology. Nature Genetics 25:25-29.
Timmis JN, Ayliffe MA, Huang CY, Martin W. 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat Rev Genet 5:123-135.
Treadaway MF. 1980. Polymer-optical brightener combinations in transparent film form useful as glazing materials capable of modifying plant growth rate. In: Google Patents.
Treangen TJ, Rocha EPC. 2011. Horizontal Transfer, Not Duplication, Drives the Expansion of Protein Families in Prokaryotes. PLoS Genet 7:e1001284.
Troxler RF, Bogorad L. 1966. Studies on the Formation of Phycocyanin, Porphyrins, and a Blue Phycobilin by Wild-Type and Mutant Strains of Cyanidium Caldarium. Plant Physiology 41:491-499.
Tsuji Y, Suzuki I, Shiraiwa Y. 2012. Enzymological Evidence for the Function of a Plastid-Located Pyruvate Carboxylase in the Haptophyte alga Emiliania huxleyi: A Novel Pathway for the Production of C4 Compounds. Plant and Cell Physiology 53:1043-1052.
Tyler BM, Tripathy S, Zhang X, Dehal P, Jiang RHY, Aerts A, Arredondo FD, Baxter L, Bensasson D, Beynon JL, et al. 2006. Phytophthora Genome Sequences Uncover Evolutionary Origins and Mechanisms of Pathogenesis. Science 313:1261-1266.
Vasileuskaya Z, Oster U, Beck CF. 2005. Mg-Protoporphyrin IX and Heme Control HEMA, the Gene Encoding the First Specific Step of Tetrapyrrole Biosynthesis, in Chlamydomonas reinhardtii. Eukaryotic Cell 4:1620-1628.
Viola R, Nyvall P, Pedersen M. 2001. The unique features of starch metabolism in red algae. Proceedings of the Royal Society of London Series B-Biological Sciences 268:1417-1422.
Voet D, Voet JG. 2004. Biochemistry: John Wiley & Sons.
Weber APM, Oesterhelt C, Gross W, Brautigam A, Imboden LA, Krassovskaya I, Linka N, Truchina J, Schneidereit J, Voll H, et al. 2004. EST-analysis of the thermo-acidophilic red microalga Galdieria sulphuraria reveals potential for lipid A biosynthesis and unveils the pathway of carbon export from rhodoplasts. Plant Molecular Biology 55:17-32.
Whitaker JW, McConkey GA, Westhead DR. 2009. The transferome of metabolic genes explored: analysis of the horizontal transfer of enzyme encoding genes in unicellular eukaryotes. Genome Biology 10.
Williams P. 2007. Quorum sensing, communication and cross-kingdom signalling in the bacterial world. Microbiology 153:3923-3938.
Woehle C, Dagan T, Martin WF, Gould SB. 2011. Red and problematic green phylogenetic signals among thousands of nuclear genes from the photosynthetic and apicomplexa-related Chromera velia. Genome Biology and Evolution.
Wogulis M, Chew ER, Donohoue PD, Wilson DK. 2008. Identification of Formyl Kynurenine Formamidase and Kynurenine Aminotransferase from Saccharomyces cerevisiae Using

Crystallographic, Bioinformatic and Biochemical Evidence‡. Biochemistry 47:1608-1621.
Würth C, Grabolle M, Pauli J, Spieles M, Resch-Genger U. 2013. Relative and absolute determination of fluorescence quantum yields of transparent samples. Nat. Protocols 8:1535-1550.
Yoon HS, Hackett JD, Pinto G, Bhattacharya D. 2002. The single, ancient origin of chromist plastids. Proceedings of the National Academy of Sciences of the United States of America 99:15507-15512.
Yoshimura E, Nagasaka S, Sato Y, Satake K, Mori S. 1999. Extraordinary high aluminium tolerance of the acidophilic thermophilic alga, Cyanidium caldarium. Soil Science and Plant Nutrition 45:721-724.
Yuasa HJ, Ball HJ. 2012. The evolution of three types of indoleamine 2,3 dioxygenases in fungi with distinct molecular and biochemical characteristics. Gene 504:64-74.
Yuasa HJ, Ball HJ. 2011. Molecular Evolution and Characterization of Fungal Indoleamine 2,3-Dioxygenases. Journal of Molecular Evolution 72:160-168.
Yuasa HJ, Ushigoe A, Ball HJ. 2011. Molecular evolution of bacterial indoleamine 2,3-dioxygenase. Gene 485:22-31.
Zdobnov EM, Apweiler R. 2001. InterProScan – an integration platform for the signature-recognition methods in InterPro. Bioinformatics 17:847-848.
Zhu MM, Skraly FA, Cameron DC. 2001. Accumulation of Methylglyoxal in Anaerobically Grown Escherichia coli and Its Detoxification by Expression of the Pseudomonas putida Glyoxalase I Gene. Metabolic Engineering 3:218-225.

APPENDICES

Author / Year	Enzyme	Supporting Information From Literature	Genome
Oesterhelt 2008	L-Ascorbate Peroxidase	EF589723 (GSuAPX01), EF589721 (GSuAPX02)	Gasu_33520, Gasu_32860
	Cytochrome C Peroxidase	EF589722 (GSuCcP01)	Gasu_42210
	Peroxidase	EF589724 (GSPrx01), EF589725 (GSPrx02), EF589726 (GSPrx03), EF589727 (GSPrx04)	Gasu_17800, Gasu_17790, Gasu_50480, Gasu_50490
Oesterhelt 2002	Glucokinase	Enzymatic Activity Detected	Gasu_04750
	Galactokinase	Enzymatic Activity Detected	Gasu_09360
	Glycerokinase	Enzymatic Activity Detected	Gasu_03170, Gasu_60010 (Glycerol Kinase)
	Hexokinase	Enzymatic Activity Detected	Not Supported By Genome
	Xylulokinase	Enzymatic Activity Detected	Gasu_46540
	Pyruvate, Orthophosphate Dikinase (PPDK)	Enzymatic Activity Detected	Gasu_42070
McCoy 2009	Sarcosine Dimethylglycine Dimethyltransferase	Assayed Enzymatically & Crystallized	Gasu_06500, Gasu_07580, Gasu_07590
Marquardt 1997, 2001	Light Harvesting Complexes	Immunoprecipitated & Determined Spectroscopically	Protein Cluster 32 (Light-Harvesting Complex Proteins)
Gross 1999 & Reichert 2003	Fructose-1,6-Bisphosphatase	Molecular Mass: 165 kDa; Enzymatic Assay	Protein Cluster 87
Seckbach 1992	Catalase	Enzymatic Assay	Gasu_11340
	Glycolate Oxidase	Enzymatic Activity Detected	Protein Cluster 109 (S-2-Hydroxy-Acid Oxidase)
	Glutamate-Glyoxylate Aminotransferase	Enzymatic Activity Detected	Gasu_04830, Gasu_31580 (Alanine-Glyoxylate Transaminase / Serine-Pyruvate Aminotransferase)
	Serine-Glyoxylate Aminotransferase	Enzymatic Activity Detected	Gasu_01160
	Hydroxypyruvate Reductase	Enzymatic Activity Detected	Gasu_47310 (Glyoxylate Reductase / Hydroxypyruvate Reductase)

Tischendorf 2007	Triose Phosphate Isomerase	Enzymatic Activity Detected	Gasu_29840
	Glyceraldehyde-3-Phosphate Dehydrogenase (NAD Dependent)	Enzymatic Activity Detected	Gasu_04650
	Glyceraldehyde-3-Phosphate Dehydrogenase (NADP Dependent)	Enzymatic Activity Detected	Gasu_43210, Gasu_48130
	3-Phosphoglycerate Kinase	Enzymatic Activity Detected	Gasu_00270, Gasu_34520
	Fructose-1,6-Bisphosphate Aldolase	Enzymatic Activity Detected	Protein Cluster 42
	Fructose-1,6-Bisphosphatase	Enzymatic Activity Detected	Protein Cluster 87
	Malate Dehydrogenase	none	Protein Cluster 84, Gasu_24070, Gasu_28900
	Glucosephosphate Isomerase	Enzymatic Activity Detected	Gasu_20100, Gasu_55050 (Glucose-6-Phosphate Isomerase)
	Glucose-6-Phosphate Dehydrogenase	Enzymatic Activity Detected	Gasu_04100 ,Gasu_42750 (Glucose-6-Phosphate 1-Dehydrogenase)
	6-Phosphogluconate Dehydrogenase	Enzymatic Activity Detected	Gasu_02840, Gasu_15770
	Phosphoglucomutase	Enzymatic Activity Detected	Gasu_08170, Gasu_46280, Gasu_28550
	Ribulose-5-Phosphate Epimerase	Enzymatic Activity Detected	Gasu_18200 (Ribulose-Phosphate 3-Epimerase
Weber 2004	Hexokinase	none	Not Supported By Genome
	Ribose-5-Phosphate Isomerase	none	Gasu_37650, Gasu_51780
	Glucose-6-Phosphate Dehydrogenase	EST present	Gasu_04100 ,Gasu_42750 (Glucose-6-Phosphate 1-Dehydrogenase)
	Transaldolase (Pentose Phosphate Pathway)	none	Gasu_24580
	Fumarase (Citric Acid Cycle)	none	Gasu_00620 (Fumarate Hydratase)
	Phosphoglucomutase	none	Gasu_08170, Gasu_46280, Gasu_28550
	Phosphoglucoisomerase	none	Gasu_20100, Gasu_55050 (Glucose-6-Phosphate Isomerase)
	Fructokinase	none	Gasu_01610, Gasu_04420
	Glucokinase	none	Gasu_04750

	UDP-Glucose Pyrophosphorylase	A4_36F11	Gasu_00860 (UTP-Glucose-1-Phosphate Uridylyltransferase)
	ADP-Glucose Pyrophosphorylase	none	Not Supported By Genome
	Floridoside-Phosphate Synthase	none	Not Supported By Genome
	Floridoside-Phosphate Phosphatase	none	Not Supported By Genome
	Galactose-1-Phosphate Uridylyltransferase	HET_36B03, HET_11G09	Gasu_52840
	UDP-Glucose Epimerase	none	Gasu_12030 (UDP-glucose 4-epimerase)
	α-Galactosidases	A4_22C06, HET_24G12	Gasu_33800, Gasu_48280, Gasu_60270
	Phosphoglycolate Phosphatase	HET_43H04	Gasu_54160
	Glycolate Oxidase	A4_38D07, HET_37H12	Protein Cluster 109 (S-2-Hydroxy-Acid Oxidase)
	Catalase	A4_33F10	Gasu_11340
	Serine-Glyoxylate Aminotransferase	A4_32A08	Gasu_01160 (Serine-Glyoxylate Transaminase)
	Serine Hydroxymethyltransferase	A4_32A08, A4_36C10, A4_10D04, A4_39-E01	Protein Cluster 207
	Glutamine Synthetase	A4_33A10, A4_33F03, HET_10AO, HET_24C05	Gasu_37790
	Glycerate Kinase	none	Gasu_01520
	Biotin Carboxylase	A4_40C10	Gasu_03050 (Acetyl-CoA Carboxylase / Biotin Carboxylase)
	β-Carboxyltransferase (Acetyl-CoA Carboxylase subunit)	HET_25A08	Gasu_40060
	3-Ketoacyl-ACP Reductase	Gasu_01440	Gasu_60850 (3-Oxoacyl-[Acyl-Carrier-Protein] Reductase)
	Enoyl-ACP Reductase	HET_03C09	Gasu_47760 (Enoyl-[Acyl-Carrier Protein] Reductase I)
	Fatty Acid Desaturases	EST present	Gasu_25760, Gasu_32800
	Acyl-CoA Synthetase	EST present	Protein Cluster 41, Gasu_01790, Gasu_03440, Gasu_35980
	Thiolases	EST present	Gasu_26380, Gasu_55920 (Acetyl-CoA C-acetyltransferase)
	Acyl-CoA Oxidase	EST present	Gasu_21670, Gasu_50940
	Acetyl-CoA Thiolase	EST present	Gasu_26380, Gasu_55920 (Acetyl-CoA C-acetyltransferase)

	Enoyl-CoA Hydratase	none	Gasu_06240 (3-Hydroxyacyl-CoA Dehydrogenase / Enoyl-CoA Hydratase /3-Hydroxybutyryl-CoA Epimerase / Enoyl-CoA Isomerase)
	3-Hydroxyacyl-CoA Dehydrogenase	none	Gasu_06240 (3-Hydroxyacyl-CoA Dehydrogenase / Enoyl-CoA Hydratase /3-Hydroxybutyryl-CoA Epimerase / Enoyl-CoA Isomerase)
	Phosphatidylserine Decarboxylase	HET_44E08	Gasu_07890
	Ethanolamine Kinase	HET_10E09	Gasu_18300 (Choline / Ethanolamine Kinase)
	Ethanolamine Phosphate Cytidylyltransferase	A4_10G06	Gasu_26040, Gasu_30330, Gasu_25270 (Phosphatidate Cytidylyltransferase)
	Phosphoethanolamine Methyl-Transferase	A4_09F12	Not Supported By Genome
	Phosphatidylglycerophosphate Synthase (PGP Synthase)	GS06870, two ESTs	Gasu_18340, Gasu_01450 (CDP-Diacylglycerol-Glycerol-3-Phosphate 3-Phosphatidyltransferase)
	Cardiolipin Synthase	HET_11H01	Gasu_07980
	Inositol-3-Phosphate Synthase	HET_12A09	Gasu_11420, Gasu_30560 (Myo-Inositol-1-Phosphate Synthase)
	Phosphatidylinositol 4-Kinase (PI-4-Kinase)	HET_10C06	Gasu_17510, Gasu_36730
	Serine Palmitoyltransferase	HET_31C05	Gasu_24950, Gasu_30970
	Sphingolipid-D4 Desaturase	GS00710, two ESTs	Gasu_65890 (Sphingolipid Delta-4 Desaturase)
	LpxA (UDP-N-Acetylglucosamine Acyltransferase)	EST present	Gasu_47350, Gasu_63160 (Acyl-[ACP]-UDP-N-Acetylglucosamine O-Acyltransferase)
	LpxB (Lipid-A-Disaccharide Synthase)	EST present	Gasu_47910, Gasu_06440, Gasu_47780,
	LpxC (UDP-3-0-[3-Hydroxymyristoyl] N-Acetylglucosamine Deacetylase)	EST present	Gasu_03310 (UDP-3-O-[3-Hydroxymyristoyl] N-acetylglucosamine Deacetylase)

	LpxD (UDP-3-0-[3-Hydroxymyristoyl] Glucosamine N-Acetyltransferase)	EST present	Gasu_49020 (UDP-3-O-[3-Hydroxymyristoyl] Glucosamine N-acyltransferase)
	LpxK (Tetraacyldisaccharide 4'-Kinase)	EST present	Gasu_45040 (Tetraacyldisaccharide 4'-Kinase)
	WaaA (3-Deoxy-D-Manno-Octulosonic-Acid Transferase)	EST present	Gasu_00550 (3-Deoxy-D-Manno-Octulosonic-Acid Transferase)
	Polyphosphate Kinase	EST present	Gasu_32760
	Glycerol Kinase	EST present	Gasu_03170, Gasu_60010
	Glucokinase	Sequence Identifier	Gasu_04750
	Galactokinase	Sequence Identifier	Gasu_09360
	Fructokinase	Sequence Identifier	Gasu_01610, Gasu_04420
	Xylulokinase / Ribulokinase	Sequence Identifier	Gasu_46540 / Gasu_19690, Gasu_20140 (Phosphoribulokinase)
	Ribokinase	Sequence Identifier	Gasu_33210
	Glycerol Kinase	GS02230, A4_40D08	Gasu_03170, Gasu_60010
	Phosphoglucomutase	Sequence Identifier	Gasu_08170, Gasu_46280, Gasu_28550
	Phosphomannomutase	A4_10D11	Gasu_47860
	Glucose-6-Phosphate Isomerase	Sequence Identifier	Gasu_20100, Gasu_55050
	Mannose-6-Phosphate Isomerase	Sequence Identifier	Gasu_38450
	Sorbitol Dehydrogenase	Sequence Identifier	Gasu_40800 (L-Iditol 2-Dehydrogenase)
	Glycogenin-Glucosyltransferase (aka Glycogenin)	HET_19E06	Gasu_06400 (Glycogenin-like Protein)
	Glycogenin	Sequence Identifier	Gasu_06400 (Glycogenin-like Protein)
	Glycogen Synthase	HET_25C3	Gasu_14030 (Glycogen Synthase)
	Glycogen Phosphorylase	A4_22A10, HET_36G09	Gasu_26320, Gasu_53650 (Glycogen Phosphorylase)
	4-α-Glucanotransferase D-enzyme	Sequence Identifier	Gasu_34050
	(1,4)-α-Glucan Branching Enzyme	Sequence Identifier	Gasu_50970
	Isoamylase Type Debranching Enzyme	Sequence Identifier	Gasu_05830, Gasu_10920
	Isoamylase	A4_41H04	Gasu_05830, Gasu_10920
	α-Amylase	A4_6H07	Gasu_21330,

			Gasu_32260, Gasu_14640
	β-Amylase	A4_14E02	Gasu_04150, Gasu_29320, Gasu_58550
	Amylopullulanase	Sequence Identifier	Not Supported By Genome (Might be a type of α-amylase)
	Glucan-(1,4)-α-Glucosidase	Sequence Identifier	Gasu_25530, Gasu_25520, Gasu_00130, Gasu_06640 (α-Glucosidase)
	Fucosyltransferase	Sequence Identifier	Gasu_23840
	Galactosyltransferase	Sequence Identifier	Gasu_09650
	Glucosyltransferase	Sequence Identifier	Gasu_29940 (α-1,3-Glucosyltransferase), Gasu_44200 (α-1,4-Glucosyltransferase)
	Mannosyltransferase	GS11580, A4_34E06	Gasu_08940, Gasu_42280 (α-1,2-Mannosyltransferase), Gasu_08400, Gasu_57500 (α-1,3-Mannosyltransferase), Gasu_12290 (β-1,4-Mannosyltransferase), Gasu_17110, Gasu_59940, Gasu_59850 (α-1,6-Mannosyltransferase)
	Xylanase	Sequence Identifier	Not Supported By Genome
	Nucleoside-Diphosphate-Sugar Epimerase	HET_37D08	Gasu_16250, Gasu_43480 (Nucleoside-Diphosphate Kinase)
	UDP-Glucose-4 Epimerase	A4_14C11	Gasu_04050 (Aldose 1-Epimerase)
	UDP-Galactose-4 Epimerase*	Sequence Identifier	Gasu_04050 (Aldose 1-Epimerase)
	UTP-Glucose-1-Phosphate Uridylyltransferase	A4_36F11	Gasu_00860 (UTP-Glucose-1-Phosphate Uridylyltransferase)
	UTP-Galactose-1-Phosphate Uridylyltransferase	HET_36B03, HET_11G09	Gasu_52840 (Galactose-1-Phosphate Uridylyltransferase)
	α-Galactosidase	HET_24G12, A4_22C06	Gasu_33800, Gasu_48280, Gasu_60270
	Secreted α-Galactosidase	Sequence Identifier	Unable To Determine
	β-Galactosidase	Sequence Identifier	Protein Cluster 89
	α-Glucosidase	A4_5D11, A4_32D09	Gasu_25520, Gasu_25530, Gasu_00130, Gasu_06640
	Mannosyl-Oligosaccharide	A4_6F10, A4_5D11	Gasu_23130

	Glucosidase		
	β-Mannosidase	Sequence Identifier	Gasu_22860
	APS-Reductase	Sequence Identifier	Gasu_09200
	PAPS-Reductase	Sequence Identifier	Not Supported By Genome
	Trehalose Synthase	A4_6H07	Gasu_13550, Gasu_55660 (Maltose α-D-Glucosyltransferase aka Trelahose Synthase)
	Trehalose-Phosphate Synthase	A4_23G07, A4_8G11, HET_40F04, A4_10B08, A4_17F02, HET_40F04, HET_42A08, A4_14F07	Protein Cluster 46 (α-α-Trehalose-Phosphate Synthase)
	Trehalose Phosphatase	HET_19A04	Gasu_25840, Gasu_55870
	Trehalase	HET_20C04	Gasu_29340 (α-α-Trehalase)
	Sucrose-Phosphate Synthase	HET_13E07, HET_24A9, A4_11B10, A4_T7comboE05, A4_T7comboC01, A4_11E07, A4_32F05	Gasu_03810
	Sucrose-Phosphate Phosphatase	Sequence Identifier	Gasu_32680 (Sucrose Phosphatase-1 [SPP1])
Baranowski 2007	Ribulose Bisphosphate Carboxylase Oxidase	Purified & Crystallized	Gasu_40760
Gross 1995	Galactose-1-Phosphate: UDP-Glucose Uridylyltransferase	Homodimer; Molecular Weight: Between 60 & 82 kDa; Enzymatic Assay	Gasu_52840 (Galactose-1-Phosphate Uridylyltransferase)
Gross 1997	Aldose Reductase	Homodimer; Molecular Weight: ~80 kDa; Enzymatic Assay	Protein Cluster 210 (Aldehyde Reductase)
Stein 1997 & Gross 2003	Myo-Inositol Dehydrogenase	Homotetramer; Molecular Weight: 42 kDa; Enzymatic Assay; 2 Isozymes (Cytosolic & Mitochondrial)	Gasu_32250
Heilmann 1997	Mannose-6-Phosphate Isomerase	Molecular Weight: 48 kDa; Enzymatic Assay	Gasu_38450
	Fructokinase	Molecular Weight: 36 kDa; Enzymatic Assay	Gasu_01610, Gasu_04420
Oesterhelt 1996	Phosphomannomutase	Dimer; Molecular Weight: 40 kDa (subunit); Enzymatic Assay	Gasu_47860
	Phosphoglucomutase	Dimer; Molecular Weight: 40 kDa (subunit); Enzymatic Assay	Gasu_08170, Gasu_46280, Gasu_28550
Stein 1997	Xylitol Dehydrogenase	Molecular Weight: 220-295 kDa; Enzymatic Assay	Gasu_25510
	D-Arabitol Dehydrogenase	Molecular Weight: 105 kDa; Enzymatic Assay	Not Supported By Genome (probably resides within protein cluster #70)

Table A1: Comparisons between genome annotation and characterized or detected enzymes in *G. sulphuraria*. Previously reported enzymes are from experimental work and EST datasets. From left to right: Author and year, enzyme, supporting information, genome identifier (Gasu_#).

VITA

Chad M. Ternes

Candidate for the Degree of

Doctor of Philosophy

Thesis: METABOLIC EVOLUTION IN *GALDIERIA SULPHURARIA*

Major Field: Plant Science

Biographical:

Education:

Completed the requirements for the Doctor of Philosophy in Plant Science at Oklahoma State University, Stillwater, Oklahoma in May, 2015.

Completed the requirements for the Bachelor of Science in Botany at Oklahoma State University, Stillwater, Oklahoma in 2009.

Experience:

Graduate Teaching Assistant: 2009 – Present (exceptions directly below)

Graduate Research Assistant: 2010 & 2012

Professional Memberships:

N/A